

# The Prospects of Formal Philosophy

Hannes Leitgeb

University of Bristol

January 2009

# Introduction

Many philosophers still deny that mathematical methods can play a substantive role in philosophy.

Many philosophers still deny that mathematical methods can play a substantive role in philosophy.

Paradigmatic case: Kant in the *Critique of Pure Reason*.

According to Kant's *Transcendental Doctrine of Method*, philosophy cannot be developed according to the *definitions-axioms-proofs* scheme.

Many philosophers still deny that mathematical methods can play a substantive role in philosophy.

Paradigmatic case: Kant in the *Critique of Pure Reason*.

According to Kant's *Transcendental Doctrine of Method*, philosophy cannot be developed according to the *definitions-axioms-proofs* scheme.

This is because mathematics deals with pure intuitions, whereas philosophy deals with pure concepts (or so Kant says).

Many philosophers still deny that mathematical methods can play a substantive role in philosophy.

Paradigmatic case: Kant in the *Critique of Pure Reason*.

According to Kant's *Transcendental Doctrine of Method*, philosophy cannot be developed according to the *definitions-axioms-proofs* scheme.

This is because mathematics deals with pure intuitions, whereas philosophy deals with pure concepts (or so Kant says).

However,

- in the meantime, mathematics has developed into a theory of abstract structures in general,
- the progress in logic shows that the “space of concepts” has itself an intricate mathematical structure.

Claims:

When philosophical theories get sufficiently complex, they are in need of mathematics.

Claims:

When philosophical theories get sufficiently complex, they are in need of mathematics.

The necessary bridge between philosophy and mathematics is often supplied by logic broadly understood.

Claims:

When philosophical theories get sufficiently complex, they are in need of mathematics.

The necessary bridge between philosophy and mathematics is often supplied by logic broadly understood.

**Plan of the talk:** *Examples*

Claims:

When philosophical theories get sufficiently complex, they are in need of mathematics.

The necessary bridge between philosophy and mathematics is often supplied by logic broadly understood.

**Plan of the talk:** *Examples*

- Is it Possible to Reconstruct Properties from Similarity?
- Are There True but Absolutely Unprovable Statements?
- Do Connectionism and Symbolic Computationalism Exclude Each Other?
- Can We Justify Probability Theory from Closeness to the Truth?
- (If you run out of questions later:)  
Is There a Probabilistic Way Out of Semantic Paradoxes?

Claims:

When philosophical theories get sufficiently complex, they are in need of mathematics.

The necessary bridge between philosophy and mathematics is often supplied by logic broadly understood.

**Plan of the talk:** *Examples*

- Is it Possible to Reconstruct Properties from Similarity?
- Are There True but Absolutely Unprovable Statements?
- Do Connectionism and Symbolic Computationalism Exclude Each Other?
- Can We Justify Probability Theory from Closeness to the Truth?
- (If you run out of questions later:)  
Is There a Probabilistic Way Out of Semantic Paradoxes?

Started by the classics – Leibniz, Frege, Russell, Ramsey, Carnap, . . . – mathematical philosophy has recently come to life again!

# Is it Possible to Reconstruct Properties from Similarity?

G. Frege, A.N. Whitehead & B. Russell: definition by logical abstraction;  
abstracting equivalence classes from equivalence relations

# Is it Possible to Reconstruct Properties from Similarity?

G. Frege, A.N. Whitehead & B. Russell: definition by logical abstraction;  
abstracting equivalence classes from equivalence relations

B. Russell, R. Carnap: generalized abstraction;  
abstracting “similarity classes” from similarity relations

# Is it Possible to Reconstruct Properties from Similarity?

G. Frege, A.N. Whitehead & B. Russell: definition by logical abstraction;  
abstracting equivalence classes from equivalence relations

B. Russell, R. Carnap: generalized abstraction;  
abstracting “similarity classes” from similarity relations

↪ abstraction for the *empirical* domain (where transitivity often fails)

# Is it Possible to Reconstruct Properties from Similarity?

G. Frege, A.N. Whitehead & B. Russell: definition by logical abstraction;  
abstracting equivalence classes from equivalence relations

B. Russell, R. Carnap: generalized abstraction;  
abstracting “similarity classes” from similarity relations

↪ abstraction for the *empirical* domain (where transitivity often fails)

- $\langle S, \sim \rangle$  is a similarity structure on  $S$  :iff  
 $\sim \subseteq S \times S$  is reflexive and symmetric.

# Is it Possible to Reconstruct Properties from Similarity?

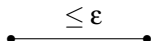
G. Frege, A.N. Whitehead & B. Russell: definition by logical abstraction;  
abstracting equivalence classes from equivalence relations

B. Russell, R. Carnap: generalized abstraction;  
abstracting “similarity classes” from similarity relations

↪ abstraction for the *empirical* domain (where transitivity often fails)

- $\langle S, \sim \rangle$  is a similarity structure on  $S$  :iff  
 $\sim \subseteq S \times S$  is reflexive and symmetric.

E.g.: *metric* similarity (for colours, . . .)



# Is it Possible to Reconstruct Properties from Similarity?

G. Frege, A.N. Whitehead & B. Russell: definition by logical abstraction;  
abstracting equivalence classes from equivalence relations

B. Russell, R. Carnap: generalized abstraction;  
abstracting “similarity classes” from similarity relations

↪ abstraction for the *empirical* domain (where transitivity often fails)

- $\langle S, \sim \rangle$  is a similarity structure on  $S$  :iff  
 $\sim \subseteq S \times S$  is reflexive and symmetric.

E.g.: *metric* similarity (for colours, . . .)

- $\langle S, P \rangle$  is a property structure on  $S$  :iff  
 $P$  is a class of subsets of  $S$ ,  $\emptyset \notin P$ , and  
for every  $x \in S$  there is an  $X \in P$ , s.t.  $x \in X$ .

# Is it Possible to Reconstruct Properties from Similarity?

G. Frege, A.N. Whitehead & B. Russell: definition by logical abstraction;  
abstracting equivalence classes from equivalence relations

B. Russell, R. Carnap: generalized abstraction;  
abstracting “similarity classes” from similarity relations

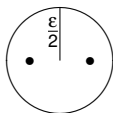
↪ abstraction for the *empirical* domain (where transitivity often fails)

- $\langle S, \sim \rangle$  is a similarity structure on  $S$  :iff  
 $\sim \subseteq S \times S$  is reflexive and symmetric.

E.g.: *metric* similarity (for colours, . . .)

- $\langle S, P \rangle$  is a property structure on  $S$  :iff  
 $P$  is a class of subsets of  $S$ ,  $\emptyset \notin P$ , and  
for every  $x \in S$  there is an  $X \in P$ , s.t.  $x \in X$ .

E.g.: (colour) spheres



Every property structure on  $S$  determines a similarity structure on  $S$ :  
(cf. Leibniz: “Peter is similar to Paul” reduces to “Peter is  $A$  now and Paul is  $A$  now”)

### Definition (Determined similarity structure)

$\langle S, \sim_P \rangle$  is determined by  $\langle S, P \rangle$  :iff

for all  $x, y \in S$ :  $x \sim_P y$  iff there is an  $X \in P$ , such that  $x, y \in X$ .

Every property structure on  $S$  determines a similarity structure on  $S$ :  
(cf. Leibniz: “Peter is similar to Paul” reduces to “Peter is  $A$  now and Paul is  $A$  now”)

### Definition (Determined similarity structure)

$\langle S, \sim_P \rangle$  is determined by  $\langle S, P \rangle$  :iff

for all  $x, y \in S$ :  $x \sim_P y$  iff there is an  $X \in P$ , such that  $x, y \in X$ .

Every similarity structure on  $S$  determines a property structure on  $S$ :

$X \subseteq S$  is a *clique* of  $\langle S, \sim \rangle$  :iff for all  $x, y \in X$ :  $x \sim y$ .

Every property structure on  $S$  determines a similarity structure on  $S$ :  
(cf. Leibniz: “Peter is similar to Paul” reduces to “Peter is  $A$  now and Paul is  $A$  now”)

### Definition (Determined similarity structure)

$\langle S, \sim_P \rangle$  is determined by  $\langle S, P \rangle$  :iff

for all  $x, y \in S$ :  $x \sim_P y$  iff there is an  $X \in P$ , such that  $x, y \in X$ .

Every similarity structure on  $S$  determines a property structure on  $S$ :

$X \subseteq S$  is a *clique* of  $\langle S, \sim \rangle$  :iff for all  $x, y \in X$ :  $x \sim y$ .

### Definition (Determined property structure)

$\langle S, P^\sim \rangle$  is determined by  $\langle S, \sim \rangle$  :iff

$P^\sim = \{X \subseteq S \mid X \text{ is a maximal clique of } \langle S, \sim \rangle\}$ .

Every property structure on  $S$  determines a similarity structure on  $S$ :  
(cf. Leibniz: “Peter is similar to Paul” reduces to “Peter is  $A$  now and Paul is  $A$  now”)

### Definition (Determined similarity structure)

$\langle S, \sim_P \rangle$  is determined by  $\langle S, P \rangle$  :iff

for all  $x, y \in S$ :  $x \sim_P y$  iff there is an  $X \in P$ , such that  $x, y \in X$ .

Every similarity structure on  $S$  determines a property structure on  $S$ :

$X \subseteq S$  is a *clique* of  $\langle S, \sim \rangle$  :iff for all  $x, y \in X$ :  $x \sim y$ .

### Definition (Determined property structure)

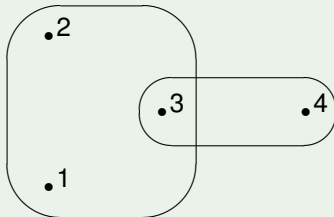
$\langle S, P^\sim \rangle$  is determined by  $\langle S, \sim \rangle$  :iff

$P^\sim = \{X \subseteq S \mid X \text{ is a maximal clique of } \langle S, \sim \rangle\}$ .

