Afterword: On Kit Fine's Reply

'Essences, Heuristics, and Metaphysical Illusions' first appeared in the journal *Metaphysics* with a reply by Kit Fine (Fine 2025), as part of a symposium on 'Essence and Modality' (Fine 1994). The symposium grew out of a session at the 2024 Pacific Division meeting of the American Philosophical Association in Portland, to celebrate the thirtieth anniversary of the publication of that paper. This afterword is a reply to Fine's thoughtful reply, which provides a welcome opportunity to take the argument of my paper further.

Fine divides his comments into four sections; so do I. For ease of comparison, my section headings are quotations of his for the corresponding sections of his paper.

1. *'A tale of two theses'*

'Essence and Modality' is usually read as arguing, or indeed *showing*, that essence is metaphysically irreducible to modality, and in particular that although Socrates is necessarily a member of singleton Socrates, he is not essentially a member of singleton Socrates. However, according to several passages in that paper, what matters is not the *truth* of such anti-reductionist claims but their *intelligibility*—though other passages do claim their truth. As I note, intelligibility is a modest claim, and hardly contentious. In his reply, Fine specifies that his concern was never just the bare intelligibility of the anti-reductionist claims, but their *reasonableness*. Even that is a low bar. Obviously, many contemporary metaphysicians make all sorts of outlandish claims without ceasing to be reasonable people by normal standards. My paper argues that the evidence for key anti-reductionist claims such as 'Socrates is not essentially a member of singleton Socrates' is much weaker than often supposed, and insufficient to undermine the abductive case for a modal theory of essence. Naturally, I hope that the community of metaphysicians will eventually come to see and accept this methodological point. But abductive theory comparison is not a cut-and-dried business, the psychology of heuristics is under-explored, and even the correct framework for evaluating semi-formal proofs is unobvious (my paper postulates heuristics and offers semi-formal proofs). As elsewhere in science, those invested in a research programme will be deeply reluctant to abandon it when it runs into trouble, and many never do. That general tendency has some advantages: it prevents research programmes from being abandoned prematurely. The best way to explore the strengths and weaknesses of hyperintensional essentialism is for its proponents to push it to the limit, even when good alternatives are available. That was presumably also the best way to explore the strengths and weaknesses of vitalism in early twentieth-century biology. From outside the research programme, hyperintensional essence looks like the *élan vital*.

2. *'Essential rot'*

Fine's second section concerns my analogue for natural numbers of his argument about singleton Socrates and Socrates' essence.[1] He focuses mainly on the semantics of complex singular terms, such as 'p(9)', where the function symbol 'p' (for predecessor) is applied to the numeral '9'. I give them the simplest, most natural, and most nearly homophonic compositional semantics, which treats them as genuine singular terms, not disguised definite descriptions. On this treatment, complex functional terms refer to their usual referents; in particular, 'p(9)' refers to the natural number 8. Indeed, it refers to 8 *directly*, in the sense that all it contributes to the compositional semantic evaluation of more complex expressions in which 'p(9)' occurs is its referent, 8.

In response, Fine's general strategy is to propose various more complicated, less natural, and less homophonic semantic treatments on which my arguments might fail. Such strategies are *always* formally available: any given argument is invalid on infinitely many complicated, unnatural, and non-homophonic semantic treatments. To my remark that there is no semantic obstacle to treating functional expressions as primitive, his initial unpromisingly generic response is: 'This may be so, but it does not mean that they directly refer to what we take to be their usual referent'. It obviously does not mean that, but the question is whether there is any evidence that they do *not* refer to what we take to be their usual referent; none has been offered.[2] Only a well-motivated alternative to my straightforward semantics poses any threat to my argument.

Next, Fine allows that 'we may provide a semantics for such expressions in which their semantic value is their usual referent', but complains:

> Williamson only provides a semantics of this sort for a language in which the means of construction are extensional; and yet we are all familiar with the fact that a semantics which is adequate for an extensional language (as when we take sentences to signify truth-values) may not be adequate when the language is taken to include non-extensional forms of expression.

Fine later amplifies this complaint by arguing that 'the most natural semantic principles governing the transition from a typed extensional language to its intensional counterpart determine that' functional expressions 'cannot be directly referential', because 'in the transition to the intensional language, there is no more reason in the case of function symbols than in the case of predicate symbols to suppose that their application will result in the rigidification of semantic value'. On Fine's approach, 'the intensional semantic value of the functional expression f(t) will be the function taking each world w into $f(w)(t(w))$, with […] the referent of f(t) depending, in general, upon the world' (Fine uses italicization to express the mapping from expressions to their referents). Thus, Fine argues, my treatment of functional expressions fails even within my favoured intensional framework.

