

Sourcehood Without a Source: The Arity Gap and the Free Will Debate

Zain Dana Harper

July 2026

Abstract

For the most part, the free will debate has been run as a search for an origin. It looks for a point at which the agent, and not the agent's own past, is responsible for what happens next. Libertarians place that point in an act of will that is not determined by anything before it. Hard incompatibilists, following Galen Strawson, argue that there is no such point, because nothing can be *causa sui*, that is, the cause of its own existence. Compatibilists move the point inside the agent's psychology and hope that the move does enough work. A libertarian holds that free will is real and needs undetermined choices. A hard incompatibilist denies that free will and real desert exist at all. For a compatibilist, free will is real and fits with determinism. One assumption unites all three parties. Call it the aseity premise. *Aseity* means self-existence: being the source of oneself, owing one's being to nothing outside oneself. The aseity premise says that free will, if it exists, must be a kind of self-origination. This paper argues that the aseity premise is a definition slipped in quietly. It was never a finding about what free will requires, and it makes the debate impossible to solve from the start. Once the premise is dropped, the question gets re-framed. *Sourcehood*, being the genuine author of one's own acts, is either impossible or *conferred*. Conferred sourcehood is inhabited, and it is addressed. In plain words: it is a standpoint given to you and lived in. You occupy it. You do not squeeze it out of yourself. The paper then names what it takes to be the single most load-bearing move in the whole debate: the **arity gap**. *Arity* is the number of argument-places a relation has. A one-place relation needs one thing to fill it. A two-place relation has room for two. P.F. Strawson's account of the reactive attitudes, read with care, shows that holding someone responsible is a two-place relation. It runs between an addresser and an addressee, each standing at a second-personal pole, that is, each in the position of an "I" speaking to a "you." The hard incompatibilist wants to reduce that relation to a one-way program of behavior-shaping, a first-person acting on a third-person. This paper argues that the reduction is blocked, and the block is a matter of logical structure. A one-place vocabulary that fixes a single target cannot fix the target of a two-place relation. The elimination of phlogiston succeeds for combustion because oxidation is a one-place property of the burning stuff. The same move cannot succeed for holding someone responsible. The second person is the second argument-place of the relation itself, and no predicate added on to a property can stand in for it. Determinism is granted in full throughout. An earlier appeal to chaos theory is explicitly withdrawn. The conclusion is that conferred sourcehood is stronger than libertarianism, because it never took on the debt that libertarianism cannot pay.

Keywords: free will, reactive attitudes, P.F. Strawson, arity gap, second-personal standpoint, Galen Strawson Basic Argument, conferred sourcehood, compatibilism, hard incompatibilism, Pereboom, Sapolsky

In plain terms

Think about someone who borrows money from you and does not pay it back.

You can respond to this in two very different ways.

In the first way, you are hurt and you blame the person. You say, “You owe me. You wronged me.” You are speaking to them. You expect that they can hear you, answer you, apologize, or make it right. This response has two people in it. There is you, and there is a “you” that you are speaking to. Both of you can act on the other.

In the second way, you stop treating the person as someone to talk to. You decide they are simply unreliable. You never lend to them again, and you take steps to protect yourself. Here there is only one person doing something. You are working *on* the other person the way you would work on a broken machine, adjusting your exposure to it. There is no moment in which the other person answers you, because their answer is not the point. The point is only to manage what they will do next.

Now bring in the free will debate. One side, the hard determinist side, says the first response is just a fancy costume worn over the second. Underneath, they say, all you are doing is shaping behavior. The hurt and the blame are decoration. Strip the decoration off, keep the useful management, and you lose nothing. You even gain something, because the decoration is what lets people be cruel to each other.

This paper answers that the first response cannot be fully described using only the words of the second. The reason is structural. The first response has a second person built into it, a source who can answer back. The second response has no place for such a person. You can see the gap most clearly in forgiveness. If you forgive the debt, you release a claim, and the other person can even refuse your forgiveness and insist the matter is not closed. If you simply write the person off as a bad bet, there is no one there to refuse anything. From the outside these two can look identical. What differs is whether there is a second person in the picture at all.

There is a second, older knot to untie first. Many people think you only truly deserve blame if you made yourself into the sort of person who would skip a debt, and made yourself make yourself, and so on all the way back. That demand is impossible. Nothing makes itself out of nothing. So the hard determinist concludes that no one ever truly deserves anything. This paper agrees that the impossible demand is impossible. Then it says the demand was the wrong test in the first place. You can owe someone an answer simply by being a member of a community of people who make claims on each other and can respond to those claims. Your standing to be addressed is given to you by the relationship. You do not have to manufacture it from nothing.

So the paper grants determinism in full. It grants that you are not the ultimate, self-made source of anything. Then it shows that this grant, large as it is, does not remove the two-person structure of holding responsible, and does not remove a real kind of authorship that is given and lived in. That authorship is conferred on you, and you do not produce it out of yourself. That given, lived-in authorship is what the paper calls conferred sourcehood.

1. The Aseity Premise and Why It Should Not Have Been There

Plainly: every side of the debate quietly demands that a free agent be the source of its own source, and that demand was never earned.

Every major position in the free will debate carries a demand that almost no one states in full. The demand is that the free agent be, in some serious sense, the *source of the source* of its own acts. Libertarians state

this demand openly and try to meet it. They appeal to agent causation, to undetermined self-forming actions (Kane 1996), and to some account of how an agent's undetermined choice still counts as the *agent's* choice and does not collapse into mere noise. Hard incompatibilists, led by Galen Strawson's Basic Argument (1994, *Philosophical Studies*; high confidence on the text and the thesis, meaning the author is confident both about what the source says and about its main claim), state the demand openly and then show that nothing can meet it. They conclude that no one is ever truly morally responsible. Compatibilists try to remove the pressure of the demand. They argue that "could have done otherwise," properly understood, does not require self-origination. Yet most compatibilist accounts, at their deepest point, put back something that behaves like a self-standing authoritative self, a self that stops the backward chain of endorsement. The aseity premise runs under the whole debate, and every position rests on it.

Here is the premise with its clothing removed. An act is free, and the agent is responsible for it in the strong sense, only if the agent is in the end the origin of the principles, values, and character from which the act flows, and is responsible for being that way, and so on, all the way down. This is Galen Strawson's own way of setting it out. It is a regress, a chain of "and before that" steps, that ends only in something impossible: *causa sui*, the self-caused self. Nietzsche rightly called this the best self-contradiction that has been conceived so far (*Beyond Good and Evil* §21; Galen Strawson cites this passage in building the Basic Argument; high confidence on where it comes from). Nothing is *causa sui*. The chain of causes passes through the agent and does not stop there, and the agent did not write the starting conditions.

