Infinite Cardinalities

PARADOX AND INFINITY
Benjamin Brast-McKie
May 14, 2024

Where to Begin...

Assumptions: Theories have to begin somewhere.

- A theory that is neutral on everything is no theory at all.
- But a conclusion cannot come from nothing.

Concepts: Can't define everything.

- Must take some concepts as primitive.
- Intuitions from patterns of use.
- Principles encode theoretical function.

Example: Consider the concept of *number*.

- Numbers are answers to 'How many?'-Questions.
- What principles might we affirm?

Finite Intuitions

Proper Subset Principle: $A \subset B \to |A| < |B|$.

- There are more mammals than lamas.
- There are more reals than rationals.

Bijection Principle: $A \simeq B \leftrightarrow |A| = |B|$.

- $A \simeq B$ means the As and Bs can be paired one-to-one with no remainders.
- We can define this without recourse to the concept of number.

Ordered Pair: $\langle a, b\rangle \coloneqq \{\{a\}, \{a, b\}\}$.

Relation: $A \times B \coloneqq \{\langle a, b\rangle : a \in A,\ b \in B\}$.

Function: A *(total) function* $f : A \to B$ is any relation $f \subseteq A \times B$ where for every $a \in A$: (1) there is some $b \in B$ where $\langle a, b\rangle \in f$; and (2) if there are some $b, c \in B$ where both $\langle a, b\rangle, \langle a, c\rangle \in f$, then $b = c$.

- If f is a function, then we may take '$f(a) = b$' to abbreviate '$\langle a, b\rangle \in f$'.

Injective: $f : A \to B$ is *injective iff* for any $a, b \in A$, if $f(a) = f(b)$, then $a = b$.

Surjective: $f : A \to B$ is *surjective iff* for all $b \in B$ there is some $a \in A$ where $f(a) = b$.

Bijection: $f : A \to B$ is *bijective iff* f is an injective and surjective function.

Equinumerous: $A \simeq B$ *iff* there is a bijection $f : A \to B$.

Infinity and Paradox?

Hilbert's Hotel: Always room for (countably many) more guests.

- Given a domain $\mathbb{D}$, $f : x \mapsto t(x)$ defines the function $\{\langle x, t(x)\rangle : x \in \mathbb{D}\}$ where '$t(x)$' is a term that may include 'x', e.g., '$x + 1$'.
- $f_m : n \mapsto n + m$ is a bijection $f_m : \mathbb{N} \to \mathbb{N}_m$ where $\mathbb{N}_m = \{k \in \mathbb{N} : k \geqslant m\}$.
- $g_m : n \mapsto n \times m$ is a bijection $g_m : \mathbb{N} \to \mathbb{N}_{(m)} = \{k \times m : k \in \mathbb{N}\}$.

(?) What about countably many countable groups of new guests?

Galileo's Roots: "Every square has its own root and every root has its own square, while no square has more than one root and no root has more than one square."

Paradox: There are many equinumerous proper subsets of infinite sets.

- By the principles above, both $|\mathbb{N}_2| < |\mathbb{N}|$ and $|\mathbb{N}_2| = |\mathbb{N}|$.
- But $x < y$ *iff* $x \leqslant y$ and $x \neq y$.
- Thus $|\mathbb{N}_2| < |\mathbb{N}|$ entails $|\mathbb{N}_2| \neq |\mathbb{N}|$: contradiction.

(?) Which principle should we give up?

The Abductive Method

Deductively Closed: Good theories include all of their implications.

Consistency: Good theories exclude contradictions.

Simplicity: Good theories are easy to understand, e.g., are finitely axiomatizable in terms of intuitively compelling and conceptually elegant concepts.

Strength: Good theories say more rather than less.

Utility: Good theories serve our aims, e.g., have useful applications.

A Metaphysical Aside

Subjectivity: Is theory choice by abduction a reflection of human psychology?

- The abductive method describes how we typically choose theories.
- Are we right to use the abductive method and why?

Realism: Is the abductive method well suited to the task of describing reality?

- We don't need to decide this before using the abductive method.
- It will help to put the method to work.

Example: Which of the principles above should we give up?

Towards a Theory of Number

Hypothesis: Suppose we were to retain the *Proper Subset Principle* (PSP).

- Then $|\mathbb{N}| > |\mathbb{N}_2| > |\mathbb{N}_3| > \ldots$ and $|\mathbb{N}| > |\mathbb{N}_{(2)}| > |\mathbb{N}_{(4)}| > \ldots$ etc.
- How are we to compare $|\mathbb{N}_{(2)}|$ and $|\mathbb{N}_{(3)}|$?

Linear Ordering: Numbers are linearly ordered by $\leqslant$ and so must satisfy the following.

`Reflexive:` $x \leqslant x$ for any number x.

`Transitive:` If $x \leqslant y$ and $y \leqslant z$, then $x \leqslant z$.

`Anti-Symmetric:` If $x \leqslant y$ and $y \leqslant x$, then $x = y$.

`Total:` Either $x \leqslant y$ or $y \leqslant x$ for any numbers x and y.

- Compare giving a theory of identity where symmetry fails.
- We wouldn't really be talking about identity.

Incomplete: Converse of PSP is false and so PSP does not define $<$.

(?) Could PSP be supplemented in some way?

(?) How would we compare the cardinality of two bags of stones?

- By trying to line them up one-to-one.

Injection Principle: $|A| \leqslant |B|$ *iff* $A \simeq C$ for some $C \subseteq B$.

- This assumes *Bijection Principle* (BP) which is in tension with PSP.
- Restricting the *Injection Principle* (IP) and BP to finite sets is *ad hoc*.
- Better to take Hilbert's Hotel to be a counterexample to PSP.

Countable Infinity

Principles: Given IP and BP, we may show that numbers are linearly ordered.

- `Anti-Symmetric` is proven by the Cantor-Schroeder-Bernstein theorem.
- `Total` is equivalent to the Axiom of Choice.

Countable Sets: A set A is *countably infinite iff* $|A| = |\mathbb{N}|$.

Identities: Bijection Principle makes many sets countably infinite.

- $|\mathbb{N}| = |\mathbb{N}_m| = |\mathbb{N}_{(m)}|$.
- $|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}|$.

(?) What about $|\mathbb{N}| = |\mathbb{R}|$?

Next Time: We will show that there are different sizes of infinity.

Infinite Cardinalities

PARADOX AND INFINITY
Benjamin Brast-McKie
March 20, 2025

Cardinality Principles

Bijection Principle: $|A| = |B|$ *iff* $A \simeq B$.

`Reflexive:` $A \simeq A$.

`Symmetric:` if $A \simeq B$, then $B \simeq A$.

* What's an inverse of a relation?
* Do functions always have inverses?
* Observe: the inverse of a bijection is a bijection.

`Transitive:` if $A \simeq B$ and $B \simeq C$, then $A \simeq C$.

Observe: We get equivalence classes but no ordering.

Injection Principle: $|A| \leqslant |B|$ *iff* $A \simeq C$ for some $C \subseteq B$.

`Reflexive:` $|A| \leqslant |A|$.

`Transitive:` if $|A| \leqslant |B|$ and $|B| \leqslant |C|$, then $|A| \leqslant |C|$.

`Anti-Symmetric:` if $|A| \leqslant |B|$ and $|B| \leqslant |A|$, then $|A| = |B|$.

Cantor-Schroeder-Bernstein Theorem: If there are injective functions $f : A \to B$ and $g : B \to A$, then there is a bijection $h : A \to B$.

`Total:` $|A| \leqslant |B|$ or $|B| \leqslant |A|$. (Requires the Axiom of Choice)

Could Define: $|A| = |B|$ *iff* $|A| \leqslant |B|$ and $|B| \leqslant |A|$.

$|A| < |B|$ *iff* $|A| \leqslant |B|$ and $|B| \nleqslant |A|$.

Countably Infinite

Countable: A set A is *countable iff* $|A| \leqslant |\mathbb{N}|$.

Infinite: A set A is *infinite iff* $|\mathbb{N}| \leqslant |A|$.

- $\mathbb{N}_m$ is countably infinite since $f(n) = n + m$ is a bijection.
- $\mathbb{N}_{(m)}$ is countably infinite since $f(n) = n \times m$ is a bijection.
- $\mathbb{Z}$ is countably infinite since there is a bijection $f(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even} \\ \frac{-(n+1)}{2} & \text{otherwise.} \end{cases}$.
- The positive rational numbers $\mathbb{Q}^+$ are countably infinite since:
  - There is an injection from $\mathbb{Q}^+$ to $\mathbb{N}^2$.
  - And $f(\langle n, m \rangle) = 2^n \cdot 3^m$ is an injection from $\mathbb{N}^2$ to $\mathbb{N}$.
  - Hence $\mathbb{Q}^+$ is countable, and so $\mathbb{Q}$ is also countable.
  - Infinite since identity is an injection from $\mathbb{N}$ to $\mathbb{Q}$.

Real Numbers

Real Interval: The real interval $(0,1)$ is uncountably infinite.

1. $|\mathbb{N}_2| \leqslant |(0,1)|$ since $f(x) = 1/x$ is an injection $f : \mathbb{N}_2 \to (0,1)$.
2. $|\mathbb{N}_2| \neq |(0,1)|$ by Cantor's diagonal argument.
3. Thus $|\mathbb{N}_1| < |(0,1)|$.
4. Observe that $g(x) = \pi(x - 1/2)$ is a bijection $g : (0,1) \to (-\pi/2, \pi/2)$.
5. Additionally $\texttt{tan} : (-\pi/2, \pi/2) \to \mathbb{R}$ is a bijection.
6. By the bijection principle, $|(0,1)| = |(-\pi/2, \pi/2)| = |\mathbb{R}|$.
7. Thus $|\mathbb{N}_2| < |\mathbb{R}|$ where $|\mathbb{N}_2| = |\mathbb{N}|$, so $|\mathbb{N}| < |\mathbb{R}|$.

Cantor's Theorem

Theorem: $|A| < |\wp(A)|$ for any set A where $\wp(A) = \{X : X \subseteq A\}$.

1. $|A| \leqslant |\wp(A)|$ since $f(a) = \{a\}$ is an injection.
2. Assume there is a bijection $f : A \to \wp(A)$.
3. Let $D = \{a \in A : a \notin f(a)\}$.
4. Since $D \subseteq A$, we know that $D \in \wp(A)$.
5. Since f is surjective, $f(d) = D$ for some $d \in A$.
6. But $d \in f(d)$ *iff* $d \in D$ *iff* $d \notin f(d)$.
7. This has the form $P \leftrightarrow \neg P$ which is equivalent to $P \wedge \neg P$.
8. Thus there is no bijection $f : A \to \wp(A)$, and so $|A| \neq |\wp(A)|$.
9. Given the above, $|A| < |\wp(A)|$.

Corollary

Universal Set: There is no set of all sets.

1. Suppose there were a set U of all sets.
2. Since every $X \in \wp(U)$ is a set, $\wp(U) \subseteq U$.
3. So $f(x) = x$ is an injection $f : \wp(U) \to U$.
4. Thus $|\wp(U)| \leqslant |U|$.
5. Moreover, $g(x) = \{x\}$ is an injection $g : U \to \wp(U)$.
6. So $|U| \leqslant |\wp(U)|$.
7. Thus $|U| = |\wp(U)|$.
8. By Cantor's Theorem, $|U| < |\wp(U)|$, so $|U| \neq |\wp(U)|$.
9. Hence there is no set U of all sets, so no set of everything!

