On a question about weak semantic conservativeness

Ewy Tsun

November 18, 2026

Abstract

We settle a question about weak semantic conservativeness raised by Stephen Mackereth.

Neologicists (amongst others) have explored a number of ways of articulating precisely a *conservativeness* condition on stipulative definitions, of which Hume's Principle is often taken to be an example.[1] In [4], Mackereth usefully distinguishes between three notions: *deductive*, *strong semantic*, and *weak semantic* conservativeness. Our aim in this note is to settle a question of Mackereth's concerning the third of these notions of conservativeness.

To this end, let us first recall the definition of weak semantic conservativeness given in [4].

Definition (Weak semantic conservativeness)**.** Let T be a theory in a language $\mathcal{L}$, and let Δ be a set of sentences in a language $\mathcal{L}' \supset \mathcal{L}$. Then Δ is *weakly semantically conservative over* T iff every $\mathcal{L}$-formula that is a consequence of $T \cup \Delta$ is already a consequence of T alone.

Note that we are concerned here with consequence in *full* (as opposed to Henkin) second-order semantics: that is, we shall be concerned only with models $\mathcal{M}$ in which n-adic second-order variables range over the full power-set $\mathcal{P}(D^n)$, where D is the domain (of individuals) of $\mathcal{M}$.

We can now state the question posed by Mackereth in [4, p. 14]:

Problem. *If* Δ_1 *and* Δ_2 *are each weakly semantically conservative over* T *(in SOL), and introduce disjoint sets of new vocabulary, does it follow that they are* jointly *weakly semantically conservative over* T*?*

[1] See, for example, Hale and Wright [3]. However, Mackereth [4, fn. 12] notes that in more recent writings, Wright may have moved away from the claim that Hume's Principle is a stipulative definition (of cardinal number).

In general (that is, for arbitrary $T, \mathcal{L}$), the answer to this question turns out to be negative. More precisely:

Theorem 1. *There is a language $\mathcal{L}$ (extending the language of pure SOL, $\mathcal{L}_{\mathrm{SOL}}$) and an $\mathcal{L}$-theory T, and sets of sentences Δ_1, Δ_2 which introduce disjoint sets of new vocabulary to $\mathcal{L}$, such that Δ_1 and Δ_2 are individually but not jointly weakly semantically conservative over T.*

Proof. Let $\mathcal{L}$ be $\mathcal{L}_{\mathrm{SOL}}$ augmented by a countable infinity of propositional constants $\{p_n : n \in \mathbb{N}\}$,[2] an individual constant z and a unary function symbol X, and let T consist of a single sentence stating that the universe, together with z and X, is isomorphic to $(\mathbb{N}, 0, S)$. Then every model $\mathcal{M}$ of T is determined (up to isomorphism) by the function $\mathrm{val}_{\mathcal{M}} : \mathbb{N} \to \{0, 1\}$ which gives the truth-value of each p_n in $\mathcal{M}$.

Now let $\mathcal{L}_1 = \mathcal{L} \cup \{Q_1\}$ and $\mathcal{L}_2 = \mathcal{L} \cup \{Q_2\}$, where Q_1, Q_2 are distinct unary predicates. We take Δ_1 to be the set of $\mathcal{L}_1$-sentences

$$\{Q_1(X^n(z)) \leftrightarrow p_n : n \in \mathbb{N}\} \cup \{\text{`the extension of } Q_1 \text{ is finite'}\}$$

and similarly take Δ_2 to be the set of $\mathcal{L}_2$-sentences

$$\{Q_2(X^n(z)) \leftrightarrow \neg p_n : n \in \mathbb{N}\} \cup \{\text{`the extension of } Q_2 \text{ is finite'}\}.^3$$

Thus $\mathcal{M}$ expands to a model of $T \cup \Delta_1$ (resp. $T \cup \Delta_2$) iff $\mathrm{val}_{\mathcal{M}}$ is eventually 0 (resp. eventually 1). Since no function $\mathbb{N} \to \{0, 1\}$ can be both eventually 0 and eventually 1, we find that $T \cup \Delta_1 \cup \Delta_2$ has no model, and therefore that $\Delta_1 \cup \Delta_2$ is not weakly semantically conservative over T.

