§ THE DIALECTIC REVIEW

Volume III · Autumn 2025

ISSUE 06

What We Owe Each Other

Autumn 2025 · 11 pieces

Eight essays and a dialogue on obligation, gratitude, and testimony — the first issue to carry a formal Counterargument alongside its essays.

ISSN 3049-7170

thedialecticreview.org

CONTENTS

ETHICS 4

The Ethics of Almost Keeping a Promise

Fatima Zahra Belkacem — Atlas International School, Casablanca

ETHICS 7

Why Gratitude Isn't a Debt

Marcus Webb — Elmwood Grammar School, Bristol

POLITICAL PHILOSOPHY 10

Can a Democracy Be Too Democratic?

Theo Lindqvist — Uppsala International School

THE COUNTERARGUMENT 14

Ulysses Chose His Ropes; Most Citizens Did Not

Maya Okonkwo — Riverside Academy, Lagos

ETHICS 16

Do Future People Have Rights?

Lucas Ferreira — Colégio Santo Inácio, São Paulo

THE COUNTERARGUMENT 20

A Right Held by a Role Is Not Much of a Right

Iris Chen — Raffles Institution, Singapore

EPISTEMOLOGY 22

Testimony and the Right to Be Believed

Simone Aubert — Lycée Chateaudun, Lyon

METAPHYSICS 25

What Do We Owe a Future Self?

Yusuf Aydin — Anatolian Science High School, Ankara

AESTHETICS 28

The Aesthetics of Apology

Camille Laurent — Académie du Nord, Lille

PHILOSOPHY OF SCIENCE 31

Can a Model Owe Its Data Anything?

Divya Menon — Cotton Tree International School, Chennai

EPISTEMOLOGY · DIALOGUE 34

On Believing What You Cannot Prove

Dorotea Novak & Priya Raman — Joint submission — Gymnázium Jana Nerudy, Prague & Millfield International, Colombo

ETHICS

The Ethics of Almost Keeping a Promise

Partial performance of a promise is usually judged as a lesser wrong than breaking it outright. This essay asks whether that judgment survives scrutiny.

Fatima Zahra Belkacem Atlas International School, Casablanca

A friend promises to help you move on Saturday, arrives, but leaves after an hour, citing a headache, having moved perhaps a third of what was planned. Compare this to a friend who simply never shows up. Nearly everyone judges the first friend less harshly than the second, and this essay does not dispute that the judgment is broadly correct. What it disputes is the usual explanation for it — that partial performance is a fraction of a promise kept, so that a third of the promise fulfilled earns a third less blame. I want to argue that partial performance is doing something structurally different from full performance, not a discounted version of it, and that recognizing this changes what we owe each other when we cannot, or do not, follow through completely.

The natural view treats a promise as something like a debt with a quantity attached: promise to move ten boxes, move three, and you have paid thirty percent of what you owed, with seventy percent still outstanding as a debt or, if abandoned, as a wrong proportional to what remains unpaid. This model explains our intuition reasonably well in cases like the moving example, where the promised task is genuinely divisible into independent units, and doing some of it

produces exactly the same value, pro-rated, as doing all of it. Three boxes moved are three boxes moved, whether or not the other seven ever get moved by anyone.

But consider a different promise: to give a eulogy, and to leave, overcome, after the first paragraph. Here the debt model fails badly, because a fragment of a eulogy is not a fraction of the value of a complete one — it may be worse than no eulogy at all, an incomplete gesture that draws attention to the promise-breaking in a way silence would not have. Or consider a promise to keep a secret for a lifetime, broken after twenty years: this is not eighty percent of the promise unkept in any sense that helps the person whose secret is now out. In both cases, what was promised was not a quantity to be pro-rated but a threshold condition — the eulogy delivered, the secret never told — and partial performance does not approach the threshold, it simply fails to meet it, sometimes while additionally creating harms (the awkwardness of the abandoned eulogy, the twenty years of trust now retroactively complicated) that full non-performance would not have created.

This suggests that promises come in at least two structurally different kinds, and that our forgiving judgment of the moving-day friend is not evidence for a general principle of proportional forgiveness, but evidence that the moving promise happened to be the divisible kind, where partial performance really does produce partial value. The mistake is generalizing from the divisible case to promises generally, which lets people who have made threshold-type promises — to finish a project, to show up for an entire event, to never repeat something told in confidence — claim partial credit for partial performance that, in the relevant respect, is not partial at all. Showing up to the first half of your sister's wedding and leaving before the vows is not half

of attending her wedding; for a threshold promise like “I will be there”, meaning at the moment that mattered, it is a failure to be there, dressed in the language of an honest effort.

What we owe each other, on this view, includes a duty most promise-keeping advice skips: to be honest, at the time of promising, about which kind of promise is being made, since the discount that partial performance deserves depends entirely on this classification and the two parties may not agree on it. The friend who promises to help you move should know, and probably does know, that this is a divisible promise where an hour of real effort is worth something even if the day is cut short. The friend who promises to be at your surgery when you wake up should recognize that this is a threshold promise where nothing short of actually being there discharges it, and that offering an apology calibrated to “I did try” misunderstands what trying was ever going to be worth in a promise of this shape. Getting this classification right, honestly and in advance rather than defensively after the fact, is itself part of what integrity in promise-making requires — not just the intention to follow through, but an accurate sense of what following through, partially, would actually be worth to the person you promised.