Fine's argument would be compelling if intensionalizing an extensional semantics involved relativizing the extensions of all types to a world parameter. Indeed, his use of the term '$t(w)$' shows that he envisages relativizing the extension of the singular term t to the world parameter w. To apply the approach consistently, one must relativize the values of variables as well as constants to the world parameter, so the semantic value of a constant of any given type is always a value the variables of that type can take. Carnap did just that in

Meaning and Necessity (Carnap 1947). The effect is to treat the values of individual variables as *individual concepts*, mappings from worlds to individuals, and first-order quantification as quantification over individual concepts. Puzzlingly, Fine omits to mention that such an approach to the semantics of intensional languages was largely abandoned, with good reason, soon after Carnap first proposed it, following Kripke's classic work on the semantics of quantified modal logic (Kripke 1963). For if the quantifier ranges over individual concepts, then $\Box\exists x\, Fx$ entails $\exists x\Box Fx$ (for extensional F), since when $\Box\exists x\, Fx$ is true, some individual concept maps each world w to something in the extension of F at w. That entailment fails on a normal understanding of quantified modal discourse. For example, let Fx be 'either x is rich or nothing is rich'. Necessarily, something is such that either it is rich or nothing is rich (by standard first-order logic). But it does not follow that something is such that necessarily either it is rich or nothing is rich; there is no such object. In general, even if F has a nonempty extension at each world, the intersection of all those extensions may be empty. Carnap's quantification over individual concepts behaves more like second-order quantification over properties that necessarily have an extension of cardinality one. Since Kripke, both first-order variables and first-order constants in modal logic are normally interpreted as simply taking individuals as their semantic values, with no relativization to a world parameter; thus, all expressions of the same type as first-order variables are rigid designators (Williamson 2013: 210-15 gives details). For proper generality, the variables range over all available semantic values for singular terms, including complex singular terms such as the functional expression 'p(9)'. They all designate individuals rigidly.

Superficially, what Fine presents as the obvious way to intensionalize extensional semantics is natural enough, but it soon turned out to be an obstacle to progress in quantified modal logic, and for the last sixty years has had little more than curiosity value. Relativization to worlds has turned out to be appropriate for the type of sentences, and for types derived from it, such as types of predicate, but inappropriate for the basic type of individuals. My arguments work with the now-standard semantic framework for singular terms, including functional expressions.

Naturally, one can introduce a more complex type of expression whose semantic values do map worlds to individuals, and may be non-rigid. To take Fine's example, one can treat '#(F)' that way: its semantic value maps each world w to the cardinality of the extension of the predicate F in w, so '#(F)' behaves like a definite description and is typically non-rigid. Since predicates are typically non-rigid, that treatment of '#' is natural enough—though in some contexts it may be appropriate to interpret '#' differently, fixing the reference of '#(F)' directly and rigidly to the cardinality of F's actual extension.

Fine also mentions the use of variables like 'v' and 'a' in physics, writing:

> These may surely be taken to be abbreviations for functional expressions, such as 'v(x, t) or 'a(x, t)', to indicate the velocity or acceleration of a body x at time t. But under the most natural reading of such expressions, they will not in general be rigid designators—[…] the velocity or acceleration of a body will in general depend upon the circumstances—and so these expressions cannot be taken to be directly referential.

The tacit parameters are those on which velocity and acceleration relevantly depend: what body it is, what time it is. But sometimes modal dependency is relevant too: the physicist may be comparing the velocity or acceleration of x at t in a physically possible history h of the system with the velocity or acceleration of x at t in an alternative physically possible history h*. Then, by the same abbreviatory policy as Fine invokes, ‘a’ and ‘v’ should be taken to be abbreviations for functional expressions such as ‘v(x, t, h)’ or ‘a(x, t, h)’, in which the modal parameter is explicit too: no principled reason has been offered for treating the modal and temporal dimensions differently in this respect. Once we treat the modal parameter as implicit, just as Fine treats the temporal parameter as implicit, rigidity is restored, since the velocity or acceleration of a given body at a given time in a given history is not itself contingent. Thus, when Fine’s policy for interpreting physical variables is applied consistently, it shows just how to understand them as in effect directly referential functional terms, in line with my account of functional terms in arithmetic. The treatment of functional terms in mathematical notation as directly referential is robust.