Here is what this paper proposes. All of that is correct, and it settles exactly the wrong question. The argument from premises 1 through 5 of the Basic Argument is sound. It follows that no one is the *ultimate, self-originating* source of their acts. The Basic Argument does not reach past that. From those premises alone, it settles nothing about whether there is *sourcehood* of any kind, whether there is responsibility, or whether there is free will worth wanting. The argument can reach that further conclusion only by carrying the aseity premise forward into the conclusion as an unargued definition. That last step is valid only if responsibility and sourcehood *mean* ultimate self-origination. They need not mean that. The aseity demand is packed into the words "true" and "ultimate" as Galen Strawson uses them, and that packing does all the work in the move from the sound premises to the conclusion that nothing matters. Strip the demand out, and the premises establish exactly what they establish, that the will is not *causa sui*, without establishing anything about whether conferred, inhabited, second-personally-constituted sourcehood is real.

This is not the usual deflationary compatibilist move, the move that simply lowers the bar for "free will" and then declares the game won. The claim here is structural. It does not redefine free will. The claim is that *aseity was never the right criterion*. The problem with aseity runs deeper than its going unmet. It is the wrong test for responsibility. The grounds for that structural claim come from cross-traditional evidence and from the arity argument developed in Section 4. For now, notice only this: the aseity demand was always optional. The concept of responsibility does not contain it by definition. What deliberation feels like from the inside does not force it on us. Several philosophical traditions that never granted the will aseity show the same thing. In those traditions one can hold agents fully responsible and treat agency as fully real, and sourcehood there is *conferred*, with no requirement that it be *originated*.

2. Determinism, Granted in Full, Is Orthogonal

Plainly: determinism shows you are not the ultimate source, and that is granted here. It leaves the separate question of whether the act ran through you untouched.

This paper grants the consequence argument (van Inwagen 1983, *An Essay on Free Will*; high confidence).

The consequence argument runs like this. If determinism is true, our acts are the consequences of the laws of nature and the remote past. We had no choice about the laws, and we had no choice about the past. So we have no choice about our acts. The argument is sound given its premises, and those premises are not contested here. Determinism is taken as given throughout. (*Determinism* is the view that the state of the world at any moment, together with the laws of nature, fixes every later state.)

The claim here is not that determinism is false. Determinism is true, and the consequence argument, read correctly, is itself a no-aseity argument. It shows that the agent is not the ultimate source, and that the past and the laws write the future. What the consequence argument does not show, and cannot show, is that the only possible sourcehood is ultimate self-origination. Here two things come apart. The argument removes *leeway*, the set of could-have-done-otherwise alternatives. Sourcehood it leaves untouched. Sourcehood is the question of whether the act that actually happened flows through the agent as a genuine source, or bypasses the agent the way an efficient cause passes through inert matter. Those are different questions. Frankfurt's cases (1969, *Journal of Philosophy*; high confidence), whatever their final success against the Principle of Alternate Possibilities (the principle that a person is responsible for an act only if they could have done otherwise), got the distinction right. For responsibility, what matters is that the act in the actual sequence ran through the agent's own agency. Whether the agent could have done otherwise is a separate matter. Sourcehood is a question about the actual sequence, and the alternatives are beside it.

One further matter deserves an explicit burial. This is the appeal to chaos theory and indeterminism as any kind of support for free will. The body of work from which this paper grows once used the Lorenz attractor and sensitive dependence on initial conditions to give the orthogonality thesis "mathematical teeth." That claim equivocates. It slides between epistemic unpredictability, which is about what observers can foresee, and constitutive irreducibility, which is about what a system is made of. The claim is withdrawn here. A trajectory that is both determined and irreducible in a chaotic system is not, by being irreducible, a *source*. That a hurricane's path cannot be predicted from its micro-scale starting conditions is a fact about observers and their computing limits. It is not the fact that the hurricane authored anything. Unpredictability is not authorship. The chaos argument is withdrawn in full.

What the chaos mathematics *does* legitimately establish is narrower and more modest. It shows that one inference is invalid: the inference from "the act is determined" to "the act dissolves into its substrate and can be fully described at the micro-level." That is the inference from determinism to *eliminative reduction*, the claim that the higher-level description can be thrown away in favor of the lowest-level one. Deterministic systems can be computationally irreducible (Wolfram; moderate confidence on the attribution, high confidence on the phenomenon). *Computationally irreducible* means there is no shortcut: the only way to find out what the system does is to run it step by step. When that holds, the macro-level description cannot be eliminated in favor of the micro-level description, even under full determinism. This takes away the hard determinist's *reductive* premise. It does not give any positive content to sourcehood. The positive content comes from elsewhere, from the argument of Sections 4 through 6. The chaos result clears the space that the reductionist claimed was already filled. It does not fill that space with sourcehood. The paper keeps that exact distinction throughout.

3. Three Traditions That Never Owed the Debt

Plainly: three traditions with no shared metaphysics each held people fully responsible without ever demanding that a person be self-made, which shows the demand was optional.

Conferred sourcehood is not an escape hatch that a philosopher invented to dodge a dilemma made of West-

ern assumptions. The structure of ownership-without-origination has been reached on its own in several traditions. Each faced the problem of freedom under total determination without ever granting the aseity premise.

Ash'arī kasb. Classical Ash'arī *kalām*, which is Islamic rational theology, worked out a doctrine from al-Ash'arī (died 935) through al-Bāqillānī, al-Juwaynī, and al-Ghazālī (high confidence on the existence and core of the doctrine; moderate confidence on the finer scholastic disputes over the mechanism of *kasb*, which were argued within the school). The doctrine holds that God creates (*khalq*) every act, and the human being *acquires* (*iktisāb*) it. God is the only true cause. The human is not the ontological originator of the act and does not bring the act into being. Yet the act is *the agent's* for the purpose of responsibility, and it is so precisely because of the acquisition, the taking-up of the act from the standpoint of the person whose act it is. This is conferred sourcehood in its strict form: total causal dependence together with genuine ownership at the locus, achieved with no self-origination on the agent's part. The Mu'tazila, a rival school, pressed for real causal power in the human. The *falāsifa*, the philosophers in the Greek tradition, charged that *kasb* is a distinction without a difference. The Ash'arī position refuses both horns. It keeps ownership without origination, and it was reached seven centuries before the words “compatibilism” and “incompatibilism” were coined.

Yoruba ori. In Yoruba thought (Wande Abimbola; Segun Gbadegesin, *African Philosophy*, 1991; Barry Hallen; moderate confidence on the philosophical interpretation, which varies by source; high confidence on the broad structure), *ori*, the inner head, the bearer of a person's destiny, is received before birth and then lived out. The destiny is conferred. The lived life is its habitation. *Ìwà*, which means character, along with effort, stays real and answerable within the destiny that was conferred. This structure was reached fully outside the Abrahamic frame and with no reference to the terms of the Western debate. The agreement between the two is not explained by a shared inheritance.

The tied camel. A saying of the Prophet, transmitted in al-Tirmidhī (*i'qil ba'irak wa tawakkal*; moderate confidence on the exact chain and grading; high confidence that the saying is canonical and widely transmitted in this sense), means “tie your camel and then rely on God.” The saying does not counsel passivity, as if a decreed world meant one should do nothing. It counsels full action: act with your own hands, to the very end, as if the outcome were entirely up to you, and within that very acting the decree is at work. The acting does not compete with the decree. It is *how* one inhabits a conferred standpoint inside a determined order.