To see that, on the other hand, each of Δ_1 and Δ_2 *is* weakly semantically conservative over T, suppose that $\mathcal{M}$ is a model of $T \cup \{\neg\phi\}$, where no p_n occurs in ϕ with $n \geq k$. Then letting val agree with $\mathrm{val}_{\mathcal{M}}$ on $\{0, \ldots, k-1\}$ and setting $\mathrm{val}(n) = 0$ (resp. $\mathrm{val}(n) = 1$) for $n \geq k$ yields a model of Δ_1 (resp. Δ_2) which (still) satisfies $T \cup \{\neg\phi\}$, as required. □

We now consider the special case where $\mathcal{L}$ is the language of pure SOL. Let us first note that, interestingly, for *finitely axiomatisable* theories T in pure SOL, there is a positive answer to Mackereth's question:

Theorem 2. *Let T be a finitely axiomatisable theory in $\mathcal{L}_{\mathrm{SOL}}$. If Δ_1 and Δ_2 introduce disjoint sets of new vocabulary and are each weakly semantically conservative over T, then they are jointly weakly semantically conservative over T.*

[2] We could instead work with a countable infinity of unary predicates P_n, and let T in addition say of (the extension of) each predicate that it is either empty or universal.

[3] Where, as usual, $X^0(z) := z$ and $X^{n+1}(z) := X(X^n(z))$.

Proof. We first observe that if Δ is weakly semantically conservative over T and σ is a categorical sentence of $\mathcal{L}_{\mathrm{SOL}}$ such that $\{\sigma\} \vDash T$, then the unique (up to isomorphism) model $\mathcal{M}$ of σ expands to a model of $T \cup \Delta$. For given any model of $T \cup \Delta \cup \{\sigma\}$ (there must be at least one: since $T \nvDash \neg\sigma$, conservativeness assures us that $T \cup \Delta \nvDash \neg\sigma$), its $\mathcal{L}_{\mathrm{SOL}}$-reduct is isomorphic to $\mathcal{M}$, and we simply push the expansion through the isomorphism.

Suppose that $\phi = \tau \wedge \psi$ is satisfiable, where τ is the conjunction of any finite axiomatisation of T. Then there is a categorical sentence σ such that $\{\sigma\} \vDash \phi$: for example, take $\sigma = \phi \wedge \forall X\big((\exists x Xx \wedge \neg X \approx U) \rightarrow \neg\phi^X\big)$,[4] where U denotes the universal concept ($Ux \leftrightarrow x = x$) and ϕ^X is ϕ with all (first-order) quantifiers relativised to X. (The models of σ are then exactly the models of ϕ of least cardinality, any two of which have equinumerous domains and so are isomorphic as models of pure SOL.) By the above observation, any model $\mathcal{M}$ of σ, and so of $T \cup \{\psi\}$, expands to a model of Δ_1 and a model of Δ_2; and since the (new) vocabularies introduced are disjoint, the expansions can be combined to give a model of $T \cup \{\psi\} \cup \Delta_1 \cup \Delta_2$. □

However, this result turns out not to hold in general – we have an analogue of Theorem 1 for the special case where $\mathcal{L} = \mathcal{L}_{\mathrm{SOL}}$.