Of course, intensional semantics is not Fine’s own preferred framework: he discusses it only in an attempt to get my treatment of functional expressions into trouble on my home territory. He sketches a simple hyperintensional account closer to his heart, on which the semantics assigns each expression an ‘objectual content’, a set of objects, where ‘the objectual content of a complex expression is the sum [union] of the objectual contents of its constituent expressions’. Fine treats an expression’s objectual content as a component of its semantic value, together with its usual referent. He treats the objectual content of a numeral as just the singleton of its usual referent.

Let us use the term ‘involves’ for the relation between an objectual content and its members. For example, 5 − 3 involves 5, because 5 is a member of the objectual content of the numeral ‘5’, which is a subset of the objectual content of the functional term ‘5 − 3’, so 5 is a member of the objectual content of ‘5 − 3’. But 2 does not involve 5, because the objectual content of the numeral ‘2’ is just the singleton of 2. On Fine’s semantics, involving is not a metalinguistic relation, since it depends only on the semantic values of the input terms, which in turn determine both their objectual contents and their usual referents. Thus, although 5 − 3 involves 5, 2 does not involve 5. But then shouldn’t we conclude by Leibniz’s Law that 5 − 3 ≠ 2? Fine’s approach is in danger of concluding that elementary arithmetic is false, strictly speaking, because there is a subtle metaphysical difference between 5 − 3 and 2. No doubt, with enough complications, one could wriggle out of the problem, but to get anywhere near ‘proving’ that $5 - 3 \neq 2$ is a strong hint that something has gone badly wrong.

In partial defence of objectual content, Fine says ‘it might be thought that if we are able to bind into the contexts occupied by the expressions $E_1, E_2, \ldots, E_n$ [in the argument places of a function symbol], it is essential that they still have objectual import’, a suggestion he attributes to a referee for his paper. But consider arithmetical sentences like $\forall x \forall y (x + y = y + x)$, where the quantifiers bind into the argument places of the function symbol ‘+’. A standard quasi-homophonic semantics for a first-order language with primitive function symbols has no trouble handling such sentences, with the usual apparatus of recursively determined semantic values for all relevant expressions relative to assignments of standard values to variables. Such appeals to ‘objectual import’ are redundant.

In short, Fine’s discussion reveals no problems for my treatment of functional terms in arithmetic and elsewhere as directly referential. That is the simplest, most natural semantics for such terms; its upshot is that some of our essentialist judgments made with such terms are mistaken.

3. *'Does the rot spread?'*

Phenomenologically, the question 'Is Socrates is essentially a member of singleton Socrates?' feels similar to the question 'Is 8 essentially *p*(9)?' My paper argues that our mistaken negative answer to the latter is naturally explained by our reliance on a heuristic that also predicts our negative answer to the former. Consequently, we should distrust both judgments. Fine's third section proposes various alternative explanations, to stop the rot spreading from 8 and 9 to Socrates and singleton Socrates.

What explains our errors is in large part psychological, and at present a matter of conjecture. Still, some psychological conjectures are more plausible than others. In the paper, I conjecture that we assess the truth-value of the claim that being *F* is 'essential' to an object *x* by assessing the goodness of '*x* is *F*' as a (partial) answer to the question 'What is *x*?', which includes applying a crude relevance filter, operating at the linguistic surface: if '*x* is *F*' is filtered out as irrelevant, being *F* is judged not essential to *x*.[3] Such a filter is quick and easy to use and has obvious advantages as a heuristic, since it helps us avoid wasting time on many bad answers to our questions, even though it sometimes eliminates correct answers too. It is plausible that we often rely on such heuristics. For the sake of argument, Fine grants that we tend to judge mistakenly that 8 is not essentially 9 – 1, but proposes explanations of our error that would not generalize to the judgment that Socrates is not essentially a member of singleton Socrates. Are his alternative explanations psychologically plausible?

On Fine's first suggestion, we judge that 8 is not essentially 9 – 1 because we misinterpret the functional term '9 – 1' as a definite description. Since his Socrates case does not employ functional terms, it is not vulnerable to such an error. The trouble with this explanation is that definite descriptions are far more cumbersome than functional terms to work with: compare '9 – 1' with the long-winded phrase 'the result of subtracting 1 from 9'. If one replaces arithmetical questions of moderate difficulty with their elaborate translations with definite descriptions in a natural language, the results rapidly become burdensome. Such paraphrases of the easy into the difficult go in the opposite direction from efficient heuristics, which simplify ruthlessly. Fine does not explain why we should tend to make life difficult for ourselves in this way. Thus, his first suggestion is psychologically implausible.

