Course: Topics in Philosophy of Mathematics
(Lecture Notes: 2023.02)

An Introduction to Gupta's Acceptable Models

Ming Hsiung

Abstract This article is a lecture note I wrote for my philosophy of mathematics course. Its main task is to explain the main ideas of Gupta's acceptable model proposed in his paper [J. Philos. Logic 11(1), 1–60, 1982]. I aim to provide detailed information on a result established by Gupta. On the one hand, I hope this explanation can be helpful for those who are learning Gupta's acceptable model, and on the other hand, I also hope to provide a guide for beginners to interpret literature.

Keywords Acceptable Model · Paradox · Quotation Names · Tarski's Schema · Truth Predicate

1 Introduction

After Tarski [10] gives a diagnosis for the causes of semantic paradoxes in the theories of truth, it is widely believed that if a language satisfies the following three conditions simultaneously, it must lead to inconsistency: (1) it is syntactically rich enough that its own syntax can be defined through certain means (such as Gödel's arithmetization method); (2) it is semantically closed, that is, it contains the truth predicate for the very language; and (3) it is classical two-valued in logic.

However, this is not correct —at least in the sense of "literally", according to a proof that Gupta provides in his paper [1] in which he propose an original theory of truth known as the revision theory of truth. Gupta claims explicitly "It is sometimes said that Tarski has shown that a language containing arithmetic cannot contain its own truth concept. This is at best a very misleading claim. Taken literally it is false." (ibid., p. 15) Then, he introduces the so-called acceptable models, proving "if the syntactic resources are weak, as they are for L in so far as the fragment containing the truth predicate is concerned, then there is no contradiction even if the language contains its own truth concept. So there is no difficulty in L's having the concept 'true-in-L'." (ibid.)

In the history of truth theories, Gupta's idea of acceptable models is fundamental for the construction of the classical two-valued models validating Tarski's schema. In some sense, it is something like Kripke [7]'s well-known thought of the construction of the Kleene's three-valued models which also make Tarski's schema true. Gupta & Belnap [3, pp. 201ff] explicate the theory of acceptable models and Kremer [5, p. 353] develops this theory further.

This manuscript is not a research paper, but rather something like a lecture note. Here we will only present the main ideas and fundamental results of Gupta's acceptable model. When I first read Gupta's 1982 paper, I found the concept of acceptable model and the relevant results difficult to understand, and

E-mail: mingshone@163.com
Orcid ID: 0000-0001-9037-2024
School of Philosophy and Social Development South China Normal University, Guangzhou 510631, People's Republic of China

felt it necessary to add some details to better understand Gupta's argument. As a set of lecture notes, I hope to demonstrate to my graduate students how to restate the main ideas of the original author, supplementing any overlooked or omitted information, and presenting every detail of the relevant results as much as possible, as if I were writing a chapter in an introductory textbook.

In writing this lecture note, I also gained some new insights from Gupta's acceptable model. For example, truth theorists often assume that we can set up a language L and its "expansion" (in other words, extension) L^+ such that L contains all quotation names of the form 'A', where A is a sentence in L^+. It seems that Gupta and most of his successors take this assumption for granted. However, on second thought, I would like to say that this assumption is weird.[1] As an expansion of L, L^+ clearly depends on L. Therefore, we can only determine which sentences are in L^+ after we have given this expansion of L. In that case, how can language L itself already contain quotation names for all sentences in L^+? It appears that there is a terrible circularity in the setup of the languages by the truth theorists.

As far as I know, only in [11, pp. 44-45], Yaqūb provides a detailed explanation of how a language can contain quotation names for all sentences in itself. In this article, I will also present a way to dispel the above concerns regarding the language setup based on a notion from Stern [9, p. 37] (see Definition 1). I believe that by supplementing these details that Gupta and his successors overlooked or disregarded, we demonstrate that presenting the original author's ideas themselves also requires careful consideration and argumentation.

I will present the main ideas and results on Gupta's acceptable models based on the first-order arithmetic language since this is the base language that the truth theorists employ most often. This language automatically meets the condition we mentioned earlier about the richness in syntax. The exposition in this article certainly applies to any other first-order language that are sufficiently rich in syntax as well.

2 Languages and Models

Recall: Let $\mathscr{L}$ be the first-order language for the Peano arithmetic. The primitive symbols of $\mathscr{L}$ are as follows.

Variables: x, y, z, x_0, x_1, y_2, $\cdots$

Connectives: $\neg$, $\wedge$

Quantifier: $\forall$

Identity Symbol: $=$ (which is a logical predicate)

Constant: 0

Predicate Symbols: $<$

Function Symbols: S, $+$, $\cdot$

Subsidiary symbols:) , (

So, $\mathscr{L}$ can be written as the tuple

$$\mathscr{L} = \langle <, S, +, \cdot, 0 \rangle .$$

Other connectives and quantifier, such as $\vee$, $\rightarrow$, and $\exists$, can be defined as usual.

Let $\mathscr{L}_{QT}$ be the language obtained from $\mathscr{L}$ by adding a unary predicate symbol T and two subsidiary symbols $\lfloor$ and $\rfloor$ (quotation marks). That is,

$$\mathscr{L}_{QT} = \langle <, S, +, \cdot, 0, \lfloor\,\rfloor, T \rangle .$$

Definition 1 We define the **terms** and **formulas** of $\mathscr{L}_{QT}$ by a simultaneous induction.[2]

[1] It is well-known that L does be able to contain the quotation numbers of sentences in an expansion of L, provided that L is a language for Peano arithmetic and the quotation names 'A' is taken as the numeral for the Gödel number of A. However, this is not the case that we discuss currently. We must emphasize that here, L is an arbitrary formal language, which may be irrelevant to arithmetic in any sense. And the quotation names are used just as their literal meaning.

[2] See Stern [9, p. 37].

