The Inexpressibility of Validity

Julien Murzi

August 30, 2013

Tarski’s Undefinability of Truth Theorem comes in two versions: that no consistent theory which interprets Robinson’s Arithmetic (Q) can prove all instances of the T-Scheme

\[(T) \quad \text{Tr}(\langle \phi \rangle) \leftrightarrow \phi,\]

and hence define truth; and that no such theory, if sound, can even express truth.\(^1\) In this note, I prove corresponding limitative results for validity. While Peano Arithmetic already has the resources to define a predicate expressing logical validity, as Jeff Ketland (2012) has recently pointed out, no theory which interprets Q closed under the standard structural rules can define nor express validity, on pain of triviality. The results put pressure on the widespread view that there is an asymmetry between truth and validity, viz. that while the former cannot be defined within the language, the latter can.\(^2\) I argue that Vann McGee’s and Hartry Field’s arguments for the asymmetry view are problematic.

\(^{1}\)Definitions of definability and expressibility will be provided in due course. For the time being, suffices to say that a theory defines truth iff it proves every instance of T, and that it expresses truth iff every such instance is true.

\(^{2}\)See McGee (1991, pp. 44-5) and Field (2008). Field writes that while “[t]ruth can’t possibly be given an extensionally correct definition within the language, ... validity presumably can” (Field, 2008, p. 127). What he means by that is that one can define the set of validities, without necessarily thereby defining the concept of validity. Kreisel’s Squeezing Argument may then reassure us that standard model-theoretic definitions of first-order logical validity are extensionally adequate (Kreisel,
1 Logical validity and validity

We’re all familiar with the standard Tarskian definition of logical validity:

(Tarski) $\phi$ is logically valid iff $\phi$ is true under all reinterpretations of its non-logical constituents.

Now let $\text{PA}$ be an axiomatisation of Peano Arithmetic, with language $\mathcal{L}_A$. Idealising quite a bit, we may take $\text{PA}$ to be a (simplified) model of our maximal theory at a certain time, i.e. a model of the closure of the set of sentences we (at least) implicitly accept at a certain time under the rules we (at least) implicitly accept at that time (Field, 1994, p. 401-5). We can then put our model to work, and ask: what can it say about logical validity?

Jeff Ketland (2012) has recently pointed out that $\text{PA}$ already has the resources to derive the following principles about a predicate $\text{Val}(x)$ intuitively reading ‘$x$ is logically valid’ (so that $\text{Val}(n)$ is true iff $n$ is the code of a valid $\mathcal{L}_A$-sentence):

1. **(V-Intro)** Given a logical derivation of $\phi$, infer $\text{Val}(\neg \phi)$;
2. **(V-Out)** $\text{Val}(\neg \phi)$ $\rightarrow$ $\phi$;
3. **(V-Imp)** $\text{Val}(\phi \rightarrow \psi)$ $\rightarrow$ ($\text{Val}(\phi)$ $\rightarrow$ $\text{Val}(\psi)$).

The above principles can be easily generalised by means of a two-place predicate $\text{Val}(x, y)$ expressing argument-validity, where $x$ is to be replaced by (the code of) the conjunction of the premises of a given argument. Not only can $\text{PA}$, and hence our maximal theory, talk about numbers (and its own syntax): it can also talk, and prove intuitive principles, about logical validity.

To be sure, the extension of ‘logically valid’ will depend on one’s view about what counts as logical. While arguably $\wedge$, $\lor$, $\rightarrow$, $\neg$, $\exists$, $\forall$ and $=$ are logical and terms such as ‘water’ aren’t, the status of semantic predicates such as ‘true’ and ‘valid’ is less clear. According to Graham Priest (2007, p. 193), the T-Scheme (T) “ought ... to be considered part of logic”, essentially on the grounds that its instances are analytic of ‘true’. The view is controversial, however. As Roy Cook (2012, p. 235) 1967); see also Field (2008, Ch. 2.2). McGee’s and Field’s view express current logical orthodoxy, although the unorthodox view defended here, viz. that truth and validity are both paradox-prone semantic notions, is by no means new. For a representative sample of the literature, see e.g. Priest and Routley (1982), Read (1979), Read (2001) and, more recently, Shapiro (2011) and Beall and Murzi (2013).

3. Logical derivation’ here means ‘derivable in the logic of $\text{PA}$’, viz. classical first-order logic.
4. Here and throughout, $\neg \phi$ indicates the numerical code of $\phi$, relative to some fixed coding scheme.
has recently observed, if “logical truth requires that uniform substitution instances of logical truths [be] logical truths”, some instances of (T) are not logically valid.

The issue, though, is largely terminological. Thus, McGee recommends that we use a broader notion of logical validity, while acknowledging that the notion doesn’t coincide with pure logical validity:

we must employ a richer notion of logical necessity, according to which there are certain sentences whose truth is so basic to our way of thinking and talking that they have the same epistemic status as logical validities, even though they are not actually logically valid. (McGee, 1991, p. 43)

In a similar spirit, Field suggests that we use “the term ‘valid’ ... in a very broad sense, one which counts ... a large amount of set theory and the basic principles of truth and satisfaction as valid” (Field, 2007, p. 99). I’ll distinguish, then, between logical validity and validity (tout court). I mention two reasons why the distinction must be made.

