An Explanation of the Veridical Uniformity Universal

Shane Steinert-Threlkeld
S.N.M.Steinert-Threlkeld@uva.nl

Accepted with minor revisions at Journal of Semantics.

Abstract

A semantic universal, which we here dub the Veridical Uniformity Universal, has recently been argued to hold of responsive verbs (those that take both declarative and interrogative complements). This paper offers a preliminary explanation of this universal: verbs satisfying it are easier to learn than those that do not. This claim is supported by a computational experiment using artificial neural networks, mirroring a recent proposal for explaining semantic universals of quantifiers. This preliminary study opens up many avenues for future work on explaining semantic universals more generally, which are discussed in the conclusion.

This paper begins to develop an explanation for a semantic universal concerning so-called responsive verbs. These are verbs – such as know, forget, and be wrong – which can take both declarative and interrogative complements.1

(1) a. Carlos knows that Meica won the race.
   b. Carlos knows who won the race.

A natural question to ask about such verbs is whether – and, if so, how – the declarative-embedding sense (as exemplified in (1a)) is related to the interrogative-embedding sense (as exemplified in (1b)) of each verb.

A number of recent authors – chief among them Spector and Egré 2015 and Theiler, Roelofsen, and Aloni 2018 – have proposed a generalization that we will here call the Veridical Uniformity Universal. A rough statement, to be made precise later: every responsive verb is veridically uniform, in that it is veridical with respect to declarative complements if and only if it is veridical with respect to interrogative complements. To see that know is veridically uniform, first look at declarative complements. The inference in (2)2 shows that know is veridical with respect to declarative complements.

(2) a. Carlos knows that Meica won the race.
   b. \(\leadsto\) Meica won the race.

1The terminology comes from Lahiri 2002.
2Here, \(\leadsto\) is intended to be a form of entailment including both presupposition and literal entailment, but not things like conversational implicature.
Similarly, the inference in (3) shows that know is veridical with respect to interrogative complements.\textsuperscript{3}

(3) \begin{align*}
\text{a. Carlos knows who won the race.} \\
\text{b. Meica won the race.} \\
\text{c. } \neg \text{Carlos knows that Meica won the race.}
\end{align*}

The \textsc{universal} states that for every responsive verb, the inference patterns just observed for know will either both be valid or both be invalid. For an example that will be discussed in more detail later, the verb be certain satisfies the \textsc{universal} because both of the inference patterns are invalid. This imposes a kind of uniformity on responsive verb meanings, by placing a constraint on the relationship between the interrogative-embedding sense and the declarative-embedding sense. In this sense, it can be seen as a \textit{semantic universal}: a constraint on meanings realized across natural languages.

As a semantic universal, the \textsc{veridical uniformity universal} calls out for an explanation. This short contribution develops an explanation, building on the recent work by Steinert-Threlkeld and Szymanik 2018 on semantic universals for quantifiers. In particular, we argue the following: veridically uniform verbs (i.e. those satisfying the \textsc{universal}) are \textit{easier to learn} than verbs that are not veridically uniform. The universal then follows because we implicitly assume that languages tend to lexicalize meanings that are easier to learn. We provide evidence for the learnability claim by training an artificial neural network to learn the meanings of various responsive verbs and finding that the veridically uniform ones are indeed easier for it to learn.

The paper is structured as follows. Section 1 introduces the semantic framework for clause-embedding verbs due to Theiler, Roelofsen, and Aloni 2017. Against this backdrop, Section 2 provides an explicit definition of both notions of veridicality and a precise statement of the \textsc{universal}. It also introduces artificial verbs with non-uniform meanings. In Section 3, we outline the shape of our explanation of the \textsc{veridical uniformity universal} in terms of learnability. Section 4 contains our computational experimental results. Finally, we conclude in Section 5 with a discussion of the result and of many open avenues for future work.

1 A Uniform Semantics for Clause Embedding

We will couch our discussion in terms of the uniform treatment of clause-embedding due to Theiler, Roelofsen, and Aloni 2017. This treatment takes place in the setting of \textit{inquisitive semantics},\textsuperscript{4} where both declaratives and interrogatives have the same semantic type. While many of the main issues discussed in the present paper do not depend on this choice, we pause to note a few \textit{prima facie} advantages.\textsuperscript{5}

Firstly, since declaratives and interrogatives have the same type, responsive verbs can be given single lexical entries. While an approach that gives separate lexical entries for the two types of complement may work,\textsuperscript{6} a simple and unified semantic entry for such

\textsuperscript{3}There are restrictions on the type of interrogative complement to be used in this inference pattern. These will be discussed in Section 2.1.

\textsuperscript{4}See Ciardelli, Groenendijk, and Roelofsen 2018; Ciardelli, Roelofsen, and Theiler 2017; Ciardelli, Groenendijk, and Roelofsen 2013; Groenendijk 2009; Mascarenhas 2009.