Special case: equivalence classes  $\Leftrightarrow$  equivalence relations ✓

## Example

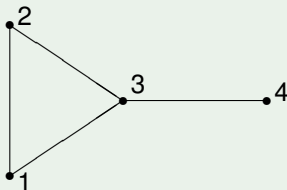
(Faithful, full)  $S_1 = \{1, 2, 3, 4\}$ ,  $P_1 = \{\{1, 2, 3\}, \{3, 4\}\}$ :



Given:  $\langle S_1, P_1 \rangle$

## Example

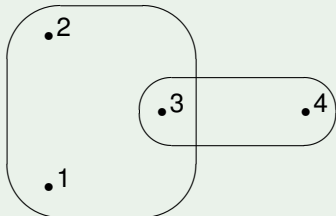
(Faithful, full)  $S_1 = \{1, 2, 3, 4\}$ ,  $P_1 = \{\{1, 2, 3\}, \{3, 4\}\}$ :



Determined:  $\langle S_1, \sim_{P_1} \rangle$

## Example

(Faithful, full)  $S_1 = \{1, 2, 3, 4\}$ ,  $P_1 = \{\{1, 2, 3\}, \{3, 4\}\}$ :

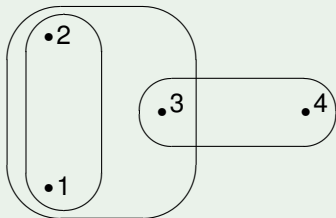


Determined:  $\langle S_1, P^{\sim P_1} \rangle = \langle S_1, P_1 \rangle \checkmark$

But this method of abstraction does not work in *all* cases...  
(as N. Goodman and others observed)

## Example

(Faithful, not full)  $S_2 = \{1, 2, 3, 4\}$ ,  $P_2 = \{\{1, 2\}, \{1, 2, 3\}, \{3, 4\}\}$ :

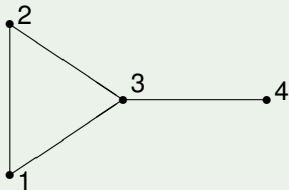


Given:  $\langle S_2, P_2 \rangle$

But this method of abstraction does not work in *all* cases...  
(as N. Goodman and others observed)

## Example

(Faithful, not full)  $S_2 = \{1, 2, 3, 4\}$ ,  $P_2 = \{\{1, 2\}, \{1, 2, 3\}, \{3, 4\}\}$ :

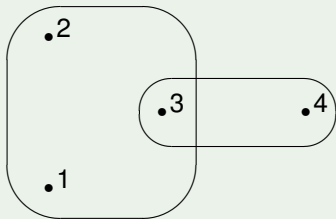


Determined:  $\langle S_2, \sim_{P_2} \rangle$

But this method of abstraction does not work in *all* cases...  
(as N. Goodman and others observed)

## Example

(Faithful, not full)  $S_2 = \{1, 2, 3, 4\}$ ,  $P_2 = \{\{1, 2\}, \{1, 2, 3\}, \{3, 4\}\}$ :

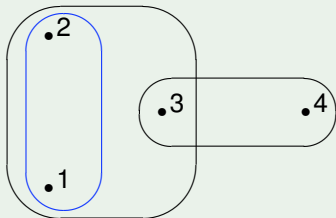


Determined:  $\langle S_2, P^{\sim P_2} \rangle$

But this method of abstraction does not work in *all* cases...  
(as N. Goodman and others observed)

## Example

(Faithful, not full)  $S_2 = \{1, 2, 3, 4\}$ ,  $P_2 = \{\{1, 2\}, \{1, 2, 3\}, \{3, 4\}\}$ :

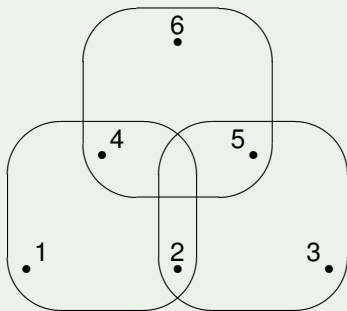


$$\langle S_2, P^{\sim P_2} \rangle \neq \langle S_2, P_2 \rangle$$

## Example

(Full, not faithful)

$$S_3 = \{1, 2, 3, 4, 5, 6\}, P_3 = \{\{1, 2, 4\}, \{2, 3, 5\}, \{4, 5, 6\}\}:$$

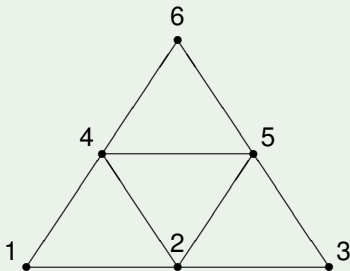


Given:  $\langle S_3, P_3 \rangle$

## Example

(Full, not faithful)

$$S_3 = \{1, 2, 3, 4, 5, 6\}, P_3 = \{\{1, 2, 4\}, \{2, 3, 5\}, \{4, 5, 6\}\}:$$

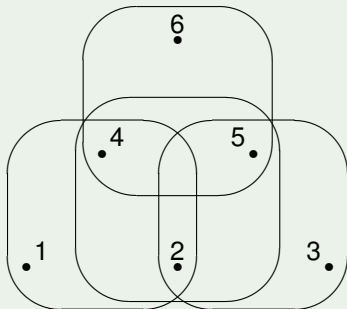


Determined:  $\langle S_3, \sim_{P_3} \rangle$

## Example

(Full, not faithful)

$$S_3 = \{1, 2, 3, 4, 5, 6\}, P_3 = \{\{1, 2, 4\}, \{2, 3, 5\}, \{4, 5, 6\}\}:$$

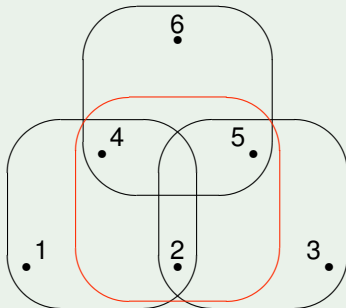


Determined:  $\langle S_3, P^{\sim P_3} \rangle$

## Example

(Full, not faithful)

$S_3 = \{1, 2, 3, 4, 5, 6\}$ ,  $P_3 = \{\{1, 2, 4\}, \{2, 3, 5\}, \{4, 5, 6\}\}$ :

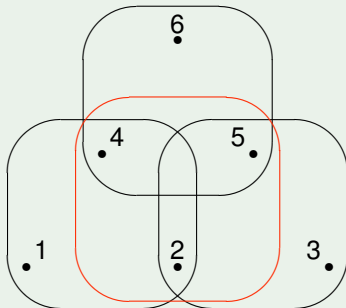


$\langle S_3, P^{\sim P_3} \rangle \neq \langle S_3, P_3 \rangle$

## Example

(Full, not faithful)

$$S_3 = \{1, 2, 3, 4, 5, 6\}, P_3 = \{\{1, 2, 4\}, \{2, 3, 5\}, \{4, 5, 6\}\}$$



$$\langle S_3, P^{\sim P_3} \rangle \neq \langle S_3, P_3 \rangle$$

QUESTION: If similarity is determined by properties, under which conditions can the latter be reconstructed from the former?

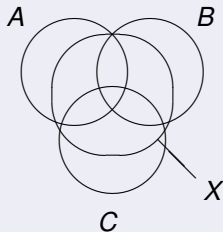
↪ *Hypergraph theory!*

↪ *Hypergraph theory!*

## Theorem (Gilmore; cf. Berge 1989)

Let  $\langle S, \sim_P \rangle$  be determined by  $\langle S, P \rangle$ , with  $S$  finite:

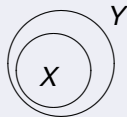
- $\langle S, \sim_P \rangle$  is faithful with respect to  $\langle S, P \rangle$  iff  
(a) for all  $A, B, C \in P$  there is an  $X \in P$ , such that  
 $(A \cap B) \cup (A \cap C) \cup (B \cap C) \subseteq X$



## Theorem (Gilmore; cf. Berge 1989)

Let  $\langle S, \sim_P \rangle$  be determined by  $\langle S, P \rangle$ , with  $S$  finite:

- 1  $\langle S, \sim_P \rangle$  is faithful with respect to  $\langle S, P \rangle$  iff
  - (a) for all  $A, B, C \in P$  there is an  $X \in P$ , such that  $(A \cap B) \cup (A \cap C) \cup (B \cap C) \subseteq X$
- 2 If  $\langle S, \sim_P \rangle$  is full with respect to  $\langle S, P \rangle$ , then
  - (b) there are no  $X, Y \in P$ , such that  $X \subsetneq Y$ .



## Theorem (Gilmore; cf. Berge 1989)

Let  $\langle S, \sim_P \rangle$  be determined by  $\langle S, P \rangle$ , with  $S$  finite:

- 1  $\langle S, \sim_P \rangle$  is faithful with respect to  $\langle S, P \rangle$  iff  
(a) for all  $A, B, C \in P$  there is an  $X \in P$ , such that  
 $(A \cap B) \cup (A \cap C) \cup (B \cap C) \subseteq X$
- 2 If  $\langle S, \sim_P \rangle$  is full with respect to  $\langle S, P \rangle$ , then  
(b) there are no  $X, Y \in P$ , such that  $X \subsetneq Y$ .
- 3  $\langle S, \sim_P \rangle$  is faithful & full with respect to  $\langle S, P \rangle$  iff (a)+(b).

## Theorem (Gilmore; cf. Berge 1989)

Let  $\langle S, \sim_P \rangle$  be determined by  $\langle S, P \rangle$ , with  $S$  finite:

- 1  $\langle S, \sim_P \rangle$  is faithful with respect to  $\langle S, P \rangle$  iff  
(a) for all  $A, B, C \in P$  there is an  $X \in P$ , such that  
 $(A \cap B) \cup (A \cap C) \cup (B \cap C) \subseteq X$
- 2 If  $\langle S, \sim_P \rangle$  is full with respect to  $\langle S, P \rangle$ , then  
(b) there are no  $X, Y \in P$ , such that  $X \subsetneq Y$ .
- 3  $\langle S, \sim_P \rangle$  is faithful & full with respect to  $\langle S, P \rangle$  iff (a)+(b).