ETHICS

Why Gratitude Isn't a Debt

Treating gratitude as a ledger to be balanced misdescribes the obligation it actually creates — and makes it easier to discharge than it should be.

Marcus Webb Elmwood Grammar School, Bristol

It is common to describe gratitude using the language of debt: we say we "owe" someone for a kindness, that we want to "repay" a favor, that an unreciprocated gift leaves us "indebted." The vocabulary is so natural that it can seem like more than a metaphor — a genuine description of the moral structure involved. This essay argues that the debt model gets the structure of gratitude wrong in a way that matters, because debts are discharged by equivalent transfer, and gratitude is not.

A financial debt has a defining feature: once repaid in full, the relationship it created is over. If I owe you fifty dollars and pay you fifty dollars, we are, as far as that transaction goes, even; nothing further binds us, and continuing to treat me as though I still owed you something would be a mistake on your part, not a virtue. This is precisely what feels wrong about applying the same structure to gratitude. If a friend cares for you through a serious illness, and you later repay them with an equivalent quantity of care during their own illness, the debt model implies you are now even — the account is settled, nothing further is owed, and the friendship can, as far as this specific transaction goes, be considered closed. But this is not how gratitude actually seems to work among people who take it seriously. The repayment does not close anything; if anything, it deepens the

relationship's ongoing character, and continuing to feel grateful for the original care, even after having reciprocated in kind, is not seen as a failure to understand that the debt has been paid — it is seen as basic decency.

The reason, I think, is that gratitude is not a transfer of equivalent value but a recognition of a certain kind of regard — the fact that someone attended to your needs when they were not obligated to, at some cost to themselves, because they valued you specifically. Repaying this in kind does not cancel the original fact; it adds a second instance of the same kind of regard, and the appropriate response to two instances of being cared for is not "we are even" but a relationship that has now demonstrated itself twice over. Debts are paid off; regard is not the kind of thing that gets paid off, because the value it responds to — being cared about, specifically and non-instrumentally — does not diminish or transfer when reciprocated. It accumulates, or at least persists, in a way a financial balance does not.

This distinction has a practical edge that the debt model actively obscures. If gratitude were a debt, the fastest way to be free of an uncomfortable obligation would be to repay it as quickly and completely as possible, converting an ongoing relational fact into a closed transaction — and people who are uncomfortable with owing anyone anything often do exactly this, reciprocating hastily and treating the reciprocation as license to withdraw. The debt model gives this maneuver a kind of moral cover: "I paid them back, we're square." But if gratitude tracks regard rather than value transferred, hasty even-ing-up of this kind is not completing an obligation, it is refusing one — using the language of repayment to avoid the actual thing gratitude asks for, which is not a matching transfer but a willingness to remain in a relationship characterized by ongoing, non-transactional regard for the other person's wellbeing.

None of this means gratitude has no natural resolution, or that a person who has been helped owes endless, escalating deference. The resolution gratitude looks for is not zeroing a balance but a change in how two people stand toward each other — from strangers, or acquaintances, to people who have shown they will attend to each other's needs without being asked. Once that fact is established, gratitude has done its work, not by canceling itself out through repayment, but by successfully creating the kind of relationship where future care, in either direction, no longer needs the language of debt and repayment at all. That absence of accounting, not a completed one, is what gratitude was aiming at from the start.

POLITICAL PHILOSOPHY

Can a Democracy Be Too Democratic?

A defence of institutional friction against the demand for perfect legislative responsiveness.

Theo Lindqvist Uppsala International School

There is a recurring complaint made against modern democracies, from across the political spectrum, that goes roughly like this: the public wants X, by a clear and stable majority, and yet the system — through courts, upper chambers, supermajority requirements, independent agencies, or sheer procedural friction — fails to deliver X, or delivers it slowly, watered down, or not at all. The complaint treats this gap between popular will and policy outcome as a defect to be minimized: a truly democratic system, on this view, would translate majority preference into law as directly and quickly as possible, and anything that slows or blocks this translation needs a special justification to survive. This essay defends the opposite starting point: that some friction between popular will and immediate policy outcome is not a defect democracies tolerate despite their commitment to self-rule, but a condition of that commitment being coherent over time at all.

The case for maximal responsiveness rests on treating the momentary majority as the final authority on what the polity wants. But a polity is not a single decision made once; it is a sequence of decisions made by a continuously changing population, most of whom will be bound by today's decision long after today's majority has moved on, died, or changed its mind. A system that translates

every current majority preference immediately into binding law is not more responsive to "the people" in any coherent sense — it is maximally responsive to whichever subset of the people happens to constitute a majority on a given Tuesday, at the expense of the same polity's own preferences on other Tuesdays, including future Tuesdays whose preferences cannot yet be registered because the people who will hold them are not yet voting, or not yet born.