First, it might be argued that there are clear examples of arguments that are valid, albeit not logically so. A first (admittedly controversial) example is given by the ω-rule:

0 has property F
1 has property F,
2 has property F,
...

Every number has property F.

Many would think that the rule is intuitively valid. Yet, the rule is invalid in first-order logic. Less controversially, other examples of valid but not logically valid arguments include analytic validities such as the following:

---

6First-order logic is compact: an argument is valid in first-order logic if and only if some finite sub-argument is valid.
7There at least two ways of validating the ω-rule. First, one could allow infinitary quantifiers, such as “there are infinitely many x such that ...”. Then, \{F(0), F(1), F(2), ...\} would imply that there are infinitely many Fs. The resulting system would be semantically incomplete, but it might be thought that this is not a problem. Relatedly, one could simply point out that the ω-rule is valid in second-order logic (with standard semantics). Here I don’t have space to discuss either option, and simply observe that both are themselves controversial. For a criticism of infinitary quantifiers, see Hanson (1997, pp. 391-2).
\[
\frac{x \text{ is a brother}}{x \text{ is male}} \quad \Phi(0) \quad \forall n(\Phi(n) \rightarrow \Phi(S(n)))),
\]

where \(S(x)\) expresses the successor function. To be sure, one might point out that such rules would be logically valid if we held fixed the interpretation of 'brother', 'male', 'successor' and numerals. However, the dialectic here is a familiar one: intuitively invalid inferences such as \langle \text{Leslie was a US president} \therefore \text{Leslie was a man} \rangle, and intuitively invalid sentences such as 'There are at least two numbers', would thereby be declared logically valid (Etchemendy, 1990).\(^8\) Whether rules such as

\[
\frac{x \text{ is water}}{x \text{ is } H_2O}
\]

are valid is perhaps less clear, but the examples, I hope, suffice to establish my point: logical validity is one kind of validity (see also Priest, 2006a).\(^9\)

Second, as hinted in the above McGee quote, validity arguably has a role to play in our epistemic lives. Gil Harman (1986, p. 18) suggests that ordinary reasoning is partly governed by the following principles:

Recognised Implication Principle. One has a reason to believe \(\phi\) if one recognises that \(\phi\) is implied by one’s view.

Recognised Inconsistency Principle. One has a reason to avoid believing things one recognises to be inconsistent.

Similarly Field advocates the existence of a connection between validity and correct reasoning, this time framed in terms of degrees of belief. Where \(P(\phi)\) refers to one’s degrees of belief in \(\phi\), Field’s principle reads:

\[
\text{(F) If it’s obvious that } \phi_1, ..., \phi_n \text{ together entail } \psi, \text{ then one ought to impose the constraint that } P(\psi) \text{ is to be at least } P(\phi_1) + ... + P(\phi_n) - (n - 1), \text{ in any circumstance where } \phi_1, ..., \phi_n \text{ are in question.} \quad \text{(Field, 2009, p. 259)}
\]

\(^8\)Again, I don’t have space here to fully defend the claim that Tarski’s account of validity cannot handle analytic validities. This is a large issue related to the even larger issue whether, as forcefully argued by John Etchemendy (1990, 2008), Tarski’s account undergenerates. For more discussion on analytic validity and the issue of under generation, see e.g. Priest (1995, p. 288) and Etchemendy (2008, p. 278 and ff).

\(^9\)For more on the slide between the foregoing examples and clear cases of logical validity, see e.g. Lycan (1989) and Sagi (2013).
Roughly: if $\phi$ entails $\psi$, then one’s degree of belief in $\psi$ should be no lower than one’s degree of belief in $\phi$.

Both Harman’s and Field’s principles are, I think, plausible. But neither principle is especially concerned with logical validity. Plainly, ordinary speakers typically don’t distinguish between valid sentences and logically valid sentences (Harman, 1986, p. 17). Hence, if the principles apply at all, they apply to validity. As Harman puts it, “since there seems to be nothing special about logical implications and inconsistencies, ... there seems to be no significant way in which logic might be specially relevant to reasoning” (Harman, 2009, p. 334). Logical validity is not specially relevant to reasoning. But validity arguably is.

What, then, is validity? One standard view sees consequence as a kind of modality. For some understanding of ‘possible’, a set of premises entails a given conclusion iff it is impossible that all the premises are true and the conclusion false:

*Truth-preservation view.* An argument $\langle \Gamma : \beta \rangle$ is valid iff it is impossible that every member of $\Gamma$ is true but $\beta$ is false.