Axioms and Intuitions

Continuum Hypothesis: There is no set A where $|\mathbb{N}| < |A| < |\mathbb{R}|$.

Independent: Adding CH or its negation to ZFC is consistent if ZFC is consistent.

- Is it up to us to choose?
- Neither intuition nor mathematical practice seems to decide the issue.

Compare: Gödel showed that ZFC is consistent if ZF is consistent.

Axiom of Choice: Every set of sets X has a function f where $f(Y) \in Y$ for all $Y \in X$.

Well-Ordering Theorem: Every set X can be well-ordered (its subsets all have least elements).

- AC and WOT are equivalent, intuitive, and extremely useful.
- Not so for CH!

The Higher Infinite

PARADOX AND INFINITY
Benjamin Brast-McKie
May 13, 2025

The Continuum Hypothesis

Sizes of Infinity: We have seen that $|\mathbb{N}| < |\wp(\mathbb{N})|$ where $|\wp(\mathbb{N})| = |\mathbb{R}|$.

- $B = \{.b_0b_1b_2\ldots : b_i \in \{0,1\} \text{ for all } i \in \mathbb{N}\}$.
- $|\wp(\mathbb{N})| = |B| = |B/\{.\overline{0}, .\overline{1}\}| = |(0,1)| = |(-\pi/2, \pi/2)| = |\mathbb{R}|$.
- $|S| = |S \cup A|$ whenever $|\mathbb{N}| \leqslant |S|$ and $|A| \leqslant |\mathbb{N}|$.

Continuum Hypothesis: There is no set A where $|\mathbb{N}| < |A| < |\mathbb{R}|$.

Independence: Adding CH or its negation to ZFC is consistent if ZFC is consistent.

- $ZFC + CH$ is consistent if ZFC is consistent (Kurt Gödel 1940).
- $ZFC + \neg CH$ is consistent if ZFC is consistent (Paul Cohen 1963).

Convention: Is it up to us to choose which we include?

- Neither intuition nor mathematical practice seems to decide the issue.
- Platonism, conventionalism, and pragmatism.

The Axiom of Choice

Axiom of Choice: Every set of sets X has a function f where $f(Y) \in Y$ for all $Y \in X$.

- Gödel (1938) showed that ZFC is consistent if ZF is consistent.
- Cohen (1963) showed that ZF¬C is consistent if ZF is consistent.
- How does AC compare to CH?

Well-Ordering Theorem: Every set X can be well-ordered (its subsets all have least elements).

- AC and WOT are equivalent, intuitive, and extremely useful.
- `Totality`: $|A| \leqslant |B|$ or $|B| \leqslant |A|$ for all sets A and B.
- That `Totality` is equivalent to the WOT is good reason to accept AC.

Orderings

Weak Total Ordering: $\langle X, \leqslant \rangle$ reflexive, anti-symmetric, transitive, and total.

Strict Total Ordering: $\langle X, < \rangle$ asymmetric, transitive, and total.

- The irreflexive kernel of WTO is STO; reflexive closure is the inverse.

Total Well-Ordering: A WTO/STO where every subset has a least element.

The Ordinals

Something from Nothing: $\varnothing,\ \{\varnothing\},\ \{\varnothing,\{\varnothing\}\},\ \{\varnothing,\{\varnothing\},\{\varnothing,\{\varnothing\}\}\},\ldots$

Infinite Succession: $0, 0', 0'', 0''', \ldots$ where taking $n' = n + 1$ makes these look familiar.

Successor: $\alpha' = \alpha \cup \{\alpha\}$.

Successor Ordinal: α is a *successor ordinal iff* $\alpha = \beta'$ for some ordinal β.

- Every ordinal has a successor and contains all of its predecessors.
- And its predecessors contain their predecessors, and so on.

Set-Transitive: For any ordinal α, if $\beta \in \alpha$ and $\gamma \in \beta$, then $\gamma \in \alpha$.

Ordering: $\alpha <_o \beta \coloneqq \alpha \in \beta$.

Question: Are the successor ordinals all of the ordinals there are?

Omega: Let $\omega = \{0, 0', 0''', \ldots\}$ be the smallest set to contain 0 that is closed under the successor operation, i.e, $\alpha' \in \omega$ whenever $\alpha \in \omega$.

- ω is not a successor ordinal.

Question: Is ω an ordinal? What's an ordinal?

Ordinal: α is an *ordinal iff* α is set-transitive and well-ordered by $<_o$.

Key Ideas: Ordinals contain their predecessors and always bottom out.

- Not all ordinals have a greatest predecessor, i.e, are successor ordinals.

Limit Ordinal: α is a *limit ordinal iff* α is an ordinal that is not a successor ordinal.

Continuation: $0, 0', 0'', 0''', \ldots, \omega, \omega', \omega'', \omega''', \ldots$ where $'$ is defined as before.

Question: How shall we write the next limit ordinal?

- $\omega + \omega = \omega \times 0''$ but $0'' \times \omega = \omega$ and $\omega + 0'' \neq 0'' + \omega$.
- $|\omega + \omega| = |\omega|$.

Well-Order Types

Cantor Ordinals: Consider $c,\ \{c\},\ \{c,\{c\}\},\ \{c,\{c\},\{c,\{c\}\}\},\ldots$ where 'c' names Cantor.

Question: Couldn't we repeat all the same tricks, substituting 'c' for '$\varnothing$'?

- What it Dedekind gets jealous and wants a hierarchy? Then Hilbert...

Well-Order Type: Every ordinal in any hierarchy is a well-ordered set.

Isomorphism: Let $\alpha \cong \beta$ *iff* there is a bijection $f : \alpha \to \beta$ such that $\gamma <_a \delta$ just in case $f(\gamma) <_b f(\delta)$ for all $\gamma, \delta \in \alpha$ where $<_a$ orders α and $<_b$ orders β.

Ordinals: The ordinals represent their own well-order type.

The Higher Infinite

PARADOX AND INFINITY
Benjamin Brast-McKie
February 14, 2024

The Ordinals

Cardinalities: Observe that $|\omega| = |\omega + 1| = |\omega + 2| = \ldots = |\omega + \omega| = \ldots$

- This is just Hilbert's hotel all over again.

Structure: But $\omega \neq \omega + 1 \neq \omega + 2 \neq \ldots \neq \omega + \omega \neq \ldots$

- Ordinals have more structure than is encoded by their cardinalities.
- 2 represents the class of two things where one is *after* the other.

Order: The ordinals have order structure— they are *well-ordered*.

- α and β have the same order-type *iff* $\langle \alpha, <_o \rangle \cong \langle \beta, <_o \rangle$.
- Even though $\omega \simeq \omega + 1$, we don't get that $\langle \omega, <_o \rangle \cong \langle \omega + 1, <_o \rangle$.
- Moreover, $1 + \omega = \omega$ even though $\omega + 1 \neq \omega$.

Sequences: If we care about numbers, why introduce the ordinals at all?

- Isn't this a bate and switch?
- Why do we need transfinite order structure?

Ordinals and Algorithms

Algorithms: Programs/instructions to add 1 some (ordinal) number of times.

- Roughly: '3' says 'add 1 three times'.
- Note that we use 'three' to say what '3' means.

First Pass: $3 + 2 = 0''' + 0'' = 0'''' + 0' = 0''''' + 0 = 0''''' = 5$.

- $\alpha + \beta' = \alpha' + \beta$.
- Doesn't generalize since $\omega + \omega = \omega' +$?

Adding Limit Ordinals: How should we think about adding 1 $\omega + \omega$ times?

- What if we could parallel process (add)?
- Then we would get $\omega + \omega = \omega$.
- Need to preserve order and "parallel processing" ignores order.

Simpler Case: $\omega + 1 \neq 1 + \omega$.

- Think of the simplest case you can to exhibit some feature.
- Not always the first case you might think of.

Conclusion: If we ignore order, we are back to cardinality.

Ordinal Arithmetic

Addition: $\alpha + 0 = \alpha$
$\alpha + \beta' = (\alpha + \beta)'$
$\alpha + \lambda = \bigcup\{\alpha + \beta : \beta <_o \lambda\}$ where $\lambda \neq 0$ is a limit ordinal.

Examples: $3 + 2 = 3 + 1' = (3 + 1)' = (3 + 0')' = (3 + 0)'' = 3'' = 5$.
$\omega + \omega = \bigcup\{\omega + 0, \omega + 1, \omega + 2, \ldots\} = \{0, 1, 2, \ldots, \omega, \omega + 1, \omega + 2, \ldots\}$.
$1 + \omega = \bigcup\{1 + 0, 1 + 1, 1 + 2, \ldots\} = \{0, 1, 2, \ldots\} = \omega$.

Multiplication: $\alpha \times 0 = 0$
$\alpha \times \beta' = (\alpha \times \beta) + \alpha$
$\alpha \times \lambda = \bigcup\{\alpha \times \beta : \beta <_o \lambda\}$ where $\lambda \neq 0$ is a limit ordinal.

Examples: $\omega \times 2 = (\omega \times 1) + \omega = ((\omega \times 0) + \omega) + \omega = (0 + \omega) + \omega = \omega + \omega$.
$\omega \times \omega = \bigcup\{\omega \times 0, \omega \times 1, \omega \times 2, \ldots\}$

$$= \begin{cases} 0, 1, \ldots \\ \omega, \omega + 1, \ldots \\ \omega + \omega, (\omega + \omega) + 1, \ldots \\ \quad \vdots \quad , \quad \vdots \quad , \quad \vdots \quad , \end{cases}$$

"Small" Cardinals

Recall: $|\mathbb{N}| < |\wp(\mathbb{N})| < |\wp^2(\mathbb{N})| < \ldots < |U|$.

- Remember that '$<$' means 'injection but no bijection' here.
- We can use ω to define U.

Definition: Let $\mathfrak{B}_\alpha = \begin{cases} \mathbb{N} & \text{if } \alpha = 0 \\ \wp(\mathfrak{B}_\beta) & \text{if } \alpha = \beta' \\ \bigcup\{\mathfrak{B}_\gamma : \gamma <_o \alpha\} & \text{otherwise.} \end{cases}$

Example: $\mathfrak{B}_\omega = \bigcup\{\mathbb{N}, \wp(\mathbb{N}), \wp^2(\mathbb{N}), \ldots\} = U$.

Uncountable Ordinals

Countable Ordinals: We have only seen countable ordinals so far.

Beth: Let $\beth_\alpha = \beta$ *iff* $|\beta| = |\mathfrak{B}_\alpha|$ and $\beta <_o \gamma$ for all γ where $|\gamma| = |\mathfrak{B}_\alpha|$.

- $\beth_0 = \omega$ since $|\omega| = |\mathfrak{B}_0| = |\mathbb{N}|$.
- We can then make big ordinals to make even bigger cardinals, etc.

Motivations: Why care about all of this?

- Because math (is awesome).
- We are probing the limits of what is thinkable, not useful.

Omega Sequences

PARADOX AND INFINITY
Benjamin Brast-McKie
February 20, 2024

Ordinals into Obscurity?