Three traditions, no shared metaphysics, each facing freedom under total determination, each arriving on its own at ownership-without-origination. None of them took the libertarian demand, and none dissolved responsibility the way the hard incompatibilist does. The peculiar feature of the Western debate is that it built the aseity demand into the *definition* of free will, and then discovered, with Galen Strawson and Robert Sapolsky, that nothing can meet it. Faced with that, it concluded “no free will” where it could have concluded “wrong definition.” The witnesses show that the definition was always optional. Conferred sourcehood is stronger than libertarianism, because it never took on the debt that libertarianism cannot pay. It never required the agent to be *causa sui*.

4. The Arity Gap: The Core Argument

Plainly: holding someone responsible has two people in it, an addresser and an addressee, and no one-person description of behavior-shaping can capture that second person.

4.1 P.F. Strawson's Move, Read Precisely

Plainly: P.F. Strawson showed that holding responsible is an address between two people. It is not a conclusion drawn from a theory about causes.

P.F. Strawson, in “Freedom and Resentment” (1962, *Proceedings of the British Academy*; high confidence on the text and the thesis), made a move whose boldness is still underrated in most readings of it. The traditional question was: does determinism make the reactive attitudes unjustified? He declined to answer that question on its own terms. His argument was that the reactive attitudes are not conclusions drawn from a prior theory about causation. The *reactive attitudes* are the ordinary responses we have to the quality of another person's will toward us: resentment, gratitude, indignation, hurt feelings, forgiveness, and the rest. They are built into taking part in shared human life. There is also an *objective attitude*, which is viewing another person as an object to be managed, treated, cured, or trained. In the objective attitude one handles the person and does not address them. We can take the objective attitude now and then: toward very young children, toward people in severe psychosis, or in moments of deliberate detachment. But the commitment to interpersonal relationship that grounds the reactive attitudes is not the sort of thing a metaphysical thesis can pry loose.

The standard reading takes this as a naturalized, nearly quietist compatibilism. On that reading, responsibility just *is* the practice of holding responsible, and the practice cannot be revised by metaphysics. (*Quietist* here means the view that the metaphysical question should be set aside without being answered.) This reading is not exactly wrong. Still, it misses what makes the move structurally interesting. It is also the reading that makes Strawson look easy to refute to a hard incompatibilist, because the quietist version seems to change the subject without answering the question. The precise reading, worked out with care by Stephen Darwall (*The Second-Person Standpoint*, 2006; high confidence), is this. The reactive attitudes encode a *second-personal* relation. To hold you responsible is to address a demand to you, from inside a relationship of mutual accountability, and in doing so to presuppose that you have the standing to address me in turn. The “you” in “I hold *you* responsible” does not name an object in the third person. It is the second party of an address. This does not repeat the quietist point that practices resist revision. It is a claim about the logical form of holding responsible.

4.2 The Hard Incompatibilist's Best Move

Plainly: the hard determinist says blame is just a costume over behavior-shaping, so drop the costume, keep the shaping, and you lose nothing and gain kindness.

Robert Sapolsky (*Determined: A Science of Life Without Free Will*, 2023; high confidence on the book's thesis) and Derk Pereboom (*Living Without Free Will*, 2001; *Free Will, Agency, and Meaning in Life*, 2014; high confidence on both) argue as follows. Once you trace the full causal history of any act, through neurons, hormones, development, culture, and evolution, there is no room left for desert, and so no warrant for the backward-looking reactive attitudes. (*Desert* is the idea that a person has earned praise or blame and so deserves it. A *backward-looking* attitude responds to what was done. A *forward-looking* one aims at what happens next.) Their positive proposal is a forward-looking package. Keep the *protective* functions: quarantine of dangerous people on the public-health model, deterrence, moral formation, and reconciliation. Discard basic desert. On this picture, “I hold you responsible” is a folk-psychological wrapper around a behavior-shaping signal. The resentment is the emotional costume on the signal. Take off the costume, and nothing of normative value is lost. Something is even gained, because the costume is what licenses cruelty.

This is a serious proposal, and it deserves a serious reply. The reply here does not claim that the practice is sacred, and it does not claim that metaphysics cannot touch the question. It does not rest on the folk psychology being obviously correct. The reply is structural. **The reduction fails because it equivocates on**

arity.

4.3 The Arity Argument, Stated Precisely

Plainly: behavior-shaping is a one-way, one-person operation. Holding responsible registers the other as someone who can answer, and that second place cannot be added back with a one-place description.

A forward-looking behavior-shaping relation has, in its logical form, a **first-person to third-person** shape. An agent, whether me, the state, or an institution, acts *upon* a system, namely you, modeled as a bundle of changeable dispositions, in order to alter its future outputs. The relation is fully specified without any address. It has no moment in which I treat you as a *source who could answer me*, who has the standing to say “you have wronged me” back to me, and whose acceptance or refusal of my demand is itself the point. In behavior-shaping your answer is only a lever for the next adjustment.

The reactive attitudes are not emotional decorations on a control signal. Each one is a *registration of the other as an addressable source*. When I resent you, I am not merely predicting your future behavior and steering it. I may resent you exactly when steering is hopeless and prediction is certain. I am holding open a relation in which your *will toward me* is the object of my response. Forgiveness makes this plain. To forgive is not to end a deterrence program. Deterrence and forgiveness come apart. You can deter someone you have not forgiven, and you can forgive someone while still deterring them. To forgive is to release a *claim* that presupposed you were a source against whom a claim could lie in the first place.

So the hard incompatibilist’s proposal to replace the reactive attitudes with the forward-looking package does not reduce one thing to another. It replaces a **two-place relation**, an addresser and an addressee, each holding a second-personal pole, with a **one-place predicate**, namely, this agent has a disposition-profile that these interventions can modify. (A *predicate* is a property attributed to a single thing, such as “is flammable.”) The second person does not sit in the sentence as a predicate added on to a third-person property. It is the *second argument-place of the relation itself*. Eliminate it, and you have not removed a gloss from something more basic. What you have removed is one of the two relata, the two things the relation holds between, and you are now describing something with a different structure.

This is the arity gap: **a one-place, target-fixing vocabulary cannot fix the target of a two-place relation.** (To *target-fix* a description is to pin down exactly what it is about, without remainder.)

4.4 The Phlogiston Objection and Its Failure

Plainly: eliminating phlogiston worked because combustion has a one-thing replacement description. Holding responsible has no such replacement, because its target includes the other person as someone who can answer.

The most searching version of the hard incompatibilist objection is the phlogiston analogy. It runs like this. I am not simply swapping the reactive attitudes for something else. I am identifying what the reactive attitudes were *tracking*, which is behavior-shaping signals, and I am eliminating only the desert-laden wrapper, keeping the function intact. This is exactly how the elimination of phlogiston worked. (*Phlogiston* was a substance once thought to be released in burning. The theory was replaced by the account of oxidation.) The elimination identified what phlogiston-talk was tracking, namely combustion, and replaced the theory, keeping the phenomenon. I am doing the same. Your “second-personal remainder” is just the name you give to your refusal to finish the reduction.