Proof: By induction over  $|S|$ .

↪ *Hypergraph theory!*

## Theorem (Gilmore; cf. Berge 1989)

Let  $\langle S, \sim_P \rangle$  be determined by  $\langle S, P \rangle$ , with  $S$  finite:

- 1  $\langle S, \sim_P \rangle$  is faithful with respect to  $\langle S, P \rangle$  iff  
(a) for all  $A, B, C \in P$  there is an  $X \in P$ , such that  
 $(A \cap B) \cup (A \cap C) \cup (B \cap C) \subseteq X$
- 2 If  $\langle S, \sim_P \rangle$  is full with respect to  $\langle S, P \rangle$ , then  
(b) there are no  $X, Y \in P$ , such that  $X \subsetneq Y$ .
- 3  $\langle S, \sim_P \rangle$  is faithful & full with respect to  $\langle S, P \rangle$  iff (a)+(b).

Proof: By induction over  $|S|$ .

Hence: Carnap, *The Logical Structure of the World*: ✓

Russell, *Our Knowledge of the External World*: ✓ (Helly's Theorem)

## Theorem (Gilmore; cf. Berge 1989)

Let  $\langle S, \sim_P \rangle$  be determined by  $\langle S, P \rangle$ , with  $S$  finite:

- 1  $\langle S, \sim_P \rangle$  is faithful with respect to  $\langle S, P \rangle$  iff  
(a) for all  $A, B, C \in P$  there is an  $X \in P$ , such that  
 $(A \cap B) \cup (A \cap C) \cup (B \cap C) \subseteq X$
- 2 If  $\langle S, \sim_P \rangle$  is full with respect to  $\langle S, P \rangle$ , then  
(b) there are no  $X, Y \in P$ , such that  $X \subsetneq Y$ .
- 3  $\langle S, \sim_P \rangle$  is faithful & full with respect to  $\langle S, P \rangle$  iff (a)+(b).

Proof: By induction over  $|S|$ .

Hence: Carnap, *The Logical Structure of the World*: ✓

Russell, *Our Knowledge of the External World*: ✓ (Helly's Theorem)

If determination starts with similarity, then “reconstruction” works.

## Theorem (Gilmore; cf. Berge 1989)

Let  $\langle S, \sim_P \rangle$  be determined by  $\langle S, P \rangle$ , with  $S$  finite:

- 1  $\langle S, \sim_P \rangle$  is faithful with respect to  $\langle S, P \rangle$  iff  
(a) for all  $A, B, C \in P$  there is an  $X \in P$ , such that  
 $(A \cap B) \cup (A \cap C) \cup (B \cap C) \subseteq X$
- 2 If  $\langle S, \sim_P \rangle$  is full with respect to  $\langle S, P \rangle$ , then  
(b) there are no  $X, Y \in P$ , such that  $X \subsetneq Y$ .
- 3  $\langle S, \sim_P \rangle$  is faithful & full with respect to  $\langle S, P \rangle$  iff (a)+(b).

Proof: By induction over  $|S|$ .

Hence: Carnap, *The Logical Structure of the World*: ✓

Russell, *Our Knowledge of the External World*: ✓ (Helly's Theorem)

If determination starts with similarity, then "reconstruction" works.

If  $n = |S|$  and  $|P| > \binom{n}{\lfloor n/2 \rfloor}$  then  $\langle S, \sim_P \rangle$  is not full. (Sperner's Theorem)

## Theorem (Gilmore; cf. Berge 1989)

Let  $\langle S, \sim_P \rangle$  be determined by  $\langle S, P \rangle$ , with  $S$  finite:

- 1  $\langle S, \sim_P \rangle$  is faithful with respect to  $\langle S, P \rangle$  iff  
(a) for all  $A, B, C \in P$  there is an  $X \in P$ , such that  
 $(A \cap B) \cup (A \cap C) \cup (B \cap C) \subseteq X$
- 2 If  $\langle S, \sim_P \rangle$  is full with respect to  $\langle S, P \rangle$ , then  
(b) there are no  $X, Y \in P$ , such that  $X \subsetneq Y$ .
- 3  $\langle S, \sim_P \rangle$  is faithful & full with respect to  $\langle S, P \rangle$  iff (a)+(b).

Proof: By induction over  $|S|$ .

Hence: Carnap, *The Logical Structure of the World*: ✓

Russell, *Our Knowledge of the External World*: ✓ (Helly's Theorem)

If determination starts with similarity, then "reconstruction" works.

If  $n = |S|$  and  $|P| > \binom{n}{\lfloor n/2 \rfloor}$  then  $\langle S, \sim_P \rangle$  is not full. (Sperner's Theorem)

(See Leitgeb 2007, JPL. In work: Monograph on a new *Logischer Aufbau*.)

# Are There True but Absolutely Unprovable Statements?

Equivalently: Are there absolutely undecidable statements? (Gödel 1951)

# Are There True but Absolutely Unprovable Statements?

Equivalently: Are there absolutely undecidable statements? (Gödel 1951)

According to Hilbert's famous *non ignorabimus* claim: No!

# Are There True but Absolutely Unprovable Statements?

Equivalently: Are there absolutely undecidable statements? (Gödel 1951)

According to Hilbert's famous *non ignorabimus* claim: No!

The question is *not* settled by the Incompleteness Theorems, although:

- If the set of absolutely provable statements is recursively enumerable, then truth exceeds absolute provability extensionally (by the First Incompleteness Theorem).

# Are There True but Absolutely Unprovable Statements?

Equivalently: Are there absolutely undecidable statements? (Gödel 1951)

According to Hilbert's famous *non ignorabimus* claim: No!

The question is *not* settled by the Incompleteness Theorems, although:

- If the set of absolutely provable statements is recursively enumerable, then truth exceeds absolute provability extensionally (by the First Incompleteness Theorem).
- If the set of absolutely provable statements is recursively enumerable, then it is not absolutely provable of a particular Turing machine that it enumerates all and only absolutely provable statements (by the Second Incompleteness Theorem).

# Are There True but Absolutely Unprovable Statements?

Equivalently: Are there absolutely undecidable statements? (Gödel 1951)

According to Hilbert's famous *non ignorabimus* claim: No!

The question is *not* settled by the Incompleteness Theorems, although:

- If the set of absolutely provable statements is recursively enumerable, then truth exceeds absolute provability extensionally (by the First Incompleteness Theorem).
- If the set of absolutely provable statements is recursively enumerable, then it is not absolutely provable of a particular Turing machine that it enumerates all and only absolutely provable statements (by the Second Incompleteness Theorem).

Let us assume for the moment the answer is 'yes': how could we then argue in favor of the existence claim

$$\text{HG } \exists p(p \wedge \neg \Box p)$$

(where  $\Box$  expresses 'it is absolutely provable that')

Either (i) by inductive evidence or (ii) by proof.

Either (i) by inductive evidence or (ii) by proof.

Let us focus on (ii):

– We certainly cannot prove HG by proving one instance of HG, i.e., a statement of the form

$$A \wedge \neg \Box A$$

Either (i) by inductive evidence or (ii) by proof.

Let us focus on (ii):

– We certainly cannot prove HG by proving one instance of HG, i.e., a statement of the form

$$A \wedge \neg \Box A$$

since

$$\Box(A \wedge \neg \Box A)$$

is inconsistent in the modal logic S4.

– But it might be possible to prove weaker claims which still entail HG, such as

- $(A \wedge \neg \Box A) \vee (B \wedge \neg \Box B)$
- $(A \wedge \neg \Box A) \vee (\neg A \wedge \neg \Box \neg A)$
- $\neg \Box A \wedge \neg \Box \neg A$

which are indeed consistent in S4.

– But it might be possible to prove weaker claims which still entail HG, such as

- $(A \wedge \neg \Box A) \vee (B \wedge \neg \Box B)$
- $(A \wedge \neg \Box A) \vee (\neg A \wedge \neg \Box \neg A)$
- $\neg \Box A \wedge \neg \Box \neg A$

which are indeed consistent in S4.

Here is an idea of how to do so (this is joint work with L. Horsten, 2007):

- (i) Formalize a version of the Church-Turing Thesis (CT) by means of the absolute provability operator (Shapiro 1985, Reinhardt 1986);
- (ii) add this formalization (ECT) to S4 and background mathematics;
- (iii) derive some statement such as the above in the resulting system.

- Language:

- $S(x)$ :  $x$  is a purely mathematical sentence
- $Proof(x, y)$ :  $x$  is an absolute proof of  $y$  (for  $y$  mathematical)
- $\Box A$ : it is absolutely provable that  $A$  (for  $A$  arbitrary)
- + “Sufficient” syntax to express:  
 $TM(x)$ :  $x$  is a Turing machine, etc.

- Language:

- $S(x)$ :  $x$  is a purely mathematical sentence
- $Proof(x, y)$ :  $x$  is an absolute proof of  $y$  (for  $y$  mathematical)
- $\Box A$ : it is absolutely provable that  $A$  (for  $A$  arbitrary)
- + “Sufficient” syntax to express:  
 $TM(x)$ :  $x$  is a Turing machine, etc.

- Principles:

T  $\Box A \rightarrow A$

K  $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

- Language:

- $S(x)$ :  $x$  is a purely mathematical sentence
- $Proof(x, y)$ :  $x$  is an absolute proof of  $y$  (for  $y$  mathematical)
- $\Box A$ : it is absolutely provable that  $A$  (for  $A$  arbitrary)
- + “Sufficient” syntax to express:  
 $TM(x)$ :  $x$  is a Turing machine, etc.