Definition: Let $\mathfrak{B}_\alpha = \begin{cases} \mathbb{N} & \text{if } \alpha = 0 \\ \wp(\mathfrak{B}_\beta) & \text{if } \alpha = \beta' \\ \bigcup\{\mathfrak{B}_\gamma : \gamma <_o \alpha\} & \text{otherwise.} \end{cases}$

Obscurity: But all of this is only studied within set theory.

- Neither standard mathematics nor the sciences need ordinals beyond ω nor cardinalities beyond the continuum.
- But there are puzzles that arise even for ω sequences.
- Simplest case of the infinite worth exploring.

Motivations: Why care about all of this?

- Because we can.
- Because it's awesome (in the religious sense).
- We are probing the limits of what is thinkable, not useful.

Zeno's Analysis

Dichotomy Paradox: "That which is in locomotion must arrive at the half-way stage before it arrives at the goal." – Aristotle, Physics VI:9, 239b10

Infinite Task: Doing infinitely many tasks, each taking a non-zero amount of time.

- Some infinite tasks are cannot be performed in a finite amount of time.

Question: What about infinite tasks with strictly decreasing times for each task?

- The harmonic series is a counterexample: $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges.
- $H = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \ldots \geqslant 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{6} + \frac{1}{6} + \ldots = \frac{1}{2} + H$.

Super Task: An infinite task that is performed in finite time.

- Example: walking across the room since $\sum_{n=1}^{\infty} \frac{1}{2^n}$ converges to 1.
- $f(1) = \frac{1}{2}, f(2) = \frac{3}{4}, f(3) = \frac{7}{8}, f(4) = \frac{15}{16}, \ldots$, so $f(n) = 1 - \frac{1}{2^n}$.
- For any $\epsilon > 0$, there is some $n \in \mathbb{N}$ where $|1 - f(m)| < \epsilon$ for any $m > n$.

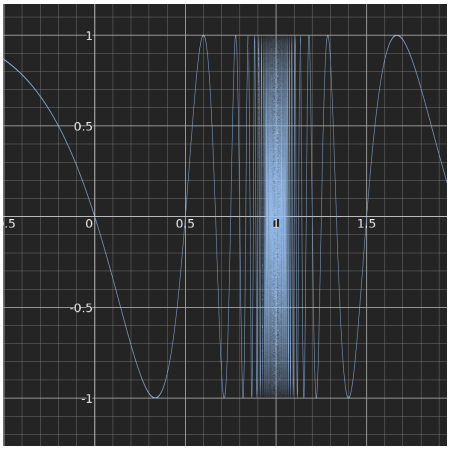
Paradox: So there are super tasks, though this would have surprised Zeno.

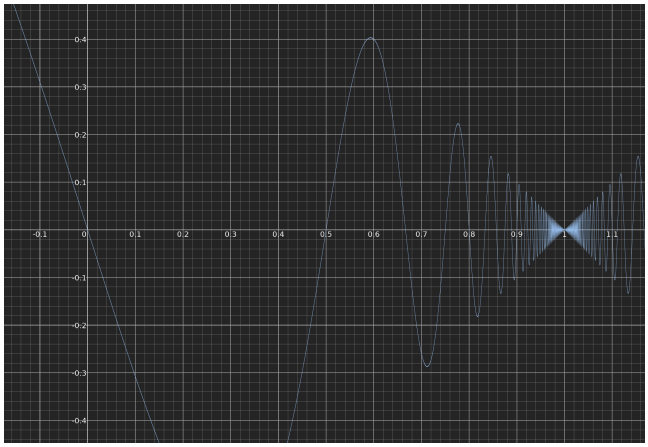
- It took the development of analysis in 17th century to solve.
- Analysis was not put on solid foundations until the 19th century by Bolzano, Cauchy, and Weierstrass.
- Poster child of illuminating paradoxes, but not all are like this.

Thomson's Lamp

Descriptions: 60s (off); 30s (on); 15s (off); ...

- Compare $f(n) = \texttt{sin}(\frac{\pi}{1-x})$ to $g(n) = \texttt{sin}(\frac{\pi}{1-x})(1-x)$.





- No continuous function satisfying the description is defined at $n = 1$.
- Subsequences of a convergent sequence converge to the same limit as the original, so neither the above nor $1, -1, 1, -1, \ldots$ converge.

Demon's Game

Setup: "As long as only finitely many of you say *aye*, each of you will receive as many gold coins as there are people who said *aye*."

- Assumes everyone is "optimally rational" and cannot collaborate.
- Is maximizing really "optimally rational" in this scenario?

Individual Version: "If you answer *aye* at most finitely many times, you will receive as many gold coins as the *aye*-answers that you give."

- Assumes no diachronic collaboration between time-slices.
- It is wrong to assume that the will is unable to persist across times.

Video Rental: $5 to rent, $2 late fee, but it is always worth it to Daniel to pay the fee.

- This is not a paradox, just a problem that Daniel has.
- Also a problem for a theory of rationality that takes Daniel to be ideally rational given his preferences at each time.

Buridan's Ass: Compare infinite ever larger bales of hay to the duplicate bale case.

- The ass does not starve, but not because of its sins against rationality.
- We can choose between duplicates/an arbitrary cut-off point.

Omega Sequences

PARADOX AND INFINITY
Benjamin Brast-McKie
February 21, 2024

Paradox Grading Rubric

Grading: Let 0 be low and 10 be high for the following qualities:

- Informative/illuminating.
- Intelligible/salient/compelling on a first take.
- Difficult to analyze/make progress.
- Examined/well-studied.
- Controversial/unsolved.
- Dangerous/risk of devouring one's career.
- Fundamental/influential for other areas.

Filthy Liars

Blackboard: The only sentence on the blackboard in room 32-141 is false.

Self-Reference: This sentence is false.

Pairs: (A) B is true. (B) A is false.

Truth Predicate: $\ulcorner T(\ulcorner A \urcorner) \urcorner$ is a sentence for every sentences A.

Truth Schema: $T(\ulcorner A \urcorner) \leftrightarrow A$.

Fixed Point Lemma: If $\varphi(x)$ has at most x free , $PA \vdash \varphi^\star \leftrightarrow \varphi(\varphi^\star)$ for some $\varphi^\star$.

Liar: Letting $\varphi(x) = \neg T(x)$, then $\vdash L \leftrightarrow \neg T(\ulcorner L \urcorner)$ where $\varphi^\star = L$.

- If $T(\ulcorner L \urcorner)$, then L by the *Truth Schema*.
- So $\neg T(\ulcorner L \urcorner)$ by *Liar*, and so $\bot$.
- If $\neg T(\ulcorner L \urcorner)$, then L by *Liar*.
- So $T(L)$ by *Truth Schema*, hence $\bot$.
- So $T(\ulcorner L \urcorner) \vee \neg T(\ulcorner L \urcorner) \vdash \bot$, and so $LEM \vdash \bot$.

Classical Logic

Excluded Middle: $\vdash A \vee \neg A$ (Every sentence is either true or not true).

Ex Falso Quidlebet: $P, \neg P \vdash Q$.

- EFQ follows from *Disjunction Introduction* and *Disjunctive Syllogism*.
- Dialetheists like Priest even give up *Modus Ponens*.

Self Reference

Claim: The problem arises from self-reference.

Insufficient: Some self-reference is OK.

- I hope this letter finds you well.
- This sentence contains five words.

Not Necessary: The sentence $\neg T(L)$ does not refer to itself.

Infinite Liar

Yablo: Even more trouble without self-reference.

- s_k: s_n is false for every $n > k$.
- s_k is true *iff* s_n is false for every $n > k$.

Finite Sequences: No paradox arises since we reach a semantic bottom.

- But if that bottom is always deferred, we are in trouble.

Solution?

Question: What would a solution look like?

- Restrict language to avoid paradoxes, i.e., deriving contradictions.
- On its own, even a successful restriction would be *ad hoc*.
- Also need to explain why that restriction is in place.

Metalanguages

Truth: We can only define truth for $\mathcal{L}_n$ in a metalanguage $\mathcal{L}_{n+1}$ for $\mathcal{L}$.

- Assume $\mathcal{L}_0$ contains no truth-predicates.
- $\mathcal{L}_{n+1}$ may include $A \leftrightarrow T_n(\ulcorner A \urcorner)$ for each sentence A of $\mathcal{L}_n$.
- T_n does not apply to sentences in $\mathcal{L}_n$, and so no paradox.

Question: Is this a good response for natural languages?

- Truth would then be radically polysemous.
- What if we define super-true as true in any sense?

Self Reference

PARADOX AND INFINITY
Benjamin Brast-McKie
February 26, 2024

From Cantor to Russell

Cantor's Theorem: Recall the proof that $|A| \neq |\wp(A)|$.

- Assume there is a bijection $f : A \to \wp(A)$.
- Let $D = \{a \in A : a \notin f(a)\}$.
- Since $D \subseteq A$, we know that $D \in \wp(A)$.
- Since f is surjective, $f(d) = D$ for some $d \in A$.
- But $d \in f(d)$ *iff* $d \in D$ *iff* $d \notin f(d)$.
- This has the form $P \leftrightarrow \neg P$ which is equivalent to $P \wedge \neg P$.
- Thus there is no bijection $f : A \to \wp(A)$, and so $|A| \neq |\wp(A)|$.

Universal Set: There is no set of all sets.

- Suppose there were a set U of all sets.
- Consider the identity map $f : U \to U$.
- Let $R = \{a \in U : a \notin f(a)\}$.
- Since $R \in U$, we may ask whether $R \in R$.
- But $R \in R$ *iff* $R \notin f(R)$ *iff* $R \notin R$.
- Hence there is no set U of all sets.

Burali-Forti Paradox

Ordinals: There is no set of all ordinals.

- Suppose there were a set Ω of all ordinals.
- Ω is set-transitive: if $x \in \Omega$ and $y \in x$, then $y \in \Omega$.
- Ω is well-ordered: if $X \subseteq \Omega$, then some $y <_o x$ for all $x \in X$.
  - If x and y are ordinals, then $x <_o y$ or $y <_o x$.
  - Ordinals contain all of their predecessors.
- So Ω is an ordinal, and hence $\Omega \in \Omega$, and so $\Omega <_o \Omega$.
- But $x \not<_o x$ for any ordinal x.
- Or, observe that $\Omega <_o \Omega'$ where $\Omega' = \Omega \cup \{\Omega\}$.
- Hence Ω does not include all ordinals.

Properties Paradox

Horse: The property *being a horse* is not a horse, i.e., does not instantiate itself.

Property: The property *being a property* is a property, i.e., instantiates itself.

Paradox: Let P be the property of not instantiate itself, i.e., $P(X) \coloneqq \neg X(X)$.

- But then $P(P)$ *iff* $\neg P(P)$.
- $\exists Y[\forall Z(Z = Y \leftrightarrow \forall X[Z(X) \leftrightarrow \neg X(X)]) \wedge Y = P]$.

Universal Liar

Liar: The proposition that *Liar* expresses is false.

- If the *Liar* is true, then by its own lights it is false.
- If the *Liar* is false, then by its own lights it is true.

Analysis: $\exists \varphi(\forall \psi[\texttt{Expresses}(\mathit{Liar}, \psi) \leftrightarrow \varphi = \psi] \wedge \neg \varphi)$.

Nonexistence?