Constitutional constraints, judicial review, and supermajority requirements can be understood, on this view, not as anti-democratic barriers imposed against popular will, but as devices by which a polity binds its future majorities on behalf of the values it takes to be more stable than any single election — in the same way an individual might set up a delayed transfer of funds specifically to prevent a future, differently-motivated version of themselves from spending it impulsively. Jon Elster's image of constitutions as "Ulysses bound to the mast" captures this precisely: Ulysses does not experience the ropes as an infringement of his freedom imposed by an external party, because he is the one who ordered them tied, precisely because he predicted that a future, differently-motivated version of himself — hearing the sirens — would want something his more stable, prior judgment had already rejected.[1] A polity that entrenches certain rights or procedures against ordinary majority repeal is doing something structurally similar: expressing a preference, at the constitutional level, that is meant to survive the polity's own future changes of mind on matters it has judged too important, or too easily distorted by momentary passion, to leave to ordinary majorities.

Ulysses does not experience the ropes as an infringement, because he is the one who ordered them tied.

The obvious danger is that this argument, taken too far, becomes a blank check for entrenched minorities to block anything a current majority wants, dressed up as protecting the polity from its future self. This is a real risk, and the argument here is not a defense of any particular friction — an unelected upper chamber with an absolute veto is not thereby justified merely because friction exists. The distinction that matters is between friction that protects a specific, previously-endorsed higher-order commitment (a bill of rights ratified by a broader process than ordinary legislation, a supermajority requirement for changing the rules of the game itself) and friction that simply protects the interests of whoever currently benefits from the status quo, with no connection to any commitment the polity itself endorsed at a more considered moment. The first kind of friction is democracy constraining itself for reasons a democracy can state and defend. The second is an obstruction wearing procedural clothing, and conflating the two — which the maximal-responsiveness view is forced to do, since it treats all friction as equally suspect — makes it harder, not easier, to tell them apart.

A democracy with zero friction is not a purer democracy; it is a series of unrelated, potentially incoherent majoritarian snapshots with no mechanism for the polity to commit itself to anything across time, including to the very idea that future majorities should be bound by rules the present majority did not personally choose. Some friction, calibrated to protect considered, higher-order commitments rather

than incumbent advantage, is not democracy's enemy. It is closer to what makes democracy, across a lifetime rather than a single election, a coherent project at all.

NOTES

1. Jon Elster, *Ulysses and the Sirens: Studies in Rationality and Irrationality* (1979).

THE COUNTERARGUMENT

Ulysses Chose His Ropes; Most Citizens Did Not

A response to "Can a Democracy Be Too Democratic?"

The Ulysses analogy only works if the same person ties the knot and later strains against it. Constitutional friction rarely has this structure, and the difference is the whole problem.

Maya Okonkwo Riverside Academy, Lagos

Lindqvist's analogy is elegant and, I think, precisely where it fails. Ulysses is bound by ropes he himself ordered tied, by his own hand, in full knowledge of what he was doing and why. The friction is self-imposed, temporary, and specifically targeted at a predictable failure of his own future judgment that he, personally, anticipated and endorsed constraining. Almost none of the constitutional friction Lindqvist wants to defend has this structure, and the gap between the analogy and the reality is not a minor detail — it is the entire question of whether the friction is democratic self-binding or something that only wears self-binding's clothing.

Consider who actually "tied the knot" in most existing constitutional orders: a convention of delegates, often centuries ago, elected (if elected at all) by a franchise that excluded most of the population later bound by their decisions, amending their own work through supermajority thresholds specifically designed to make revision by later, more inclusive majorities extremely difficult. This is not Ulysses binding himself. It is closer to Ulysses' grandfather binding Ulysses, without asking him, and then requiring Ulysses to assemble two-thirds of the crew before the ropes can even be loosened to be re-

examined. Lindqvist's essay needs the binding to be an act of the same continuous agent across time — "the polity binds its future majorities on behalf of the values it takes to be more stable" — but a polity is not a single continuous agent in the sense this requires; it is a sequence of substantially different electorates, and treating an eighteenth-century electorate's preferences as the earlier, wiser phase of the same self that a twenty-first-century electorate must now defer to assumes exactly the kind of trans-generational personal identity that political philosophy has never been able to establish for a polity the way it can, roughly, for one person across a single lifetime.

This is not an argument for zero friction, and Lindqvist is right that some friction can be genuinely self-binding — a right ratified by a recent, broad, still-living majority, revisable by a process that majority itself endorsed, functions much more like Ulysses' own ropes. But this describes a narrow subset of the friction actually operating in most democracies, and the more common case — a rule imposed by a narrower, older, less representative body, entrenched against revision by a broader, more current one — is not democracy binding itself against its own future irrationality. It is one generation, or one faction within it, binding all subsequent generations without their consent, and calling this Ulysses and the mast flatters an arrangement that, in its actual historical origins, usually looks like something closer to the opposite.

ETHICS

Do Future People Have Rights?

The puzzle behind climate obligations owed to people who do not yet exist to hold them.