(1) All variables and the constant 0 are terms (of $\mathscr{L}_{QT}$).
(2) If t_1 and t_2 are terms, St_1, $(t_1 + t_2)$, and $t_1 \cdot t_2$ are also terms.
(3) If t_1 and t_2 are terms, $T(t_1)$, $t_1 < t_2$, and $t_1 = t_2$ are formulas (of $\mathscr{L}_{QT}$).
(4) If A and B are formulas, $\neg A$ and $(A \wedge B)$ are formulas.
(5) If A is a formula and v is a variable, $\forall v A$ are formulas.
(6) If A is a formula, $\lfloor A \rfloor$ is a term.

Free variables are defined as usual (for instance, the free variables of $\forall v A$ are precisely those in A minus v, and the free variables of $T(t_1)$ are precisely those in t_1) except that for any formula A, $\lfloor A \rfloor$ has no free variable. A formula without any free variable is called a **sentence**. For a sentence A, we use $T\lfloor A \rfloor$ as a shorthand of $T(\lfloor A \rfloor)$.

The following notion will be useful.

Definition 2 We define a function ρ on the set of terms and formulas of $\mathscr{L}_{QT}$ as follows.[3]

(1) If t is a variable or the constant 0, $\rho(t) = 1$.
(2) $\rho(St) = \rho(t)$, $\rho(t_1 + t_2) = \max(\rho(t_1), \rho(t_2))$, and $\rho(t_1 \cdot t_2) = \max(\rho(t_1), \rho(t_2))$.
(3) $\rho(T(t)) = \rho(t)$, $\rho(t_1 < t_2) = \max(\rho(t_1), \rho(t_2))$, and $\rho(t_1 = t_2) = \max(\rho(t_1), \rho(t_2))$.
(4) $\rho(\neg A) = \rho(A)$ and $\rho(A \wedge B) = \max(\rho(A), \rho(B))$.
(5) $\rho(\forall v A) = \rho(A)$.
(6) $\rho(\lfloor A \rfloor) = \rho(A) + 1$.

$\rho(t)$ and $\rho(A)$ is called the **degree** of t and A respectively.

See appendix for information about how Gupta treats with the quotation names.

We now introduce the two languages that we really focus on. To highlight these two languages, we give them in the following definition.

Definition 3 First, let $\mathscr{L}^*_{QT}$ be the language obtained from $\mathscr{L}$ by adding a unary predicate symbol T and the constant symbols $\lfloor A \rfloor$ for all sentence A of $\mathscr{L}_{QT}$. Let use $A \in \mathscr{L}_{QT}$ to denote that A is a sentence in $\mathscr{L}_{QT}$. So, we have

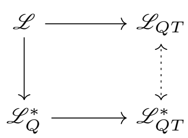
$$\mathscr{L}^*_{QT} = \langle <, S, +, \cdot, 0, \{\lfloor A \rfloor \mid A \in \mathscr{L}_{QT}\}, T \rangle .$$

Next, let $\mathscr{L}^*_{Q}$ be the T-free fragment of $\mathscr{L}^*_{QT}$, that is, the language obtained from $\mathscr{L}^*_{QT}$ by deleting T:

$$\mathscr{L}^*_{Q} = \langle <, S, +, \cdot, 0, \{\lfloor A \rfloor \mid A \in \mathscr{L}_{QT}\} \rangle .$$

We now have four languages: $\mathscr{L}$, $\mathscr{L}_{QT}$, $\mathscr{L}^*_{QT}$, and $\mathscr{L}^*_{Q}$. We illustrate their relation by the commutative diagram 1. Note that an arrow from a language to another denotes that the latter is an expansion of the former, and the dotted arrow between $\mathscr{L}_{QT}$ and $\mathscr{L}^*_{QT}$ denotes that they have the same sentences.

Table 1 Relation of the four languages



The terms and formulas of $\mathscr{L}^*_{QT}$ (and $\mathscr{L}^*_{Q}$) can be formed routinely (the simultaneous induction is no longer needed). Note that although $\mathscr{L}^*_{QT}$ and $\mathscr{L}_{QT}$ have different primitive symbols, they have the same sentences. We leave the proof of the following lemma to the reader.

[3] See Gupta [1, p. 11], Gupta & Belnap [3, pp. 201-202], Stern [9, p. 37]. Note that the degree that we define starts from 1 rather than 0. It will makes our presentation of the main lemma (Lemma 2) more elegant.

Lemma 1 *The sentences of $\mathscr{L}^*_{QT}$ are precisely those of $\mathscr{L}_{QT}$.*

Another crucial observation is that the small language $\mathscr{L}^*_Q$ contains all quotation names for the sentences in the large language $\mathscr{L}^*_{QT}$. Also, note that $\mathscr{L}^*_Q$ is also sufficiently rich in syntax because it is an expansion of $\mathscr{L}$. So, we realize a tacit assumption of truth theorists that we mention in the introduction section: a language can have all quotation names for sentences in an expansion of this language, and moreover, these quotation names can be introduced as primitive symbols.

From now on, we will only consider the languages $\mathscr{L}^*_Q$ and $\mathscr{L}^*_{QT}$ unless otherwise claimed. We use Θ to denote the set of all sentences in $\mathscr{L}^*_{QT}$.

Definition 4 Let D be a set including Θ. A **model for** $\mathscr{L}^*_Q$, namely M, is a tuple

$$\left\langle D, S^M, +^M, \cdot^M, 0^M, <^M, \left\{\lfloor A\rfloor^M \mid A \in \mathscr{L}^*_{QT}\right\}\right\rangle$$

where S^M is a unary function on D, $+^M$ and $\cdot^M$ are binary function on D, 0^M is an element of D, $<^M$ is a binary relation on D. Moreover, for any $A \in \mathscr{L}^*_{QT}$, we always stipulate that $\lfloor A\rfloor^{\mathfrak{M}} = A$.[4]

Definition 5 A **model for** $\mathscr{L}^*_{QT}$, namely $\mathfrak{M}$, is a model for $\mathscr{L}^*_Q$ plus the interpretation of T. For simplicity, we will use the notation $\mathfrak{M} = \langle M, X\rangle$ (or $M + X$), where X is the interpretation of T, that is, $T^{\mathfrak{M}} = X$.