This objection is the best one in the literature. It must be answered on its own terms. Sidestepping will not do.

The answer is that the phlogiston analogy contains its own refutation. Phlogiston elimination succeeds because there is a *target-fixing description* of combustion, namely oxidation, that is available in the third person, that captures everything phlogiston-talk was tracking, and against which the old theoretical gloss can be measured and found empty. Oxidation is a one-place property of the burning stuff. The proposed elimination of the reactive attitudes owes the same thing: a third-personal, forward-looking, target-fixing description of what holding responsible was tracking, against which the second-personal “gloss” can be measured and found to add nothing. **The hard incompatibilist does not have this description, and the reason is structural. It is not a gap in current science.**

Several phenomena reduce cleanly: map-coloring, combustion, the apparent motion of the geocentric sky. Each of these is a one-place phenomenon, or one that is already third-personal, and each has a target-fixing description available without any reference to the second person. The second-personal relation is the one case the move cannot reach. Its target, the addressee *as an addressee*, the other as a source who can answer, is a constitutive part of what is being tracked. You cannot strip it off as a theoretical commitment and keep the phenomenon.

The hard incompatibilist has a reduction-plus-elimination *schema*, a general recipe. He does not have an *instance* of it for the two-place case. And the two-place character of holding responsible is not incidental. It belongs to what holding responsible *is*.

4.5 The Solipsism Test: Proof of Floating Free

Plainly: imagine a universe with exactly one mind. The need for an inhabited standpoint is fully met there, yet holding responsible is impossible, which proves the second person is something extra.

The claim is that the second-personal remainder, call it (R), is something over and above what any one-place account of agency can supply. (R) does not merely restate the need for an inhabited standpoint. This claim needs a precise demonstration. An assertion will not settle it. Here is the test.

Consider a solipsistic universe that contains exactly one inhabited standpoint. (*Solipsistic* here means containing only a single mind.) In this universe, the necessity-of-locus claim is met in full. There is a standpoint, agency happens there, and the inhabited locus is the only place it could occur. But (R), the second-personal remainder, is *false* in this universe, because there is no one to address. There is no second pole, and no addressee whose uptake, whose taking-in of the address, is a success condition for anything. Forgiveness, resentment, and the whole register of second-personal relations have no objects and cannot occur.

A proposition that can be *false* in a universe where the necessity-of-locus claim is *true* is not entailed by that claim. It says something strictly extra. The arity gap, the gap between the one-place necessity of a standpoint and the two-place structure of second-personal address, is a structural gap. It is not a verbal one. (R) floats free of the residue.

4.6 Answerability Without Basic Desert

Plainly: proving the relation has two places does not yet prove the addressee deserves blame. This part shows the standing to be addressed comes from the community of address. It does not come from self-origination, so the Basic Argument does not reach it.

The arity argument, as stated through Section 4.5, establishes that holding responsible is an irreducibly two-place relation, and that no one-place, forward-looking vocabulary can target-fix its addressee-pole. What the argument has not yet established, and what Galen Strawson’s Basic Argument directly challenges, is

whether either pole of that two-place relation carries *basic desert*. Basic desert is the warranted, backward-looking attribution of blame or credit, grounded in the agent's own self-authored character. The gap is precise. Proving that the relation has two places is not yet proving that the addressee *deserves*, in the regress-vulnerable sense, the demand directed at them. Until that gap is closed, the Basic Argument still bites. Even a fully second-personal address could be directed at someone who, by the logic of premises 1 through 5, cannot be ultimately responsible for being the kind of person who merits it. This subsection closes the gap. More precisely, it reroutes around the gap by separating the normative work the arity argument does from the normative work it has, so far, been silently borrowing.

The partition. The philosophical literature on moral responsibility has, in recent decades, produced a taxonomy that the arity argument can join, with no obligation to defeat it. Gary Watson, in "Two Faces of Responsibility" (1996), distinguishes *attributability* from *accountability*. Attributability is the question of whether an act expresses the agent's own evaluative commitments and is the agent's own act. Accountability is the question of whether the agent is properly subject to demands, claims, and reactive responses. These come apart. Attributability settles whose act it is, and accountability settles who stands inside the practice of mutual address. David Shoemaker (2011) refines this into a triad: *attributability*, *answerability*, and *accountability*. Here *answerability* concerns the agent's standing to be called to account, to be asked for reasons, and to be entered into the exchange of justification. Accountability concerns full reactive engagement, including resentment and indignation. The triad matters here because answerability, as Shoemaker uses it, does not obviously require the agent to be the ultimate self-originating source of the character from which the act flowed. It requires only that the agent be the kind of thing that can be addressed with a demand for reasons and can respond, so that the second-personal address has the right kind of uptake available to it.

Stephen Darwall's account of the *second-personal standpoint* (*The Second-Person Standpoint*, 2006) supplies the underlying mechanism. Second-personal authority, the standing to address a demand, is not derived from the addressee's metaphysical self-origination. The address-community constitutes it. I have the authority to demand of you because we stand in a relation of mutual accountability, each recognizing in the other a fellow member of the community of reasons-givers, each entitled to make claims the other has the standing to answer. The authority is *relational and conferred*. It rests on no one's aseity. This is the crucial asymmetry. Galen Strawson's Basic Argument (1994) targets desert grounded in *how one is*, traced back to *how one became that way*, and finally to a self-caused self. But Darwall's second-personal authority does not inherit its normative standing from that chain at all. Its standing is earned in the address-community, that is, in the mutual recognition of second-personal standing by participants who can take up demands. It is not borrowed from the agent's self-authorship of their own character.

The demand-to-blame inference. The argument can now take the form of a positive inference, going beyond the partition. When I address a second-personal demand to you, such as "you wronged me" or "you owe me an account," the normative standing of that demand does not trace through your causal history to an impossible self-caused origin. It traces through your membership in the community of addressable sources. The demand is warranted because you are the kind of thing that can receive and respond to demands: a being with uptake-capacity, with responsiveness to reasons, and with standing inside the mutual-accountability community that P.F. Strawson (1962) described. Your originating yourself has nothing to do with it. The Basic Argument's regress applies to *desert grounded in self-origination*: to be truly responsible for what one does, one must be truly responsible for how one is, and so on back to the impossible *causa sui*. It does not apply to *answerability grounded in second-personal standing*, because that standing is not derived from a backward-looking chain through self-authorship. The present structure of address-relations constitutes it. There is no regress to run. A prior fact about the agent's self-causation does not justify the standing. What justifies it is a standing feature of the address-community in which both parties take part.