Lucas Ferreira Colégio Santo Inácio, São Paulo

Suppose a policy choice made today will determine, decades from now, which of two possible populations comes into existence — not merely which people among an already-existing population will be affected, but literally which people will exist at all, since different policies lead to different people meeting, marrying, and having children at different times, cascading into entirely different sets of future individuals. Now suppose one of the two possible populations will inherit a badly damaged climate, and the other will not, but crucially, the badly-off population, given that it exists, would still, on balance, judge its lives worth living — hard, constrained, but not so miserable as to prefer non-existence. Standard rights talk asks: whose rights does the harmful policy violate? And this essay argues that the honest answer is: nobody's, in the sense rights talk usually requires — which is not a reason to stop worrying about the harm, but a reason to see that our concept of rights was not built for this problem, and cannot do the moral work climate ethics wants from it.

This is a version of what Derek Parfit called the non-identity problem, and it is worth being precise about why it defeats the ordinary rights-based complaint.[1] A right is standardly understood as something that protects an interest of an existing or determinately-identified future person against being made worse off

than they otherwise would have been. But in the case described, there is no one who is made worse off by the harmful policy, because the specific individuals who will exist in the damaged-climate scenario would not have existed at all under the alternative policy — a different set of people, born to different parents at different times, would exist instead. You cannot make a particular person worse off by an act that is a necessary condition of that very person's existence, because the only alternative to the harm, for that specific person, is that person's own non-existence, and (given that their life is still worth living) non-existence is not better for them than the harmed life they actually get to have. The policy harms no one in particular, even though it is, in an obvious ordinary sense, a worse outcome than the alternative.

The temptation is to route around this with an impersonal standard — the policy makes the world worse, full stop, aggregate wellbeing considered, without needing any particular victim to ground the complaint. This works as a consequentialist verdict, but it costs rights-based ethics exactly the vocabulary that made it distinctive and, for many purposes, more demanding than pure aggregation: the idea that some acts are wrong specifically because they wrong *someone*, violate a claim *held by* a particular party, and that this wrongness is not reducible to, or fully captured by, a utilitarian tally of total welfare. If the best available complaint against the climate-damaging policy is "total welfare would have been higher otherwise," we have quietly abandoned rights talk in favor of the aggregative framework it was often meant to supplement or check, and lost the resources to explain why this feels like a wrong done *to* the damaged-climate population specifically, rather than merely a regrettable arithmetic fact about the universe.

I want to suggest a different route, one that does not require every future person to have a right violated in order to explain what is wrong here. Some obligations are owed not to particular individuals but to a role or a position — the position of "whoever exists in the future and inherits what we leave behind," regardless of which specific people come to occupy it. This is close to how we already treat obligations to institutions across changing membership: a university's obligation to maintain its library is not owed to any particular future student, all of whom are as yet undetermined and substitutable, but to the position "future student" as such, which some determinate population will eventually occupy no matter which specific individuals they turn out to be. Climate obligations to the future can be understood the same way: not as rights held by determinate future rights-holders, which the non-identity problem shows we cannot coherently posit, but as a duty attached to the role that some future population, whoever they turn out to be, will occupy — a duty that does not require solving the metaphysics of personal identity across non-identity cases, because it was never owed to individuals in the first place.

This reframing has a cost, which is honesty requires admitting: role-based duties are usually considered weaker, more defeasible, and less action-guiding than duties grounded in an actual person's claim against you, precisely because no specific party can come forward and say "you wronged me." But a weaker duty that we can actually state coherently is more useful than a stronger one — a violated right — that turns out, on the non-identity problem's own terms, not to exist. Climate ethics does not need every future person to have a right in order to have a real and demanding obligation on its hands. It needs only the much more modest, and considerably more defensible, claim that whoever eventually occupies the future is owed something by whoever currently occupies the position of deciding what that future will contain.

NOTES

1. Derek Parfit, *Reasons and Persons* (1984), part IV.

THE COUNTERARGUMENT

A Right Held by a Role Is Not Much of a Right

A response to "Do Future People Have Rights?"

Ferreira's role-based duty is a clever patch, but it quietly gives climate policy exactly the discretion that rights language exists to remove.

Iris Chen Raffles Institution, Singapore

Ferreira is right that the non-identity problem defeats a naive rights complaint, and right that abandoning rights language for pure aggregation costs something real. But his proposed replacement — a duty owed to a role, not a person — solves the metaphysical problem by creating a practical one that his essay does not fully confront: role-based duties, precisely because no specific claimant can ever come forward, are far more easily discounted, traded off, and deprioritized against the vivid, present claims of people who very much can come forward.

This is not a hypothetical worry. Compare a duty to an identifiable present person — a specific worker whose pension is being cut — to a duty to "whoever occupies the future." The first generates political urgency precisely because someone, right now, can protest, litigate, and vote against it. The second generates a duty that every political actor can acknowledge in principle while deprioritizing in practice, because there is no one currently alive whose standing is damaged by the deprioritizing, and the university-library analogy Ferreira relies on actually demonstrates this weakness rather than solving it: universities routinely underfund long-term maintenance in favor of present-cohort priorities precisely because the future-student role

has no seat on the current budget committee, and no mechanism forces anyone to answer to it directly. If this is the model climate obligation is supposed to run on, it predicts, and arguably licenses, the exact pattern of institutional neglect Ferreira's essay is trying to argue against.