Theorem 3. *There is a theory T in $\mathcal{L}_{\mathrm{SOL}}$, and sets of sentences Δ_1, Δ_2 which introduce disjoint sets of new vocabulary, such that Δ_1 and Δ_2 are individually but not jointly weakly semantically conservative over T.*

Proof. (I am grateful to Stephen Mackereth for supplying the present argument.) As observed in [1, p. 25], there are theories T in pure SOL which are semantically complete (for every $\mathcal{L}_{\mathrm{SOL}}$-sentence ϕ, $T \vDash \phi$ or $T \vDash \neg\phi$) but not categorical.[5] Let T be any such theory. It then follows from completeness that weak semantic conservativeness over T reduces to joint satisfiability with T. Let κ be the smallest cardinality of a model of T. (Note that κ is infinite, since finite cardinalities are characterisable.) Since T is not categorical, it must also have models of cardinality $> \kappa$. So, it suffices to find sets Δ_1, Δ_2 such that $T \cup \Delta_1$ is satisfiable and only has models of cardinality κ, while $T \cup \Delta_2$ is satisfiable and only has models of cardinality $> \kappa$.

To this end, we adapt a construction in [4, fn. 16]. Let Sat' be the third-order relation which holds of (B, N, X, V, Z, u) iff B is an infinite concept,

[4] '$X \approx Y$' abbreviates 'there is a bijection between X and Y'.

[5] This follows from a quick cardinality argument. Since $\mathcal{L}_{\mathrm{SOL}}$ is countable, there are at most $2^{\aleph_0}$ complete $\mathcal{L}_{\mathrm{SOL}}$-theories. But for each nonzero cardinal κ, the theory $\mathrm{Th}(\kappa)$ (:= the set of $\mathcal{L}_{\mathrm{SOL}}$-sentences true in a domain of size κ) is complete, so we must have $\mathrm{Th}(\kappa) = \mathrm{Th}(\lambda)$ for some distinct κ, λ.

N, X, V, Z are relations on B, (N, X) is a copy of $(\mathbb{N}, S)$, V codes finite sequences of objects in B as objects in B, and under this coding u is the Gödel number of an $\mathcal{L}_{\mathrm{SOL}}$-formula satisfied (in the domain B) by Z. Given a new third-order constant $\mathfrak{S}$ (of appropriate type), Craig's construction in [2] (relativised as in [4, fn. 16]) yields a formula $\psi(\mathfrak{S})$ which says that $\mathfrak{S}$ is Sat'. Now let $C(\mathfrak{S})$ be the condition '$\neg\exists B$ ($B \not\approx U$ and for every $\mathcal{L}_{\mathrm{SOL}}$-sentence ϕ, $B \vDash \phi$ iff $U \vDash \phi$)', where '$B \vDash \phi$' is expressed by means of $\mathfrak{S}$ in the obvious way. We take distinct third-order constants $\mathfrak{S}_1, \mathfrak{S}_2$ and put $D_1 := \psi(\mathfrak{S}_1) \wedge C(\mathfrak{S}_1)$, $D_2 := \psi(\mathfrak{S}_2) \wedge \neg C(\mathfrak{S}_2)$. Then a model of T expands to a model of D_1 iff it is of least cardinality κ, and expands to a model of D_2 iff it is of cardinality $> \kappa$. Therefore, as required, each of $\Delta_1 := \{D_1\}$ and $\Delta_2 := \{D_2\}$ is jointly satisfiable with T while $\Delta_1 \cup \Delta_2$ is not.[6] □

References

[1] S. Awodey and E.H. Reck (2002), Completeness and categoricity, part I: Nineteenth-century axiomatics to twentieth-century metalogic, *History and Philosophy of Logic* 23(1): 1–30.

[2] W. Craig (1965), Satisfaction for n-th order languages defined in n-th order languages, *Journal of Symbolic Logic* 30(1): 13–25.

[3] B. Hale and C. Wright (2001), *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*, Oxford: Clarendon Press.

[4] S. Mackereth (2026), Neo-logicism and conservativeness, *Journal of Philosophy* 123(1): 5–34.

[6] Thanks to Stephen Mackereth not only for supplying this proof of Theorem 3, but also for the reference to Awodey and Reck [1] and for suggestions which improved the presentation of our proofs. Thanks also to Claude (Opus 5.5) for assistance with proof-reading and for pointing me to the present strengthening of an old version of Theorem 2 (which had T empty in place of T finitely axiomatisable).