The view accommodates the examples of pp. 3-4, but has recently come under attack, chiefly on the grounds that the very claim that valid arguments preserve truth (i) entails Curry-driven triviality and (ii) is inconsistent with Gödel’s Second Incompleteness Theorem (Priest, 2006a,b; Field, 2008, 2009; Beall, 2009)\textsuperscript{10}. However, the results to be presented below don’t require that valid arguments be truth-preserving—they are indeed compatible with views that reject this assumption. One such view is endorsed by Field (2008, 2009). According to Field, validity is a primitive notion whose role is to constrain belief, and degrees of belief, via principles such as ($F$):

*Normative view.* validity normatively constrains our degrees of belief.

Now for some well-known, and less well-known, theorems.

### 2 Validity Curry

While $V$-Intro is adequate for a predicate expressing logical validity, it doesn’t seem adequate for a predicate expressing validity. Suppose $\phi$ has been derived by means

\textsuperscript{10}For a critical discussion of Field’s and Beall’s arguments, see Shapiro (2011) and Murzi and Shapiro (forthcoming).
of rules that are themselves valid. Then, we would like to be able to say that \( \phi \) is itself valid. That is, the following rule should be in place:

\[
(V\text{-Intro}^*) \text{ Given a valid derivation of } \phi, \text{ infer } \text{Val}(\lnot \phi).\]

The rule is intuitively valid. On the truth-preservation view, it states that sentences proved only by means of necessarily truth-preserving rules are themselves necessary. On the normative view, it effectively says that, if we have proved \( \phi \) by means of valid rules, then, if we have thereby come to believe \( \phi \) to degree 1, then we should also believe \( \text{Val}(\lnot \phi) \) to degree 1. Unlike \( V\text{-Intro} \) and \( V\text{-Out} \), however, \( V\text{-Intro}^* \) and \( V\text{-Out} \) (with ‘\( \text{Val}(x) \)’ now taken to express validity) spell trouble.

**Theorem 1** (Myhill 1960, Kaplan and Montague 1960). Let \( T \) be a theory which (i) interprets \( Q \), (ii) proves all instances of \( V\text{-Out} \) and (iii) is closed under \( V\text{-Intro}^* \). Then, \( T \) is inconsistent.

**Proof.** We reason in \( T \). The Diagonal Lemma yields a sentence \( \kappa \) such that \( \kappa \iff \neg \text{Val}(\lnot \kappa) \). We assume \( \text{Val}(\lnot \kappa) \), and derive \( \neg \text{Val}(\lnot \kappa) \) by applications of \( V\text{-Out} \) and \( \leftrightarrow\text{-E} \). Hence, \( \neg \text{Val}(\lnot \kappa) \). But, then, \( \kappa \) follows on no assumptions, and must therefore be valid. Contradiction.\(^{12}\)

In effect, the above reasoning—the Knower Paradox—is but a stronger Liar: it is the same paradox, except that the left-to-right direction of \( (T) \) is replaced by the weaker \( V\text{-Intro}^* \).\(^{13}\) Standard treatments of the Liar paradox still apply, however. On paracomplete treatments, one may not validly infer \( \neg \text{Val}(\lnot \kappa) \) from a derivation of absurdity from \( \text{Val}(\lnot \kappa) \), i.e. the rule of negation introduction is rejected as (logically) invalid (see e.g. Kripke, 1975; Brady, 2006; Field, 2008; Horsten, 2009). On paraconsistent treatments, the above argument shows the existence of a true contradiction, viz. \( \text{Val}(\lnot \kappa) \land \neg \text{Val}(\lnot \kappa) \) (see e.g. Priest, 2006b; Beall, 2009).

\(^{11}\)This is, in effect, a notational variant of the unrestricted Rule of Necessitation: given a derivation of \( \phi \), infer \( \text{Tr}(\lnot \phi) \).

\(^{12}\)The reasoning displayed in this proof is sometimes referred to as the Pseudo-Scotus Paradox. It might be objected that \( Q \)’s axioms and \( V\text{-Intro}^* \) are logically invalid. See e.g. Field (2008, p. 304 and p. 306), Ketland (2012) and Cook (2013). However, the assumption that they are valid is plausible on any interesting weakening of the notion of logical validity. For more discussion, see Murzi and Shapiro (forthcoming).

\(^{13}\)To see this, note that both paradoxes have the same general form: from some \( \phi \) such that \( \vdash \phi \iff \lnot \Box(\lnot \phi) \), one proves \( \lnot \Box(\lnot \phi) \), and hence \( \phi \), on the assumption that \( \Box(\lnot \phi) \). Then, \( \Box(\lnot \phi) \) follows on no assumptions via either the T-Scheme or Necessitation, depending on whether \( \Box \) is interpreted as, respectively, ‘true’ or ‘valid’.
Now suppose we’d like to attribute validity not only to sentences, but also to arguments. We’d then introduce in the language (at least) a two-place validity predicate \( \text{Val}(x, y) \) governed by the following rules:

\[
\begin{align*}
\text{(VP)} & \quad \text{Given a valid derivation of } \psi \text{ from } \phi, \text{ infer } \text{Val}(\neg \phi \land \neg \psi).\\
\text{(VD)} & \quad \text{From } \phi \text{ and } \text{Val}(\neg \phi \land \neg \psi), \text{ infer } \psi.
\end{align*}
\]