The better response to the non-identity problem, I think, is not to weaken the right into a role-based duty but to relocate the complaint slightly: the wrong is not done to the specific future individuals (who cannot be harmed relative to their own non-existence, as Ferreira correctly shows) but to us, now, in our capacity as the only people who can act as trustees for a future that cannot yet advocate for itself, and who have a duty grounded in *our own agency and foreseeable causal power* rather than in any right possessed by the eventual population. This keeps the demandingness of rights language — it is not a debatable, tradeable role-interest, but a direct duty binding on identifiable present agents who can be told, specifically, that they are the ones failing it — without requiring the incoherent claim that a not-yet-existing, non-identity-affected person has been wronged. Ferreira diagnosed the problem correctly. But the fix needs to strengthen the obligation on the present actor, not soften the entitlement of the future one, or it will be treated, correctly, as optional the moment it competes with anyone who can actually speak up for themselves.

EPISTEMOLOGY

Testimony and the Right to Be Believed

We often speak as though a sincere speaker has a right to be believed. This essay argues the right, if it exists, is much narrower and more conditional than everyday moral talk assumes.

Simone Aubert Lycée Chateaudun, Lyon

When someone tells you something sincerely, and you doubt them without specific evidence, they will often say something like "are you calling me a liar?" — treating disbelief itself as an accusation, a kind of insult independent of whether the doubt turns out to be justified. This reaction assumes something like a default right to be believed: that sincere assertion creates, by itself, a presumption in the speaker's favor that a hearer needs positive reason to override. This essay examines that assumption and argues that it holds only within a much narrower range of cases than everyday moral intuition extends it to — specifically, cases where the hearer has independent reason to think the speaker's reliability on this topic is unimpaired, which is not a default condition but an achievement that has to be established.

The strongest argument for a default right to be believed is a kind of testimonial version of the golden rule: since we all depend constantly on others believing us, and since demanding independent verification for every assertion would make ordinary communication and cooperation impossible, we have a shared interest in a background norm that grants sincere speakers a presumption of credibility, defeasible by specific counter-evidence but not requiring

positive evidence to establish in the first place. This is broadly right as a description of how testimonial exchange has to work for a functioning society, and it explains why demanding a receipt for every casual claim a friend makes would be not just impractical but genuinely insulting, treating an ordinary social interaction as an interrogation.

But notice what does the actual work in this justification: not the fact of sincerity itself, but the background assumption that the speaker's testimony on this kind of topic is, statistically and structurally, reliable enough that extending the presumption serves everyone's interests, including the hearer's. This assumption is not equally true across all domains and all speakers, and where it breaks down, the golden-rule justification for the presumption breaks down with it. A patient describing their own pain to a doctor is, on the relevant topic — their own first-person experience — in about as reliable a position as any witness can be, which is why medical ethics has moved, correctly, toward crediting patient reports of pain far more than clinical practice once did. But an eyewitness describing a stranger's face under stress, at night, briefly glimpsed, is testifying about a domain — unfamiliar-face recognition under poor conditions — where decades of research show even sincere, confident testimony is wrong at rates that should trouble any legal system that treats confident eyewitness identification as strong default evidence.[1] The sincerity is identical in both cases. The reliability, and therefore the presumption's justification, is not.

This means the right to be believed, where it exists, is topic-relative and reliability-conditioned rather than a general entitlement that attaches to the speech act of sincere assertion as such. Recognizing this is not an insult to speakers; it is a more accurate map of where testimonial trust is actually doing epistemic work versus where it is being extended by social habit into domains that cannot support it. It

also reframes what looks like disrespect in cases where it is not: a juror instructed to weigh eyewitness testimony against its known unreliability for cross-racial identification under stress is not calling the witness a liar, any more than a doctor asking a second opinion on a scan is calling the first radiologist incompetent — both are responding, correctly, to variation in domain reliability that has nothing to do with the sincerity of the person in front of them.

The uncomfortable implication is that some speakers, through no fault of their sincerity, occupy positions — as witnesses to unfamiliar events, as reporters on topics outside their expertise, as narrators of their own past under conditions known to distort memory — where the default presumption simply should not apply as strongly as it does for other speakers on other topics, and treating this as an insult, rather than an accurate epistemic assessment, actually makes the presumption less trustworthy everywhere, by refusing to let it track the one variable — domain reliability — that ever justified extending it in the first place.

NOTES

1. See the extensive literature summarized in the U.S. National Research Council's 2014 report *Identifying the Culprit: Assessing Eyewitness Identification*, which reviews decades of findings on eyewitness reliability under suboptimal viewing conditions.

METAPHYSICS

What Do We Owe a Future Self?

If personal identity over time is less robust than we assume, the obligations we think we owe our future selves need a different foundation.

Yusuf Aydin Anatolian Science High School, Ankara

We routinely make sacrifices for our future selves — saving money we could spend now, studying for an exam years before a career depends on it, exercising for a body we will inhabit decades from now — and we justify these sacrifices with a simple thought: that future self is still *me*, so caring for them is just an extension of ordinary self-interest, no more mysterious than caring about my own welfare tomorrow morning. This essay takes seriously a line of argument, most associated with Derek Parfit, that this simple thought is less secure than it feels, and asks what happens to the obligation to our future selves once the metaphysical guarantee behind it is weakened.