Response: Isn't the most natural response to just deny that there is a set R, or property P, or proposition expressed by *Liar*.

Ad Hoc: Need to explain why there is no such set, property, or proposition.

Proposition: Why doesn't *Liar* express a proposition?

- Can't simply appeal to paradox to explain its nonexistence.

Properties: Why isn't there such a property as P?

- Seems like most properties have this property, e.g., *being a horse*.

Sets: Why isn't there a Russell set R?

- All sets do not belong to themselves, and there is no set of all sets.

Vicious Circle Principle

Diagnosis: "No totality can contain members defined in terms of itself."

- Want something that explains all of the "reflexive paradoxes."

Take Two: "Whatever contains an apparent variable must not be a possible value of that variable."

- $R \coloneqq \{x : x \notin x\}$ i.e., $\exists X(R = Y \wedge \forall Y[Y = X \leftrightarrow \forall z(z \in Y \leftrightarrow z \notin z)])$.

Types: "Whatever contains an apparent variable must be of a different type from the possible values of that variable…"

Self Reference

PARADOX AND INFINITY
Benjamin Brast-McKie
February 28, 2024

An Untyped Language

Vicious Circle Priciple: "Whatever contains an apparent variable must not be a possible value of that variable."

Language: Names $c_1, c_2, \ldots \in C$, variables $x_1, x_2, \ldots \in V$, predicates $R_1, R_2, \ldots \in P$, operators $\neg, \vee, \wedge, \rightarrow, \leftrightarrow, \forall\alpha, \exists\alpha$ where α is any variable.

Formulas: The set of *formulas* F is defined recursively:

- $R(\alpha_1, \ldots, \alpha_n)$ is a formula in F if $R \in P$ and $\alpha_1, \ldots, \alpha_n \in C \cup V$.
- $z(\varphi_1, \ldots, \varphi_n)$ is a formula in F if $z \in V$ and $\varphi_1, \ldots, \varphi_n \in C \cup V \cup F$.
- $\neg\varphi, \varphi \vee \psi, \varphi \wedge \psi, \ldots, \forall\alpha\varphi, \exists\alpha\varphi$ are formulas in F if $\varphi, \psi \in F$ and $\alpha \in V$.
- Nothing else is a formula in F.

Example: In the following examples z is of higher type:

- Mathematical induction shows that for any property z, if $z(0)$ and $z(x')$ whenever $z(x)$, then $z(x)$ for all numbers x.
- $\forall z(z \vee \neg z)$.

Self Reference: Observe that $\neg z(z)$ is a formula, but it will not have a type.

Ramified Theory of Types

Simple Types: The simple types will be defined recursively.

- 0 is the simple type of *individuals*.
- If $t_1, \ldots, t_n$ are simple types, then $(t_1, \ldots, t_n)$ is a simple type.
- Nothing else is a simple type.

Example: 'Kim' is type 0, 'is running' is type (0), and 'Kim is running' is type $()$.

Ramified: Also defined recursively:

- 0^0 is a ramified type.
- If $t_1^{o_n}, \ldots, t_n^{o_n}$ are ramified types, $o \in \mathbb{N}$, and $o > \mathtt{max}\{o_1, \ldots, o_n\}$, then $(t_1^{o_1}, \ldots, t_n^{o_n})^o$ is a ramified type where $o \geqslant 0$ if $n = 0$.
- Nothing else is a ramified type.

Example: 'Kim' is type 0^0, 'is running' is type $(0^0)^1$, and 'Kim is running' is type $()^1$.

Predicative Types: t^o is *predicative* if o is as low as it can be, c.f., $(0^0)^2$.

A Typed Language

Atomic Formulas: $R(c_1, \ldots, c_n)$ is *atomic* if R is a predicate and $c_1, \ldots, c_n$ are constants.

Typed Expresions: The expressions of the language will be typed recursively.

- $c : 0^0$ if c is a constant.
- $\varphi : ()^0$ if φ is atomic.
- If $\varphi : (t_1^{o_1}, \ldots, t_n^{o_n})^a$ and $\psi : (d_1^{r_1}, \ldots, d_n^{r_n})^b$, then:
  - $\neg\varphi : (t_1^{o_1}, \ldots, t_n^{o_n})^a$.
  - $\varphi \vee \psi : (t_1^{o_1}, \ldots, t_n^{o_n}, d_1^{r_1}, \ldots, d_n^{r_n})^{\max\{a,b\}}$.
  - $\vdots$
- If $(t_1^{o_1}, \ldots, t_n^{o_n})^a$ is predicative and $\varphi_1 : t_1^{o_1}, \ldots, \varphi_n : t_n^{o_n}$ for expressions $\varphi_1, \ldots, \varphi_n$, then $z(\varphi_1, \ldots, \varphi_n) : ((t_1^{o_1}, \ldots, t_n^{o_n})^a, t_1^{o_1}, \ldots, t_n^{o_n})^{a+1}$.
- If $\varphi : (t_1^{o_1}, \ldots, t_n^{o_n})^a$, then $\forall x : t_i^{o_i}\varphi : (t_1^{o_1}, \ldots, t_n^{o_n})^a$.
- There is more that we won't get into…

Typed Formuals: A *typed formula* is a typed expression that is a formula.

- $\neg z(z)$ is a formula, but cannot be typed.
- $\forall z(z \vee \neg z)$ can be typed.

Type Restrictions: "Whatever contains an apparent variable must be of a different type from the possible values of that variable…"

Axiom of Reducibility

Identity of Indiscernibles: $x = y$ *iff* $\forall z[z(x) \leftrightarrow z(y)]$.

Stratification: For each order $n \in \mathbb{N}$, we may articulate a version of the principle above: $x = y$ *iff* $\forall z : (0^0)^n[z(x) \leftrightarrow z(y)]$.

Planets Example: Do Hesperus and Phosphorus have the same first-order properties?

- But what if they differ on second-order properties?
- Maybe someone loves Hesperus but no one loves Phosphorus.

Induction Example: "A finite number is one which possesses *all* properties possessed by 0 and by the successors of all numbers possessing them."

- If something holds of all first-order properties, why think it holds for all higher-order properties?

Axiom of Reducibility: "[E]very propositional function is equivalent, for all its values, to some predicative function." (pp. 242-3)

Translation: Every typed formula is logically equivalent to some formula with a predicative type, and so: $x = y$ *iff* $\forall z : (0^0)^1[z(x) \leftrightarrow z(y)]$.

Type Theory

PARADOX AND INFINITY
Benjamin Brast-McKie
March 4, 2024

Reals from the Rationals

Construction: Dedekind aimed to construct the real numbers to ground analysis.

Definition: A *real number* is any nonempty proper subset $X \subset \mathbb{Q}$ where:

1. X is downward closed, i.e., $x \in X$ whenever $y \in X$ and $x < y$.
2. X has no greatest element, i.e., for every $x \in X$, there is some $y \in X$ where $x < y$.

Roots: Let $\sqrt{2} = \{x \in \mathbb{Q} : x^2 < 2\}$.

Higher-Order: The real numbers are of higher-order than the rationals.

Subsumption: The rationals are *subsumed* by the reals $\frac{2}{5} = \{x \in \mathbb{Q} : x < \frac{2}{5}\}$.

Greatest Lower Bounds

Ordering: For $x, y \in \mathbb{R}$, $x \leqslant y$ *iff* $x \subseteq y$.

Lower Bound: x is a *lower bound* of $X \subseteq R$ *iff* $x \leqslant y$ for all $y \in X$.

Greatest Lower Bound: x is a *greatest lower bound* of $X \subseteq R$ *iff* x is a lower bound of X and $x \geqslant y$ for any lower bound of y of X.

- The definition of the glb of X quantifies over real numbers in X.
- Mathematics is full of definitions of properties of every higher-order.
- We want to quantify over all properties of real numbers at once.

Properties: '$P(6)$' vs. '$6 \in \pi$' where $\pi = \{x \in \mathbb{N} : P(x)\}$.

Type Restrictions: "Whatever contains an apparent variable must be of a different type from the possible values of that variable…"

Predicative Properties

Typed Expresions: Recall the recursive clauses from last time:

- If $(t_1^{o_1}, \ldots, t_n^{o_n})^a$ is predicative and $\varphi_1 : t_1^{o_1}, \ldots, \varphi_n : t_n^{o_n}$ for expressions $\varphi_1, \ldots, \varphi_n$, then $z(\varphi_1, \ldots, \varphi_n) : ((t_1^{o_1}, \ldots, t_n^{o_n})^a, t_1^{o_1}, \ldots, t_n^{o_n})^{a+1}$.
- If $\varphi : (t_1^{o_1}, \ldots, t_n^{o_n})^a$, then $\forall x : t_i^{o_i} \varphi : (t_1^{o_1}, t_{i-1}^{o_{i-1}}, \ldots, t_{i+1}^{o_{i+1}}, t_n^{o_n})^a$.

Stratification: Quantification is stratified by ramified types.

Predicative Types: t^o is *predicative* if o is as low as it can be, c.f., $(0^0)^1$ and $(0^0)^2$.

Axiom of Reducibility

Axiom of Reducibility: "[E]very propositional function is equivalent, for all its values, to some predicative function." (pp. 242-3)

- Want all properties of real numbers to be on a par.
- Russell assumes there always are equivalent predicative properties.
- What is the property *being the glb of X* equivalent to?
- Neither Russell nor anyone else can say.

Construction: This undermines the spirit of Dedekind's project.

- The aim is to construct the numbers (ultimately from the empty set).
- But Russell was a logicist, not an constructivist/intuitionist.
- Logicists have a much more realist conception of mathematics where logic describes objective universal principles (the laws of thought) and mathematics reduces to logic.
- Nevertheless, one might worry that the AR is not logical, i.e., it is thinkable (consistent) for it to be false.

Mathematics or Metaphysics

Planets Example: Let Hesperus and Phosphorus have the same first-order properties.

- Hesperus is shining *iff* Phosphorus is shining, etc.
- Could they differ in higher-order properties?
- Maybe someone loves Hesperus but no one loves Phosphorus?
- There is a relation someone bears to Hesperus but not to Phosphorus?
- Given that Hesperus is Phosphorus, this might seem unlikely.

Iron Spheres: Consider two iron spheres that have all the same properties.

- Could there be a higher-order property upon which they differ?

Logic: Logic is concerned with what must be the case.

- By 'must' we mean: it would be contradictory for it to not be the case.
- Compare the axioms of set theory which are contingent.
- Must the axiom of reducibility hold?
- Many have thought it contingent (e.g., Ramsey and Wittgenstein).

Logicists: Ensuring AR is a truth of logic is required for logicism.

- Reducing math to logic is an ambitious program.
- Neither logicists and intuitionists can easily accommodate AR.

Type Theory

PARADOX AND INFINITY
Benjamin Brast-McKie
March 6, 2024

Against Ramification

Orders: Ramsey rejects Russell's divisions of the types into orders.

Autological: A predicate is *autological iff* it expresses a property it has.

Heterological: A predicate is *heterological iff* it is not autological.

- Whereas 'short' is autological, 'long' is heterological.
- Is 'heterological' heterological?
- Use/mention distinction.