The inference that runs from second-personal address to warranted blame can then be stated. If you stand

in the address-community, if you have uptake-capacity for demands, and if you acted from an evaluative orientation that is yours in Watson's attributability sense, then holding you to account, resenting, demanding, and attributing blame, is warranted by those three conditions taken together, none of which requires self-origination. The Basic Argument's premises 1 through 5 remain sound throughout. They establish that you did not cause yourself. What they do not establish is that the second-personal demand directed at you lacks normative standing, because that standing was not borrowed from self-origination in the first place.

R.J. Wallace (*Responsibility and the Moral Sentiments*, 1994) offers a complementary route through the fairness of holding expectations. On Wallace's account, reactive attitudes are fair when the agent has the general capacities for reflective self-control, that is, the ability to grasp moral reasons and be moved by them. Fairness, here, is the operative normative category, and it takes its standing from the *present relational* fact of capacity-possession. It does not take it from the historical fact of self-authorship. The regress does not bite, because Wallace's condition is synchronic and relational, with nothing diachronic or self-grounding in it. (*Synchronic* means holding at a single time. *Diachronic* means holding across time.)

The honest residual bound. The partition and the demand-to-blame inference are defended here. They are not proven. Three limits must be stated with precision.

First, the route through Watson's attributability and Wallace's fairness is cleaner for this argument than the route through Shoemaker's full accountability triad, because Shoemaker explicitly places *resentment* on the accountability side of his taxonomy, the side that carries the backward-looking reactive register most exposed to the desert charge. If the arity argument wishes to travel fully within Shoemaker's framework, it must accept that the accountability face of the triad, meaning resentment, indignation, and blame in its full backward-looking mode, may not be fully insulated from the Basic Argument by the answerability partition alone. The safer claim is this: *answerability* and the *attributability* conditions together warrant the demand-to-account and a range of reactive responses. Whether they warrant resentment at the full backward-looking intensity that Strawson's Section 8 invokes is a question the partition leaves open.

Second, the demand-to-blame inference moves from "second-personal standing is constituted in the address-community" to "blame is warranted." This move is defensible. Blame directed at an addressable source who acted from their own evaluative orientation, within a community of mutual accountability, is not an arbitrary imposition. But it is not demonstrated that this exhausts what "warranted" blame requires. The hard incompatibilist will press the point: even granting that the address is structurally appropriate, *proportionality* and *desert* in the full sense require more. The answer is that the arity argument does not need to win that point in full. It needs only to show that the second-personal demand carries normative standing that the Basic Argument cannot dissolve, and that is what the demand-to-blame inference establishes.

Third, the argument applies reflexively to this paper's own standing. If second-personal authority is constituted in address-communities and cannot be self-bestowed, then the claim here to have resolved the desert gap is itself subject to uptake and refusal by the community of philosophical inquirers. The inference is offered to that community. Its standing is conferred by the quality of the argument. No authority the paper exercises over the Basic Argument by mere assertion confers it. That reflexive application is not a weakness. It shows the argument is consistent: the no-aseity structure runs through the argument itself. High confidence on the partition and the mechanism it relies on. Moderate confidence on the completeness of the demand-to-blame inference as a full reply to the desert charge. The honest bound is that the inference is currently unfalsified, short of being proven unbiteable.

5. The Basic Argument, Granted and Redirected

Plainly: the Basic Argument proves nothing causes itself, and that is granted. The only false step is treating that as proof that no sourcehood of any kind exists.

Galen Strawson's Basic Argument (*Philosophical Studies*, 1994; also *Freedom and Belief*, 1986; high confidence) may be stated as follows.

1. To be truly morally responsible for what one does, one must be truly responsible for *how one is*, that is, for the mental nature and the principles from which the act flows.
2. To be truly responsible for how one is, one must have brought it about that one is that way.
3. To have brought *that* about, one must already have been a certain way, and have been responsible for *that* prior way.
4. This regresses infinitely. To stop it, one must be *causa sui*, ultimately self-originating.
5. Nothing can be *causa sui*.
6. Therefore no one is ever truly, ultimately morally responsible.

This paper agrees with premises 1 through 5 entirely. The aseity-free framework has already denied, across existence, status, meaning, and agency, that anything is self-standing. To exempt the will alone would be to grant it an exemption that nothing else receives. The Basic Argument is the no-aseity thesis applied specifically to the will. Self-causation *is* aseity, aseity is impossible, and the Basic Argument has the proof of impossibility right. Van Inwagen's consequence argument is granted on the same grounds. It too is a no-aseity argument in another register, showing that the agent is not the ultimate source.

The point of disagreement is precise and confined to one inference: from the true content of premise 6, *no ultimate, self-originating responsibility*, to the conclusion *no responsibility, no sourcehood of any kind*. That inference is valid only if responsibility and sourcehood mean ultimate self-origination. The aseity demand is packed into the words "true" and "ultimate" as Galen Strawson uses them. Strip the demand out, and premises 1 through 5 establish what they establish, that the will is not *causa sui*, and establish *nothing* about whether there is conferred, inhabited, second-personally-constituted sourcehood. Conferred sourcehood was never a candidate for self-origination, so it is untouched by the proof that self-origination is impossible.

The Basic Argument refutes libertarianism, which *did* claim self-origination. It refutes Kane's event-causal libertarianism (*The Significance of Free Will*, 1996; high confidence), which tried to buy back aseity within a naturalistic frame through undetermined self-forming actions. Even granting Kane the indeterminism, the luck objection (Mele, *Free Will and Luck*, 2006; high confidence) bites. An undetermined choice, at the moment of undetermination, is not more the agent's own. If anything it is less the agent's own, because nothing about the agent settles it. This is the deep reason aseity is the wrong place to look: *self-origination, even if secured, would not give ownership*. It would give a coin-flip that the agent happens to be standing next to. Ownership is conferred and inhabited. It is not originated. Kane's heroic effort to secure aseity, and its vulnerability to luck, is perhaps the strongest available evidence that sourcehood was never going to be found in self-origination.

6. Pereboom's Four Cases and the Structure of Bypass

Plainly: the difference between a manipulated puppet and an ordinary person is whether sourcehood was built through address or built around it. The visible manipulator is not what makes the difference.

Pereboom's four-case argument (*Living Without Free Will*, 2001; high confidence) is the sharpest blade in the hard incompatibilist armory. It works like this. In Case 1, Professor Plum has been created moments

ago by neuroscientists who manipulate him in real time to reason and act egoistically, which issues in a murder. He meets every standard compatibilist condition: reasons-responsiveness, second-order endorsement, and guidance control. (*Reasons-responsiveness* is the capacity to register reasons and act on them. *Second-order endorsement* is wanting to have the wants one acts on. *Guidance control* is the agent's own mechanism steering the act.) Yet, intuitively, Plum is not responsible. He is a puppet, whether or not his internal states meet the compatibilist criteria. Cases 2, 3, and 4 progressively remove the external manipulators: first prior programming, then a deterministic upbringing engineered to produce these dispositions, then ordinary determinism. Each case, Pereboom argues, is relevantly like the one before it. If Case 1 produces no responsibility, and Case 4 is relevantly similar, then no determined agent is responsible.