Parfit's challenge to personal identity begins with thought experiments about gradual replacement and psychological continuity — teleporters, brain splits, slow neuron-by-neuron replacement — designed to show that our confidence in a single, all-or-nothing fact of personal identity across time outruns what any theory of identity can actually deliver. What survives these thought experiments, Parfit argued, is not identity as an all-or-nothing further fact, but psychological continuity and connectedness as a matter of degree: the more of my current memories, intentions, personality traits, and psychological dispositions persist into a later person, the more that

later person is connected to me, but this connectedness can come in degrees, and can fade gradually rather than switching off at a determinate moment.

If this is right, the relationship between me-now and me-in-forty-years is not identity in the strong sense that would automatically justify treating their welfare exactly as I treat my own welfare tomorrow. The person who will exist in forty years may share only a fraction of my current memories, almost none of my current daily concerns, possibly different values shaped by decades of experience I cannot predict. Parfit's own conclusion was not that this severs the obligation, but that it changes its foundation: the reason to care for my future self is not that identity guarantees it is *still me* in the deep sense that renders the question of obligation moot, but that a high degree of psychological connectedness, though it comes apart from strict identity, is still a morally relevant relation — similar in kind, though generally stronger in degree, to the relation I have with people I love, whose welfare I also have reason to care about without needing to be numerically identical to them.

This reframing has a concrete implication that most advice about long-term planning ignores: the strength of an obligation to a future self should scale with the degree of psychological connectedness actually expected, not with a flat assumption of identity that treats a self forty years away exactly like a self one day away. This helps explain a pattern that pure self-interest struggles with: people are often far more willing to sacrifice for a near future self (tomorrow's exam) than a distant one (retirement decades off), and the standard explanation — short-term thinking, weakness of will — treats this as a simple failure of rationality. Parfit's framework suggests an alternative reading: the distant future self really is, in the relevant psychological sense, less connected to the present self than the near future self is, so a somewhat weaker motivational pull toward the

distant self is not straightforwardly irrational, even though modern life, with its genuinely long financial and health horizons, requires us to act as though the pull should be stronger than it naturally is.

This does not license abandoning long-term self-care as philosophically unfounded — psychological connectedness, even weakened, is still a real and substantial relation, stronger by far than my relation to a stranger, and the practical case for prudent long-term planning survives Parfit's argument largely intact. What changes is the justification: not an unshakeable metaphysical identity that makes the question of "why should I care about them" a category error, but a real, gradient relation that gives good reason to care a great deal, while also explaining, honestly, why that care does not always feel as immediate or as motivationally forceful as caring about the self that will wake up tomorrow morning.

AESTHETICS

The Aesthetics of Apology

A well-crafted apology and a sincere one are not always the same thing, and the gap between them is a genuinely aesthetic problem, not just an ethical one.

Camille Laurent Académie du Nord, Lille

Public apologies are routinely criticized on two separate grounds that get run together in ordinary commentary: that they are insincere, and that they are badly made — clumsy, evasive, tonally wrong, using the passive voice to obscure agency ("mistakes were made"). This essay argues that these are genuinely distinct failures, one ethical and one aesthetic, and that conflating them causes us to misjudge apologies in both directions: crediting sincere apologies that are aesthetically botched, and discounting insincere apologies that happen to be well-crafted, when in fact craft and sincerity vary independently and both matter, for different reasons, to whether an apology succeeds.

Start with why craft should matter at all, if sincerity is the real ethical core of an apology. The answer is that an apology is not merely an internal state (regret, remorse) that happens to get expressed; it is a communicative act whose entire function is to be received, understood, and to produce a specific effect in another person — a sense of being acknowledged, of the harm being accurately named, of the wrongdoer having genuinely understood what was done and to whom. A sincerely-felt but badly-constructed apology can fail at this communicative task even while succeeding at the internal, ethical one: a person who deeply regrets what they did but expresses it with

vague language, wrong emphasis, or an accidental focus on their own suffering rather than the victim's ("this has been a really hard time for me too") may produce, in the person they are apologizing to, exactly the reaction — that they have not been heard — that a well-crafted insincere apology, ironically, avoids.

This is where the aesthetic dimension becomes visible as a genuinely separate axis. Apologies have a form, and the form has conventions that a speaker can execute well or badly independent of what they actually feel: naming the specific harm rather than a vague category of wrongdoing, using active rather than passive constructions to keep the apologizer as the grammatical subject of the wrongdoing, sequencing acknowledgment before any explanation (since explanation offered first reads as excuse), and calibrating length and register to the severity of the harm, since an apology too brief for a serious wrong reads as dismissive regardless of the feeling behind it, and one too elaborate for a minor one reads as performative. These are craft judgments, of exactly the kind we make about a poem or a speech, and a wrongdoer with real remorse but poor command of the form can produce an apology that is aesthetically clumsy in ways that actively undermine the sincere content it is trying to carry — the same way a sincere love letter, badly written, can fail to move its recipient despite the real feeling behind it.