More formally:

\[
\begin{align*}
\text{(VP)} & \quad \frac{\phi \vdash \psi}{\Gamma \vdash \text{Val}(\neg \phi \land \neg \psi)} \\
\text{(VD)} & \quad \frac{\Gamma \vdash \text{Val}(\neg \phi \land \neg \psi), \Delta \vdash \phi}{\Gamma, \Delta \vdash \psi}.
\end{align*}
\]

VP is, again, intuitively valid. On the truth-preservation view, it states that, if \( \psi \) has been derived from \( \phi \) by means of truth-preserving rules, then the argument \( \langle \phi : \psi \rangle \) is necessarily truth-preserving. On the normative view, it effectively says that, if \( \psi \) has been derived from \( \phi \) by means of valid rules, then the argument \( \langle \phi : \psi \rangle \) is valid, and hence \( P(\phi) \leq P(\psi) \). VD is also compelling on both views. On the truth-preservation view, it implies that, if we’re in a position to assert that the argument \( \langle \phi : \psi \rangle \) is truth-preserving, then we may infer \( \psi \) given \( \phi \). On the normative view, it tells us that, if we’re in a position to assert that the argument \( \langle \phi : \psi \rangle \) is validity, then \( P(\phi) \leq P(\psi) \).

Note that VP and VD are generalisations of, respectively, V-Intro* and V-Out. To see this, it is sufficient to instantiate VP and VD using a constant \( T \) expressing valid truth. Instantiating VP yields a notational variant of V-Intro*, rewritten using our two place predicate \( \text{Val}(x, y) \) in place of a one-place validity predicate \( \text{Val}(x) \):

\[
\text{(V-Intro**)} \quad \frac{T \vdash \psi}{\Gamma \vdash \text{Val}(\neg T \land \neg \psi)}.
\]

Likewise, instantiating VD thus

\[
\frac{\Gamma \vdash \text{Val}(\neg T \land \neg \psi), T \vdash T}{\Gamma, T \vdash \psi}
\]

yields a notational variant of a rule corresponding to the \( T \) axiom for a necessity operator, i.e. V-Out*:

\[
\text{(V-Out*)} \quad \text{Val}(\neg T \land \neg \psi), T \vdash \psi.
\]

It is therefore no surprise that VP and VD also spell trouble. What is surprising is the minimal logical resources required to show this: one only needs to assume the validity of the standardly accepted structural rules, viz. Identity, Contraction and Cut:
The proof is a validity-involving version of Curry’s Paradox: Validity Curry, or v-Curry, for short.¹⁴

**Theorem 2.** Let T be any theory which interprets Q and is closed under VP, VD, Id and SContr. Then, T is trivial.¹⁵

**Proof.** The Diagonal Lemma yields a sentence π, which intuitively says of itself, up to equivalence, that it validly entails that you will win the lottery:

\[ \vdash_T \pi \leftrightarrow \text{Val}(\neg \neg \pi, \bot) \]

Let Σ now be the following derivation of the further theorem Val(\neg \neg \pi, \bot):

\[ \pi \vdash_T \pi \vdash_T \pi \leftrightarrow \text{Val}(\neg \neg \pi, \bot) \]
\[ \pi \vdash_T \text{Val}(\neg \neg \pi, \bot) \]
\[ \vdash \pi \vdash_T \bot \quad \text{SContr} \]
\[ \vdash \pi \vdash_T \bot \quad \text{VD} \]
\[ \vdash \pi, \pi \vdash_T \bot \quad \text{VD} \]
\[ \vdash \pi \vdash_T \text{Val}(\neg \neg \pi, \bot) \]

Using Σ, we can then ‘prove’ that you will win the lottery:

\[ \Sigma \quad \vdash_T \pi \leftrightarrow \text{Val}(\neg \neg \pi, \bot) \]
\[ \vdash_T \text{Val}(\neg \neg \pi, \bot) \]
\[ \vdash_T \pi \quad \text{VD} \]
\[ \vdash_T \bot \]

In short: validity, just like truth, is plagued by paradox. More precisely, since VD and VD are generalisations of the validity rules that yield the Knower Paradox, the v-Curry Paradox is a generalisation of the Knower Paradox, which, recall, is but a stronger Liar.

**Parenthetical note.** Theorem 2 can be strengthened. While the theorem assumes that Q be valid in the target sense, this assumption is not strictly needed, as Cook (2013) has recently pointed out.


¹⁵It is worth noting that Cut is effectively built in our formulation of VD. Hence, should the latter rule be formulated differently, Cut may need to be mentioned alongside Id and SContr in the statement of Theorem 2 below.
Theorem 3 (Cook). If VP and VD are valid, Q is trivial.