- Principles:

T  $\Box A \rightarrow A$

K  $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

*Proof*- $\Box S(\ulcorner A \urcorner) \rightarrow (\Box A \leftrightarrow \exists y Proof(y, \ulcorner A \urcorner))$

- Language:

- $S(x)$ :  $x$  is a purely mathematical sentence
- $Proof(x, y)$ :  $x$  is an absolute proof of  $y$  (for  $y$  mathematical)
- $\Box A$ : it is absolutely provable that  $A$  (for  $A$  arbitrary)
- + “Sufficient” syntax to express:  
 $TM(x)$ :  $x$  is a Turing machine, etc.

- Principles:

T  $\Box A \rightarrow A$

K  $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

*Proof*- $\Box$   $S(\ulcorner A \urcorner) \rightarrow (\Box A \leftrightarrow \exists y Proof(y, \ulcorner A \urcorner))$

Nec 
$$\frac{\vdash A}{\vdash \Box A}$$

- Language:

- $S(x)$ :  $x$  is a purely mathematical sentence
- $Proof(x, y)$ :  $x$  is an absolute proof of  $y$  (for  $y$  mathematical)
- $\Box A$ : it is absolutely provable that  $A$  (for  $A$  arbitrary)
- + “Sufficient” syntax to express:  
 $TM(x)$ :  $x$  is a Turing machine, etc.

- Principles:

$$T \quad \Box A \rightarrow A$$

$$K \quad \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$$

$$Proof\text{-}\Box \quad S(\ulcorner A \urcorner) \rightarrow (\Box A \leftrightarrow \exists y Proof(y, \ulcorner A \urcorner))$$

$$Nec \quad \frac{\vdash A}{\vdash \Box A}$$

$$ECT \quad \Box \forall x \exists y \Box \varphi(x, y) \rightarrow \exists e [TM(e) \wedge \forall x \varphi(x, e(x))]$$

- Language:

- $S(x)$ :  $x$  is a purely mathematical sentence
- $Proof(x, y)$ :  $x$  is an absolute proof of  $y$  (for  $y$  mathematical)
- $\Box A$ : it is absolutely provable that  $A$  (for  $A$  arbitrary)
- + “Sufficient” syntax to express:  
 $TM(x)$ :  $x$  is a Turing machine, etc.

- Principles:

T  $\Box A \rightarrow A$

K  $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

*Proof*- $\Box$   $S(\ulcorner A \urcorner) \rightarrow (\Box A \leftrightarrow \exists y Proof(y, \ulcorner A \urcorner))$

Nec 
$$\frac{\vdash A}{\vdash \Box A}$$

ECT  $\Box \forall x \exists y \Box \varphi(x, y) \rightarrow \exists e [TM(e) \wedge \forall x \varphi(x, e(x))]$

- + “Sufficient” mathematics to prove the undecidability of first-order arithmetical truth (Gödel, Tarski)

- Language:

- $S(x)$ :  $x$  is a purely mathematical sentence
- $Proof(x, y)$ :  $x$  is an absolute proof of  $y$  (for  $y$  mathematical)
- $\Box A$ : it is absolutely provable that  $A$  (for  $A$  arbitrary)
- + “Sufficient” syntax to express:  
 $TM(x)$ :  $x$  is a Turing machine, etc.

- Principles:

T  $\Box A \rightarrow A$

K  $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

*Proof*- $\Box$   $S(\ulcorner A \urcorner) \rightarrow (\Box A \leftrightarrow \exists y Proof(y, \ulcorner A \urcorner))$

Nec  $\frac{\vdash A}{\vdash \Box A}$

ECT  $\Box \forall x \exists y \Box \varphi(x, y) \rightarrow \exists e [TM(e) \wedge \forall x \varphi(x, e(x))]$

- + “Sufficient” mathematics to prove the undecidability of first-order arithmetical truth (Gödel, Tarski)

### Theorem (Horsten & Leitgeb, unpubl.)

Principles  $\vdash \exists y [S(y) \wedge \forall x (\neg Proof(x, y) \wedge \neg Proof(x, \neg y))] \vee$   
 $(\neg \Box A_1 \wedge \neg \Box \neg A_1) \vee (\neg \Box A_2 \wedge \neg \Box \neg A_2) \vee \dots \vee (\neg \Box A_n \wedge \neg \Box \neg A_n)$

So does this show that

$$\text{HG } \exists p(p \wedge \neg \Box p)$$

is vindicated?

So does this show that

$$\text{HG } \exists p(p \wedge \neg \Box p)$$

is vindicated?

Not really: All it shows is that *either ECT is false or HG is the case*.  
(cf. Gödel's 1951 dichotomy; this is *not* Lucas-Penrose style!)

So does this show that

$$\text{HG } \exists p(p \wedge \neg \Box p)$$

is vindicated?

Not really: All it shows is that *either ECT is false or HG is the case*.  
(cf. Gödel's 1951 dichotomy; this is *not* Lucas-Penrose style!)

In this case, the obvious reaction will be to argue that ECT is false, and that ECT therefore cannot be an adequate representation of CT.

Indeed, given different presuppositions, one can derive the opposite result:

- Assume that our  $\square$ -language includes propositional quantifiers and Hilbert-style epsilon terms for propositions, i.e., for every formula  $A[p]$  with a propositional variable  $p$ , there is a formula  $\varepsilon pA[p]$ , such that

$$\exists pA[p] \leftrightarrow A[\varepsilon pA[p]]$$

is a logical truth.

Indeed, given different presuppositions, one can derive the opposite result:

- Assume that our  $\Box$ -language includes propositional quantifiers and Hilbert-style epsilon terms for propositions, i.e., for every formula  $A[p]$  with a propositional variable  $p$ , there is a formula  $\varepsilon p A[p]$ , such that

$$\exists p A[p] \leftrightarrow A[\varepsilon p A[p]]$$

is a logical truth.

- In this calculus, combined with a system of modal logic, we get:

Theorem (Leitgeb 2009, *New Waves in Philosophy of Mathematics*)

KT +  $\varepsilon p$ -calculus  $\vdash \neg \exists p (p \wedge \neg \Box p)$

Indeed, given different presuppositions, one can derive the opposite result:

- Assume that our  $\Box$ -language includes propositional quantifiers and Hilbert-style epsilon terms for propositions, i.e., for every formula  $A[p]$  with a propositional variable  $p$ , there is a formula  $\varepsilon pA[p]$ , such that

$$\exists pA[p] \leftrightarrow A[\varepsilon pA[p]]$$

is a logical truth.

- In this calculus, combined with a system of modal logic, we get:

Theorem (Leitgeb 2009, *New Waves in Philosophy of Mathematics*)

KT +  $\varepsilon p$ -calculus  $\vdash \neg \exists p(p \wedge \neg \Box p)$

Once again: More likely to be a flaw in the system than a proper result?

Indeed, given different presuppositions, one can derive the opposite result:

- Assume that our  $\Box$ -language includes propositional quantifiers and Hilbert-style epsilon terms for propositions, i.e., for every formula  $A[p]$  with a propositional variable  $p$ , there is a formula  $\varepsilon pA[p]$ , such that

$$\exists pA[p] \leftrightarrow A[\varepsilon pA[p]]$$

is a logical truth.

- In this calculus, combined with a system of modal logic, we get:

Theorem (Leitgeb 2009, *New Waves in Philosophy of Mathematics*)

KT +  $\varepsilon p$ -calculus  $\vdash \neg \exists p(p \wedge \neg \Box p)$

Once again: More likely to be a flaw in the system than a proper result?

Hence: The H(oly) G(rail) in philosophy of mathematics is still waiting to be found, but mathematical methods help us in the search.

The recent emphasis in philosophy of mathematics on the study of actual *mathematical practice* is not in opposition to *logical reconstruction*.

# Do Connectionism and Symbolic Computationalism Exclude Each Other?

Here is a problem in the philosophy of mind:

# Do Connectionism and Symbolic Computationalism Exclude Each Other?

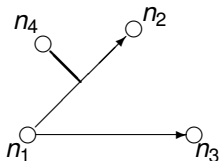
Here is a problem in the philosophy of mind:

- On the one hand, (human, animal, robot?) brains seem to be physical systems which can be described in terms of differential or difference equations, i.e., as *dynamical systems*.

# Do Connectionism and Symbolic Computationalism Exclude Each Other?

Here is a problem in the philosophy of mind:

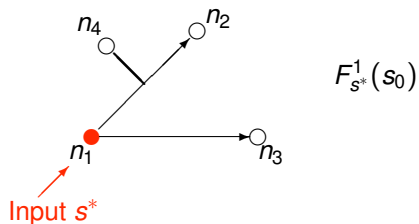
- On the one hand, (human, animal, robot?) brains seem to be physical systems which can be described in terms of differential or difference equations, i.e., as *dynamical systems*.



# Do Connectionism and Symbolic Computationalism Exclude Each Other?

Here is a problem in the philosophy of mind:

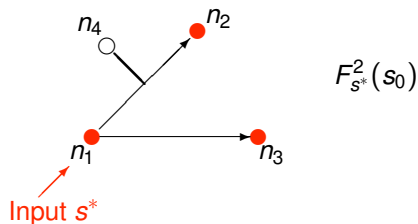
- On the one hand, (human, animal, robot?) brains seem to be physical systems which can be described in terms of differential or difference equations, i.e., as *dynamical systems*.



# Do Connectionism and Symbolic Computationalism Exclude Each Other?

Here is a problem in the philosophy of mind:

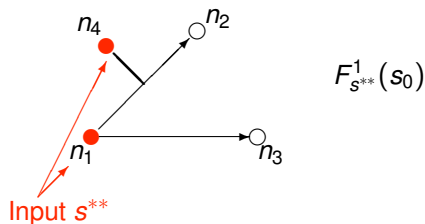
- On the one hand, (human, animal, robot?) brains seem to be physical systems which can be described in terms of differential or difference equations, i.e., as *dynamical systems*.