In the following, when we say a model, we always mean a model for $\mathscr{L}^*_{QT}$. For a model $\mathfrak{M} = \langle M, X\rangle$, we will say that M is the **ground model** of $\mathfrak{M}$.

Definition 6 Let $\mathfrak{M} = \langle M, X\rangle$ be a model given in Definition 5. An **assignment** on the model $\mathfrak{M}$ (or on M) is a function from the set of variables to the domain D.

For any $a \in D$, we define that $\sigma[a/x]$ is the assignment which is same as σ except that $\sigma[a/x](x) = a$. More specifically,

$$\sigma[a/x](y) = \begin{cases} a, & \text{if } y = x; \\ \sigma[a/x](y), & \text{otherwise.} \end{cases}$$

Definition 7 Let $\mathfrak{M} = \langle M, X\rangle$ be a model and σ be an assignment as given in Definition 5 and 6. We define $t[X, \sigma]$ (the **interpretation** of a term t). For brevity, we have suppressed the parameters of $\mathfrak{M}$ except X.

(1) $t[X, \sigma] = \sigma(t)$ if t is a variable,
$t[X, \sigma] = 0^M$ if t is the constant 0, and
$\lfloor A\rfloor[X, \sigma] = \lfloor A\rfloor^{\mathfrak{M}}$ (that is, A) for any $A \in \Theta$.
(2) $(St)[X, \sigma] = S^M(t[X, \sigma])$,
$(t_1 + t_2)[X, \sigma] = t_1[X, \sigma] +^M t_2[X, \sigma]$, and
$(t_1 \cdot t_2)[X, \sigma] = t_1[X, \sigma] \cdot^M t_2[X, \sigma]$.

If X is evident from the context, we will use the routine notation, for instance, $t[a/x_1, b/x_2]$ for $t[X, \sigma]$ where t is the term $t(x_1, x_2)$, and σ is such that $\sigma(x_1) = a$ and $\sigma(x_2) = b$.

Definition 8 Let $\mathfrak{M} = \langle M, X\rangle$ be a model and σ be an assignment as given in Definition 5 and 6. We define $X, \sigma \models A$ (the **valuation** of a formula A).

(1) $X, \sigma \models T(t)$, iff $t[X, \sigma] \in X$,
$X, \sigma \models t_1 < t_2$, iff $\langle t_1[X, \sigma], t_2[X, \sigma]\rangle \in <^M$,
$X, \sigma \models t_1 = t_2$, iff $t_1[X, \sigma] = t_2[X, \sigma]$.
(2) $X, \sigma \models \neg A$, iff $X, \sigma \models A$ fails, and
$X, \sigma \models (A \wedge B)$, iff both $X, \sigma \models A$ and $X, \sigma \models B$ holds.
(3) $X, \sigma \models \forall v A$, iff for all $a \in D$, $X, \sigma[a/x] \models A$.

As above, for a formula $A(x_1, x_2)$, we will use the notation $X \models A[a/x_1, b/x_2]$ for $X, \sigma \models A$, where σ is such that $\sigma(x_1) = a$ and $\sigma(x_2) = b$.

[4] See, for instance, Gupta [1, p. 9].

3 Acceptable Models and Revision Sequences

The following definition is the central concept in discussion. Note that in this definition, we do not need the stipulation we set up in Item (2) below Definition 5.

Definition 9 Let $\mathfrak{M} = \langle M, X \rangle$ be a model. We define **Θ-neutrality in the ground model** M for all non-logical symbols.[5]

(1) The symbol 0 is Θ-neutral in M, if its interpretation is not in Θ, i.e., $0^M \notin \Theta$.
(2) The predicate $<$ is Θ-neutral in M, if for any $d \in \mathbb{N} \cup \Theta$ and any $A, B \in \Theta$,
$d <^M A$, iff $d <^M B$, and
$A <^M d$, iff $B <^M d$.
(3) The function symbol S is Θ-neutral in M, if the range of S is disjoint from Θ, and for any $A, B \in \Theta$, $S^M(A) = S^M(B)$.
(4) The function symbol $+$ is Θ-neutral in M, if the range of $+$ is disjoint from Θ, and for any $d \in \mathbb{N} \cup \Theta$ and any $A, B \in \Theta$,
$d +^M A = d +^M B$, and
$A +^M d = B +^M d$.
(5) The function symbol $\cdot$ is Θ-neutral in M, if the range of $\cdot$ is disjoint from Θ, and for any $d \in \mathbb{N} \cup \Theta$ and any $A, B \in \Theta$,
$d \cdot^M A = d \cdot^M B$, and
$A \cdot^M d = B \cdot^M d$.

A ground model is called **Θ-acceptable**, if all non-logical symbols are Θ-neutral in this model.[6]

The following examples are used to illustrate the idea behind the requirements set up in the previous definition.

Example 1 Let $\mathfrak{M} = \langle M, X \rangle$ be a model such that $0^M = \neg T(0)$. Let $L = \neg T(0)$. Then

$$\begin{aligned} X \models L, &\text{ iff } X \not\models T(0) && \text{by Definition 8 (4)} \\ &\text{iff } 0^M \notin X && \text{by Definition 8 (3)} \\ &\text{iff } L \notin X. \end{aligned}$$

On the other hand, we have $X \models T\lfloor L \rfloor$, iff $L \in X$. Thus, there is no X such that $\langle M, X \rangle \models T\lfloor L \rfloor \leftrightarrow L$. The sentence L is the liar sentence in $\mathscr{L}^*_{QT}$.

Example 2 Let $\mathfrak{M} = \langle M, X \rangle$ be a model such that for any $a \in D$, $a <^M 0^M$, iff $a = \forall x\,(x < 0 \rightarrow \neg T(x))$. Let $L = \forall x\,(x < 0 \rightarrow \neg T(x))$. Then

$$\begin{aligned} X \models L, &\text{ iff } X \models \forall x\,(x < 0 \rightarrow \neg T(x)) \\ &\text{iff for any } a \in D, \text{ if } a <^M 0^M, \text{ then } X \models \neg T(\bar{a}/x) && \text{by Definition 8 (5)} \\ &\text{iff } L \notin X. \end{aligned}$$

It is clear that $X \models T\lfloor L \rfloor$, iff $L \in X$. Thus, there is no X such that $\langle M, X \rangle \models T\lfloor L \rfloor \leftrightarrow L$. The sentence L is the liar sentence in $\mathscr{L}^*_{QT}$.