Solution: In the style of *Principia Mathematica* we have:

- $H(w)$ *iff* there is a property P where w expresses P and $\neg P(w)$.
- The variable 'P' ranges over first-order properties which H is not.
- "This theory of a hierarchy of orders of functions of individuals escapes the contradictions; but it lands us in an almost equally serious difficulty, for it invalidates many important mathematical arguments…"

Simple Type Theory

Logical: Ramsey accepts Russell's solution to the logical paradoxes.

Simple Types: Recall the function application clause from the simple theory of types:

- If $\varphi_1 : t_1, \ldots, \varphi_n : t_n$, then $z(\varphi_1, \ldots, \varphi_n) : ((t_1, \ldots, t_n), t_1, \ldots, t_n)$.
- Observe that z is of *higher type* than its arguments.

Self-Application: A predicate, "cannot significantly take itself as argument…"

- Neither $P(P)$ nor $\neg P(P)$ can be simply typed, and so meaningless.
- This case differs from *heterological* since no semantic terms occur.

Question: Is there any self application in $R \coloneqq \{x : \neg \mathtt{In}(x, x)\}$?

Sets: Russell identifies classes (sets) with functions (properties).

- '$x \in y$' is notational variant of '$y(x)$'.
- Thus '$x \in x$' abbreviates '$x(x)$' which cannot be simply typed.

Question: Why not respond by faulting the identification of sets with properties?

- Blocking this identification further undoes Russell's solution.
- But it does not solve the paradox.
- Ramsey thought we should leave this part of Russell's solution.

Dividing the Paradoxes

Russell: Solves all the paradoxes by accepting RTT + AR.

Ramsey: Rejects AR, and also rejects RTT to preserve mathematics.

- Ramsey divides the paradoxes into the logical and semantic.
- The semantic paradoxes include naming, expressing, meaning, etc.
- The semantic paradoxes concern the interpretation of language.
- Ramsey thought they deserved their own solution.

Tarski: Recall Tarski's levels of languages.

- Could distinguish between true_1, true_2, etc.
- The semantic terms for a language cannot belong to that language.
- The object/metalanguage distinction is widely accepted.

Truth: What of a theory of truth, meaning, reference, etc.?

- The object/metalanguage distinction doesn't provide a theory of truth.
- Nor does it provide a theory of meaning, or reference, etc.
- Lots remains to be done to understand how language works.

Higher-Order Logic

Mathematics: Truth theories remain controversial but mathematics is preserved.

- The logical paradoxes are solved.
- We also get the theory of simple types as a result.

Higher-Order: Since talk of orders has been banished, we may reclaim the term.

- Today, 'order' refers to the type of the variable bound by a quantifier.
- The quantifiers of *first-order logic* bind variables in name position.
- The quantifiers of *higher-order logics* bind variables of higher type.

Quine: Thought higher-order logic was "set theory in sheep's clothing."

- Instead of recasting sets as properties, *first-orderists* go the other way.
- But then how are we to escape Russell's paradox (this is for next week).

Types: Consider the following defense of higher-order logic:

- Properties, relations, etc., are not first-order objects.
- If we recognize type distinctions, why only first-order quantifiers?
- If there are some higher-order things, why not quantify over them?

Controvercy: Quine's shadow remains long in philosophy but not computer science.

Set Theory

PARADOX AND INFINITY
Benjamin Brast-McKie
March 11, 2024

Motivations

Dialectic: Recall Ramsey's simple theory of types.

- Sets are replaced with properties.
- '$x \in x$' is treated as '$x(x)$' which cannot be typed.
- If '$\in$' is intelligible in its own right, the paradox remains.
- Set theory is more intuitive than simple type theory.
- Worth developing in place of or alongside type theory.

Naive Set Theory

Language: $\forall, \exists, \neg, \vee, \wedge, \rightarrow, \leftrightarrow, =, \in, S, x_1, x_2, \ldots, (,)$ are primitive symbols.

- For purposes of illustration we may add other predicates and names.
- But strictly speaking, the language of set theory is very austere.

Comprehension: Every open sentence with one free variable corresponds to a set.

- $\exists y \forall x(x \in y \leftrightarrow \varphi)$ where 'y' does not occur in 'φ'.
- From open sentences to predicates *vs.* definite descriptions.
- Sets are objects not properties.

Question: Why assume uniqueness.

Extensinoality: Sets are defined by their members.

- $\forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y$.
- The set of fish that walk is identical to the set of pigs that fly.
- Properties need not be identified with sets.
- Set theory restricts attention to the extensions of predicates.

Russell's Paradox

Russell Set: The open sentence '$x \notin x$' can be used to define a set.

- $R \coloneqq \{x : x \notin x\}$, i.e., $\exists y \forall x(x \in y \leftrightarrow x \notin x)$.
- $R \in R$ *iff* $R \notin R$.
- Restricting comprehension looks *ad hoc*.

The Iterative Conception of Set

Intuitions: No reason to expect our intuitions to be univocal.

- Lots of reasons to expect otherwise, e.g., *same number as*, etc.
- First impressions often have to be revised.

Extensinoality: Compare the following metaphysical theses.

- *What it is to be* (identical to) a set is to have the members it has.
- *What it is for* a set to exist is for its members to exist.
- Sets *ontologically depend* on their members: part of what it is for a set to exist is for all of its members to exist.
- Put otherwise: for a set to exist it is *necessary for* its members to exist.
- But then sets can't be members of themselves.

Sufficiency: Is the existence of some entities *sufficient for* a set to exist?

- Do you have to "put a lasso around" some entities to make a set?
- Intuitionists of a certain stripe might claim so.
- What about the empty set? Is it a product of our conceptual exertion.
- Platonists reject this, taking sets to exist objectively and necessarily.
- Existence of the members to be sufficient for the existence of a set.

Separation Axiom

Construction: Sets are constructed not in time, but in nature.

- The ingredients precede the product in constitution.
- Given any things, we have a set from which we can build new sets.

Comprehension: $\forall z \exists y \forall x (x \in y \leftrightarrow (y \in z \wedge \varphi))$, i.e., some $k \coloneqq \{x \in z : \varphi\}$ for any set z.

- What of Russell's paradox?
- For any set z, there is a set $R_z \coloneqq \{x \in z : x \notin x\}$.

Indefinite Extensibility: Whence the contradiction?

- Assume $R_z \in z$ for contradiction.
- Then $R_z \in R_z$ *iff* $R_z \in z$ and $R_z \notin R_z$ *iff* $R_z \notin R_z$.
- So $R_z \notin z$.
- Letting $z' \coloneqq z \cup \{R_z\}$, then $R_{z'} \notin z'$, so $z'' \coloneqq z' \cup R_{z'}$, etc.

Solution: Consider the following conclusions.

- There is no universal set.
- No set belongs to itself and so $R_z = z$ for any set z.
- Every set is indefinitely extensible.

Set Theory

PARADOX AND INFINITY
Benjamin Brast-McKie
March 13, 2024

Axioms and Theorems

Restriction: We will restrict the quantifiers to sets.

SEPARATION: $\forall z \exists y \forall x (x \in y \leftrightarrow (x \in z \wedge \varphi))$ where 'y' does not occur in 'φ'.

- Given any set z, there is a set of z's members that are φ.

Question: Could there be more than one subset of φs?

EXTENSIONALITY: $\forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y$.

- Extensionality guarantees uniqueness.
- Whereas *Extensionality* is an axiom, *Separation* is an *axiom schema*.

Question: Could there be a universal set U?

- Assume there is a universal set U where $\forall x (x \in U)$.
- $\exists y \forall x (x \in y \leftrightarrow (x \in U \wedge \varphi))$ from *Separation* with U for z.
- $\exists y \forall x (x \in y \leftrightarrow \varphi)$ since $x \in U$ for all x.
- $\exists y \forall x (x \in y \leftrightarrow x \notin x)$ by replacing φ with $x \notin x$.
- $\forall x (x \in r \leftrightarrow x \notin x)$ by existential elimination.
- $r \in r \leftrightarrow r \notin r$ by instantiating x with r.
- Hence $\neg \exists y \forall x (x \in y)$ is a theorem.

Theorems: We don't need an axiom to rule out U.

Question: What other axioms do we need to describe the concept?

Zermelo's Theory of Sets

NULL SET: $\exists y \forall x (x \notin y)$.

- There is a set with no members.

PAIRS: $\forall z \forall w \exists y \forall x (x \in y \leftrightarrow (x = z \vee x = w))$.

- For any sets x and w, there is a set whose only members are x and w.

UNIONS: $\forall z \exists y \forall x (x \in y \leftrightarrow \exists w (x \in w \wedge w \in z))$.

- For any set z, there is a set of all members of members of z.

Subset Definition: $x \subseteq z \coloneqq \forall w (w \in x \rightarrow w \in z)$

- Every member of x is a member of z.

POWER SET: $\forall z \exists y \forall x (x \in y \leftrightarrow x \subseteq z)$.

- For any set z, there is a set y of all subsets of z.

Infinite Sets

Question: Does anything guarantee that there are infinite sets?

- If we want there to be infinite sets, how would we guarantee this?

Contains the Null Set: $\varnothing \in y \coloneqq \exists x(x \in y \wedge \forall z(z \notin x))$.

Successor: $z = x' \coloneqq \forall y(y \in z \leftrightarrow (y \in x \vee y = x))$.[1]

INFINITY: $\exists y[\varnothing \in y \wedge \forall x(x \in y \rightarrow \exists z(z \in y \wedge z = x'))]$.

REGULARITY: $\exists x\varphi \rightarrow \exists x(\varphi \wedge \forall y(y \in x \rightarrow \neg\varphi[y/x]))$ where φ does not contain 'y' and $\varphi[y/x]$ is the result of replacing all occurrences of 'x' in φ with 'y'.

- If some set is such that φ, there is "smallest set" x that is such that φ.

Example: Letting φ be '$\exists z(z \in x)$', there is a set that only contains the empty set.

- $\exists x\exists z(z \in x) \rightarrow \exists x(\exists z(z \in x) \wedge \forall y(y \in x \rightarrow \forall z(z \notin y)))$.

Example: Assume there is a set that belongs to itself, i.e., $\exists x(x \in x)$.

- $\exists x(x \in x \wedge \forall y(y \in x \rightarrow y \notin y))$ by *Regularity*.
- $r \in r$ and $\forall y(y \in r \rightarrow y \notin y)$ by conjunction and existential elimination.
- $r \in r \rightarrow r \notin r$ by universal elimination.
- $r \notin r$ by conditional elimination.
- Hence $\neg\exists x(x \in x)$ is a theorem.

Stage Theory

Motivation: Why believe these axioms and not some others?

Iterative Conception: Because they conform to an iterative conception of set.

Stages: Here are the stage axioms.

- No stage is earlier than itself.
- Earlier than is transitive.
- Earlier than is connected/total.
- There is an earliest stage.
- Every stage has a next stage.

Formation: Here are the formation axioms.

- There is a limit stage which does not have a latest predecessor.
- Every set is formed at a unique stage.
- Every member of a set is formed earlier than that set.
- If the members of a set are formed before a stage the set is formed at that stage.

[1] Better to define the successor function $'$.

Absolute Generality

PARADOX AND INFINITY
Benjamin Brast-McKie
March 18, 2024

Restricted Quantifiers

Coffee: There is no more coffee.

- Of course there is some coffee somewhere.

Bags: I have packed everything.

- But not everything in existence.