# Do Connectionism and Symbolic Computationalism Exclude Each Other?

Here is a problem in the philosophy of mind:

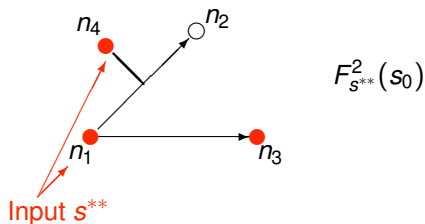
- On the one hand, (human, animal, robot?) brains seem to be physical systems which can be described in terms of differential or difference equations, i.e., as *dynamical systems*.



# Do Connectionism and Symbolic Computationalism Exclude Each Other?

Here is a problem in the philosophy of mind:

- On the one hand, (human, animal, robot?) brains seem to be physical systems which can be described in terms of differential or difference equations, i.e., as *dynamical systems*.



# Do Connectionism and Symbolic Computationalism Exclude Each Other?

Here is a problem in the philosophy of mind:

- On the one hand, (human, animal, robot?) brains seem to be physical systems which can be described in terms of differential or difference equations, i.e., as *dynamical systems*.
- On the other hand, these brains seem to have beliefs, draw inferences, and so forth, the contents of which can be expressed by *sentences*:

# Do Connectionism and Symbolic Computationalism Exclude Each Other?

Here is a problem in the philosophy of mind:

- On the one hand, (human, animal, robot?) brains seem to be physical systems which can be described in terms of differential or difference equations, i.e., as *dynamical systems*.
- On the other hand, these brains seem to have beliefs, draw inferences, and so forth, the contents of which can be expressed by *sentences*:

$x$  believes that  $\neg\phi$

$x$  infers that  $\phi \vee \psi$  from  $\phi$

# Do Connectionism and Symbolic Computationalism Exclude Each Other?

Here is a problem in the philosophy of mind:

- On the one hand, (human, animal, robot?) brains seem to be physical systems which can be described in terms of differential or difference equations, i.e., as *dynamical systems*.
- On the other hand, these brains seem to have beliefs, draw inferences, and so forth, the contents of which can be expressed by *sentences*:

$x$  believes that  $\neg\phi$

$x$  infers that  $\phi \vee \psi$  from  $\phi$

Working assumption: We take these to be distinct but *compatible* perspectives on the same phenomenon.

# Do Connectionism and Symbolic Computationalism Exclude Each Other?

Here is a problem in the philosophy of mind:

- On the one hand, (human, animal, robot?) brains seem to be physical systems which can be described in terms of differential or difference equations, i.e., as *dynamical systems*.
- On the other hand, these brains seem to have beliefs, draw inferences, and so forth, the contents of which can be expressed by *sentences*:

$x$  believes that  $\neg\phi$

$x$  infers that  $\phi \vee \psi$  from  $\phi$

Working assumption: We take these to be distinct but *compatible* perspectives on the same phenomenon.

So we have to associate *system states* with *propositions*:

system states carry information that can be expressed linguistically!

What might a theory of such *interpreted dynamical systems* look like?

## Definition

A triple  $\mathcal{S} = \langle S, ns, \leq \rangle$  is an ordered discrete dynamical system :iff

## Definition

A triple  $\mathcal{S} = \langle S, ns, \leq \rangle$  is an ordered discrete dynamical system :iff

- 1  $S$  is a non-empty set (the set of states)

## Definition

A triple  $\mathcal{S} = \langle S, ns, \leq \rangle$  is an ordered discrete dynamical system :iff

- 1  $S$  is a non-empty set (the set of states)
- 2  $ns : S \rightarrow S$  (the internal next-state function)

## Definition

A triple  $\mathcal{S} = \langle S, ns, \leq \rangle$  is an ordered discrete dynamical system :iff

- 1  $S$  is a non-empty set (the set of states)
- 2  $ns : S \rightarrow S$  (the internal next-state function)
- 3  $\leq \subseteq S \times S$  is a partial order (the information ordering) on  $S$ ,

## Definition

A triple  $\mathcal{S} = \langle S, ns, \leq \rangle$  is an ordered discrete dynamical system :iff

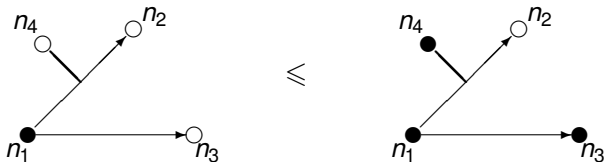
- 1  $S$  is a non-empty set (the set of states)
- 2  $ns : S \rightarrow S$  (the internal next-state function)
- 3  $\leq \subseteq S \times S$  is a partial order (the information ordering) on  $S$ , s.t.  
for all  $s, s' \in S$  there is a supremum  $sup(s, s') \in S$  with respect to  $\leq$ .

## Definition

A triple  $\mathcal{S} = \langle S, ns, \leq \rangle$  is an ordered discrete dynamical system :iff

- 1  $S$  is a non-empty set (the set of states)
- 2  $ns : S \rightarrow S$  (the internal next-state function)
- 3  $\leq \subseteq S \times S$  is a partial order (the information ordering) on  $S$ , s.t. for all  $s, s' \in S$  there is a supremum  $sup(s, s') \in S$  with respect to  $\leq$ .

E.g.:  $S = \{s \mid s : N \rightarrow \{0, 1\}\}$  (with  $N = \{n_1, n_2, n_3, n_4\}$ )



Supremum of two states: the *union* of their patterns of activation

## Definition

A triple  $\mathcal{S} = \langle S, ns, \leq \rangle$  is an ordered discrete dynamical system :iff

- 1  $S$  is a non-empty set (the set of states)
- 2  $ns : S \rightarrow S$  (the internal next-state function)
- 3  $\leq \subseteq S \times S$  is a partial order (the information ordering) on  $S$ , s.t.  
for all  $s, s' \in S$  there is a supremum  $sup(s, s') \in S$  with respect to  $\leq$ .

The internal dynamics of such systems is given by the iteration  
 $ns(= ns^1), ns^2, ns^3, \dots$  of the mapping  $ns$ .

## Definition

A triple  $\mathcal{S} = \langle S, ns, \leq \rangle$  is an ordered discrete dynamical system :iff

- 1  $S$  is a non-empty set (the set of states)
- 2  $ns : S \rightarrow S$  (the internal next-state function)
- 3  $\leq \subseteq S \times S$  is a partial order (the information ordering) on  $S$ , s.t. for all  $s, s' \in S$  there is a supremum  $sup(s, s') \in S$  with respect to  $\leq$ .

The internal dynamics of such systems is given by the iteration  $ns(= ns^1), ns^2, ns^3, \dots$  of the mapping  $ns$ .

Now we add an input which is regarded to activate a fixed state  $s^* \in S$ :

## Definition

A triple  $\mathcal{S} = \langle S, ns, \leq \rangle$  is an ordered discrete dynamical system :iff

- 1  $S$  is a non-empty set (the set of states)
- 2  $ns : S \rightarrow S$  (the internal next-state function)
- 3  $\leq \subseteq S \times S$  is a partial order (the information ordering) on  $S$ , s.t. for all  $s, s' \in S$  there is a supremum  $sup(s, s') \in S$  with respect to  $\leq$ .

The internal dynamics of such systems is given by the iteration  $ns(= ns^1), ns^2, ns^3, \dots$  of the mapping  $ns$ .

Now we add an input which is regarded to activate a fixed state  $s^* \in S$ : the “next” state of the system is given by the superposition of  $s^*$  with the next internal state  $ns(s)$ , i.e., we actually iterate

$$F_{s^*}(s) := sup(s^*, ns(s))$$

## Definition

A triple  $\mathcal{S} = \langle S, ns, \leq \rangle$  is an ordered discrete dynamical system :iff

- 1  $S$  is a non-empty set (the set of states)
- 2  $ns : S \rightarrow S$  (the internal next-state function)
- 3  $\leq \subseteq S \times S$  is a partial order (the information ordering) on  $S$ , s.t. for all  $s, s' \in S$  there is a supremum  $sup(s, s') \in S$  with respect to  $\leq$ .

The internal dynamics of such systems is given by the iteration  $ns(= ns^1), ns^2, ns^3, \dots$  of the mapping  $ns$ .

Now we add an input which is regarded to activate a fixed state  $s^* \in S$ : the “next” state of the system is given by the superposition of  $s^*$  with the next internal state  $ns(s)$ , i.e., we actually iterate

$$F_{s^*}(s) := sup(s^*, ns(s))$$

Stable states  $s_{stab}$  of  $F_{s^*}$  are the “answers” which the system gives to  $s^*$ .

Finally, we assign formulas to states of these systems.

Finally, we assign formulas to states of these systems.

Let  $\mathcal{L}$  be a propositional language:

Finally, we assign formulas to states of these systems.

Let  $\mathcal{L}$  be a propositional language:

### Definition

A quadruple  $\mathcal{S}_{\mathcal{J}} = \langle S, ns, \leq, \mathcal{J} \rangle$  is an interpreted ordered system :iff

Finally, we assign formulas to states of these systems.

Let  $\mathcal{L}$  be a propositional language:

### Definition

A quadruple  $\mathcal{S}_{\mathcal{J}} = \langle S, ns, \leq, \mathcal{J} \rangle$  is an interpreted ordered system :iff

- 1  $\langle S, ns, \leq \rangle$  is an ordered discrete dynamical system

Finally, we assign formulas to states of these systems.

Let  $\mathcal{L}$  be a propositional language:

## Definition

A quadruple  $\mathcal{S}_{\mathcal{I}} = \langle S, ns, \leq, \mathcal{I} \rangle$  is an interpreted ordered system :iff

- 1  $\langle S, ns, \leq \rangle$  is an ordered discrete dynamical system
- 2  $\mathcal{I} : \mathcal{L} \rightarrow S$  (the interpretation mapping) has “nice” properties, such as

Finally, we assign formulas to states of these systems.