Proof. Let Q* be the conjunction of Q’s axioms. Using the Diagonal Lemma, we can derive

\[ Q^* \vdash \kappa \leftrightarrow \text{Val}(\kappa \land Q^* \rightarrow \bot) \]

We then ‘prove’ \( \text{Val}(\kappa \land Q^* \rightarrow \bot) \) and, from this, conclude \( Q^* \vdash \bot \) \hfill \Box

Beyond Id, SContr and Cut,\(^{16}\) all is needed for deriving the Validity Curry is that VP and VD be valid. It immediately follows from Theorem 2 and 3 that theories of semantic paradox which retain the standard structural rules—including the standard paracomplete and paraconsistent ones—still validate the \( \nu \)-Curry Paradox. Such theories can in general express sentence validity, but they’re unable to express argument validity. For more discussion, see Zardini (2011), Beall and Murzi (2013) and Murzi (2012). End note. **

3 Validity and expressibility

The results of §2 suggest natural generalisations of Tarski’s undefinability and inexpressibility of truth theorems. I’ll closely follow, mutatis mutandis, Peter Smith’s presentation of Tarski’s original results (Smith, 2007, pp. 180-2). As above, we’ll take some (possibly not recursively axiomatisable) theory \( T \) which interprets PA to represent our maximal theory at a certain time, and hence the consequence relation of our language at that time. That is, ignoring for simplicity’s sake the relativisation to time, we’ll effectively assume that \( T \)’s consequence relation is the consequence relation of our maximal theory, so that \( T \) is, so to speak, validity sound and complete. In short: \( \phi \vdash_T \psi \) iff \( \phi \) entails \( \psi \).

We can now show that \( T \) can neither define, nor express, validity. To this end, we make use of the two following (standard) definitions:

Definition 4 (Definability). \( \Phi(x,y) \) defines a binary relation \( R \) in \( T \) iff: \( mRn \) iff \( \vdash_T \Phi(m,n) \), where \( m \) and \( n \) are names of, respectively, \( m \) and \( m^{\prime} \).\(^{17}\)

Definition 5 (Expressibility). \( \Phi(x,y) \) expresses a binary relation \( R \) iff: \( mRn \) iff \( \Phi(m,n) \) is true.

---

\(^{16}\)This last rule is in effect built into VD. See infra fn. 15.

\(^{17}\)Definability, as defined in Definition 4, is sometimes also referred to as weak representability.
Now let Valid be a numerical relation such that Valid\((m, n)\) is true iff \(m\) and \(n\) are the Gödel numbers of formulae \(\phi\), \(\psi\) of \(L_A\) such that \(\phi\) entails \(\psi\). Suppose an open wff Val\((x, y)\) belonging to an arithmetical language \(L'_A\) including \(L_A\) expresses this numerical relation Valid. Then, for any \(L_A\) sentences \(\phi, \psi, \phi\) entails \(\psi\) iff Valid\((\phi, \psi)\) is true. That is, for any \(L_A\) sentences \(\phi, \psi, \phi\) entails \(\psi\) iff Val\((\phi, \psi)\) is true. In turn, this motivates the following definitions.

**Definition 6** (Formal validity predicate). An open \(L'_A\)-wff Valid\((x, y)\) is a formal validity-predicate for \(L_A\) iff Valid\((x, y)\) expresses validity, i.e. iff for every \(L_A\)-sentence \(\phi, \psi, \phi\) entails \(\psi\) iff Val\((\phi, \psi)\) is true.

**Definition 7** (Validity theory). A theory \(T\) (with language \(L'_A\) which includes \(L_A\)) is a validity theory for \(L_A\) iff, for some \(L'_A\)-wff Val\((x, y)\), \(\phi\) entails \(\psi\) iff \(\vdash_T\) Val\((\phi, \psi)\). A validity theory for \(L_A\) is also a validity definition for \(L_A\).

Note that, given our idealising assumption that \(T\) is validity sound and complete, it follows from Definitions 5 and 6 that Val\((x, y)\) is a formal validity predicate iff Val\((x, y)\) expresses \(T\)’s consequence relation. More formally, Val\((x, y)\) is a formal validity predicate iff: Val\((\phi, \psi)\) is true iff \(\phi \vdash_T \psi\). Furthermore, on the foregoing assumption, \(T\) is a validity theory iff the following holds: \(\phi \vdash_T \psi\) iff \(\vdash_T\) Val\((\phi, \psi)\). That is, if \(T\) is validity sound and complete, \(T\) is a validity theory iff it is closed under the naïve validity rules:

\[
\frac{\phi \vdash_T \psi}{\vdash_T \text{Val}(\phi, \psi)} \quad \frac{\Gamma \vdash_T \text{Val}(\phi, \psi)}{\Gamma, \Delta \vdash_T \phi}.
\]

We’re now ready to introduce our main results.

Can \(T\) define its own consequence relation, i.e. can \(T\) contain its own validity definition? This question is answered in the negative by the following theorem.