Example 3 Fix a sentence A. Let $\mathfrak{M} = \langle M, X \rangle$ be a model such that for any $a \in D$, $S^M(A) = a$, iff $a = \forall x\,(S\lfloor A \rfloor = x \rightarrow \neg T(x))$. Let $L = \forall x\,(S\lfloor A \rfloor = x \rightarrow \neg T(x))$. Then

$$\begin{aligned} X \models L, &\text{ iff } X \models \forall x\,(S\lfloor A \rfloor = x \rightarrow \neg T(x)) \\ &\text{iff for any } a \in D, \text{ if } S^M(A) = a, \text{ then } X \models \neg T(\bar{a}/x) && \text{by Definition 8 (5)} \\ &\text{iff } L \notin X. \end{aligned}$$

[5] See Gupta [1, p. 10], Gupta & Belnap [3, p. 74], Schweizer [8, p. 12], Kremer [5, p. 353], Stern [9, p. 38].
[6] Gupta [1, p. 186]. In Stern [9, p. 38], they are also called proper premodels.

Then, there is no X such that $\langle M, X\rangle \models T\lfloor L\rfloor \leftrightarrow L$. Again, the sentence L is the liar sentence in $\mathscr{L}^*_{QT}$.

Definition 10 Let $\mathfrak{M} = \langle M, X\rangle$ be a model. We define X_α for all ordinal α inductively.

(1) $X_0 = X$.
(2) If $\alpha = \beta + 1$, $X_\alpha = \{A \mid X_\beta \models A\}$.
(3) If α is a limit ordinal, $X_\alpha = \bigcup_{\beta<\alpha} \bigcap_{\gamma\geq\beta} X_\gamma$.

The sequence of the sets X_α $(\alpha \leq \omega)$ is called a **revision sequence** on M.[7]

Example 4 Consider the sentence $L = \neg T(0)$ in Example 1. We have known that for any X, $X \models L$, iff $L \notin X$. Let $X_0 = X$ be the empty set. Then, $X_0 \models L$, and so, $L \in X_1$. Thus, $X_1 \not\models L$. Generally, for any $n \geq 0$, if n is even, $X_n \models L$; if n is odd, $X_n \not\models L$. Therefore, $L \notin X_\omega$.

Notation: $T^n(x)$ is a shorthand for $T(T(\ldots T(x)\ldots))$ where T occurs n times.

Example 5 Let $A = T^n(\lfloor 0 = 0\rfloor)$ and $X_0 = X = \varnothing$. Then $X_0 \models 0 = 0$, and so $0 = 0 \in X_n$ for all $n \geq 1$. In general, $X_n \models T^n(\lfloor 0 = 0\rfloor)$, and so $T^n(\lfloor 0 = 0\rfloor) \in X_k$ for all $k \geq n+1$. $n+1$ is the smallest number k such that $T^n(\lfloor 0 = 0\rfloor)$ appears in X_k.

Example 6 Let $\mathfrak{M} = \langle M, X\rangle$ be a model such that for any $a \in D$, $a <^M 0^M$, iff $a = T^n(\lfloor 0 = 0\rfloor)$ for some $n \geq 0$. Let $L = \forall x\,(x < 0 \to T(x))$. Then

$$\begin{aligned} X \models A, &\text{ iff } X \models \forall x\,(x < 0 \to T(x)) \\ &\text{ iff for any } a \in D, \text{ if } a <^M 0^M, \text{ then } X \models T(\bar{a}/x) \\ &\text{ iff for any } n \geq 0,\ X \models T^{n+1}(\lfloor 0 = 0\rfloor) \end{aligned}$$

Now, let $X_0 = X = \varnothing$. By the previous example, we can see that for any $n \geq 0$, $T^n(\lfloor 0 = 0\rfloor) \in X_\omega$, and so $X_\omega \models T^n(\lfloor 0 = 0\rfloor)$. We obtain $X_\omega \models A$.

At the same time, we note that for any $n \geq 0$, $A \notin X_n$, and so $A \notin X_\omega$. However, since $X_\omega \models A$, $A \in X_{\omega+1}$. Consequently, $\omega + 1$ is the least stage of revision sequence where the sentence A becomes true.

4 Isomorphisms

In this section, we recall the isomorphisms between models, which will be used in proving Lemma 2. For any first order language, we can define the notion of isomorphism to any two models of this language. For simplicity, we will only consider those between models for $\mathscr{L}^*_{QT}$.

Definition 11 For $i = 0, 1$, fix models

$$\mathfrak{M}_i = \left\langle D_i, S^{\mathfrak{M}_i}, +^{\mathfrak{M}_i}, \cdot^{\mathfrak{M}_i}, 0^{\mathfrak{M}_i}, <^{\mathfrak{M}_i}, \left\{\lfloor A\rfloor^{\mathfrak{M}_i} \mid A \in \mathscr{L}^*_{QT}\right\}, T^{\mathfrak{M}_i}\right\rangle.$$

An **isomorphism** between $\mathfrak{M}_0$ and $\mathfrak{M}_1$ is a bijection φ from D_0 to D_1, such that

(1) $\varphi\left(0^{\mathfrak{M}_0}\right) = 0^{\mathfrak{M}_1}$. And for any $A \in \mathscr{L}^*_{QT}$, $\varphi(A) = A$ (Note that $A^{\mathfrak{M}_0} = 0^{\mathfrak{M}_1} = A$).
(2) For any $a, b \in D_0$,
$\varphi\left(S^{\mathfrak{M}_0}(a)\right) = S^{\mathfrak{M}_1}(\varphi(a))$,
$\varphi\left(a +^{\mathfrak{M}_0} b\right) = \varphi(a) +^{\mathfrak{M}_1} \varphi(b)$, and
$\varphi\left(a \cdot^{\mathfrak{M}_0} b\right) = \varphi(a) \cdot^{\mathfrak{M}_1} \varphi(b)$
(3) For any $a, b \in D_0$,
$a <^{\mathfrak{M}_0} b$, iff $\varphi(a) <^{\mathfrak{M}_1} \varphi(b)$,
$a \in T^{\mathfrak{M}_0}$, iff $\varphi(a) \in T^{\mathfrak{M}_1}$.