Let  $\mathcal{L}$  be a propositional language:

## Definition

A quadruple  $\mathcal{S}_{\mathcal{I}} = \langle S, ns, \leq, \mathcal{I} \rangle$  is an interpreted ordered system :iff

- 1  $\langle S, ns, \leq \rangle$  is an ordered discrete dynamical system
- 2  $\mathcal{I} : \mathcal{L} \rightarrow S$  (the interpretation mapping) has “nice” properties, such as
  - for all  $\varphi, \psi \in \mathcal{L}$ :  $\mathcal{I}(\varphi \wedge \psi) = \sup(\mathcal{I}(\varphi), \mathcal{I}(\psi))$

Finally, we assign formulas to states of these systems.

Let  $\mathcal{L}$  be a propositional language:

## Definition

A quadruple  $\mathcal{S}_{\mathcal{I}} = \langle S, ns, \leq, \mathcal{I} \rangle$  is an interpreted ordered system :iff

- 1  $\langle S, ns, \leq \rangle$  is an ordered discrete dynamical system
- 2  $\mathcal{I} : \mathcal{L} \rightarrow S$  (the interpretation mapping) has “nice” properties, such as
  - for all  $\varphi, \psi \in \mathcal{L}$ :  $\mathcal{I}(\varphi \wedge \psi) = \sup(\mathcal{I}(\varphi), \mathcal{I}(\psi))$
  - for every  $\varphi \in \mathcal{L}$ : there is a unique  $\mathcal{I}(\varphi)$ -stable state

Finally, we assign formulas to states of these systems.

Let  $\mathcal{L}$  be a propositional language:

## Definition

A quadruple  $\mathcal{S}_{\mathcal{J}} = \langle S, ns, \leq, \mathcal{J} \rangle$  is an interpreted ordered system :iff

- 1  $\langle S, ns, \leq \rangle$  is an ordered discrete dynamical system
- 2  $\mathcal{J} : \mathcal{L} \rightarrow S$  (the interpretation mapping) has “nice” properties, such as
  - for all  $\varphi, \psi \in \mathcal{L}$ :  $\mathcal{J}(\varphi \wedge \psi) = \sup(\mathcal{J}(\varphi), \mathcal{J}(\psi))$
  - for every  $\varphi \in \mathcal{L}$ : there is a unique  $\mathcal{J}(\varphi)$ -stable state
  - $\vdots$

## Definition

- 1  $\mathcal{S}_{\mathcal{J}} \models \varphi \Rightarrow \psi$  :iff  
if  $s_{stab}$  is the unique  $\mathcal{J}(\varphi)$ -stable state, then  $\mathcal{J}(\psi) \leq s_{stab}$

Finally, we assign formulas to states of these systems.

Let  $\mathcal{L}$  be a propositional language:

## Definition

A quadruple  $\mathcal{S}_{\mathcal{I}} = \langle S, ns, \leq, \mathcal{I} \rangle$  is an interpreted ordered system :iff

- 1  $\langle S, ns, \leq \rangle$  is an ordered discrete dynamical system
- 2  $\mathcal{I} : \mathcal{L} \rightarrow S$  (the interpretation mapping) has “nice” properties, such as
  - for all  $\varphi, \psi \in \mathcal{L}$ :  $\mathcal{I}(\varphi \wedge \psi) = \text{sup}(\mathcal{I}(\varphi), \mathcal{I}(\psi))$
  - for every  $\varphi \in \mathcal{L}$ : there is a unique  $\mathcal{I}(\varphi)$ -stable state
  - $\vdots$

## Definition

- 1  $\mathcal{S}_{\mathcal{I}} \models \varphi \Rightarrow \psi$  :iff  
if  $s_{\text{stab}}$  is the unique  $\mathcal{I}(\varphi)$ -stable state, then  $\mathcal{I}(\psi) \leq s_{\text{stab}}$
- 2  $\mathcal{T}\mathcal{H}_{\Rightarrow}(\mathcal{S}_{\mathcal{I}}) = \{ \varphi \Rightarrow \psi \mid \mathcal{S}_{\mathcal{I}} \models \varphi \Rightarrow \psi \}$   
(the conditional theory corresponding to  $\mathcal{S}_{\mathcal{I}}$ ).

## Theorem (Representation)

For all  $\mathcal{S}_{\mathcal{J}}$ :  $\mathcal{T}\mathcal{H} \Rightarrow (\mathcal{S}_{\mathcal{J}})$  is closed under the following rules:

- $\frac{}{\varphi \Rightarrow \varphi}$  (*Reflexivity*)
- $\frac{\vdash \varphi \leftrightarrow \psi, \varphi \Rightarrow \rho}{\psi \Rightarrow \rho}$  (*Left Equivalence*)
- $\frac{\varphi \Rightarrow \psi, \vdash \psi \rightarrow \rho}{\varphi \Rightarrow \rho}$  (*Right Weakening*)
- $\frac{\varphi \Rightarrow \psi, \varphi \wedge \psi \Rightarrow \rho}{\varphi \Rightarrow \rho}$  (*Cautious Cut*)
- $\frac{\varphi \Rightarrow \psi, \varphi \Rightarrow \rho}{\varphi \wedge \psi \Rightarrow \rho}$  (*Cautious Monotonicity*)

## Theorem (Representation)

For all  $\mathcal{S}_{\mathcal{J}}$ :  $\mathcal{T}\mathcal{H} \Rightarrow (\mathcal{S}_{\mathcal{J}})$  is closed under the following rules:

- $\frac{}{\varphi \Rightarrow \varphi}$  (Reflexivity)
- $\frac{\vdash \varphi \leftrightarrow \psi, \varphi \Rightarrow \rho}{\psi \Rightarrow \rho}$  (Left Equivalence)
- $\frac{\varphi \Rightarrow \psi, \vdash \psi \rightarrow \rho}{\varphi \Rightarrow \rho}$  (Right Weakening)
- $\frac{\varphi \Rightarrow \psi, \varphi \wedge \psi \Rightarrow \rho}{\varphi \Rightarrow \rho}$  (Cautious Cut)
- $\frac{\varphi \Rightarrow \psi, \varphi \Rightarrow \rho}{\varphi \wedge \psi \Rightarrow \rho}$  (Cautious Monotonicity)

Furthermore, for all consistent  $\mathcal{T}\mathcal{H}$  that are closed under these rules there is an interpreted system  $\mathcal{S}_{\mathcal{J}}$ , such that  $\mathcal{T}\mathcal{H} = \mathcal{T}\mathcal{H} \Rightarrow (\mathcal{S}_{\mathcal{J}})$ .

## Theorem (Representation)

For all  $\mathcal{S}_{\mathcal{J}}$ :  $\mathcal{T}\mathcal{H} \Rightarrow (\mathcal{S}_{\mathcal{J}})$  is closed under the following rules:

- $\frac{}{\varphi \Rightarrow \varphi}$  (Reflexivity)
- $\frac{\vdash \varphi \leftrightarrow \psi, \varphi \Rightarrow \rho}{\psi \Rightarrow \rho}$  (Left Equivalence)
- $\frac{\varphi \Rightarrow \psi, \vdash \psi \rightarrow \rho}{\varphi \Rightarrow \rho}$  (Right Weakening)
- $\frac{\varphi \Rightarrow \psi, \varphi \wedge \psi \Rightarrow \rho}{\varphi \Rightarrow \rho}$  (Cautious Cut)
- $\frac{\varphi \Rightarrow \psi, \varphi \Rightarrow \rho}{\varphi \wedge \psi \Rightarrow \rho}$  (Cautious Monotonicity)

Furthermore, for all consistent  $\mathcal{T}\mathcal{H}$  that are closed under these rules there is an interpreted system  $\mathcal{S}_{\mathcal{J}}$ , such that  $\mathcal{T}\mathcal{H} = \mathcal{T}\mathcal{H} \Rightarrow (\mathcal{S}_{\mathcal{J}})$ .

Proof: (i) induction over formulas; (ii) construction of systems.

The following rules do *not* hold in all interpreted systems:

- $$\frac{\varphi \Rightarrow \rho}{\varphi \wedge \psi \Rightarrow \rho} \text{ (Monotonicity)}$$
- $$\frac{\varphi \Rightarrow \psi, \psi \Rightarrow \rho}{\varphi \Rightarrow \rho} \text{ (Transitivity)}$$
- $$\frac{\varphi \Rightarrow \psi}{\neg \psi \Rightarrow \neg \varphi} \text{ (Contraposition)}$$

The following rules do *not* hold in all interpreted systems:

- $$\frac{\varphi \Rightarrow \rho}{\varphi \wedge \psi \Rightarrow \rho} \text{ (Monotonicity)}$$
- $$\frac{\varphi \Rightarrow \psi, \psi \Rightarrow \rho}{\varphi \Rightarrow \rho} \text{ (Transitivity)}$$
- $$\frac{\varphi \Rightarrow \psi}{\neg \psi \Rightarrow \neg \varphi} \text{ (Contraposition)}$$

It turns out that the logics that are adequate for interpreted systems are systems of *non-monotonic* logic. These are logics governing

if  $\varphi$  then (normally/indicative)  $\psi$

The following rules do *not* hold in all interpreted systems:

- $\frac{\varphi \Rightarrow \rho}{\varphi \wedge \psi \Rightarrow \rho}$  (Monotonicity)
- $\frac{\varphi \Rightarrow \psi, \psi \Rightarrow \rho}{\varphi \Rightarrow \rho}$  (Transitivity)
- $\frac{\varphi \Rightarrow \psi}{\neg\psi \Rightarrow \neg\varphi}$  (Contraposition)

It turns out that the logics that are adequate for interpreted systems are systems of *non-monotonic* logic. These are logics governing

if  $\varphi$  then (normally/indicative)  $\psi$

Hence: If *neural networks with distributed representations* are looked at in the right way, then they are found to reason according to *rational symbolic rules*.