**Theorem 8** (Indefinability of \(\vdash_T\) in \(T\)). Let \(T\) be any non-trivial theory which interprets \(Q\) with language \(L_V\) including a fresh predicate Val\((x, y)\). Then, \(T\) cannot define its own consequence relation.

**Proof.** Since \(T\) is at least as strong as \(Q\), the Diagonal Lemma applies, and gives us a sentence \(\pi\)—a v-Curry sentence!—such that

\[
\vdash_T \pi \iff \text{Val}(\pi, \bot).
\]

However, since \(T\) defines its own consequence relation, \(T\) validates \(\pi\) and \(\bot\). Hence, we can run the v-Curry reasoning in \(T\), and conclude \(\vdash_T \bot\), which contradicts our assumption that \(T\) isn’t trivial.

\(\square\)
It immediately follows that, if $T$ is validity sound and complete, then $T$ cannot define validity.

**Corollary 9** (Indefinability of validity). *Let $T$ be as above. Then, if $T$ is validity sound and complete, $T$ cannot define validity.*

**Proof.** This follows at once from Theorem 8 and the observation that, if $T$ is validity sound and complete, $T$ defines validity iff it validates VP and VD. □

Can at least $T$ express validity? This question must also receive a negative answer, provided Q is sound. The proof makes use in the metalanguage of disquotational truth-predicate, i.e. we assume that, for all $\phi \in L_V$, $\phi$ is true iff $\phi$, where ‘true’ $\not\in L_V$. It also assumes that our meta-theory validates the standard structural rules, and that the metalanguage entailment connective we’ve been using so far satisfies (a version of) conditional proof and modus ponens, i.e. that, for all $\phi \in L_V$, if there is a derivation of $\psi$ from $\phi$ in the metatheory, then $\phi$ entails $\psi$, and if $\phi$ and $\phi$ entails $\psi$, then $\psi$, where ‘entails’ $\not\in L_V$.

**Theorem 10** (Inexpressibility of validity). *Let $T$ be any non-trivial theory which interprets Q, with language $L_V$. Then, if Q is sound, T cannot contain a predicate expressing validity.*

**Proof.** Let $T$ be a theory that interprets Q, and suppose there is a $L_V$ predicate $\text{Val}(x, y)$ expressing validity. We then have:

1. $\vdash_T \pi \leftrightarrow \text{Val}(\uparrow \pi \uparrow, \uparrow \bot \uparrow)$.

However, (the subtheory of $T$ which interprets) Q is ex hypothesi sound. So

2. $\pi \leftrightarrow \text{Val}(\uparrow \pi \uparrow, \uparrow \bot \uparrow)$.

Since $\text{Val}(x, y)$ expresses validity, the following holds:

3. $\pi$ entails $\bot$ iff $\text{Val}(\uparrow \pi \uparrow, \uparrow \bot \uparrow)$ is true.

We can now reason the standard Curry way. We assume $\pi$, and derive that $\pi$ entails $\bot$ from 2. and 3. Assuming again $\pi$, we derive $\bot$ and, discharging both occurrences of $\pi$, we deduce that $\pi$ entails $\bot$. From 2. and 3., we conclude $\bot$. □

---

18While I’m assuming the metalanguage to be classical, and hence take the truth-predicate to be typed, the proof would also go through using a type-free truth-predicate in a non-classical metalanguage.
Notice that the multiple discharge of \( \pi \) is effectively equivalent to assuming that SContr holds (see e.g. Negri and von Plato, 2001, Ch. 8).

We’re now confronted with a familiar dilemma: either we invalidate one of the naïve validity principles at work in the foregoing proofs, or, if \( T \) is to either define or express validity, \( T \)’s logic must be (radically) weakened: one among Id, SContr and Cut must go.

4 Incompleteness and classical validity

It might be objected that the foregoing results are inconsistent with both Gödel’s Second Incompleteness Theorem and Löb’s Theorem and that, for this reasons, our definitions of a formal validity predicate and of a validity theory must be incorrect. The Second Incompleteness Theorem states that no consistent recursively axiomatised theory \( T \) which interprets \( \mathbb{Q} \) and is strong enough to prove the Hilbert-Bernays conditions for a predicate \( \text{Prov}_T \) expressing provability-in-\( T \) can prove its own consistency.\(^{19}\) Löb’s Theorem is a formalised version of Curry’s Paradox, which, as it turns out, entails the Second Incompleteness Theorem. Where \( T \) is as above, Löb’s Theorem states that, if \( \vdash_T \text{Prov}_T(\neg \phi) \rightarrow \phi \), then \( \vdash_T \phi \).

In a nutshell, one might use Gödel’s and Löb’s results to argue against VD thus. Consider the following version of the Second Incompleteness Theorem.

**Theorem 11.** Let \( T \) be any recursively axiomatised theory that interprets \( \mathbb{Q} \), and let \( \text{Prov}_T(x) \) be a provability predicate for \( T \). Suppose \( T \) asserts that theorems of \( T \) are valid and that validity is factive:

1. \( \text{Prov}_T(\neg \phi) \rightarrow \text{Val}(\neg \phi) \);
2. \( \text{Val}(\neg \phi) \rightarrow \phi \).