\textsuperscript{5}See Uegaki 2018 for an overview of semantic treatments of responsive verbs.

\textsuperscript{6}As in, for example, George 2011. Though their approach is considered reductive and not ambiguous, Spector and Egré 2015 can also be seen as providing two entries per responsive verb, since they take the
verbs is to be preferred. Secondly, this treatment of responsive verbs makes an analogy between the Veridical Uniformity Universal and semantic universals in other domains more explicit. For example, on the standard story, while determiners denote generalized quantifiers, only a very restricted subset of such quantifiers – the monotone ones, the conservative ones – are denoted by any determiner in any language. Analogously, a uniform treatment of responsive verbs allows the Universal to be formulated in a similar way: only a certain subset of possible verb meanings are denoted in natural language. Finally, a natural extension of this approach has additional empirical advantages in its ability to treat varying degrees of exhaustivity and false-answer sensitivity.

Both declaratives and interrogatives denote downward-closed sets of propositions, where a proposition is a set of possible worlds. In other words, they are of type \( T := \langle \langle s, t \rangle, t \rangle, t \rangle \). Given such a meaning \( P \), we will write \( \text{info}(P) = \bigcup P \) for the informative content of \( P \).

The alternatives of \( P \) are its maximal elements: \( \text{alt}(P) = \{ p \in P : \neg \exists q \in P \text{ s.t. } p \subset q \} \). Interrogatives are traditionally distinguished by being inquisitive, i.e. by having more than one alternative. Declaratives and interrogatives differ in the type of object of type \( T \) that they denote. Declaratives have only one alternative, denoting the downward-closure of the set of possible worlds where the sentence is classically true. And if the declarative is non-tautologous and non-contradictory, we have \( \emptyset \neq \text{info}(P) \neq W \). By contrast, interrogatives will have more than one alternative but will be non-informative (i.e. \( \text{info}(P) = W \)).

To specify attitude verbs in this setting, we assume that each agent \( a \) has a set \( \text{DOX}_w^a \) of worlds compatible with their beliefs at world \( w \). Verbs will be of type \( \langle T, \langle e, T \rangle \rangle \): they take in an inquisitive sentence meaning (the complement) and an agent, and return a new inquisitive meaning. By definition and as desired, however, the resulting \( T \)-type meanings will always have a single alternative, since they are the meanings of declarative sentences. As an example, consider the following entry for be certain, due to Theiler, Roelofsen, and Aloni 2017.

\[
\text{be-certain} := \lambda P_T \lambda a_e \lambda p_{\langle s, t \rangle}. \forall w \in p : \text{DOX}_w^a \in P
\]

This entry says that Carlos is certain that Meica won the race is true at a world \( w \) just in case Carlos’ belief worlds at \( w \) resolve the denotation of that Meica won the race. This latter will be the downward-closure of the set of worlds in which Meica did win the race. So the sentence will be true just in case Carlos believes that Meica won the race. Note that this does not require that Meica in fact won the race in \( w \), merely that Carlos believes it to be so. Similarly, for Carlos is certain who won the race to be true, he has to believe of some person – not necessarily the true winner – that that person won the race, since those are the alternatives of the embedded interrogative.

---

7 See Barwise and Cooper 1981; Peters and Westerståhl 2006; Szymanik 2016; Steinert-Threlkeld and Szymanik 2018.
8 See Klinedinst and Rothschild 2011 for intermediate exhaustivity, George 2011; George 2013 for false-answer sensitivity, and Theiler, Roelofsen, and Aloni 2018 for the refined version of the semantics presented here that can capture both.
9 See Ciardelli, Roelofsen, and Theiler 2017.
10 Note that here and in what follows, I am freely switching between sets and their characteristic functions.
2 The Veridical Uniformity Universal

In this section, we will extract a general definition of veridicality with respect to declarative and interrogative complements, which will allow us to explicitly state the VERIDICAL UNIFORMITY UNIVERSAL. Then, we will show that both know and be-certain are veridically uniform and introduce two novel, hypothetical verb meanings that fail to be uniform. Table 1 at the end of the section summarizes these developments.

2.1 Explicitly Defined

Veridicality with respect to declarative complements has a standard and easy-to-state formulation. In particular, we just need to explicitly state that the inference from Ving a declarative complement to the truth of that complement is valid, as we saw exemplified in (2).\footnote{We note that we are here overloading the symbol $P$: in this section, it stands for an arbitrary declarative sentence while before it was a variable of type $T$. Context makes it clear which sense is intended.}

\begin{equation}
\begin{array}{l}
\text{(5) A responsive verb } V \text{ is veridical with respect to declarative complements if and only if for every agent } A \text{ and declarative sentence } P, \text{ the following inference is valid:} \\
\text{a. } A \text{ V's that } P. \\
\text{b. Therefore, } P. \\
\end{array}
\end{equation}

In other words: if $A$ V's that $P$ is true at a world $w$, then so too is $P$.