Restriction: The domain of quantification is, somehow, restricted by context.

- Context does not change the meaning of 'there is' or 'everything'.
- The quantifiers have the same semantics, just the domain shifts.

Unrestricted Quantifiers

Biology: All humans are mortal.

Identity: Everything is self-identical.

Sets: No set is a member of itself.

- It appears that we can quantify over *everything*.
- If not, quantified claims are not as informative as they could be.

Question: Are we able to quantify over everything, at least sometimes?

Russellian Doubts

Reductio: Consider the following argument against unrestricted quantification:

1. Assume that we can quantify over all sets.
2. The domain of quantification must include all sets.
3. Quantifier domains are sets.
4. So there is a universal set of all sets.
5. We can derive naive comprehension from separation.
6. A contradiction follows by Russell's paradox.
7. Hence we cannot quantify over all sets.
8. Sets are things.
9. Thus we cannot quantify over everything.

Domain Free Quantification

All-in-One: Cartwright rejects (3) above.

- Can we quantify over everything though there is no set of everything?
- "There is no set that has as members all and only those things that are not members of themselves. But the things that are not members of themselves can simultaneously be the values of the variables of a first-order language; so at any rate I claim." —Cartwright (1994, p. 3)

Absolute Generality: Why can't x be instantiated by w in $\forall x(x \in w \leftrightarrow x \notin x)$?

- Because there is no set w according to ZFC.
- Rather, by *Separation*, we get $\forall x(x \in w' \leftrightarrow (x \in z \wedge x \notin x))$ for some z.
- But there is no universal set, and so at most $w' = z$.
- But if we can quantify over everything, what justifies *Separation*?

Relatively Unrestricted Quantification

Indefinite Extensibility: In quantifying unrestrictedly, new entities can always be defined.

- In particular, w falls outside the domain of the quantifier.
- Naive comprehension may be preserved, but w cannot instantiate x.

Self-Defeating: It is not possible to quantify over everything.

- Hence I am not quantifying over everything.
- So there is something that I am not quantifying over.
- But this is self-defeating, and so false.

Context Domains: Not everything is quantified over in context c.

- For the reasons above, this theses is false in c.

Context Principle: For any c, there is some c' where something quantified over in c' is not also quantified over in c.

- So not everything is quantified over in c, which is false in c.
- But c was an arbitrary context, so the *Context Principle* is false in any c.

Show Don't Tell: The relativist might claim only to be able to always shift the context.

- Start in c, Russell's paradox moves us to c' with broader quantifiers.
- Should we trust a theory that we can't state?

Absolutism: Claiming that we can quantify over everything is not self-defeating.

- We may quantify over everything but are still beholden to *Separation*.
- Not as many sets exist as we might think we are able to naively define.

Absolute Generality

PARADOX AND INFINITY
Benjamin Brast-McKie
March 20, 2024

Absolutism

Gloss: It is possible to quantify over absolutely everything.

- By 'possible' we mean 'there is an interpretation $I, \hat{a}$ of the language'.

Minimal Language: Consider a language $\mathcal{L}$ with just '$\forall$', 'x', and 'F' as primitive symbols.

- The extension $I(F) \subseteq D_I$ interprets the predicate 'F'.
- $\hat{a}$ is a variable assignment for $\mathcal{L}$ *iff* $\hat{a}(x) \in D_I$.
- 'Fx' is true in $I, \hat{a}$ *iff* $\hat{a}(x) \in I(F)$.
- '$\forall x Fx$' is true in $I, \hat{a}$ *iff* 'Fx' is true in $I, \hat{b}$ for every x-variant $\hat{b}$ of $\hat{a}$.

Restatement: For anything y there is an interpretation I of $\mathcal{L}$ where $y \in D_I$.

- Everything belongs to the domain of quantification D_I.
- $\forall y \exists D(y \in D)$.
- But this is just the claim that there is a universal set.

Direct Method: We must learn to understand absolute quantification directly.

- Model theory does not provide an adequate account.
- At most we can say: $\forall x \exists y(x = y)$.
- But this is trivial, failing to communicate the substance of absolutism.

No Domain Theory

Absolute Plurality: There is a plurality of everything, i.e., $\exists xx \forall y(y \prec xx)$.

- "They lifted the piano," "They won the championship," etc.
- We read '$y \prec xx$' as 'y is *one of* the xxs'.

Quinian Doubts: Some have thought that $\exists xx\varphi$ is just shorthand for $\exists x(S(x) \wedge \varphi)$.

- Similarly, '$y \prec xx$' is just shorthand for '$y \in x$'.
- But this flattens the absolutists current attempt to state their thesis.
- Instead the absolutist must take plural quantifiers to be primitive.
- We do seem to have plural quantifiers in English.

Plural Separation: For any*things*, there is a set of those that are such that φ.

- Formally: $\forall xx \exists y \forall x(x \in y \leftrightarrow x \prec xx \wedge \varphi)$ where 'y' doesn't occur in φ.
- Naive comprehension follows: $\exists y \forall x(x \in y \leftrightarrow \varphi)$.

Plural Comprehension: $\exists yy \forall x(x \prec yy \leftrightarrow \varphi)$ where 'yy' does not occur in φ.

Indefinitely Extensible?

Extensions: Predicates are interpreted by assigning them to sets.

- But how are we to interpret '`set`'?
- Suppose $I(\texttt{set}) \subseteq D$.
- But since $I(\texttt{set}) \notin I(\texttt{set})$, there is a set not in the extension of '`set`'.

Relativism: Claims that we can always extend any extension of the predicate '`set`'.

- $I \subseteq J$ *iff* $I(\kappa) \subseteq J(\kappa)$ for any predicate κ.
- $I \subset J$ *iff* $I \subseteq J$ and $J \nsubseteq I$.
- For any I of $\mathcal{L}$, there is some J of $\mathcal{L}$ where $I \subset J$.
- Every extension of '`set`' has a broader extension.

Absolutism: Instead of sets, suppose extensions are taken to be pluralities.

- The intended extension of '`set`' is the plurality of all sets.
- $\exists yy \forall x(x \prec yy \leftrightarrow \texttt{set}(x))$.
- The relativist may claim that this fails to capture indefinite extensibility.
- The absolutist is happy to avoid indefinite extensibility.

Time Travel

PARADOX AND INFINITY
Benjamin Brast-McKie
April 1, 2024

Time Travel

Question: What is it to travel in time?

Personal Duration: The journey's duration in personal time: Δ_p.

External Time: The time for Earth's frame of reference: t_i.

External Duration: The difference in the external start and end times: $\Delta_w \coloneqq t_e - t_b$.

Trival Time Travel: Occurs when $\Delta_w \neq 0$.

Non-Trival Time Travel: Occurs when $\Delta_p \neq \Delta_w$ (e.g., $0 < \Delta_p < \Delta_w$).

- If Bob travels close to the speed of light.

Extraordinary Time Travel: Occurs when Δ_w is negative or much greater than Δ_p.

Example: t_b is 10am EST, April 1st, 2024 and t_e is 10am EST, April 1st, 1924.

- Then $\Delta_w = -100$ years (i.e., 100 years in the past).

Metaphysical Possibility

Question: Is it possible to travel in time (in an extraordinary way)?

Practical Possibility: Is it possible for me to do a double backflip (on Earth's surface)?

Nomological Possibility: Is it nomologically possible for me to do a double backflip?

Metaphysically Possibility: Is it metaphysically possible to travel faster than the speed of light?

Objective Modality: Each modality concerns a range of objective *ways for things to be*.

- Metaphysical modality is the maximal objective modality.

Epistemic Modality: Is it possible that $a^n + b^n = c^n$ for some $a, b, c \in \mathbb{N}^+$ and $n > 2$?

- Fermat's last theorem was proven to be true (1995).
- Moreover, it is not possible for Fermat's last theorem to be false.
- Nevertheless, it may be epistemically possible for the uniformed.
- Or consider the epistemic possibility that $2,641 \times 31 \neq 81,971$.

The Possibility of Time Travel

Assume: It is practically impossible to travel back in time.

Agnostic: It is nomologically possible to travel back in time.

Question: Is it metaphysically possible to travel back in time?

World Travel

Actuality: Suppose that nobody has ever arrived from a future time.

- Deep in an MIT laboratory, Michele finishes her time machine.
- She gets in, eager to zip off into the distant past.
- Is it possible to travel into the past?

Branching Worlds: Is it possible to change the past?

- Instead of traveling to the actual past, one has traveled to another past.
- In what sense is traveling to a branching world count as time travel?

Take Two: Assume we restrict attention to time travel within one time-line.

- Michele can't travel back in time if she hasn't already arrived.
- If she has already arrived, she must travel back to those times.

Open Future: Is the future open if time travelers have already arrived?

- At least it is not as open as it might otherwise be assumed to be.
- But traveling back in time may be assumed to be entirely fixed.

Possibility: We don't just want to ask if our open future includes any time travel.

- We are asking if there are any worlds at all that include time travel.

Grandfather Paradox

Paradox: Tim travels to a time before his grandfather and grandmother met.

Question Can Tim kill his grandfather?

- If so, then neither Tim's father, nor Tim would have been born.
- So Tim wouldn't have traveled back in time, nor killed his grandfather.
- But how could Tim fail if appropriately poised to kill his grandfather?
- It would seem that Tim both can and cannot kill his grandfather.
- Perhaps this shows that time travel is not possible after all.

Equivocation: Or perhaps Tim can time travel, but only do exactly what he did.

- Considering everything, Tim cannot kill his grandfather.
- But 'can' is context sensitive, only taking some things into account.
- There are contexts which do not take everything into account where Tim *can* kill his grandfather.
- Tim has the necessary skills, position, timing, etc., he just doesn't do it.

Determined: Perhaps Tim can kill his grandfather even though it is impossible.

- Time travel has been defined in such a way that the future is closed.
- But as we will see next time, nothing forces this choice.

Time Travel

PARADOX AND INFINITY
Benjamin Brast-McKie
April 2, 2024

Future Possibility

Open Future: There is a reading of 'possible' that is about the future.

- Metaphysical modality is much broader than that reading.
- Future possibility reading is a subset of nomological possibility.
- The laws in our world may be incompatible with time travel.
- We intend to ask if there is any possible world in which someone travels into the actual past of that world.

Grandfather Paradox

Paradox: Tim travels to a time before his grandfather and grandmother met.

- Tim intends to kill his grandfather and is poised to do so.

Question: Can Tim kill his grandfather?

- If so, then neither Tim's parent, nor Tim would have been born.
- So Tim wouldn't have traveled back in time, nor killed his grandfather.
- But how could Tim fail if appropriately poised to kill his grandfather?
- It would seem that Tim both can and cannot kill his grandfather.
- Perhaps this shows that time travel is not possible after all.

Equivocation: Lewis takes this argument to equivocate on 'can'.

- Considering everything, Tim cannot kill his grandfather.
- But 'can' is context sensitive, only taking some things into account.
- There are contexts which do not take everything into account where Tim *can* kill his grandfather.
- Tim has the necessary skills, position, timing, etc., he just doesn't do it.

Example: Holding some facts fixed, I can speak Finnish.

- Holding others fixed, I cannot speak Finnish.
- Thus there are contexts in which Tim can kill his grandfather.
- Nevertheless, it is impossible for Tim to kill his grandfather.