[7] See Gupta [1, p. 10], Herzberger [4, p. 68].

Theorem 1 *Let φ be an isomorphism between $\mathfrak{M}_0$ and $\mathfrak{M}_1$ given as above. For any formula $A(x_1, \dots, x_n)$ of $\mathscr{L}^*_{QT}$ and for any assignment σ on $\mathfrak{M}_0$,*

$$\mathfrak{M}_0, \sigma \models A, \textit{ iff. } \mathfrak{M}_1, \sigma[\varphi(\sigma(x_1))/x_1, \dots, \varphi(\sigma(x_n))/x_n] \models A.$$

In particular, for any sentence A, $\mathfrak{M}_0 \models A$, iff $\mathfrak{M}_1 \models A$.

The proof of the above theorem is essentially the same as that of Theorem 2, for which we will provide a detailed proof.

Definition 12 Let $\mathfrak{M}_0$ and $\mathfrak{M}_1$ be given as above. We stipulate any non-sentence element of D_0 has degree 0. An ***n*-degree isomorphism** between $\mathfrak{M}_0$ and $\mathfrak{M}_1$ is a bijection φ from D_0 to D_1, such that Items (1) to (3) hold only for those elements of degrees $\leq n$.

Theorem 2 *Let φ be an n-degree isomorphism between $\mathfrak{M}_0$ and $\mathfrak{M}_1$ given as above. For any formula $A(x_1, \dots, x_n)$ of degree $\leq n$ and for any assignment σ on $\mathfrak{M}_0$,*

$$\mathfrak{M}_0, \sigma \models A, \textit{ iff. } \mathfrak{M}_1, \sigma[\varphi(\sigma(x_1))/x_1, \dots, \varphi(\sigma(x_n))/x_n] \models A.$$

In particular, for any sentence A of degree $\leq n$, $\mathfrak{M}_0 \models A$, iff $\mathfrak{M}_1 \models A$.

Proof First, by an easy induction on the structure of t, we can prove that for any term $t(x_1, \dots, x_n)$ with $\rho(t) \leq n$, $\varphi(t[\sigma]) = t[\varphi \circ \sigma]$, that is,

$$\varphi(t[a_1/x_1, \dots, a_n/x_n]) = t[\varphi(a_1)/x_1, \dots, \varphi(a_n)/x_n].$$

Next, we prove by induction on the structure of A that

$$\mathfrak{M}_0 \models A[a_1/x_1, \dots, a_n/x_n], \text{ iff. } \mathfrak{M}_1 \models A[\varphi(a_1)/x_1, \dots, \varphi(a_n)/x_n].$$

We consider the following cases:

(i) When A is atomic, we need to consider three sub-cases: A is $t_1 = t_2$, $t_1 < t_2$, or $T(t)$. We only prove the first sub-case.

$$\begin{aligned}\mathfrak{M}_0 \models A[a_1/x_1, \dots, a_n/x_n], &\text{ iff } t_1[a_1/x_1, \dots, a_n/x_n] = t_2[a_1/x_1, \dots, a_n/x_n] \\ &\text{ iff } \varphi(t_1[a_1/x_1, \dots, a_n/x_n]) = \varphi(t_2[a_1/x_1, \dots, a_n/x_n]) \\ &\text{ iff } t_1[\varphi(a_1)/x_1, \dots, \varphi(a_n)/x_n] = t_2[\varphi(a_1)/x_1, \dots, \varphi(a_n)/x_n] \\ &\text{ iff } \mathfrak{M}_1 \models A[\varphi(a_1)/x_1, \dots, \varphi(a_n)/x_n].\end{aligned}$$

Note that the second "iff" holds because φ is a injective, and the third is due to the result on t that we just prove.

(ii) When A is $\neg B$ or $B \vee C$, the case is easy. We omit the details.

(iii) When A is $\forall x B(x, x_1, \dots, x_n)$, then

$$\begin{aligned}\mathfrak{M}_0 \models A[a_1/x_1, \dots, a_n/x_n], &\text{ iff } \mathfrak{M}_0 \models B[a/x, a_1/x_1, \dots, a_n/x_n] \text{ for all } a \in D \\ &\text{ iff } \mathfrak{M}_1 \models B[\varphi(a)/x, \varphi(a_1)/x_1, \dots, \varphi(a_n)/x_n] \text{ for all } a \in D \\ &\text{ iff } \mathfrak{M}_1 \models B[b/x, \varphi(a_1)/x_1, \dots, \varphi(a_n)/x_n] \text{ for all } b \in D \\ &\text{ iff } \mathfrak{M}_1 \models A[\varphi(a_1)/x_1, \dots, \varphi(a_n)/x_n].\end{aligned}$$

Note that the second "iff" is due to the induction hypothesis, and the third holds because φ is a surjection.

5 Main Results

The following lemma establishes a basic property of revision sequences on a Θ-acceptable model. It is this lemma by which Gupta [1, p. 11] proves that there is a model validating (all instances of) Tarski's schema even though the corresponding language to the model is sufficiently rich, semantically closed and evaluated by the classical two-valued schema.

Lemma 2 *Let M be a Θ-acceptable ground model M and X_α ($\alpha \geq 0$) be a revision sequences on M.*

(1) *For any closed term t and any ordinal $\alpha, \beta \geq 0$, $t[X_\alpha] = t[X_\beta]$.*
(2) *For any sentence A and any ordinal $\alpha \geq \rho(A)+1$, $A \in X_{\rho(A)+1}$, iff $A \in X_\alpha$.*[8]

Proof of Lemma 2

(1) We prove the result by induction on the structure of t.