Some care has to be taken in generalizing from the example (3) to a general definition of veridicality for declarative complements. In particular, Spector and Egré 2015 (footnote 7) provide the following definition: $V$ is veridical with respect to interrogative complements iff the inference from $A$ V's $Q$ to $A$ V's that $P$ is true, where $P$ expresses the true and complete answer to $Q$. Theiler, Roelofsen, and Aloni 2018 point out that this definition wrongly characterizes know as non-veridical, due to so-called weakly exhaustive readings.

To rule out mis-classifications like this one, we will focus on questions where differences in exhaustive strength do not arise. We will say that a question $Q$ is exhaustive if and only if every world belongs to exactly one alternative of $Q$.\footnote{Theiler, Roelofsen, and Aloni 2018 call such questions exhaustively-neutral, since they only exhibit one level of exhaustivity, regardless of how they are interpreted. The interested reader can consult that paper for more precise definitions.}

There are two types of questions that are guaranteed to be exhaustive: polar questions such as whether Carlos won the Race and wh-questions which query about a property that is guaranteed to have a unique satisfier in each world. Since races only have one winner, who won the race is an example of the latter. We can now state the definition:

\begin{equation}
\begin{array}{l}
\text{(6) A responsive verb } V \text{ is veridical with respect to interrogative complements if and only if for every agent } A \text{ and exhaustive question } Q, \text{ and sentence } P \text{ expressing a true complete answer to } Q, \text{ the following inference is valid:} \\
\text{a. } A \text{ V's } Q. \\
\text{b. } P. \\
\text{c. Therefore, } A \text{ V's that } P. \\
\end{array}
\end{equation}

An exhaustive question has alternatives that denote a partition of logical space. Such questions are quite special: the true and complete answer in a world $w$ simply is the cell of the partition containing $w$. 

Using these two notions, we can now state the Universal. Say that a responsive verb $V$ is veridically uniform just in case it is veridical with respect to interrogative complements if and only if it is veridical with respect to declarative complements. That is: it’s either veridical with respect to both kinds of complement or with respect to neither kind of complement.

**The Veridical Uniformity Universal:** All responsive verbs are veridically uniform.

This Universal is a *semantic universal*: it makes a claim about the meaning systems of all natural languages. It has the following form: while syntactic objects of a certain kind (responsive verbs) have meanings of a certain kind, they only denote a restricted subset of those meanings (the veridically uniform ones).

A virtue of the present framework for responsive verbs is that the Universal can be stated in such a form. Before a discussion of how we will explain this semantic universal, we first exhibit examples of uniform and non-uniform responsive verbs.

### 2.2 **know** and **be-certain** are Veridically Uniform

As a paradigm veridically uniform verb, consider **know**. That it’s veridically uniform can be seen by the validity of the inference patterns exhibited in (2) and (3). A light modification of the entry (4) for **be-certain** can provide an entry for **know**:

(4) \[ \text{be-certain} := \lambda P \lambda a \lambda p \lambda (s,t). \forall w \in p : \text{DOX}_w^a \in P \]

(7) \[ \text{know} := \lambda P \lambda a \lambda p. \forall w \in p : \text{DOX}_w^a \in P \text{ and } w \in \text{DOX}_w^a \]

The clause $w \in \text{DOX}_w^a$ requires belief in a true resolution of the complement of **know**, which is exactly what was absent in **be-certain**. We note that in a complete treatment of this verb, this condition should be treated as a presupposition instead of a component of the asserted content. We return to the issue of presuppositions in the concluding Section 5.

Now, consider **be certain**, with its lexical entry (4). Our earlier conversation shows that the relevant declarative inference pattern is not valid, i.e. that **be certain** is not veridical with respect to declarative complements.

(8) a. Carlos is certain that Meica won the race.

b. $\neg$ Meica won the race.

Similarly, the previous discussion showed that the relevant interrogative inference pattern is not valid, i.e. that **be certain** is not veridical with respect to interrogative complements.

(9) a. Carlos is certain who won the race.

b. Meica won the race.

c. $\neg$ Carlos is certain that Meica won the race.

---

13 An anonymous reviewer observes that because the definition of veridicality for interrogative complements appeals only to how verbs behave with exhaustive question complements, the Universal (and the experiment presented later in this paper) cannot distinguish between verbs that agree on declarative complements and exhaustive question complements, but possibly diverge on non-exhaustive questions. Whether or not this restriction in the definition has empirical consequences will be postponed for future work.