Change: Do we mean to ask about Tim's abilities in this context sensitive sense?

- Instead we may ask: is it possible for Tim to kill his grandfather?
- The answer is already clear, but perhaps this is still frustrating to Tim.

Open Future

Determined: Tim's actions are entirely determined during his journey.

- This includes not killing his grandfather.
- But it may also include a whole lot else to which he is unaware.
- He is only able to become aware of some of the things that he can't do.
- And not just for Tim: no time traveler can kill their ancestor.
- But if there are any time travelers, then there are a lot of them.
- What explains the fact that none of them succeed is consistency.
- But it may still seem strange to Tim that however he tries he fails.

Actuality: For Lewis, each world is a space-time continuum.

- So the actual world is also a space-time continuum.
- So we all do only what we will do, though we don't know what.
- From this perspective, what is strange about time travel is that it puts us in a position to figure out what we cannot do.
- But determinism follows from the conception of possible worlds.

Open Future: Is time travel compatible with the open future?

- Consider a set of *world states* S and a *task relation* $\rightarrow$ over S.
- Let $\tau : \mathbb{Z} \rightarrow S$ be a *world history iff* $\tau(x) \rightarrow \tau(x+1)$ for all $x \in \mathbb{Z}$.
- Nothing binds us to a single world and so the future is open.
- Just because you travel to the actual past, nothing holds you there.
- But that does that mean such cases don't count as time travel?

Determinism: One cannot change *the* past or *the* future.

- But there need not be a unique past nor a unique future.

The Metaphysics of Time

PARADOX AND INFINITY
Benjamin Brast-McKie
April 8, 2024

Ameliorating Intuitions

Time: Calling something a theory of time does not make it a theory of time.

- Must fill the appropriate theoretical role, conforming to a significant extent with our intuitions.
- Compare a theory of sets that rejects extensionality.
- Or a theory of identity that rejects reflexivity.

Pre-Theory: Intuitions correspond to common ways of speaking about time.

- These ways of speaking serve our practical aims.
- Nothing as systematic as talk of sets used naively in mathematics.
- Our aim is to improve on this situation.

The B-Series

Earlier-Than: An asymmetric and transitive relation over events.

- The ordering of events by the earlier-than relation is called the B-series.

Events: Queen Anne's death; the poker is hot.

- Events may be understood roughly as instantaneous configurations.
- Does not capture a natural way of speaking about extended events.
- Could replace 'events' with 'states' or 'propositions'.

Russell: An event is past, present, or future only in relation to an event in time.

- Typically it is the event of assertion that we intend to relate.
- Queen Anne's death is past in relation to the present assertion event.
- But events are never past, present, or future *simpliciter*.

No Change: If e_1 is earlier than e_2, then e_1 is *always* earlier than e_2.

- The B-series does not change.
- Can a change be $\langle e_1, e_2 \rangle$ where e_1 is earlier than e_1?
- The poker being hot (e_1) is earlier than the poker being cool (e_2).

Space: Compare "change" over space.

- The tip of the poker is hot; the handle of the poker is not hot.
- But the poker need not change for this to be true.
- How does change over time differ from "change" over space?

The A-Series

Change: Events change from being future, then present, then past.

- A-series: *past*, *present*, *future*.
- Without the A-series their is no change at all.
- Everything in time must have each of the A-series properties.

Relational Properties: Some properties include other objects.

- *Being North of London* is a relational property (includes London).
- Non-relational properties may be called *absolute*.

Atemporal: At most, A-series properties relate events to something outside of time.

- If *past* is a relational property, it does not relate two events in time.
- Let '$P(e, x)$' read '*e is past relative to x*'.
- If x is an event, $P(e, x)$ is always or never the case, so cannot change.

Spotlight: What do the A-series properties include if not other events?

- *Was in*, *is in*, *will be in* "the spotlight."
- B-series as moving through the A-series.
- A-series as moving over the B-series.
- Film projector metaphor: was projected, is projected, will be projected.

Absolute: A-series properties may just as well be taken to be absolute.

- Either way, the A-series properties are incompatible.
- $Pe \vdash \neg Ne \wedge \neg Fe$; $Ne \vdash \neg Pe \wedge \neg Fe$; $Fe \vdash \neg Pe \wedge \neg Ne$.

Paradox

Argument 1: The A-series is essential to the reality of time.

P1 If time is real, then events change.

P2 If an event changes, then its A-series properties are what change.

P3 If an event's A-series properties change, events have A-series properties.

C1 Therefore, if time is real, then events have A-series properties.

Argument 2: Events do not have A-series properties.

P4 If an event has an A-series property, it has every A-series property.

P5 The A-series properties are incompatible.

C2 There are no events that have A-series properties.

Argument 3: Putting these first two arguments together, McTaggard concludes:

C3 Time is not real.

The Metaphysics of Time

PARADOX AND INFINITY
Benjamin Brast-McKie
April 10, 2024

Paradox

Argument 1: If time is real, then events have A-series properties.

P1 If time is real, then events change.

P2 If an event changes, then its A-series properties are what change.

P3 If an event's A-series properties change, events have A-series properties.

C1 Therefore, if time is real, then events have A-series properties.

Argument 2: Events do not have A-series properties.

P4 If an event has an A-series property, it has every A-series property.

P5 The A-series properties are incompatible.

C2 There are no events that have A-series properties.

Argument 3: Putting these first two arguments together, McTaggard concludes:

C3 Time is not real.

Being in Time

Respones: No event has every A-series property *at once*.

- If *e is* present, then *e was future* and *will be past*.
- So Fe at a past time p, and Pe at a future time f.

Repost: "But every moment, like every event, is both past, present, and future."

- So $\neg Fe$ when p is present or future, and $\neg Pe$ when f is present or past.
- The response generates the same problem, yielding a vicious regress.

Vicious: Is the regress really vicious?

- Is the contradiction ever avoided, or ever preserved?
- Compare building a set U out of members which include U.

Events: It becomes extremely artificial to speak in terms of events.

- Is Fe an event?
- Also, most events seem to occur over a duration, not at a time.

Tense: Involves a mixture of tense operators and temporal properties.

- Properties cannot be iterated, so best to stick to operators.
- Let '$\Diamond\!\!\!-\varphi$/$\Diamond\!\!\!+\varphi$' read 'It was/will be the case that φ'.

The Reality of Tense

Tense: Let φ be a sentence where e is the "event" of it being the case that φ:

- Replace Pe with $\Diamond\varphi$, replace Fe with $\Diamond\varphi$, and replace Ne with φ.

Inference Rules: In place of **P4** we may maintain $\varphi \vdash \Diamond\Diamond\varphi \wedge \Diamond\Diamond\varphi$.

- Also have $\Diamond\varphi \vdash \Diamond\Diamond\varphi \wedge \Diamond\Diamond\varphi$ and $\Diamond\varphi \vdash \Diamond\Diamond\varphi \wedge \Diamond\Diamond\varphi$.
- And $\Diamond\varphi \dashv\vdash \Diamond\Diamond\varphi$ and $\Diamond\varphi \dashv\vdash \Diamond\Diamond\varphi$.

Operators: To say $\Diamond\varphi$, $\Diamond\Diamond\varphi$, etc., is not to say that an event e has some property.

- Thus we need not say that Fe at a past time, nor Pe at a future time.
- No contradiction arises.

Semantics: Given a strict total ordering $\langle T, < \rangle$ of *times* where $x, y \in T$, consider:

- $x \vDash \Diamond\varphi$ *iff* $y \vDash \varphi$ for some $y < x$.
- $x \vDash \Diamond\varphi$ *iff* $y \vDash \varphi$ for some $y > x$.

Change: Let '$\odot\varphi$' read 'There is a change as to whether it is the case that φ'.

- $\triangle\varphi \coloneqq \Diamond\varphi \vee \varphi \vee \Diamond\varphi$ expresses that it is *sometimes* the case that φ.
- $\odot\varphi \coloneqq \triangle\varphi \wedge \triangle\neg\varphi$ expresses that things change (compare **P1**).

Does Time Flow?

Objection: The tense semantics does not capture the sense in which time flows.

- Suppose that $n \vDash \varphi \wedge \Diamond\Diamond\varphi \wedge \Diamond\Diamond\varphi$ where n is the present time.
- So $x \vDash \Diamond\varphi$ and $y \vDash \Diamond\varphi$ for some $x < n < y$.
- But these claims are permanent, i.e., they never change.

Impermenance: The metalinguistic claims about our language need not change.

- What changes are the claims made in the object language.
- Letting $\triangledown\varphi \coloneqq \neg\triangle\neg\varphi$, one might claim $\triangledown\exists p(p \wedge \neg\Diamond p \wedge \neg\Diamond p)$.
- Or consider the more radical claim $\triangledown\forall p(p \rightarrow \neg\Diamond p \wedge \neg\Diamond p)$.

Space: It would seem something similar may be said about space.

- Consider the poker where every point along it has a temperature.
- Let '$L\varphi$' and '$R\varphi$' read: 'To the left φ' and 'To the right φ'.
- If $0 \vDash 20^\circ$, then $-5 \vDash R20^\circ$ and $5 \vDash L20^\circ$.
- Thus we have not captured the difference between time and space.

Present: Whereas space has no privileged center, time has a privileged present.

- The present is what obtains, or perhaps all that exists.
- Maybe the past also has a privileged status, and is always growing.

Time and Change

PARADOX AND INFINITY
Benjamin Brast-McKie
April 16, 2024

Real Change

Grid: Consider a universe consisting of three pieces on a 3×3 grid.

- Consider three successive configurations of the grid in time.
- Compare this to three configurations of the grid separated in space.
- How does change across time differ from change across space?

Identity: The spatially separated grids are not identical.

- By contrast, the temporally separated grids are one.
- The properties of one and the same grid differ at different times.
- Call a complete configuration a *world state*.

Properties: What properties are to be included in a world state?

- A piece x is *shrew* at t *iff* either: (1) x is shaded at t and t is before 11am; or (2) x not shaded at t and t is after 11am.
- Suppose a shaded piece goes on being shaded at 11am.
- Does that piece change from being shrew to not being shrew at 11am?

Things: Consider the object which consists of the grid at different times.

- This object does not change in time, but goes on just as it is.
- Suppose we exclude temporally defined properties and things.
- We can ask what real properties every real thing has at a time.

Existence: World states determines which properties everything has.

- But suppose one grid ends and another begins at each change.
- Can still say that each grid is thus and so at each time.
- Do the things and properties that exist also change?

Change: A difference between the real properties real things have.

- Do two times differ only if there is a change between them?
- Something is some way at time t, and not that way at time t'.
- Could there be two times where the same things are all the same ways?

Real Possibility

Logical Possibility: "What is in question here is not whether it is physically possible for there to be time without change but whether this is logically or conceptually possible." — Shoemaker (p. 368, 1969)

- Important to distinguish logical possibility in the sense of consistency.
- It is consistent for the atom to be gold and to have only 6 protons.
- There is no *way for things to be* where the gold atom has only 6 protons.

Metaphysical Possibility: Broadest range of objective possibilities (ways for things to be).

- Fixing the interpretation, how must things be for the claim to be true?
- Is there any way whatsoever for things to be where the claim is true?
- Interpretational possibility concerns truth on any interpretation.