Case 1: $t = 0$. Then $t[X_\alpha] = t[X_\beta] = 0$.

Case 2: $t = \lfloor A \rfloor$. Then $t[X_\alpha] = t[X_\beta] = A$.

Case 3: $t = Su$. Then $t[X_\alpha] = S^M(u[X_\alpha]) = S^M(u[X_\beta]) = t[X_\beta]$, where the second equality is implied by the induction hypothesis.

Case 4, 5: $t = u + v$ or $t = u \cdot v$. Similar as Case 3.

To sum up, we can conclude that for any closed term t, $t[X_\alpha] = t[X_\beta]$.

(2) We prove the result by induction on the degree of A. Fix a sentence A and suppose the result holds for all sentences B with $\rho(B) < \rho(A)$. That is, the induction hypothesis is that for any B with $\rho(B) < \rho(A)$ and any ordinal $\beta \geq \rho(B)+1$, we already have $B \in X_{\rho(B)+1}$, iff $B \in X_\beta$. We will prove that for any ordinal $\alpha \geq \rho(A)+1$, $A \in X_{\rho(A)+1}$, iff $A \in X_\alpha$. Our proof is a transfinite induction on $\alpha \geq \rho(A)+1$.

Base step. $\alpha = \rho(A)+1$. Trivial.

Induction step. We consider two cases.

Case 1. α is a limit. Then, for any ordinal β with $\rho(A)+1 \leq \beta < \alpha$, by induction hypothesis, $A \in X_{\rho(A)+1}$, iff $A \in X_\beta$. Hence, we can get immediately $A \in X_{\rho(A)+1}$, iff $A \in X_\alpha$.

Case 2. $\alpha = \beta + 1$. Note that since $\alpha \geq \rho(A)+1$, iff $\beta \geq \rho(A)$, what we need to prove is that for all $\beta \geq \rho(A)$, $A \in X_{\rho(A)+1}$, iff $A \in X_{\beta+1}$. In the following, we fix arbitrarily $\beta \geq \rho(A)$.

Note that there are countably infinite many sentences of degree $\geq \rho(A) \geq 1$. Among them, there are countably infinite many ones in $X_{\rho(A)}$ and at the same time, there are also countably infinite many ones out of $X_{\rho(A)}$. The case is similar for X_β. Therefore, we can find a bijection on the set of sentences of degree $\geq \rho(A)$, namely ψ, such that for any B of degree $\geq \rho(A)$,

$$B \in X_{\rho(A)}, \text{ iff } \psi(B) \in X_\beta. \tag{1}$$

We now define a function on D as follows

$$\varphi(a) = \begin{cases} \psi(a), & \text{if } a \text{ is a sentence } B \text{ with } \rho(B) \geq \rho(A); \\ a, & \text{otherwise.} \end{cases}$$

φ is apparently a bijection on D. We claim it is a $\rho(A)$-degree isomorphism between the models $\mathfrak{M}_0 = \langle M, X_{\rho(A)} \rangle$ and $\mathfrak{M}_1 = \langle M, X_\beta \rangle$. Note that the two models have the same ground model $\mathfrak{M}$, which is Θ-acceptable.

We verify that φ meets the requirements for $\rho(A)$-degree isomorphism.

[8] This lemma is a special case of Lemma 3. The latter is the one Gupta originally gave. See Gupta [1, p. 11-14] for the details of Gupta's proof. Note that provided that the degree of a formula started from 0, we would have to state that for any $\alpha \geq \rho(A) + \mathbf{2}$, ... See also the footnote given for the definition of degree.

(1) Since 0 is Θ-neutral in $\mathfrak{M}$, $0^{\mathfrak{M}_0}$ is not a sentence. Hence,

$$\begin{aligned} \varphi\left(0^{\mathfrak{M}_0}\right) &= 0^{\mathfrak{M}_0} && 0^{\mathfrak{M}_0} \text{ is not a sentence} \\ &= 0^{M} && M \text{ is the ground model of } \mathfrak{M}_0 \\ &= 0^{\mathfrak{M}_1}. && M \text{ is the ground model of } \mathfrak{M}_1 \end{aligned}$$

Still, we need to verify that for any $B \in \mathscr{L}^{*}_{QT}$, if the quotation name $\lfloor B \rfloor$ has a degree $\leq \rho(A)$, then $\varphi\left(\lfloor B \rfloor^{\mathfrak{M}_0}\right) = \lfloor B \rfloor^{\mathfrak{M}_1}$. Indeed, suppose $\rho\left(\lfloor B \rfloor\right) \leq \rho(A)$, that is, $\rho\left(B\right) < \rho(A)$, then by definition of φ, we get immediately that $\varphi\left(B\right) = B$.

(2) For any $a, b \in D_0$,
if $a \in \Theta$ with $\rho(a) \geq \rho(A)$, then

$$\begin{aligned} \varphi\left(S^{\mathfrak{M}_0}(a)\right) &= S^{\mathfrak{M}_0}(a) && S^{\mathfrak{M}_0}(a) \notin \Theta \\ &= S^{\mathfrak{M}_1}\left(\psi(a)\right) && \psi(a) \in \Theta \text{ and } S \text{ is } \Theta\text{-neutral} \\ &= S^{\mathfrak{M}_1}\left(\varphi(a)\right). && \text{Definition of } \varphi \end{aligned}$$

Otherwise, then

$$\begin{aligned} \varphi\left(S^{\mathfrak{M}_0}(a)\right) &= S^{\mathfrak{M}_0}(a) && S^{\mathfrak{M}_0}(a) \notin \Theta \\ &= S^{\mathfrak{M}_0}\left(\varphi(a)\right) && \varphi(a) = a \\ &= S^{\mathfrak{M}_1}\left(\varphi(a)\right). && S^{\mathfrak{M}_0} = S^{M} = S^{\mathfrak{M}_1} \end{aligned}$$

In either of the two cases, we obtain $\varphi\left(S^{\mathfrak{M}_0}(a)\right) = S^{\mathfrak{M}_1}\left(\varphi(a)\right)$,
Similarly, we can prove that $\varphi\left(a +^{\mathfrak{M}_0} b)\right) = \varphi(a) +^{\mathfrak{M}_1} \varphi(b)$, and $\varphi\left(a \cdot^{\mathfrak{M}_0} b)\right) = \varphi(a) \cdot^{\mathfrak{M}_1} \varphi(b)$. We leave the details to the reader.