14 We note that predicates of relevance (Elliott et al. 2017) have been put forth as counter-examples to the Universal (Theiler, Roelofsen, and Aloni 2018). We discuss the status of counter-examples in general in Section 3 and predicates of relevance in particular in the discussion Section 5.
(9a) could be true because Carlos is certain that Hyoung – not the actual winner Meica – won the race. Nevertheless, because neither inference pattern is valid, be certain is still veridically uniform.

2.3 Examples of Non-Uniform Verbs

We now define two non-veridically-uniform verbs, exhibiting the two possible ways of failing to be uniform (being veridical with respect to one, but not the other, type of complement). To define these lexical entries, we will not explicitly define them by cases based on how many alternatives the complement has. Rather, we will define them using natural operations on inquisitive contents that nevertheless do implicitly distinguish between sentence and question denotations.

2.3.1 knopinion

Our first verb – dubbed knopinion – will have a meaning roughly equivalent to knows when it embeds an interrogative but is opinionated when it embeds a declarative. It will be defined using inquisitive negation, standardly defined as follows: \( \neg P := \{ \bigcup P \} ^{\downarrow} \), where \( \overline{X} \) denotes set complementation. If \( P \) is the meaning of a question which covers all of logical space – which are the only kind under consideration presently – then \( \neg P := \{ \emptyset \} \). If \( P \) is a declarative meaning, i.e. has only one alternative, then \( \neg P \) is the downward-closure of the complement of that alternative. Because this operator treats declaratives and interrogatives very differently, we can use it to define a verb meaning that is veridical with respect to interrogative complements but not with respect to declaratives. We give knopinion the following lexical entry.

\[
(10) \quad \text{knopinion} := \lambda P \lambda a \lambda p. \forall w \in p : w \in \text{DOX}_w \text{ and } (\text{DOX}_w \in P \text{ or } \text{DOX}_w \in \neg P)
\]

To see that this is veridical with respect to interrogative complements, consider the meaning of Carlos knopinions who won the race: knopinion(who won the race)(C). Because the complement is a question, its inquisitive negation will be \( \{ \emptyset \} \). If it’s true that Carlos knopinions who won the race, then since \( w \in \text{DOX}_w \), it follows that Carlos’ belief-worlds are not in \( \neg \text{who won the race} \). Thus, they must resolve the question who won the race; moreover, because they contain \( w \), they must do so truthfully. Therefore, Carlos knopinions that Meica won the race will also be true.

By contrast, suppose that Carlos knopinions that Meica won the race is true. It does not follow that Meica won the race: all that’s required is that Carlos correctly believes either that Meica did or that she did not. Thus, if she did not win the race and Carlos believes that, he will knopinion that Meica won the race. This shows that knopinion is not veridical with respect to declarative complements and, therefore, is not veridically uniform.

2.3.2 wondows

Our second verb – dubbed wondows – will have a meaning roughly equivalent to is uncertain when it embeds an interrogative but knows when it embeds a declarative. Its lexical entry is given below.

\[
(11) \quad \text{wondows} := \lambda P \lambda a \lambda p. \forall w \in p : w \in \text{DOX}_w \subseteq \text{info}(P) \text{ and } \text{DOX}_w \cap q \neq \emptyset \forall q \in \text{alt}(P)
\]
Two ingredients distinguish declaratives and interrogatives. The clause $\text{dox}_w^a \subseteq \text{info}(P)$ will be trivial for interrogative complements, since in that case $\text{info}(P) = W$. But it’s non-trivial for non-tautologous declaratives. Moreover, if $P$ is a declarative and has only one alternative, then the first conjunct will entail the second. But if it is interrogative, then the quantification over the alternatives of the question has a substantial effect: the agent must leave all alternatives open at the moment.

`wondows` is veridical with respect to declarative complements: the first conjunct turns out to be equivalent to `know` in this case\(^{15}\) and moreover entails the second conjunct. By contrast, consider the sentence `Carlos wondows who won the race`. This means that for every possible winner, Carlos leaves it open that that person won the race. Supposing that Meica did in fact win the race, it does not then follow that `Carlos wondows that Meica won the race`, since the latter requires knowledge. This shows that `wondows` is not veridical with respect to interrogative complements and, therefore, is not veridically uniform.

<table>
<thead>
<tr>
<th>Verb</th>
<th>Lexical Entry: $\lambda P. \lambda a. \lambda p. (\forall w \in P : \ldots)$</th>
<th>Veridical</th>
<th>Declarative</th>
<th>Interrogative</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>know</code></td>
<td>$\text{dox}_w^a \in P$ and $w \in \text{dox}_w^a$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><code>wondows</code></td>
<td>$w \in \text{dox}_w^a \subseteq \text{info}(P)$ and $\text{dox}_w^a \cap q \neq \emptyset \forall q \in \text{alt}(P)$</td>
<td>✓</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><code>kno-opinion</code></td>
<td>$w \in \text{dox}_w^a$ and $(\text{dox}_w^a \in P$ or $\text{dox}_w^a \in \neg P)$</td>
<td>x</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td><code>be-certain</code></td>
<td>$\text{dox}_w^a \in P$</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Table 1: Four verbs, exemplifying the possible profiles of veridicality.