On Fine's second suggestion, we think through the arithmetic cases in metaphysical terms. Specifically, he postulates that we rely on the correct metaphysical principle *Ontological Link*, but combine it with the mistaken semantic principle *Referential Link*:

Ontological Link	If it is essential to y that A and if the proposition expressed by A is directly about x then y ontologically depends upon x.
Referential Link	The proposition expressed by a sentence A is directly about x whenever A contains a term that directly refers to x.

But these are principles of a kind that only a highly sophisticated contemporary analytic philosopher like Kit Fine would be in a position to entertain or assume. How many laypeople grasp the metaphysical category of ontological dependence or the semantic category of direct reference? Presumably, we are supposed to grasp them 'implicitly', but how would that work? Fine writes of our 'unthinking acceptance of *Referential link*' and of how we 'infer, on the basis of *Ontological Link*', but these alleged cognitive processes seem quite imaginary for normal human beings. Thus, his second suggestion is no more psychologically plausible than the first.

In short, Fine lacks a viable explanation for our mistaken judgment that 8 is not essentially 9 – 1. He does have an argument that, whatever the explanation is, it cannot result from a heuristic operating at the linguistic surface, as my Relevance Filter does:

> [L]et us control for this form of error by expressing ourselves in such a way that irrelevant linguistic detail cannot intrude on our judgement. Thus, in the case of singleton Socrates, I say: take the *object x* that is Socrates and take the *object y* that is singleton Socrates; then it is essential to *y* to have *x* as a member, though not essential to *x* to belong to *y*. It seems to me that our judgements in this case are just as firm as when we employ singular terms within the scope of the essentiality operator.

But what cognitive difference can this change in scope make? When we come to evaluate the open sentences 'It is essential to *y* to have *x* as a member' and 'It is essential to *x* to belong to *y*', we cannot do it unless we keep in mind what *x* and *y* are—Socrates and singleton Socrates respectively. The change of scope here makes no material difference to the operation of the relevance filter. It is like this proposal to resolve a Frege puzzle: take the *object x* that is Hesperus and the *object y* that is Phosphorus, and ask yourself whether *x* is *y*. We are no further forward.

Fine comments: 'Those of us who work in the theory of essence will recognize in Williamson's notion of irrelevance a linguistic cognate to the metaphysical notion of ontological dependence'. Such theorists of essence must be careful not to postulate fine-grained metaphysical structure that is merely a projection of the pragmatics of relevance.

4. *'Methodological musings'*

Fine's last section considers the subsequent development of the metaphysical and logical research programme adumbrated in 'Essence and Modality'. As he points out, 'It took almost 20 years after the publication of E&M for it to begin to be extensively cited'. But if one asks *who* was citing it, one can easily check that from early on it was being cited by leading metaphysicians, especially of the younger generation. That fits my memory of the period: from early on—even before E&M was published—metaphysicians 'in the know' were aware that it mounted an original and serious challenge to received (intensionalist) ways of thinking. Its influence spread gradually outwards, like the ripples made by a stone thrown into a pond. Thus, the hyperintensionalist research programme has been under development, at least by an elite of specialists, for over thirty years. We are already in a position to make some educated guesses about its prospects for the long run.

The most obvious effect of a hyperintensionalist framework is to multiply distinctions. Doing so is not always progress. We are better off without a distinction between 5 – 3 and 2, though of course we need one between the *expressions* '5 – 3' and '2'. In terms of abductive theory comparison, multiplying distinctions almost always detracts from simplicity, and often from strength (since it obstructs drawing consequences from the theory).

Presumably, the abductive value of multiplying distinctions is to enable the theory to accommodate more data. Such data should be reasonably hard, otherwise one might do better by rejecting both the distinctions and the data. Thus, a critique that undermines the data a hyperintensionalist theory is meant to explain is a potentially mortal threat. Of course, my paper focused on only one kind of supporting data for hyperintensionalism, though perhaps the most salient. My hunch is that the critique can be generalized, but that conjecture remains to be tested. At any rate, in Fine's work it encounters the most rigorous and systematic case available for hyperintensionalism: a good challenge.

Notes

1 The relevant metaphysical view of the definitions of the natural numbers is mentioned sympathetically in the original paper (Fine 1994: 14).

2 Fine glosses my text as if I had proposed treating a term like ‘*p*(9)’ as primitive, but it is clearly semantically complex, with the function symbol ‘*p*’ and the numeral ‘9’ as simpler constituents; the proposal is to treat the function symbol itself as primitive.

3 Since modal *words* like ‘must’ and ‘necessary’ normally do not trigger such uses of the relevance filter, even if they in fact have the same intension as ‘essential’, my argument against Fine’s hyperintensionalist account of ‘essential’ does not generalize to my intensionalist account of ‘must’ and ‘necessary’.