(3) For any $a, b \in D_0$,
if $a, b \in \Theta$ with $\rho(a), \rho(b) \geq \rho(A)$, then

$$\begin{aligned} a <^{\mathfrak{M}_0} b, &\text{ iff } a <^{\mathfrak{M}_0} \psi(b) && \psi(b) \in \Theta \text{ and } < \text{ is } \Theta\text{-neutral} \\ &\text{iff } \psi(a) <^{\mathfrak{M}_0} \psi(b) && \psi(a) \in \Theta \text{ and } < \text{ is } \Theta\text{-neutral} \\ &\text{iff } \varphi(a) <^{\mathfrak{M}_0} \varphi(b). && \text{Definition of } \varphi \end{aligned}$$

If $a \in \Theta$ with $\rho(a) \geq \rho(A)$ and b is either out of Θ or in Θ but $\rho(b) < \rho(A)$, then

$$a <^{\mathfrak{M}_0} b, \text{ iff } \varphi(a) <^{\mathfrak{M}_0} \varphi(b). \qquad \varphi(a) = a, \varphi(b) = b$$

If $b \in \Theta$ with $\rho(a) \geq \rho(A)$ and a is either out of Θ or in Θ but $\rho(b) < \rho(A)$, we can do this case as above.
If $a, b \notin \Theta$ or $a, b \in \Theta$ but $\rho(a), \rho(b) < \rho(A)$,

$$\begin{aligned} a <^{\mathfrak{M}_0} b, &\text{ iff } a <^{\mathfrak{M}_0} \varphi(b) && \varphi(b) = b \\ &\text{iff } \psi(a) <^{\mathfrak{M}_0} \varphi(b) && \psi(a) \in \Theta \text{ and } < \text{ is } \Theta\text{-neutral} \\ &\text{iff } \varphi(a) <^{\mathfrak{M}_0} \varphi(b). && \text{Definition of } \varphi \end{aligned}$$

To sum up, we can conclude that $a <^{\mathfrak{M}_0} b$, iff $\varphi(a) <^{\mathfrak{M}_1} \varphi(b)$.
At last, we prove that $a \in T^{\mathfrak{M}_0}$, iff $\varphi(a) \in T^{\mathfrak{M}_1}$.
If $a \in \Theta$ with $\rho(a) < \rho(A)$, then let $a = B$, and so what we want is that $B \in X_{\rho(A)}$, iff $B \in X_{\beta}$. This is correct by induction hypothesis (about the degree of A), since $\rho(B) < \rho(A)$ and $\beta \geq \rho(A) \geq \rho(B) + 1$.
If $a \in \Theta$ with $\rho(a) \geq \rho(A)$, then let $a = B$, and we notice $\varphi(a) = \psi(B)$. Hence, we only need to prove $B \in X_{\rho(A)}$, iff $\psi(B) \in X_{\beta}$. This is precisely Eq. (1).
If a is not a sentence, the proof is trivial.

So far, we know that φ is an n-degree isomorphism between $\mathfrak{M}_0 = \langle M, X_{\rho(A)}\rangle$ and $\mathfrak{M}_1 = \langle M, X_\beta\rangle$. By Theorem 2, $X_{\rho(A)} \models A$, iff $X_\beta \models A$. In conclusion, $A \in X_{\rho(A)+1}$, iff $A \in X_\alpha$. □

Theorem 3 *Let $\mathfrak{M} = \langle M, X\rangle$ whose ground model is Θ-acceptable. Then for any sentence A, $X_\omega \models T\lfloor A\rfloor \leftrightarrow A$. That is to say, (all instances of) Tarski's schema are satisfied in the model $\langle M, X_\omega\rangle$.*

Proof By Lemma 2, for any sentence A and for any $\alpha \geq \rho(A) + 1$, $A \in X_{\rho(A)+1}$, iff $A \in X_\alpha$. Since $\rho(A)$ is finite, for all A and for all $\alpha \geq \omega$, $A \in X_\omega$, iff $A \in X_\alpha$. In particular, $A \in X_\omega$, iff $A \in X_{\omega+1}$. It follows immediately $X_\omega \models T\lfloor A\rfloor \leftrightarrow A$. □

6 Further Results

From the proof of Lemma 2, we can see immediately that it holds for two different revision sequences on a ground model.

Lemma 3 *Let M be a Θ-acceptable ground model M and X_α, Y_α $(\alpha \geq 0)$ be two revision sequences on M.*

(1) *For any closed term t and any number α, β, $t[X_\alpha] = t[Y_\beta]$.*
(2) *For any sentence A and any $\alpha \geq \rho(A) + 1$, $A \in X_{\rho(A)+1}$, iff $A \in Y_\alpha$.*[9]

Definition 13 A ground model M is **Thomason**, if all of the revision sequences on M eventually converge to one and the same fixed point.[10]

Theorem 4 *All Θ-acceptable ground models are Thomason.*[11]

In Definition 9, we can use any subset Σ of Θ instead of Θ itself. In that case, we give the notion of Σ-neutrality. Note that Θ-neutrality is the special case when $\Sigma = \Theta$. A ground model is **Σ-acceptable**, if all of the non-logical symbols are Σ-neutral.

Definition 14 Let M be a ground model. We say a sentence A is **stably true/false on** M, if for any $X \subseteq D$, there is an ordinal α such that for any $\beta \geq \alpha$, $A \in X_\beta / A \notin X_\beta$. A sentence is **stable on** M, if it is either stably true or stably false on M. A sentence is **unstable on** M, if it is not stable on M.[12]

The following is a generalization of Lemma 3.