(Leitgeb 2001, *Artificial Intelligence*; 2004, *Inference on the Low Level*  
2005, *Synthese* & French *Scientific American*; 2009, *Handbook of Ind. Logic*  
New project on extensions to *inductive logic* forthcoming!

# Can We Justify Probability Theory from Closeness to the Truth?

Is it possible to justify (i) the axioms of probability, and (ii) update by conditionalization, *epistemically*, i.e., in some truth-related manner?

(This is joint work with R. Pettigrew: “On an Objective Justification of Bayesianism”, Parts I & II, submitted.

See Joyce 1998, Greaves & Wallace 2006 for related work.)

# Can We Justify Probability Theory from Closeness to the Truth?

Is it possible to justify (i) the axioms of probability, and (ii) update by conditionalization, *epistemically*, i.e., in some truth-related manner?

(This is joint work with R. Pettigrew: “On an Objective Justification of Bayesianism”, Parts I & II, submitted.)

See Joyce 1998, Greaves & Wallace 2006 for related work.)

- Let  $W = \{w_1, \dots, w_n\}$  be the (finite) set of possible worlds,  $\mathcal{P}(W)$  be the corresponding set of propositions
- Let  $\text{Bel}(W)$  be the set of functions  $b : \mathcal{P}(W) \rightarrow \mathbb{R}_0^+$  (the set of quantitative “belief functions” on  $\mathcal{P}(W)$ ).

# Can We Justify Probability Theory from Closeness to the Truth?

Is it possible to justify (i) the axioms of probability, and (ii) update by conditionalization, *epistemically*, i.e., in some truth-related manner?

(This is joint work with R. Pettigrew: “On an Objective Justification of Bayesianism”, Parts I & II, submitted.)

See Joyce 1998, Greaves & Wallace 2006 for related work.)

- Let  $W = \{w_1, \dots, w_n\}$  be the (finite) set of possible worlds,  $\mathcal{P}(W)$  be the corresponding set of propositions
- Let  $\text{Bel}(W)$  be the set of functions  $b : \mathcal{P}(W) \rightarrow \mathbb{R}_0^+$  (the set of quantitative “belief functions” on  $\mathcal{P}(W)$ ).

We assume epistemic states are analyzed quantitatively:

- An agent’s epistemic state at time  $t$  may be represented by a belief function  $b_t \in \text{Bel}(W)$  which assigns to each proposition a degree of belief.

# Can We Justify Probability Theory from Closeness to the Truth?

Is it possible to justify (i) the axioms of probability, and (ii) update by conditionalization, *epistemically*, i.e., in some truth-related manner?

(This is joint work with R. Pettigrew: “On an Objective Justification of Bayesianism”, Parts I & II, submitted.

See Joyce 1998, Greaves & Wallace 2006 for related work.)

- Let  $W = \{w_1, \dots, w_n\}$  be the (finite) set of possible worlds,  $\mathcal{P}(W)$  be the corresponding set of propositions
- Let  $\text{Bel}(W)$  be the set of functions  $b : \mathcal{P}(W) \rightarrow \mathbb{R}_0^+$  (the set of quantitative “belief functions” on  $\mathcal{P}(W)$ ).

We assume epistemic states are analyzed quantitatively:

- An agent’s epistemic state at time  $t$  may be represented by a belief function  $b_t \in \text{Bel}(W)$  which assigns to each proposition a degree of belief.

Thus, we must determine, for evidence  $E \subseteq W$ , which belief functions it would be rational for an agent to have at a time when she is in possession of  $E$ .

Basic norm: Choose your degree of belief in a proposition in a way such that its expected distance from the truth is minimized – *minimize expected inaccuracy*:

Basic norm: Choose your degree of belief in a proposition in a way such that its expected distance from the truth is minimized – *minimize expected inaccuracy*:

### Definition (Expected Local Inaccuracy)

Let

- $I : \mathcal{P}(W) \times W \times \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$  be an inaccuracy function  
( $I(A, w, x)$ : inaccuracy of degree of belief  $x$  in proposition  $A$  at world  $w$ ),
- $b : \mathcal{P}(W) \rightarrow \mathbb{R}_0^+$  be a belief function,
- $x \in \mathbb{R}_0^+$  a degree of belief,
- $A, E \subseteq W$  propositions.

Then we define the *expected inaccuracy of  $x$  in proposition  $A$  by the lights of  $b$ , with respect to  $I$ , and over the set  $E$  of epistemically possible worlds* as follows:

$$\text{LExp}_b(I, A, E, x) = \sum_{w \in E} b(\{w\}) I(A, w, x)$$

Basic norm: Choose your degree of belief in a proposition in a way such that its expected distance from the truth is minimized – *minimize expected inaccuracy*:

### Definition (Expected Local Inaccuracy)

Let

- $I : \mathcal{P}(W) \times W \times \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$  be an inaccuracy function  
( $I(A, w, x)$ : inaccuracy of degree of belief  $x$  in proposition  $A$  at world  $w$ ),
- $b : \mathcal{P}(W) \rightarrow \mathbb{R}_0^+$  be a belief function,
- $x \in \mathbb{R}_0^+$  a degree of belief,
- $A, E \subseteq W$  propositions.

Then we define the *expected inaccuracy of  $x$  in proposition  $A$  by the lights of  $b$ , with respect to  $I$ , and over the set  $E$  of epistemically possible worlds* as follows:

$$\text{LExp}_b(I, A, E, x) = \sum_{w \in E} b(\{w\}) I(A, w, x)$$

But how shall we define the inaccuracy  $I(A, w, x)$  of a degree of belief  $x$  in a proposition  $A$  at a world  $w$ ?

Here is a convenient choice:

### Definition (Quadratic Inaccuracy; Brier scores)

Suppose  $\lambda \in \mathbb{R}_{>0}$ . Then, given  $A \subseteq W$ ,  $w \in W$ ,  $x \in \mathbb{R}_0^+$ , we define

$$Q_\lambda(A, w, x) =_{df.} \lambda(\chi_A(w) - x)^2$$

where  $\chi_A : W \rightarrow \{0, 1\}$  is the characteristic function of  $A$ . The family of such functions  $Q_\lambda$  is called the *quadratic family of inaccuracy functions*.

Here is a convenient choice:

### Definition (Quadratic Inaccuracy; Brier scores)

Suppose  $\lambda \in \mathbb{R}_{>0}$ . Then, given  $A \subseteq W$ ,  $w \in W$ ,  $x \in \mathbb{R}_0^+$ , we define

$$Q_\lambda(A, w, x) =_{df.} \lambda(\chi_A(w) - x)^2$$

where  $\chi_A : W \rightarrow \{0, 1\}$  is the characteristic function of  $A$ . The family of such functions  $Q_\lambda$  is called the *quadratic family of inaccuracy functions*.

Using a quadratic inaccuracy function, we get exactly the intended result:

### Theorem

Suppose  $\lambda \in \mathbb{R}_{>0}$ ,  $b \in \text{Bel}(W)$ ,  $A, E \subseteq W$ , and  $\sum_{w \in E} b(w) \neq 0$ . Then

$$\sum_{w \in E} b(\{w\}) Q_\lambda(A, w, x)$$

is minimal if, and only if,

$$x = \frac{\sum_{w \in A \cap E} b(\{w\})}{\sum_{w \in E} b(\{w\})}$$

But why *quadratic* inaccuracy functions?

Because only those avoid particular epistemic dilemmas:

But why *quadratic* inaccuracy functions?

Because only those avoid particular epistemic dilemmas:

## Theorem

The following two propositions are equivalent:

- (i)  $f : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$  is strictly increasing and continuously differentiable;  $f(0) = 0$ ; and, for all belief functions  $b \in \text{Bel}(W)$ , all  $w_j \in W$ , and all  $a_1, \dots, a_{j-1}, a_{j+1}, \dots, a_n \in \mathbb{R}_0^+$ :

$$\begin{aligned} & \frac{d}{dx} \sum_{w_i \in W} b(\{w_i\}) f(|\chi_{\{w_i\}}(w_i) - x|) \\ &= \frac{d}{dx} \sum_{w_i \in W} b(\{w_i\}) f(\|w_i - (a_1, \dots, a_{j-1}, x, a_{j+1}, \dots, a_n)\|) \end{aligned}$$

- (ii) There is a  $\lambda \in \mathbb{R}_{>0}$ , such that, for all  $x \in \mathbb{R}_0^+$ :

$$f(x) = \lambda x^2$$

Hence:

If an agent is rational only insofar as she minimizes the expected distance of her beliefs from the truth (free from epistemic dilemmas), then one can prove that every rational agent must reason according to probability theory.

Hence:

If an agent is rational only insofar as she minimizes the expected distance of her beliefs from the truth (free from epistemic dilemmas), then one can prove that every rational agent must reason according to probability theory.

Using similar methods, one can show that e.g. *Jeffrey conditionalization* does *not* minimize expected inaccuracy, and it is possible to prove which alternative method of update does.

Hence:

If an agent is rational only insofar as she minimizes the expected distance of her beliefs from the truth (free from epistemic dilemmas), then one can prove that every rational agent must reason according to probability theory.

Using similar methods, one can show that e.g. *Jeffrey conditionalization* does *not* minimize expected inaccuracy, and it is possible to prove which alternative method of update does.

Final remark:

How to justify probability theory for *indexical* or *introspective* statements?  
Similar methods in combination with insights from *modal* and *dynamic* logic!?  
(Logic and probability theory are *not* in opposition with each other!)

We found:

- Hypergraph theory can be used to determine under which conditions properties can be reconstructed from similarity.
- Mathematical logic and modal logic may tell us about what we can or cannot prove in principle.
- Dynamical systems theory can be applied to justify systems of nonmonotonic logic or conditional logic. Both together throw new light on the symbolic computationalism vs. connectionism debate.
- Classical real analysis, in combination with one basic epistemic norm, suffices to derive probability theory.
- (If you run out of questions later:)  
Functional analysis can be employed to support an epistemic account of truth for semantically closed languages.