The uncomfortable flip side is that craft can also successfully simulate the aesthetic signature of sincerity without the underlying feeling — a well-coached public apology can hit every formal convention (specific harm named, active voice, acknowledgment before explanation) while being, ethically, empty. This is why professionally drafted corporate and political apologies often feel simultaneously well-formed and hollow: the form has been executed correctly by people skilled at exactly this craft, which is precisely why the hollowness registers as a distinct, nameable feeling rather than

simply as a badly made apology — we are responding to a mismatch between competent form and absent content, which is a different aesthetic experience than responding to incompetent form with genuine content behind it.

Recognizing apology as having this two-part structure — an ethical core (genuine remorse, accurate understanding of the harm) and an aesthetic form (the craft of expressing it in a way that lands) — lets us make more precise judgments than "that felt fake" or "that was a bad apology," which currently get used interchangeably for two different failures. It also suggests something practically useful: that teaching people to apologize well, as a matter of craft — much as we might teach the craft of any other form of consequential speech — is not teaching them to fake sincerity, any more than teaching poetic form teaches insincerity in poetry. It is giving genuine remorse, which often arrives inarticulate and badly organized, the formal vehicle it needs to actually be received as what it is.

PHILOSOPHY OF SCIENCE

Can a Model Owe Its Data Anything?

Statistical models are usually described as owing accuracy to their predictions. This essay asks whether they can also be said to owe something to the people whose data trained them.

Divya Menon Cotton Tree International School, Chennai

A predictive model — a credit-scoring algorithm, a medical risk calculator, a recommendation system — is ordinarily evaluated along a single axis: does it produce accurate predictions. This essay asks about a different, less commonly framed relationship: whether a model can be said to owe anything to the people whose historical data it was trained on, independent of how well it predicts for future users, and whether "owe" is even coherent language to apply here, given that a model is not an agent in any ordinary sense.

The obvious objection arrives immediately: models are not moral agents, cannot have obligations, and any talk of what a model "owes" is a category mistake that should be redirected to the actual agents involved — the people who built, deployed, and profited from the model. This is correct as far as it goes, and nothing in this essay disputes that ultimate responsibility sits with human institutions, not the artifact itself. But redirecting the question entirely to the builders misses something the model's specific structure makes newly visible: a predictive model is, in a precise sense, a compressed record of the people in its training data, and the question of what is owed to them is not fully answered by asking whether the builders acted responsibly in a general sense — it requires asking a specific question

about the relationship between the data subjects and the artifact built from their information, which is different in kind from an ordinary product-liability question.

Consider a concrete case: a medical risk model trained on historical patient records systematically under-predicts a disease in a demographic that was, historically, under-diagnosed and under-recorded for that disease — not because the model or its builders did anything additionally wrong, but because the model faithfully learned the pattern present in its training data, which encoded decades of prior clinical neglect. The model is, in a real sense, propagating a harm done to a past population of patients into predictions that will affect a present population, some of whom may be the same people or their direct descendants in similar social positions. The harm here has an unusual temporal structure: it originates in how the training subjects were treated, is encoded silently into the model's learned parameters, and re-emerges as a harm to current users who may have had no part in generating the biased data at all. This is not simply "the builders were careless"; it is a distinctive kind of harm-transmission that a purely accuracy-focused evaluation of the model, and a purely blame-focused evaluation of its builders, both fail to fully capture.

I want to suggest that the right way to talk about this is not that the model owes something — models are not obligation-bearing agents, and the objection above is correct on this point — but that the training data subjects are owed something by whoever deploys a model built from records of them, and that this obligation is distinct from, and in addition to, the ordinary obligation to build an accurate product. It is closer to a kind of downstream stewardship duty: having taken a compressed record of real people's treatment and turned it into an artifact that will make decisions about other real people, the deployer owes the original data subjects an active effort to

identify and correct for exactly the kind of historically-encoded harm described above, not merely an effort to maximize aggregate predictive accuracy, which can be entirely indifferent to whether the errors are evenly distributed or concentrated on the population the training data already failed once.

This reframing matters because "make the model more accurate" and "honor what is owed to the people in the training data" are not the same instruction, and can conflict: a model can become more accurate in aggregate while becoming more unfair to a subpopulation whose historical mistreatment is exactly what the aggregate accuracy metric is rewarding it for learning. Naming this as a distinct obligation — owed not by the model, which cannot owe anything, but by its builders, to the people whose lives were compressed into its training data — gives us a more precise target than either "be accurate" or the vague, easily-deflected charge that "the algorithm is biased," which too often lets the actual bearers of the obligation, the humans who chose the data and shipped the model, disappear behind a noun that cannot, strictly, owe anyone anything at all.

EPISTEMOLOGY

On Believing What You Cannot Prove

Two editors argue, in alternating voice, over whether justified belief requires evidence at all.