Lemma 4 *Let M be a Σ-acceptable ground model M and X_α, Y_α $(\alpha \geq 0)$ be two revision sequences on M. Suppose $\Sigma(\subseteq \Theta)$ includes all unstable sentences on M.*

Then, for any sentence A, there is an ordinal ρ, such that for any $\alpha \geq \rho$, $A \in X_\rho$, iff $A \in Y_\alpha$.[13]

Definition 15 A is **paradoxical on** M, if for any $X \subseteq D$, neither there is an ordinal α such that for any $\beta \geq \alpha$, $A \in X_\beta$, nor is there an ordinal α such that for any $\beta \geq \alpha$, $A \notin X_\beta$. [14]

Note that if a sentence is paradoxical on M, it must be unstable on M. But, the converse is not necessary. We close our discussion of Gupta's acceptable models by leaving the following two conjectures:[15]

9 See Gupta [1, p. 11].
10 See Gupta [1, p. 42], Gupta & Belnap [3, p. 209], Kremer [5, p. 352].
11 Gupta [1, p. 42], Gupta & Belnap [3, p. 210]
12 See Gupta [1, p. 46].
13 See Gupta & Belnap [3, p. 202].
14 See Gupta [1, p. 49].
15 There are other research directions about Gupta's acceptable models. See, for instance, [2].

Conjecture 1 Let M be a Σ-acceptable ground model M for some $\Sigma(\subseteq \Theta)$ including all paradoxical sentences on M. Let X_α ($\alpha \geq 0$) be a revision sequence on M. Then, for any sentence A, there is an ordinal ρ, such that for any $\alpha \geq \rho$, $A \in X_\rho$, iff $A \in X_\alpha$.[16]

Conjecture 2 Let $\mathfrak{M} = \langle M, X \rangle$ whose ground model is Σ-acceptable for some $\Sigma(\subseteq \Theta)$ including all paradoxical sentences on M. Then for any sentence A, $X_\omega \models T\lfloor A \rfloor \leftrightarrow A$.

Appendix

The following paragraph shows how Gupta introduces the quotation names:

> "Let L be a classical first-order language that contains besides various predicates, function symbols and names, a special one-place logical predicate T and quotation marks ⌜'⌝. and ⌜'⌝. We allow in L quotation names for closed formulas (sentences) only. ...... Let M (= (D, I)) be a model of L. ... I assigns to a quotation name ⌜'*A*'⌝ the sentence." (Gupta [1, p. 9]. Note that ⌜ ⌝ is used by Gupta as the quotation marks in the meta-language (English). So, "'A'" is a counterpart of the current notation "$\lfloor A \rfloor$".)

We can see that on the one hand, Gupta takes the quotation marks as primitive symbols in the language L. On the other hand, when he set up the model for L, the quotation marks are dismissed, and they appear as components only in the quotation names. Now, the quotation names are taken as primitive symbols instead, being interpreted in the model. Clearly, this is somewhat confusing.

Along Gupta's line of thought, people usually take it for granted that quotation names can be straightforwardly brought into a language as the primitive symbols. For instance, we cite a paragraph from Kremer [6] as follows:

> "Let us consider a first-order language L, with connective &, $\vee$, and $\neg$, quantifiers $\forall$ and $\exists$, the equals sign =, variables, and some **stock of names** [my hightlight], function symbols and relation symbols. We will say that L is a truth language, if it has a distinguished predicate T and quotation marks ' and ', which will be used to form quote names: if A is a sentence of L, then 'A' is a name. Let SentL = $\{A : A$ is a sentence of $L\}$."

In this quoted text, among the stock of names, how do the quotation names 'A' come from? Kremer has no word about it, as if these quotation names had been already presented in the language as its primitive symbols. However, these names are precisely the names of the sentences in this language. How could there be these names before the formation of these sentences? And don't we need these names as component symbols to form these sentences? Clearly, there is a vicious circle here. To break this circle, We need to figure out whether the sentences depend on the quotation names, or the quotation names depend on the sentences in a language.

There is another excerpt here that also takes the introduction of quotation names as for granted.

> "Formally, suppose that L is an uninterpreted first order language. $M = \langle D, I \rangle$ is a classical model for L iff D is a nonempty set and I is a function assigning to each name of L a member of D, to each n-place function symbol of L an n-place function on D, and to each n-place relation symbol a function from D^n to $\{\mathbf{t}, \mathbf{f}\}$. Suppose that L and L^+ are first-order languages, where L^+ is L expanded with a distinguished predicate (one-place relation symbol) T, and where L has a quote name 'A' for each sentence A of L^+. Then L and L^+ are a corresponding ground language and truth language." (Kremer [5, p. 346])

References

1. Gupta, A. (1982). Truth and paradox. *Journal of Philosophical Logic*, *11*(1), 1–60.

[16] Is there any literature about this result?

2. Gupta, A. (1990). Two theorems concerning stability. In J. Dunn, & A. Gupta (Eds.) *Truth or Consequences : Essays in Honor of Nuel Belnap*, (pp. 49–60). Dordrecht: Kluwer Academic.
3. Gupta, A., & Belnap, N. (1993). *The Revision Theory of Truth*. Cambridge: MIT Press.
4. Herzberger, H. G. (1982). Notes on naive semantics. *Journal of Philosophical Logic*, *11*(1), 61–102.
5. Kremer, P. (2010). How truth behaves when there's no vicious reference. *Journal of Philosophical Logic*, *39*(4), 345–367.
6. Kremer, P. (2016). The Revision Theory of Truth. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2016 ed.
7. Kripke, S. A. (1975). Outline of a theory of truth. *Journal of Philosophy*, *72*(19), 690–712.
8. Schweizer, P. (1992). A syntactical approach to modality. *Journal of Philosophical Logic*, *21*(1), 1–31.
9. Stern, J. (2016). *Toward Predicate Approaches to Modality*. Cham: Springer.
10. Tarski, A. (1944). The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research*, *4*(3), 341–376.
11. Yaqūb, A. M. (1993). *The Liar Speaks the Truth: A Defense of the Revision Theory of Truth*. New York: Oxford University Press.