### 3 Explaining Semantic Universals

Having seen examples of veridically uniform and non-uniform responsive verbs, the pressing question becomes: why are only the former attested in natural languages? While many answers are possible (a fact to which we return in the concluding discussion), we here explore the following

**Learnability Hypothesis:** The **Veridical Uniformity Universal** holds because veridically uniform verbs are easier to learn than non-uniform verbs.

A primary motivation for exploring this **Hypothesis** comes from Steinert-Threlkeld and Szymanik 2018. They explored the same hypothesis for a range of semantic universals concerning which generalized quantifiers are expressed by determiners. Much of the recent discussion of the **Veridical Uniformity Universal**\(^{16}\) has drawn explicit parallels between it and the relevant semantic universals about quantifiers. One would hope then that whatever explains the latter could also explain the former.

The **Learnability Hypothesis** helps explain the presence of the uniformity of responsive verbs together with an implicit assumption: languages tend to lexicalize expressions that are easier for speakers of the language to learn. The assumption is needed for the following reason: the **Hypothesis** really claims that among possible verb *meanings*, those that are uniform will be easier to learn. Similarly, the experiment presented

\(^{15}\)Because $P = \text{info}(P)^\downarrow$.

\(^{16}\)See especially Theiler, Roelofsen, and Aloni 2018; Uegaki 2018.
below in support of the HYPOTHESIS looks at the ease of learning of meanings of the right kind directly. The implicit assumption links this fact about verb meanings to the lexicon: languages tend to attach words to such easy-to-learn meanings. To express the harder-to-learn meanings, we must rely on complex syntactic structures and compositional interpretation thereof. For example, it seems that has a correct belief about has the meaning knopinion. But no lexical verb expresses that meaning, precisely because it is harder to learn. It’s also possible that alleged counter-examples to the UNIVERSAL will be ripe candidates for analysis via lexical decomposition.\footnote{Thanks to an anonymous reviewer for urging me to clarify this point about the implicit assumption and to Chris Kennedy for the example has a correct belief about and the thought about decomposition. Steinert-Threlkeld and Szymanik (2018) discuss the point about lexicalization in more detail in the case of determiners since the universals there are stated explicitly in terms of what ‘simple’ (i.e. lexical) determiners denote.}

While we do not defend the implicit assumption here, two features of this package of views are worth remarking upon. First, on this picture, universals may be robust statistical generalizations and not strictly universal claims.\footnote{See Evans and Levinson 2009.} Both the learnability claim and the implicit assumption are general tendencies, not necessarily exception-less rules. Second, this approach offers the promise of explaining which particular universals are attested, not just why there are universals at all. An analogue of the Chomskyan argument for Universal Grammar (i.e. that restricting the search space for a language learner makes the task easier) would only provide an argument that one should expect there to be semantic universals.

The resulting package can be impressionistically described as follows: one can imagine a kind of heat map overlaid on the space of possible verb meanings, with redder shades meaning easier to learn and bluer shades meaning harder to learn. In other words, the language learner will be attracted to warmer regions on the map.\footnote{By contrast, the Chomskyan argument says that the learner cannot wander through the whole space, but is restricted from the outset to a sub-region of the space.} The claim embodied by the LEARNABILITY HYPOTHESIS is that the region of the space corresponding to the uniform verbs will be quite warm, quite red on the map. The claim is not that every uniform verb is easier to learn than every non-uniform verb. Similarly, the claim is not that all uniform verbs are equally easy to learn. This raises the important question: when do differences in learnability result in robust typological facts and when do they not?\footnote{Thanks to Chris Kennedy for pressing this line of thinking, which applies beyond the present case study.}

A full and precise answer to this question must await future work, integrating an explicit model of the lexicalization process. The basic idea would be that the uniform region of verb space is such a warm region that as languages evolve, they gravitate towards it and only lexicalize meanings from that region. A strong pattern of ease of learning, which does not make either universal claim above, could still have that effect.

4 Experiment

To lend support to the HYPOTHESIS, we will present a computational experiment comparing the ease of learning the four verbs discussed in Section 2.1. In particular, following the methodology of Steinert-Threlkeld and Szymanik 2018, we will train an artificial neural network\footnote{See Nielsen 2015; Goodfellow, Bengio, and Courville 2016 for textbook introductions.} to learn the meanings of the four verbs and compare how well it learns
each one. The code for replicating the experiment and the analysis may be found at https://github.com/shanest/responsive-verbs.