It follows by Löb’s Theorem that \( T \) is trivial.

**Proof.** Suppose \( T \) derives both 1. and 2. Then, since \( \text{Prov}_T(\neg \phi) \rightarrow \phi \) follows from 1. and 2. by the transitivity of \( \rightarrow \), \( T \) also derives \( \text{Prov}_T(\neg \phi) \rightarrow \phi \). By Löb’s Theorem, \( T \) derives \( \phi \). ☐

That this is in effect the Second Incompleteness Theorem can be seen by contraposing the proof while setting \( \phi := 0 \neq 1 \). We can now derive the following corollary.

---

\(^{19}\)The derivability-conditions are the predicate-analogues of the Rule of Necessitation (see infra, fn. 11) and of the K and \( 4 \) axioms of modal logic.
**Corollary 12.** Let $T$ be as above and suppose $T$ is closed under $VD$. Then, $T$ is trivial.

**Proof.** One need only notice that $VD$ entails $V$-Out (see §2). Then, the result immediately follows from Theorem 10. \qed

Have we shown that $VD$ must be rejected? I think not, but let’s first examine the main assumption on which the arguments depends.

The argument assumes that our theory of validity $T$ is **recursively axiomatisable**. McGee defends the assumption thus:

[i]f ... we identify the [validities] as those statements from whose denials one can derive a contradiction, and if we take derivability to be provability is some explicitly describable system of rules, it will follow that the set of necessary truths is a recursively enumerable set. (McGee, 1991, pp. 44-5)

In short: if $\text{Val}(\varphi)$ is true iff $\neg\varphi \vdash T \bot$, for some recursively axiomatisable $T$, then the set of validities is recursively enumerable. It then follows from Corollary 12 that $VD$ must fail. Moreover, as McGee observes, “if the set [of validities] is recursively enumerable ... [then we can] explicitly define it” (McGee, 1991, p. 45). That is, there is, in McGee’s view, a fundamental **asymmetry** between truth and validity: unlike truth, validity is definable in the language, and “[w]e must confront the paradox [the Knower Paradox] directly; there is no escaping in the metalanguage” (McGee, 1991, p. 45). Yet, it is doubtful that the claim that validity is recursively axiomatisable can be made plausible in the present context. For one thing, the claim is not available to someone who accepts the validity of arguments with infinitely many premises, such as the example of §2.\footnote{Incidentally, McGee concedes that instances of the omega-rule are at least informally valid (McGee, 2011, p. 33).} For another, McGee’s argument for the recursive axiomatisability of validity fails to convince. The argument assumes that, if $\phi$ is valid, then we can derive a contradiction from $\neg\phi$. But since to derive $\bot$ from $\neg\phi$ is to (classically) prove $\phi$, this is tantamount to assuming a form of verificationism, to the effect that validities must have proofs of a certain kind. This is implausible, however.

To begin with, Goldbach’s Conjecture could be true but (absolutely) unprovable. It would then follow that, since one wouldn’t be able to derive $\bot$ from the negation of Goldbach’s Conjecture, Goldbach’s Conjecture, unlike, say, Fermat’s Theorem, or some other provable arithmetical theorem, would be invalid in the target sense. This
seems counterintuitive, though. Since both Goldbach’s Conjecture and Fermat’s Theorem are arithmetical sentences, it would seem that either they are both valid, or neither is.\textsuperscript{21} Second, even if all mathematical truths are absolutely provable, the set of absolutely provable sentences would still not be guaranteed to be recursively enumerable. For example, perhaps every mathematical truth is absolutely provable by some possible finite being or other, even if no possible finite being can give absolute proofs of all mathematical truths.\textsuperscript{22}

In any event, the claim that validity is recursively axiomatisable is itself problematic. The claim implies that $\text{Val}(x)$ can be interpreted as $\text{Prov}_T(x)$, which in turn implies, because of Löb’s Theorem, that our theory of validity cannot assert that valid sentences are true, on pain of triviality. Yet we would seem to know that validity is a mode of truth. As William Reinhardt puts it in the case of knowledge, “we know that what is known is true, but the reason for this is that if it were not so we would not call it known” (Reinhardt, 1986, pp. 468-9). The point, I take it, carries over to validity.

McGee (1991, pp. 45-9) confronts this second issue directly. While discussing what he calls logical necessity, i.e. what we are simply calling validity, he concedes that the principles which generate the Knower Paradox are “intuitively obvious”. In particular, he acknowledges that it is natural to think, of someone who denies the necessity of all instances of the schema ‘If $\phi$ is necessary, then $\phi$’ (i.e. $V$-Out), that he “would not know what he was talking about[:] even though he uses the word ‘necessary’, he could not be talking about necessity; he must be using the word in a deviant way” (p. 48). Still, McGee’s recommendation is that we substitute $V$-Out with the weaker, but consistent

$$(L) \text{Val}(\overline{\text{Val}(\overline{\phi}) \rightarrow \phi}) \rightarrow \text{Val}(\overline{\phi}),$$

which, he claims, “our dogmatic insistence upon $V$-Out has obscured to view” (p. 49). Validity can be seen to be non-paradoxical, McGee suggests, provided we give up some “intuitively obvious” principles about validity, such as the truism that valid sentences are true.