The reply here rests on the arity argument. The break between Case 1 and Case 4 is not a difference in local dynamics. Pereboom is right that those can be matched. It is also not the mere presence of visible manipulators. The difference is whether the agent's sourcehood was *constituted through the second-personal community* or *bypassed it*. Manipulation, in the morally relevant sense, is shaping that bypasses the address. The neuroscientists of Case 1 operate *on* Plum in the first-to-third mode. They do not address him, do not demand of him, and do not stand in mutual accountability with him, such that his uptake of their address is a success condition for anything. Ordinary formation runs *through* the second-personal. A child is addressed, held responsible for the quality of its will toward others, drawn into the reactive-attitude community, and takes up principles as its own *in dialogue with other sources* (Fischer and Ravizza, *Responsibility and Control*, 1998; high confidence on this mechanism).

The opponent will press the point: "through the address-community" versus "by bypass" marks a place on a gradient, and it is not a difference in kind. It is a difference in causal route, and every transition from Case 1 to Case 4 differs in causal route. This objection is the sharpest available, and it has a precise answer. A smooth variation in the *degree* of second-personal constitution is fully consistent with a categorical difference at the limit. Responsibility is *graded* exactly because second-personal constitution is graded. Cults and coercive control are hard cases precisely because the address-community is *partly* bypassed. The theory predicts this, and is not embarrassed by it. The claim "the cases shade into each other" is granted, and it is not what the opponent needs. The opponent needs "there is no point at which second-personal constitution reaches zero while the dynamics stay matched." There is such a point. It is Case 1, real-time external authorship, where the address-community's contribution to *this act's* sourcehood is nil. A continuous function can still take the value zero. A dimmer switch is continuous, and it still has an off position.

The criterion that sorts the cases does not beg the question. It is the arity argument, fixed independently. *Taking-up* is the relation in which the agent occupies the second pole of an address, has the standing to answer and refuse, and whose uptake is part of the act's success conditions. *Installing* is the first-to-third operation in which the agent's uptake is not a success condition. This distinction is attested in linguistics, in developmental psychology, and in the philosophy of indexicals *before* any of the parties reach the free will question. (*Indexicals* are expressions like "I," "you," and "here," whose reference depends on who is speaking to whom.) Apply the distinction to ordinary child development. A child who does not take up the address is merely conditioned. Such a child is not formed, and we mark that difference in ordinary talk. We distinguish raising a child from training a dog *by* the second-personal uptake condition. Now apply it to Case 1. The manipulators get the murder whether or not Plum answers anything. That asymmetry, whether uptake is a success condition, is the criterion. The criterion is not drawn from the conclusion it sorts.

7. Compatibilism, the Regress, and What Conferred Sourcehood Adds

Plainly: ordinary compatibilism smuggles in a small self-made self to stop the backward chain; conferred sourcehood denies the chain needs a self-made stopping point at all.

Standard compatibilism, in Frankfurt's hierarchical account (1971, *Journal of Philosophy*; high confidence) and in Fischer and Ravizza's guidance control, holds that an act is free when it flows from the self in the right way: a first-order desire endorsed by a second-order volition, a reasons-responsive mechanism, owned by the agent. (A *first-order desire* is a want for something. A *second-order volition* is a want about which first-order want moves you.) This is genuine progress. The Frankfurt cases, whatever their standing against the Principle of Alternate Possibilities, correctly locate the question in the *actual sequence*, away from counterfactual alternatives.

The debt that ordinary compatibilism cannot discharge is what Watson and Velleman (cited in substance as this critique's source; high confidence on the attribution of the regress pressure) press as the regress of identification. Why is the second-order volition *authoritative*? Because the agent identifies with it. And the third-order? At some point the identification must simply settle, and that settling looks like a small aseity, a point where the buck stops *in* the self as an unexplained identifier. Ordinary compatibilism tends to install an authoritative self to halt the regress, and then does not pay for it in the currency of an account.

Conferred sourcehood parts ways here, and the parting is its whole substantive contribution to the compatibilism literature. It does not halt the regress in the self. It denies that the regress needs a self-standing terminus at all. The buck does not stop *in* the agent as an uncaused identifier. It stops at a standpoint that is itself conferred: caused, constituted, re-enacted, and holding an authority that is bestowed-and-inhabited, with nothing self-generated about it. Christine Korsgaard (*Self-Constitution*, 2009; high confidence) captures the structural point. Agency is the continuous activity of constituting oneself as author. It is not a standing property of a pre-existing self. The "self" is not the fixed point from which the endorsement comes. The self *is* the endorsing activity, re-enacted each time, grounded in nothing more self-standing than anything else in a no-aseity world. This makes conferred sourcehood something other than a version of compatibilism. It is compatible with compatibilism's best insights, and it declines compatibilism's deepest evasion.

8. Forgiveness, Resentment, and the Evidence of the Second Person

Plainly: forgiving and writing someone off can look identical from outside. Being managed hurts in a way that being addressed does not, and both facts reveal a second person the shaping account cannot capture.

Two further phenomena light up the arity argument from different angles.

Forgiveness. The hard incompatibilist can model forgiveness as a disposition-reset that restores cooperative expected value after a defection. Maintaining resentment is costly, and ending it signals that the cooperation channel is open again. This is a real and non-trivial functional story. But notice what it leaves unspecified. It predicts the behavioral and signaling profile of "I have forgiven you" without capturing the directed character of forgiveness as an act. I can reset my dispositions toward you, re-open the cooperation channel, and stop the resentment-maintenance loop *without forgiving you*. I can do this by simply ceasing to care, or by writing you off as a hazard now contained, or by a one-sided prudential recalculation in which you are never addressed at all. The behavioral and signaling profile of "I have forgiven you" and "I have written you off as no longer worth resenting" can be made identical. What differs is whether a *claim against a source was released to that source*. Forgiveness has an addressee who can *refuse* it. Dropping a maintenance cost has no addressee to refuse anything. The asymmetry is on display, and it does not rest on assertion. "I forgive you" requires

a “you” who could say “I do not accept your forgiveness” in response, and that response has normative consequences: it keeps the claim alive, it registers a standing grievance. “I have ceased to maintain the resentment-loop” has no analog for any of this. The second-personal structure is visible in the grammar of the practice.

The objective stance as status-demotion. P.F. Strawson noted that being moved to the objective attitude toward a person, which treats them as a system to be managed and no longer as a source to be addressed, feels like a demotion in their status. The hard incompatibilist can explain why it *feels bad*. On the shaping account, being moved from the participant protocol to the management protocol is a fitness-relevant status-demotion for a social primate, so of course it registers as a loss. This is a strong predictive match. Notice, though, that it predicts the *valence*, that it feels bad, without predicting the *content*, what the hurt is about. The shaping account predicts: moved to management protocol, therefore loss, therefore feels bad. It does not predict that the loss will present as a withdrawal of *standing to answer*. The report is “you are treating me as something that cannot answer you,” and it is not “you have lowered my expected cooperative value.” These are different intentional contents, and the second is the one people moved to the objective stance report: the patient who says “stop managing me and *talk to me*”; the prisoner whose grievance is “you treat me as if I cannot answer,” and not “you have predicted my behavior.” The second-personal account predicts that specific content independently of the free will question. The shaping account predicts only the valence and must treat the content as confabulation, that is, as an after-the-fact story the mind tells that does not track the real cause. That is the remainder, and it is independently attested, in linguistics, in the philosophy of indexicals, and in developmental psychology’s account of the second person, before anyone reaches the question of free will.