4.1 Methods
We generated data to train such a network as follows. We set a number \( N = 16 \) of possible worlds. For each pair of a verb and a truth-value, we generated 16000 data points. Such a point consists of either a set of worlds or a partition (corresponding to declarative and interrogative meanings), a set \( \text{DOX}_w^a \) of worlds, and one \( w \) as the world of evaluation. Partitions are generated randomly, with the number of cells (i.e. alternatives in the question) being chosen uniformly at random between 2 and 4. The set \( \text{DOX}_w^a \) and \( w \) are then generated as randomly as possible in keeping with the verb/truth-value pair in question. For each such pair, 4000 points were held out from training and used only for testing the model, in order to see how well it generalized.

Our model of learning was a feed-forward neural network consisting of four hidden layers, each of 128 units with exponential linear activation\(^{22}\) and 10% dropout\(^{23}\). A single input to the model consists of a label for the verb, a binary vector of length \( N \) representing the characteristic function of \( \text{DOX}_w^a \), a label for the world \( w \), and the partition.\(^{24}\) The partition is encoded as an \( N^2 \) length binary vector, where each world is mapped to a binary vector corresponding to the cell of the partition that to which it belongs and the resulting vectors are concatenated.\(^{25}\) The labels are the correct truth-value for the input, the network’s output is its estimate of the probability in each truth-value, and its goal is to minimize the cross-entropy between its guess and the correct label. The training set was randomized. We trained the network using the RMSProp optimizer\(^{26}\) for 15 epochs with batch size 128.\(^{27}\)

To operationalize ease of learning, we measured accuracy on the test set for each verb at the end of this fixed amount of training time. Measuring accuracy after a fixed amount of training time, and not necessarily when the network has converged and stopped learning, provides a good operationalization of ease of learning for the following reason. Since the network will have seen exactly the same amount of data for each verb, if it has learned to extract the truth-conditions from the training data and can predict the truth-values for one verb significantly better than another, then there is a strong sense in which that verb was easier to learn than the other for this model.

4.2 Results
We followed the training procedure described above for \( n = 60 \) independent trials, since both the network’s initial state and the data generation as well as the order of presentation of the training data are random. Figure 1 depicts the results.

---

\(^{22}\)This is the function \( \text{elu}(x) = \begin{cases} x & x \geq 0 \\ e^x - 1 & x < 0 \end{cases} \). See Clevert, Unterthiner, and Hochreiter 2016.

\(^{23}\)Srivastava et al. 2014.

\(^{24}\)The labels for the verb and the world are so-called one-hot vectors: vectors of the appropriate dimension (4 for the verbs, \( N \) for the world), with all zeros except for a single 1, corresponding to the value.

\(^{25}\)This resembles the \( E \) function from Theiler, Roelofsen, and Aloni 2018 in the case of partitions.

\(^{26}\)Hinton, Srivastava, and Swersky 2014.

\(^{27}\)Again see Nielsen 2015; Goodfellow, Bengio, and Courville 2016; Steinert-Threlkeld and Szymanik 2018 for explanations of all of the terms.
Figure 1: Left: accuracy on held-out test set versus number of training steps for each verb for each trial. Right: distributions of differences in final accuracy between know and the other verbs (top) and be-certain and the other verbs (bottom).

The sub-figure on the left plots the raw learning curves. The x-axis is the number of training steps, with the y-axis being accuracy on the held-out test set. As can be seen, it looks like know and be-certain consistently achieve a higher degree of accuracy than knopinion and wondows. While pair-wise dependent t-tests on the final accuracies of each pair of a uniform and a non-uniform verb confirmed that the latter are learned to a higher degree than the former (see Table 2), the distributions of the final outcomes are more informative. These are depicted in the right two sub-figures.

The top-right figure shows the distribution of per-trial differences in final accuracy between know and the three other verbs. The raw data are the tickmarks on the x-axis; the curves are Gaussian kernel density estimates of the distribution. Two features should be highlighted: many trials with be-certain have a negative difference, meaning that the latter was learned to a higher degree of accuracy. Its distribution almost appears to be centered at 0. By contrast, in every trial, know is learned to a higher degree of accuracy than both knopinion and wondows. The bottom-right sub-figure shows the same information with be-certain. In particular, this verb is also learned to a higher degree of accuracy than both knopinion and wondows in every trial.

These distributional facts provide strong prima facie support to the LEARNABILITY HYPOTHESIS: indeed it looks like the two uniform verbs are easier to learn by an artificial neural network than the two non-uniform verbs. In Appendix A, we provide a more
detailed analysis of the networks’ behavior, to see which features of the input they learn to rely on.