This reasoning also fails to convince. In effect, McGee is here merging together two different issues: the question whether validity is paradoxical, and the problem how the validity paradoxes should be solved. While it may be that $V$-Out has

\textsuperscript{21}Admittedly, this objection is more plausible when validity is equated with necessary preservation of truth. Thanks to Jack Woods for helping me appreciate this point.

\textsuperscript{22}Thanks to an anonymous referee for suggesting this point.
obscured (L) to view, the point remains that V-Out is intuitively obvious: more obvious, I submit, than the assumption that validity is recursively axiomatisable. We may eventually revert to a recursively axiomatisable theory of validity, as McGee suggests, and invalidate V-Out. But this would be a reaction to the validity paradoxes: the recursive axiomatisability of the set of validities, and the subsequent invalidation of V-Out, would not be one of our initial assumptions.

It might be insisted that validity cannot be paradoxical on the grounds that “the notion of validity is to be ... defined in set-theory” (Field, 2008, p. 298) and that, since set-theory is consistent, so must be validity. This argument clearly would not work, however. For if it were legitimate to assume that validity is model-theoretically definable in order to show that there are no paradoxes of validity, then it would also be legitimate to assume that truth is model-theoretically definable in order to show that there are no paradoxes of truth. That truth-in-\( \mathcal{L} \) and validity in \( \mathcal{L} \) can be defined in set-theory by no means imply that there are no paradoxes of truth and validity.

Still, it might be thought that there are independent reasons to assume that validity is classical, which again would imply that validity cannot be a paradoxical notion. To see this, notice that, if validity is classical, then the following disjunction would hold:

\[
(D) \quad \pi \land \text{Val}(\neg \pi, \neg \bot) \lor \neg \pi \land \neg \text{Val}(\neg \pi, \neg \bot).
\]

It is now easy to check that both disjuncts are incompatible with VD. The second entails

\[
(D^*) \quad \neg \text{Val}(\neg \text{Val}(\neg \pi, \neg \bot) \land \pi, \neg \bot),
\]

which effectively says that an instance of VD is invalid. As for the first, given VD, it immediately entails \( \bot \). Field (2008, p. 307) mentions two reasons why we should think that validity is classical:

(a) that assuming excluded middle for validity claims leads to simpler reasoning about validity;

and

(b) that it would seem to be somewhat detrimental to the role of logic as regulator of reasoning if we were unable to say that any given piece of reasoning is either valid or not valid.
He concludes that “these considerations together seem to ... make the assumption that validity claims obey excluded middle a reasonable working hypothesis” (Field, 2008, p. 307). Both (a) and (b), though, validate parallel arguments against the claim that truth is paradoxical. Ad (i), if logic is a guide for correct reasoning and if, as Field supposes, only a simple logic can serve such a purpose, then reasoning in classical or intuitionistic logic is arguably simpler than reasoning in any weakening of these logics. But it would be too quick to conclude that one direction of the T-Scheme should fail on the grounds that assuming excluded middle for truth claims leads to simpler reasoning about truth. Certainly this argument would not be available to Field, who has long been advocating a non-classical, paracomplete logic of paradox (Field, 2003, 2007, 2008). Ad (ii), a parallel argument would conclude that it would be detrimental to the role of truth as a regulator of assertion if we were unable to say that any given sentence is either true or not true. Either way, it would then follow that truth must be a classical notion and that, for this reason, at least one direction of the T-Scheme must fail. But this would hardly be a reason for thinking that the Liar reasoning isn’t paradoxical.

5 Concluding remarks

I’ve argued that we have reasons—largely epistemic ones—for countenancing a broader notion of validity, alongside logical validity. Such a notion is no more, and no less, paradoxical than truth. As John Myhill puts it, commenting on the notion of absolute provability (a kind of validity):

the situation is completely analogous to that Epimendes paradox which arises when we try to formalize the notion of truth, and does not show the notion of [absolute] provability to be any more paradoxical than the notion of truth. (Myhill, 1960, p. 470)

The paradoxes of validity are generalisations of the truth-theoretic paradoxes. The Knower Paradox is but a stronger Liar, and v-Curry is, in turn, a generalisation of the Knower, one that employs stronger, though equally compelling, validity principles, as well as weaker logical resources. Consequently, validity gives rise to stronger indefinability and inexpressibility results: the standardly assumed structure of the validity relation suffices to make validity indefinable and, if Q is sound, inexpressible.
References


Cook, R.: 2013, There is no paradox of logical validity. Forthcoming in *Logica Universalis*.