Therefore: even in philosophy, *calculemus!*

# Is There a Probabilistic Way Out of Semantic Paradoxes?

In natural language we use the predicate 'is true' (briefly:  $Tr$ ) in order to express that sentences are true.

# Is There a Probabilistic Way Out of Semantic Paradoxes?

In natural language we use the predicate 'is true' (briefly:  $Tr$ ) in order to express that sentences are true.

Now let us assume we want to set up a formal theory for  $Tr$  (just as set theory is a theory for  $\in$ ): obviously, every "law" of the form

$$Tr('α') \leftrightarrow α$$

ought to be derivable in such a theory.

# Is There a Probabilistic Way Out of Semantic Paradoxes?

In natural language we use the predicate 'is true' (briefly:  $Tr$ ) in order to express that sentences are true.

Now let us assume we want to set up a formal theory for  $Tr$  (just as set theory is a theory for  $\in$ ): obviously, every "law" of the form

$$Tr('α') \leftrightarrow α$$

ought to be derivable in such a theory.

(E.g.,  $Tr('2 = 1 + 1') \leftrightarrow 2 = 1 + 1$  should be derivable.)

# Is There a Probabilistic Way Out of Semantic Paradoxes?

In natural language we use the predicate 'is true' (briefly:  $Tr$ ) in order to express that sentences are true.

Now let us assume we want to set up a formal theory for  $Tr$  (just as set theory is a theory for  $\in$ ): obviously, every "law" of the form

$$Tr('α') \leftrightarrow α$$

ought to be derivable in such a theory.

(E.g.,  $Tr('2 = 1 + 1') \leftrightarrow 2 = 1 + 1$  should be derivable.)

However, it turns out that if we add "sufficient" arithmetic to a theory like that, then a contradiction can be derived (as observed by A. Tarski):

# Is There a Probabilistic Way Out of Semantic Paradoxes?

In natural language we use the predicate 'is true' (briefly:  $Tr$ ) in order to express that sentences are true.

Now let us assume we want to set up a formal theory for  $Tr$  (just as set theory is a theory for  $\in$ ): obviously, every "law" of the form

$$Tr('α') \leftrightarrow α$$

ought to be derivable in such a theory.

(E.g.,  $Tr('2 = 1 + 1') \leftrightarrow 2 = 1 + 1$  should be derivable.)

However, it turns out that if we add "sufficient" arithmetic to a theory like that, then a contradiction can be derived (as observed by A. Tarski):

This is because one can prove that there is a sentence  $\lambda$ , such that

$$\lambda \leftrightarrow \neg Tr('λ')$$

follows from arithmetic!

So what can we do? (This is actually a long story, but...)

So what can we do? (This is actually a long story, but. . .)

Instead of postulating that  $Tr('α')$  and  $α$  are *equivalent*, we might rather postulate that the same *degree of belief* is assigned to  $Tr('α')$  and  $α$ !

So what can we do? (This is actually a long story, but...)

Instead of postulating that  $Tr('α')$  and  $α$  are *equivalent*, we might rather postulate that the same *degree of belief* is assigned to  $Tr('α')$  and  $α$ !

Let us formalize this: let  $\mathcal{L}$  be the language of arithmetic extended by  $Tr$ .

QUESTION: Is there a function  $P : \mathcal{L} \rightarrow [0, 1]$ , such that

- $P$  satisfies the analogues of Kolmogorov's probability axioms (e.g.,  $P(\neg\alpha) = 1 - P(\alpha)$ ;  $P(\alpha) = 1$  for provable  $\alpha$ , etc.)

So what can we do? (This is actually a long story, but...)

Instead of postulating that  $Tr('α')$  and  $α$  are *equivalent*, we might rather postulate that the same *degree of belief* is assigned to  $Tr('α')$  and  $α$ !

Let us formalize this: let  $\mathcal{L}$  be the language of arithmetic extended by  $Tr$ .

QUESTION: Is there a function  $P : \mathcal{L} \rightarrow [0, 1]$ , such that

- $P$  satisfies the analogues of Kolmogorov's probability axioms (e.g.,  $P(\neg\alpha) = 1 - P(\alpha)$ ;  $P(\alpha) = 1$  for provable  $\alpha$ , etc.)
- $P$  satisfies:

$$P(Tr('α')) = P(α)$$

So what can we do? (This is actually a long story, but...)

Instead of postulating that  $Tr('α')$  and  $α$  are *equivalent*, we might rather postulate that the same *degree of belief* is assigned to  $Tr('α')$  and  $α$ !

Let us formalize this: let  $\mathcal{L}$  be the language of arithmetic extended by  $Tr$ .

QUESTION: Is there a function  $P : \mathcal{L} \rightarrow [0, 1]$ , such that

- $P$  satisfies the analogues of Kolmogorov's probability axioms (e.g.,  $P(\neg\alpha) = 1 - P(\alpha)$ ;  $P(\alpha) = 1$  for provable  $\alpha$ , etc.)
- $P$  satisfies:

$$P(Tr('α')) = P(\alpha)$$

*Well, almost!*

So what can we do? (This is actually a long story, but...)

Instead of postulating that  $Tr('α')$  and  $α$  are *equivalent*, we might rather postulate that the same *degree of belief* is assigned to  $Tr('α')$  and  $α$ !

Let us formalize this: let  $\mathcal{L}$  be the language of arithmetic extended by  $Tr$ .

QUESTION: Is there a function  $P : \mathcal{L} \rightarrow [0, 1]$ , such that

- $P$  satisfies the analogues of Kolmogorov's probability axioms (e.g.,  $P(\neg\alpha) = 1 - P(\alpha)$ ;  $P(\alpha) = 1$  for provable  $\alpha$ , etc.)
- $P$  satisfies:

$$P(Tr('α')) = P(\alpha)$$

*Well, almost!*

## Theorem

- 1 *There is no function  $P$  that satisfies all the conditions from above.*

So what can we do? (This is actually a long story, but...)

Instead of postulating that  $Tr('α')$  and  $α$  are *equivalent*, we might rather postulate that the same *degree of belief* is assigned to  $Tr('α')$  and  $α$ !

Let us formalize this: let  $\mathcal{L}$  be the language of arithmetic extended by  $Tr$ .

QUESTION: Is there a function  $P : \mathcal{L} \rightarrow [0, 1]$ , such that

- $P$  satisfies the analogues of Kolmogorov's probability axioms (e.g.,  $P(\neg\alpha) = 1 - P(\alpha)$ ;  $P(\alpha) = 1$  for provable  $\alpha$ , etc.)
- $P$  satisfies:

$$P(Tr('α')) = P(\alpha)$$

*Well, almost!*

## Theorem

- 1 *There is no function  $P$  that satisfies all the conditions from above.*
- 2 *There is a function  $P$  that satisfies all the conditions above except for the axiom of countable additivity.*

So what can we do? (This is actually a long story, but...)

Instead of postulating that  $Tr('α')$  and  $α$  are *equivalent*, we might rather postulate that the same *degree of belief* is assigned to  $Tr('α')$  and  $α$ !

Let us formalize this: let  $\mathcal{L}$  be the language of arithmetic extended by  $Tr$ .

QUESTION: Is there a function  $P : \mathcal{L} \rightarrow [0, 1]$ , such that

- $P$  satisfies the analogues of Kolmogorov's probability axioms (e.g.,  $P(\neg\alpha) = 1 - P(\alpha)$ ;  $P(\alpha) = 1$  for provable  $\alpha$ , etc.)
- $P$  satisfies:

$$P(Tr('α')) = P(\alpha)$$

*Well, almost!*

## Theorem

1. *There is no function  $P$  that satisfies all the conditions from above.*
2. *There is a function  $P$  that satisfies all the conditions above except for the axiom of countable additivity.*

Proof: 1. Construction of a clever self-referential formula. 2. Hahn-Banach.

What about our “Liar” sentence  $\lambda$ ?

What about our “Liar” sentence  $\lambda$ ?

Since  $\lambda \leftrightarrow \neg Tr(' \lambda')$  is provable,  $P(\lambda \leftrightarrow \neg Tr(' \lambda')) = 1$ .

What about our “Liar” sentence  $\lambda$ ?

Since  $\lambda \leftrightarrow \neg Tr(' \lambda')$  is provable,  $P(\lambda \leftrightarrow \neg Tr(' \lambda')) = 1$ .

This implies:

$$\begin{aligned} P(\lambda) &= P(\neg Tr(' \lambda')) \\ &= 1 - P(Tr(' \lambda')) \\ &= 1 - P(\lambda) \end{aligned}$$

What about our “Liar” sentence  $\lambda$ ?

Since  $\lambda \leftrightarrow \neg Tr(' \lambda')$  is provable,  $P(\lambda \leftrightarrow \neg Tr(' \lambda')) = 1$ .

This implies:

$$\begin{aligned} P(\lambda) &= P(\neg Tr(' \lambda')) \\ &= 1 - P(Tr(' \lambda')) \\ &= 1 - P(\lambda) \end{aligned}$$

Hence,  $P(\lambda) = P(\neg\lambda) = \frac{1}{2}$ .

What about our “Liar” sentence  $\lambda$ ?

Since  $\lambda \leftrightarrow \neg Tr(' \lambda')$  is provable,  $P(\lambda \leftrightarrow \neg Tr(' \lambda')) = 1$ .

This implies:

$$\begin{aligned} P(\lambda) &= P(\neg Tr(' \lambda')) \\ &= 1 - P(Tr(' \lambda')) \\ &= 1 - P(\lambda) \end{aligned}$$

Hence,  $P(\lambda) = P(\neg \lambda) = \frac{1}{2}$ .

Maybe the beginning of an epistemic theory of type-free truth (and even probability)?

(Leitgeb 2008, *Review of Symbolic Logic*)