9. The Occasionalism of Agency: A Narrowed Claim

Plainly: the strong old claim that authorship can change with no physical change is withdrawn. What remains is that authorship does not reduce to the agent’s state at a single instant, because it depends on relational history.

The body of work from which this paper is drawn develops a companion claim, the “occasionalism of agency,” which holds that authorship drains and returns in ways that do not always track a difference in dynamics. (*Occasionalism* is, in its classical use, the view that no created thing has real causal power. Here it names the claim that authorship can shift without an underlying dynamical shift.) That claim, as originally stated, is withdrawn. Under the framework’s own physicalism, the view that everything real is physical, every real psychological transition is a real dynamical transition. The addict who watches herself relapse, and the dissociated subject who watches herself act, are each in a real dynamical change. Frankfurt’s unwilling addict (1971; high confidence), the person moved by a will she does not endorse, is in a different dynamical state from the integrated agent. The part-led versus Self-led distinction in Internal Family Systems, if it tracks something real, tracks a real difference in network configuration. The original claim that agency-loss does not always track a dynamical difference is false, and the falsity is conceded.

What survives is a narrower proposition. Ownership of an act is not fixed by, and does not reduce to, the agent’s *local synchronic dynamical state*, its state at a single instant. Two episodes can be matched in the agent’s instantaneous dynamical profile and still differ in authorship, because authorship is constituted by relational and historical physical facts, specifically, whether the mechanism was taken up through the second-personal address-community or installed by bypass, and those facts do not show up in the instantaneous state. Ownership supervenes on the total physical facts, so there is no dualism. It does not supervene on the local dynamics, so there is no reduction to dynamics. (To say ownership *supervenes on* the physical facts is to

say there can be no difference in ownership without some difference in those facts.) This is a weaker claim, and it is stated as the weaker claim. Its only clean witness is the manipulation case, which means it depends on the arity argument and does not float on independent evidence. It is stated at moderate confidence. It depends on the second-personal remainder, and does not corroborate it.

The concrete significance of this narrowed claim is this. Pereboom's four-case argument holds the structural and dispositional conditions fixed and varies etiology, the causal history. It cannot, on its own terms, also hold fixed the *second-personal standing*, because second-personal standing is not a structural-dispositional fact. That is the gap the arity argument opens. The narrowed occasionalism argument is the observation that the gap is filled by relational and historical physical facts. Nothing supernatural fills it.

10. Conferred Sourcehood: The Verdict

Plainly: drop the aseity demand and you are not left with nothing. You are left with authorship that is given, inhabited, and two-place, and that is exactly what no one-place vocabulary can eliminate.

The argument of this paper may be assembled into a single structure.

The aseity premise, the demand that free will consist in self-origination, is a smuggled assumption. It is not an analytic necessity. Dropping it does not leave a vacuum. It leaves *conferred sourcehood*: the standpoint that is caused, constituted, addressed, and re-enacted; the authorship that is inhabited and not generated from within; the self that is, as Korsgaard has it, the continuous activity of constituting an author, and not a pre-existing object whose endorsements are read off.

Conferred sourcehood is not a deflationary position. It does not lower the bar for what responsibility requires. It challenges whether the bar was ever correctly described. The three cross-traditional witnesses show that ownership-without-origination is a robust structure, arrived at independently. It is not an evasion. The arity argument shows that the hard incompatibilist's proposed replacement of the reactive attitudes with forward-looking management is an elimination, and not a reduction. It eliminates the second argument-place of a two-place relation, and that cannot be accomplished by supplying a one-place predicate, however finely calibrated.

The proposal, stated exactly: holding responsible is a two-place relation whose second relatum is an addressee *as an addressee*, a source who can answer, against whom a claim can lie, whose refusal of forgiveness keeps the claim alive, and whose uptake of address is part of the success conditions of address. No third-personal, forward-looking, one-place description of behavior-shaping captures that second relatum. The eliminativist owes such a description. Structurally, it cannot be supplied, because the second argument-place is not a predicate to be stripped. It is a relatum, one to be either included or discarded. Discard it, and you have not refined holding responsible. You have abolished it and produced something else. That something else may be useful, and it is not the same thing.

Determinism is granted in full throughout. The consequence argument is granted, and the Basic Argument is granted through premise 5. What the determinism and the Basic Argument together establish is that the agent is not the ultimate, self-originating source of anything. That is true and important. They do not establish, and cannot establish without the aseity premise smuggled back in at the conclusion, that there is therefore no sourcehood, no responsibility, and no second-personal standing. Conferred sourcehood claims none of the things the hard incompatibilist successfully refutes. It claims only what the arity argument shows cannot be refuted by any one-place vocabulary: the irreducible two-place structure of an address, the remainder that forgiveness exhibits and that the solipsism test proves does not reduce to the necessity of a locus.

Verdict on the argument. The arity gap, the structural impossibility of target-fixing a two-place relation with a one-place vocabulary, passes the test for substantive independence. It is falsifiable. The argument fails if a complete forward-looking redescription is supplied that captures, without remainder, the second-personal differentials: forgiving-while-detering versus deterring-without-forgiving, releasing-a-claim versus dropping-a-cost, being-owed-an-account versus predicting-behavior. No such description has been supplied. The eliminativist's bullet is unbitten. It is not proven unbiteable. That distinction is maintained throughout, and it is the honest bound.

The argument would be insincere if it exempted the page it is printed on. Authority, like sourcehood, cannot be self-bestowed. A body of work that declares itself canonical by self-assertion inhabits the same aseity error the argument was written to expose. Authority is conferred and relational. It is constituted, like sourcehood, in a second-personal community that can answer back and withhold uptake. The argument applies reflexively, and it would not be serious otherwise.

References

All citations are drawn from primary sources engaged in the supporting corpus, with confidence labels as indicated.