5 Discussion

This short contribution has presented the Veridcal Uniformity Universal for responsive verbs and a preliminary explanation thereof: verbs satisfying the universal are easier to learn than those that do not. A computational experiment provided the evidence for this learnability claim.

We do not thereby claim that learnability is the only factor that explains the emergence of this or other universals. Other likely factors include communicative needs and something like descriptive complexity in a language of thought.\textsuperscript{28} We merely note that making good on the latter is a non-trivial task. Exploring the role of these other factors and the interrelations among all of them in explaining semantic universals remains a ripe area for further research.

We also do not thereby claim that the present model of learnability is in any sense the ‘best’ or ‘correct’ one. Rather, it is a promising place to start, especially given the success of similar models in explaining universals for quantifiers (Steinert-Threlkeld and Szymanik 2018).\textsuperscript{29} While we find it plausible that other models of learning will exhibit the same pattern, more work can be done in this area. And all such models stand to be integrated with explicit models of the evolution of language, to fully corroborate the claim that languages tend to lexicalize easy-to-learn meanings. It would also be good to test not just different models but also children’s ability to learn novel uniform and non-uniform verbs to see if the former are easier.\textsuperscript{30}

Other work to be done includes ‘scaling up’ the experiment presented here to a wider range of uniform and non-uniform verbs. Aside from providing a robustness check on the results of the present experiment, a much larger scale experiment can address the following worry. There are other properties that know and be-certain have but that wondows and knopinion lack; so the experiment cannot show that it is veridical uniformity that makes the former easier to learn than the latter.\textsuperscript{31} While this point is true, a larger-scale experiment that has a network learn many verbs that vary in the combination of semantic properties that they have could go some way in identifying exactly which properties make the verbs more learnable.

Another critical avenue involves extending the modeling approach here to handle presuppositions. On the one hand, this would allow for more appropriate lexical entries for verbs like know. On the other hand, it could go some way towards helping explain alleged counter-examples to the Universal: predicates of relevance.\textsuperscript{32} These are what we can call Strawson veridically uniform: they can fail to be uniform, but only in that the con-

\textsuperscript{28}See Feldman 2000; Goodman et al. 2008; Piantadosi, Tenenbaum, and Goodman 2016.

\textsuperscript{29}We note that although both studies proceed by training artificial neural networks, there are technical differences in the network architectures that prevent direct comparison of the results from the two domains. Thanks to an anonymous referee for pressing this point.

\textsuperscript{30}Compare Hunter and Lidz 2013 on conservativity for quantifiers.

\textsuperscript{31}Thanks to Wataru Uegaki for pushing this point. He highlights the following two properties. Say $V$ is choosy iff for all declarative complements and agents $a$, $V(P)(a) \neq V(\neg P)(a)$ and that $V$ is sensitive to non-alternatives iff there are complements $P$ and $P'$ and individual $a$ such that $\text{alt}(P) = \text{alt}(P')$ but $V(P)(a) \neq V(P')(a)$. Then both know and be-certain are both choosy and sensitive, while knopinion is not choosy and wondows is not sensitive.

\textsuperscript{32}See Elliott et al. 2017 and Theiler, Roelofsen, and Aloni 2018 for discussion.
clusions of the inference patterns defining veridicality can have presupposition failure. So those inference patterns will still be Strawson valid for those verbs. It would be interesting to compare the status of Strawson uniform verbs with uniform and non-uniform verbs both in the present framework and related ones. That being said, we reiterate the point from Section 3 that the present explanation does allow for there to be occasional counter-examples to universals.

In summary, then, we take the present experimental result to be a promising initial contribution that opens many doors to future work. We hope that others will be inspired to work on explaining both the Veridical Uniformity Universal and other semantic universals in terms of learnability and cognate factors.

References


33See von Fintel 1999.
A Preliminary Network Behavior Analysis

While the primary result of our experiment concerned the accuracy of a network after a fixed period of training, that measure alone does not reveal a huge amount about the behavior of the network. In this appendix, we present a very preliminary analysis of the
networks’ behavior, in an attempt to ‘look inside the black box’ and understand what factors drive the networks’ failures.34

First, we look at the confusion matrix of the network, which allows us to see what kinds of mistakes are most frequent. Because we ran 60 trials, Table 3 reports an average confusion matrix across all the trials. In Figure 2, the distributions of true positives/negatives and false positives/negatives are plotted by verb.

<table>
<thead>
<tr>
<th>label</th>
<th>all</th>
<th>know</th>
<th>be-certain</th>
<th>knopinion</th>
<th>wondows</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15412.2</td>
<td>1176.4</td>
<td>3881.1</td>
<td>261.7</td>
<td>3878.5</td>
</tr>
<tr>
<td>0</td>
<td>587.8</td>
<td>14823.7</td>
<td>118.9</td>
<td>3738.3</td>
<td>121.6</td>
</tr>
</tbody>
</table>

Table 3: Average confusion matrix across all 60 trials, in total and by verb. The rows are predicted truth-value, and the columns the actual truth value.