Dorotea Novak & Priya Raman — Joint submission — Gymnázium Jana Nerudy, Prague & Millfield International, Colombo

DOROTEA

Start simply. I believe my late grandmother, whom I never met, would have liked me. I have no evidence for this — no letters, no testimony from anyone who knew us both, nothing. And yet I do not think this belief is irrational, or that I am obligated to suspend it until evidence turns up. If you tell me this is simply an unjustified belief I happen to enjoy holding, I want to know what work "justified" is doing in your account that my belief fails to meet.

PRIYA

It fails to meet the most basic requirement there is: that a belief be responsive to reasons that make its content more likely to be true. You have offered nothing that makes it more likely your grandmother would have liked you, specifically, as opposed to any of the other things she might have felt about a person she never met. What you have is a hope wearing the grammar of a belief. That is not a small failure — it is the entire distinction epistemology exists to protect.

DOROTEA

But notice what your standard would also throw out. I believe my closest friend loves me. My evidence is extensive — years of behavior — but strictly, no evidence of that kind ever entails an inner state; it is

compatible, in the strict sense, with sophisticated long-term deception. If you demand evidence that makes the belief's truth more likely in some rigorous, quasi-statistical sense, you'll find that most of what holds a life together — trust in people, trust in one's own memory, trust that tomorrow will resemble today — survives on much thinner grounds than you are prepared to admit, once you look closely. My grandmother belief is an extreme case of something your own epistemology depends on constantly.

PRIYA

There is a difference between thin evidence and no evidence, and you are trying to erase it by pointing at cases with thin evidence and calling them equivalent to yours. My belief that my friend loves me is built on thousands of specific, observed actions that would be extremely improbable under the hypothesis that she does not — it is thin only compared to mathematical proof, not thin in any absolute sense. Your belief about your grandmother has zero data points of this kind. There is no action of hers you have ever observed. The two cases are not on a shared spectrum where I am simply pickier than you; they are different in kind, and eliding that difference is the actual move I object to.

DOROTEA

Then let me change the case, because I don't think the zero/nonzero distinction is doing what you want it to do either. Suppose I believe, without specific evidence, that the universe is intelligible — that its regularities can in principle be understood by minds like ours. This belief cannot itself be established by evidence, because every piece of evidence I would cite already presupposes that observation and inference track reality, which is the very thing in question. And yet this belief is not optional for you either; you rely on it every time you reason from evidence at all. So even your own practice depends on at

least one belief that is not itself grounded in evidence, on pain of circularity. Once you grant me that one, I want to know on what principled ground you refuse the others.

PRIYA

That one is different again, and I think you know it: the intelligibility of the universe is not a belief I hold as an isolated item I could be talked out of by counter-evidence; it is closer to a precondition for the entire activity of believing-on-evidence to make sense at all, which is why Wittgenstein called this kind of proposition a hinge — not one more belief among others, resting on the same axis as ordinary claims, but the thing the axis turns on.[1] A hinge proposition earns its special status precisely by being presupposed by every act of evidence-weighing, not by being an ordinary belief that gets a pass. Your grandmother belief presupposes nothing of the kind; the entire practice of reasoning about people's characters proceeds perfectly well without it. It is not a hinge. It is simply unsupported, and dressing it in the same language as universal intelligibility does not make it one.

DOROTEA

Fair — I will not defend the grandmother belief as a hinge, you're right that it isn't one. But notice what you have just conceded: that at least one category of belief is legitimately held without ordinary evidential support, because of the role it plays rather than the data behind it. That is already more than the strict evidentialist you opened with was willing to allow. I think the honest position is not 'all belief requires evidence' but 'different beliefs are answerable to different standards depending on what kind of belief they are and what role they play in a life' — and once you've granted that hinge propositions get a different standard, you no longer have a principled way to say in advance that no other category could ever qualify. You have to look at each case. Which is all I ever wanted from you about

my grandmother — not that you agree it's justified, but that you stop assuming the evidentialist standard was ever going to be the only one on the table.

PRIYA

I'll grant that much, though I don't think it rescues the grandmother belief specifically — it just relocates the argument to a harder, more interesting question: which beliefs actually qualify for a non-evidential standard, and which are simply comfortable. That question doesn't get answered by pointing out that the category exists. It gets answered case by case, and I still think yours loses. But I concede the category is real, and that concession is bigger than the one I walked in willing to make.

NOTES

1. Ludwig Wittgenstein, *On Certainty* (1969, published posthumously), §341–343 on "hinge" propositions that stand fast for us and are not themselves subject to ordinary doubt.

MASTHEAD

What We Owe Each Other — Issue 06

EDITOR-IN-CHIEF

Aanya Bhatt

Cathedral and John Connon School, Mumbai

DEPUTY EDITOR, ETHICS

Adaeze Nwosu

Ridgeview College, Accra

DEPUTY EDITOR, EPISTEMOLOGY

Dorotea Novak

Gymnázium Jana Nerudy, Prague

FACULTY ADVISOR, POLITICAL PHILOSOPHY

Prof. Elena Marsh

Princeton University

REVIEWING EDITOR, EPISTEMOLOGY

Dr. Amit Sharma

University of Oxford

FACULTY ADVISOR, ETHICS

Prof. Naomi Katz

University of Toronto

WORKSHOP LEAD, PHILOSOPHY OF MIND

Dr. Julian Voss

National University of Singapore