Abadi, M., Burrows, M., Lampson, B., and Plotkin, G. (1993). "A Calculus for Access Control in Distributed Systems." *ACM Transactions on Programming Languages and Systems*, 15(4), 706–734. [cited in related membrane work]

Darwall, S. (2006). *The Second-Person Standpoint: Morality, Respect, and Accountability*. Harvard University Press. [high confidence]

Fischer, J. M. and Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press. [high confidence]

Frankfurt, H. (1969). "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy*, 66(23), 829–839. [high confidence]

Frankfurt, H. (1971). "Freedom of the Will and the Concept of a Person." *Journal of Philosophy*, 68(1), 5–20. [high confidence]

Gbadegesin, S. (1991). *African Philosophy: Traditional Yoruba Philosophy and Contemporary African Realities*. Peter Lang. [moderate confidence on philosophical interpretation]

Kane, R. (1996). *The Significance of Free Will*. Oxford University Press. [high confidence]

Kim, J. (1966). "On the Psycho-Physical Identity Theory." *American Philosophical Quarterly*, 3(3), 227–235. [high confidence on the foundational event-constitution view]

Korsgaard, C. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press. [high confidence]

Mele, A. (2006). *Free Will and Luck*. Oxford University Press. [high confidence]

Nietzsche, F. *Beyond Good and Evil*, §21. [high confidence; cited as Galen Strawson's own source for "best self-contradiction"]

Pereboom, D. (2001). *Living Without Free Will*. Cambridge University Press. [high confidence]

Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life*. Oxford University Press. [high confidence]

Sapolsky, R. (2023). *Determined: A Science of Life Without Free Will*. Penguin Press. [high confidence on thesis]

Strawson, G. (1986). *Freedom and Belief*. Oxford University Press. [high confidence]

Strawson, G. (1994). “The Impossibility of Moral Responsibility.” *Philosophical Studies*, 75(1–2), 5–24. [high confidence]

Strawson, P.F. (1962). “Freedom and Resentment.” *Proceedings of the British Academy*, 48, 1–25. [high confidence on text and thesis]

van Inwagen, P. (1983). *An Essay on Free Will*. Oxford University Press. [high confidence]

<<<END ARITY-GAP-PAPER>>>

<<<BEGIN PROVENANCE-LEDGER>>>

Provenance Ledger for the Arity-Gap Paper

This ledger records, for each substantive claim in the paper, its source in the adjudicated substrate and the confidence label carried forward from that substrate. It is the paper’s firewall: every claim traces to a source, and unsourced claims do not appear.

Claim	Source in corpus	Confidence
Free will debate conducted as search for origin; aseity premise shared by all parties	part-III-free-will.md §0	high (this is the chapter’s stated organizing claim, internally coherent)
Basic Argument (G. Strawson 1994) premises 1–5 stated and granted	part-III-free-will.md §4; G. Strawson, <i>Philosophical Studies</i> 1994	high
Nietzsche <i>BGE</i> §21 as source for “best self-contradiction”	part-III-free-will.md §4; credited as G. Strawson’s cited source	high
van Inwagen consequence argument, <i>Essay on Free Will</i> 1983	part-III-free-will.md §5	high
P.F. Strawson “Freedom and Resentment,” <i>Proceedings of British Academy</i> 1962	part-III-free-will.md §1.1	high
Darwall <i>Second-Person Standpoint</i> 2006	part-III-free-will.md §1.1	high
Sapolsky <i>Determined</i> 2023 thesis	part-III-free-will.md §1.2	high
Pereboom <i>Living Without Free Will</i> 2001; <i>Free Will, Agency, and Meaning in Life</i> 2014	part-III-free-will.md §1.2 and §6	high

Claim	Source in corpus	Confidence
Second-personal remainder (R): holding-responsible is two-place and resists reduction to one-place shaping	part-III-free-will.md §1.2–1.5; MOCK-VIVA.md (“the corpus’s best independent philosophy”)	high
Phlogiston objection and its failure on arity grounds	part-III-free-will.md §1.3	high
Solipsism test proving (R) floats free of the necessity-of-locus residue	part-III-free-will.md §1.5	high
Forgiveness asymmetry: release-of-claim vs. cost-drop; forgiveness can be refused	part-III-free-will.md §1.4	high
Objective-stance demotion: predicts the valence and leaves the content unexplained	part-III-free-will.md §1.4	moderate (on the claim about what people report; the structural point about valence/content distinction is high)
Ash’arī <i>kasb</i> doctrine: al-Ash’arī d. 935, elaborated by al-Bāqillānī, al-Juwaynī, al-Ghazālī	part-III-free-will.md §3	high (existence and core); moderate (finer scholastic disputes)
Yoruba <i>orì</i> : Gbadegesin 1991, Abimbola, Hallen	part-III-free-will.md §3	high (broad structure); moderate (philosophical interpretation)
Tied-camel hadith: al-Tirmidhī	part-III-free-will.md §3	high (canonical transmission); moderate (exact chain/grading)
Frankfurt 1969 Frankfurt cases, <i>Journal of Philosophy</i>	part-III-free-will.md §5	high
Frankfurt 1971 hierarchical account, <i>Journal of Philosophy</i>	part-III-free-will.md §5	high
Watson/Velleman regress critique of compatibilist identification	part-III-free-will.md §5	high (credited in substance to both)
Korsgaard <i>Self-Constitution</i> 2009	part-III-free-will.md §2.4 and §5	high
Fischer and Ravizza <i>Responsibility and Control</i> 1998	part-III-free-will.md §2.3 and §6	high
Kane <i>Significance of Free Will</i> 1996	part-III-free-will.md §4	high
Mele <i>Free Will and Luck</i> 2006: luck objection to Kane	part-III-free-will.md §4	high
Pereboom four-case argument: manipulation continuum	part-III-free-will.md §6	high
Chaos “mathematical teeth”: withdrawn as equivocating epistemic unpredictability with constitutive irreducibility	00-CONTRIBUTION.md §3 (“carries no ‘mathematical teeth’ ”); MOCK-VIVA.md body/preamble table row 1; THESIS.md line 52 and 215	high (the retraction is explicit in the corpus; the equivocation charge is the adjudicated finding)

Claim	Source in corpus	Confidence
Occasionalism of agency (O) withdrawn; (O') retained as narrowed, (R)-dependent claim	part-III-free-will.md §2.2–2.3	high (on the withdrawal of the original); moderate (on (O') as a defensible reformulation)
Cross-domain convergence as genuine evidence; workflow-convergence as etiologically correlated internal consistency	MOCK-VIVA.md §3; CONSOLIDATION.md §1	high
MOCK-VIVA verdict: “the corpus’s single best piece of independent philosophy” re the free-will chapter	MOCK-VIVA.md §1	high (direct quotation)
Provenance firewall: AI-assisted substrate; human candidate must own and defend	CONSOLIDATION.md §4	high

Items not claimed, per the adjudications:

- The grand forced-relocation of the explanatory gap (needs thick index; returned as REDESCRIPTION by the Remainder Law clause ii; conditional and undischarged). Not present in the arity-gap paper.
- Chaos theory as providing “mathematical teeth” to the irreducibility thesis. Explicitly withdrawn in this paper and in the corpus.
- The Lorenz attractor as evidence for free will or sourcehood. Never claimed in this paper.
- Type-B physicalism rejection as proven. Declared but not proven; the membrane + thin seam + arity argument are type-B-independent and do not depend on that rejection.
- Workflow-convergence as independent confirmation. Corrected to internal consistency; not invoked in this paper.

<<<END PROVENANCE-LEDGER>>>