A few patterns can be extracted from this data. First, the network appears to learn a bias towards labeling examples as ‘true’: for every verb, there are more true positives than true negatives and more false positives than false negatives. This is somewhat surprising since the data provided to the network is balanced across verb and truth-value (there are an equal number of examples for each such pair). It suggests that the network extracts patterns from the data that cause it to overestimate how often the verbs are true. Second, while many of the distributions appear to be the same for the uniform and non-uniform verbs, the distributions of true positives and false negatives for knopinion appears to be much lower variance than those for wondows. This suggests that the network has some difficulty extracting the condition for wondows being true since its behavior with those examples is the highest variance.

Another feature to analyze is the dependence of accuracy on various semantic factors. Table 4 breaks down the accuracy by verb and by three factors – complement type, whether \( w \in \text{DOX}^a_w \), and whether \( \text{DOX}^a_w \in P \) – aggregated across all trials.

<table>
<thead>
<tr>
<th>factor</th>
<th>value</th>
<th>know</th>
<th>be-certain</th>
<th>knopinion</th>
<th>wondows</th>
</tr>
</thead>
<tbody>
<tr>
<td>complement</td>
<td>declarative</td>
<td>0.983</td>
<td>0.986</td>
<td>0.954</td>
<td>0.983</td>
</tr>
<tr>
<td></td>
<td>interrogative</td>
<td>0.923</td>
<td>0.924</td>
<td>0.921</td>
<td>0.841</td>
</tr>
<tr>
<td>( w \in \text{DOX}^a_w )</td>
<td>1</td>
<td>0.964</td>
<td>0.957</td>
<td>0.954</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.919</td>
<td>0.953</td>
<td>0.887</td>
<td>0.924</td>
</tr>
<tr>
<td>( \text{DOX}^a_w \in P )</td>
<td>1</td>
<td>0.961</td>
<td>0.966</td>
<td>0.949</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.945</td>
<td>0.943</td>
<td>0.929</td>
<td>0.922</td>
</tr>
</tbody>
</table>

Table 4: Accuracy by verb and various semantic features of the input, aggregated across all trials.

The pattern for complement type is striking: for every verb, accuracy is significantly higher on declarative complements than on interrogative complements. This suggests that the network has difficulty in extracting patterns from such complements which are, in an intuitive sense, more complex. Far and away the lowest accuracy occurs with wondows on interrogative complements. Intuitively, this truth-condition is the most ‘global’, since

---

34 Thanks to an anonymous referee for encouraging these types of analyses.
it requires non-empty intersection with every alternative of a question. These two facts suggest that the network struggles to extract such global patterns from the input.

One other interesting pattern of complement type should be highlighted. The accuracy of *know* on interrogative complements – in which case it’s equivalent to *know* – is quite comparable to that of *know*. Similarly, the accuracy of *windows* on declarative complements – in which case it’s equivalent to *knows* – is quite comparable to that of *know*. This suggests that the difficulty in learning can be ‘factored’ into the difficulty of learning the truth-conditions for interrogative and for declarative complements. This factorization highlights another interesting pattern that might help explain VERIDICAL UNIFORMITY: the difficulty for each type of complement will be more aligned for uniform verbs than for non-uniform verbs.

Turning to the feature $w \in \text{DOX}_w^a$, notice that BE-CERTAIN appears to be the only verb whose accuracy does not depend on the presence of this feature. This makes sense: it is also the only verb whose truth value does not depend on this feature. For the other three verbs, they are less accurate when $w \notin \text{DOX}_w^a$ than when it is. This is somewhat surprising: it suggests that the network has not learned that for each of these verbs, $w \notin \text{DOX}_w^a$ suffices for falsity.

Finally, it appears that the networks are more accurate when $\text{DOX}_w^a \in P$ than when not, for every verb. In this case, every verb’s truth value depends on this feature, so we should not expect much difference across verbs. That performance is worse when $\text{DOX}_w^a \notin P$ again suggests that networks are not learning that this condition suffices for falsity for be-certain and know. This also reflects the general bias towards positive answers over negative ones discussed in the context of the confusion matrices.

---

For *windows*, $\text{DOX}_w^a \subseteq \text{info}(P)$ is equivalent to $\text{DOX}_w^a \in P$ for declarative $P$. 

---

Figure 2: Distributions (Gaussian kernel density estimates) of the true/false positives/negatives by verb.
This preliminary analysis of the network’s behavior shows that there are subtle and intriguing patterns in how it treats different inputs. In the interest of space, a more detailed analysis of these features and of others will be left for future work.