Total Freezes

Universe: Suppose there is a possible world with A, B, and C regions.

- Local freezes occur every 3rd year in region A.
- Local freezes occur every 4th year in region B.
- Local freezes occur every 5th year in region C.

Total: On certain years, the freezes in different regions align.

- A and B freeze together every 12th year.
- A and C freeze together every 15th year.
- B and C freeze together every 20th year.
- A, B, and C all freeze every 60th year.

No Change: On that 60th year, does time pass without change?

- More dramatically, could there be permanently frozen worlds?
- What about worlds that occupy the same world states more than once?

Possible Worlds: Let W be the set of *world states* and T a strict total order of *times*.

- A *world evolution* is any function from $\tau : T \to W$.
- Which world evolutions are possible worlds?
- Are constant functions permitted? What about loops?

Convention: Is it a matter of convention whether there are total freezes or not?

- Which of two bodies is rotating around each other?
- The year is exactly 365 days long with a one day freeze every 4th year.
- Compare the continuum hypothesis or axiom of choice for sets.

Newcomb's Problem

PARADOX AND INFINITY
Benjamin Brast-McKie
April 22, 2024

Aporia

Opinionated: Associated with being informed or knowledgeable (a strong look).

- Can be inappropriate for sensitive, complicated, or subtle matters.
- Can also make one less sensitive to alternatives (confirmation bias).

Uninformed: Having no opinion can be due to lack of knowledge about a case.

- Knowledge is valued over ignorance, and opinion signals knowledge.
- But opinion doesn't entail knowledge, nor is it valuable itself.

Aporia: Suspension of opinion for the sake of greater sensitivity to the truth.

- Aporia is often a state that is achieved since we often begin with biases.
- Aporia can be difficult to maintain, leading back to opinion.

Obvious: Taking one's assumptions/biases to be obvious is the opposite.

- Paradoxes are one way to achieve aporia, even if only temporarily.
- Aim is to see one's views in the context of many alternatives.

Two Boxes or One?

Boxes: A small box has \$1000 and a big box could have a \$1,000,000.

- On Wednesday, you will choose between two boxes or just the big box.
- On Monday, a predictor with 99% accuracy put \$1,000,000 in the big box if you are predicted to take the big box, and nothing otherwise.
- What would an ideally rational agent do?

Two Box

Dominance: Take both boxes, of course!

- The big box is either `Full` or `Empty` and not both.
- If `Empty`, then `OneBox` gives \$0 and `TwoBox` gives \$1,000.
- If `Full`, then `OneBox` gives \$1,000,000 and `TwoBox` gives \$1,001,000.
- `TwoBox` *dominates* `OneBox` since it never leaves you worse off.

Rational: To be rational, choose a dominant strategy if there is one.

One Box

Probability: But this ignores the probabilities for the outcomes `Full` and `Empty`.

- The actions under consideration are $\mathcal{A} = \{\texttt{OneBox}, \texttt{TwoBox}\}$.
- The outcomes $\mathcal{O}_A = \{\texttt{Full}, \texttt{Empty}\}$ are exclusive and exhaustive for $A \in \mathcal{A}$.
- $EV(A) = \sum_{i \in I_A} v(S_i^A) P(S_i^A | A)$ where $\mathcal{O}_A = \{S_i^A : i \in I_A\}$ and $A \in \mathcal{A}$.
- $EV(\texttt{OneBox}) = \$1,000,000 \times .99 + \$0 \times .01 = \$990,000$.
- $EV(\texttt{TwoBox}) = \$1,001,000 \times .01 + \$1,000 \times .99 = \$11,000$.
- So `OneBox` is 90 times higher expected utility than `TwoBox`.

Maximize: To be rational, choose an $A \in \mathcal{A}$ where $EV(A) \geqslant EV(B)$ for any $B \in \mathcal{A}$ (if any).

- Expected utility is maximized by `OneBox`.
- Thus a rational agent will `OneBox`.

Double or Nothing

Bets: You are given $1,000 with the option of drawing a stone from an urn.

- The urn has 51 black stones and 49 white, all well mixed.
- Drawing black doubles your winnings, but drawing white loses all.
- Should you take draw a stone or take the $1000?

Utility: Suppose a rational agent would maximize expected utility.

- $EV(\texttt{Take}_1) = \$1,000 \times 1.00 = \$1,000$ since $\mathcal{O}_{\texttt{Take}_1} = \{T_1\}$
- $EV(\texttt{Draw}_1) = \$2,000 \times .51 + \$0 \times .49 = \$1,020$ since $\mathcal{O}_{\texttt{Draw}_1} = \{B_1, W_1\}$.
- Since $EV(\texttt{Draw}_1) > EV(\texttt{Take}_1)$, the rational agent will draw.
- $EV(\texttt{Take}_2) = \$2,000 \times 1.00 = \$2,000$ since $\mathcal{O}_{\texttt{Take}_2} = \{T_2\}$
- $EV(\texttt{Draw}_2) = \$4,000 \times .51 + \$0 \times .49 = \$2,040$ since $\mathcal{O}_{\texttt{Draw}_2} = \{B_2, W_2\}$.
- The same reasoning may be repeated indefinitely.

Risk: What are the chances that you walk away with nothing after n draws?

- $1,000 after 0 draws with probability 1: $EV(0) = \$1,000$.
- $2,000 after 1 draws with probability .51: $EV(1) = \$1,020$.
- $4,000 after 2 draws with probability .26: $EV(2) = \$1,040$.
- $EV(n) = \$1,000 \times 2^n \times .51^n \geqslant \$1,000$ for all $n \geqslant 0$.

Certanty: Fails to account for the value of certainty.

- $EV(\texttt{Take}'_1) = (\$1,000 + v(\texttt{Certain})) \times 1.00 = \$1,000 + v(\texttt{Certain})$.
- $EV(\texttt{Draw}'_1) = (\$2,000 + \frac{1}{.51} v(\texttt{Certain})) \times .51 = \$1,020 + v(\texttt{Certain})$.
- Certainty has a different kind of value than money.

Newcomb's Problem

PARADOX AND INFINITY
Benjamin Brast-McKie
April 24, 2024

Green Grass

Drought: The town of Bayes has seen a terrible drought and the grass is dying.

- $P(\texttt{WetGrass} \mid \texttt{Umbrellas}) = .99$ and $P(\texttt{DryGrass} \mid \texttt{Umbrellas}) = .01$.
- $P(\texttt{WetGrass} \mid \neg\texttt{Umbrellas}) = .1$ and $P(\texttt{DryGrass} \mid \neg\texttt{Umbrellas}) = .9$.
- Assume that $v(\texttt{WetGrass}) = 10$ and $v(\texttt{DryGrass}) = -10$.
- $EV(\texttt{Umbrellas}) = .99 \times v(\texttt{WetGrass}) + .01 \times v(\texttt{DryGrass}) = 9.8$.
- $EV(\neg\texttt{Umbrellas}) = .1 \times v(\texttt{WetGrass}) + .9 \times v(\texttt{DryGrass}) = -8$.

Solution: Upon learning of these numbers, the mayor calls for the Bayesians to go outside with their umbrellas so that the grass can get some water.

- Something has gone wrong, but what is it?

Epilepsy

Lassie: Angela has a dog Lassie which can reliably predict her seizures.

- $P(\texttt{Seizure} \mid \texttt{Bark}) = .99$ and $P(\neg\texttt{Seizure} \mid \texttt{Bark}) = .01$.
- Assume that $v(\texttt{Seizure} + \texttt{Meds}) = 0$ and $v(\texttt{Seizure} + \neg\texttt{Meds}) = -10$.
- Lassie is barking and so Angela is sure to take her medication.
- But given what was said above, has she made a mistake?

Two Conditionals

Indication: Umbrellas on the streets and Lassie's barks are reliable indicators.

- If you only knew `Umbrellas/Barks` you could make a safe bet.
- `Umbrellas` and `Barks` indicate a cause, but are not causes themselves.
- Their is a common cause of `Umbrellas/WetGrass` and `Barks/Seizure`.

Indicatives: Indicative conditionals can be used to assert conditional knowledge.

- If `Barks`, then `Seizure`.
- If `Umbrellas`, then `WetGrass`.

Subjunctives: Subjunctive conditionals can be used to track causal connections.

- If the Bayesians *were* to go out with umbrellas, the grass *would* be wet.
- If Lassie *were* to bark, then Angela would have a seizure.

Two Box

Action: Faced with two boxes, the question is what are you in a position to do.

- LIKELY: If you choose `OneBox`, the big box will be `Full`.
- UNLIKELY: If you were to choose `OneBox`, the big box would be `Full`.
- Choosing `OneBox` isn't going to fill it with money (compare `Umbrellas`).
- So you might as well take what is there, and hence `TwoBox`.

Independence: When can we use an expected utility calculation as before?

- Is the outcome causally or probabilistically dependent on the action?
- An outcome S is *counterfactually independent* of an action A *iff* either:
  1. $A \boxdot\!\!\rightarrow S$ and $\neg A \boxdot\!\!\rightarrow S$. 2. $A \boxdot\!\!\rightarrow \neg S$ and $\neg A \boxdot\!\!\rightarrow \neg S$.
- $P(\texttt{OneBox} \boxdot\!\!\rightarrow \texttt{Full}) = P(\texttt{Full})$ and $P(\texttt{TwoBox} \boxdot\!\!\rightarrow \texttt{Full}) = P(\texttt{Full})$.
- $P(\texttt{OneBox} \boxdot\!\!\rightarrow \texttt{Empty}) = P(\texttt{Empty})$ and $P(\texttt{TwoBox} \boxdot\!\!\rightarrow \texttt{Empty}) = P(\texttt{Empty})$.
- By exclusivity and exhaustivity, $P(\texttt{Empty}) = 1 - P(\texttt{Full})$.

Causal Decision Theory: To be rational, maximize expected causal utility if possible.

- Weighting utilities in proportion to their likelihood is not the problem.
- The problem is mistaking probabilistic for counterfactual dependence.
- $ECU(A) = \sum_{i \in I_A} v(S_i^A) P(A \boxdot\!\!\rightarrow S_i^A)$ instead of $EV(A) = \sum_{i \in I_A} v(S_i^A) P(S_i^A | A)$.
- $\begin{aligned} ECU(\texttt{OneBox}) &= \$1,000,000 \times P(\texttt{OneBox} \boxdot\!\!\rightarrow \texttt{Full}) + \$0 \times P(\texttt{OneBox} \boxdot\!\!\rightarrow \texttt{Empty}) \\ &= \$1,000,000 \times P(\texttt{Full}). \end{aligned}$
- $\begin{aligned} ECU(\texttt{TwoBox}) &= \$1,001,000 \times P(\texttt{TwoBox} \boxdot\!\!\rightarrow \texttt{Full}) + \$1,000 \times P(\texttt{OneBox} \boxdot\!\!\rightarrow \texttt{Empty}). \\ &= \$1,001,000 \times P(\texttt{Full}) + \$1,000 \times P(\texttt{Empty}) \\ &= \$1,001,000 \times P(\texttt{Full}) + \$1,000 \times (1 - P(\texttt{Full})) \\ &= \$1,000,000 \times P(\texttt{Full}) + \$1,000 \\ &= ECU(\texttt{OneBox}) + \$1,000 \end{aligned}$
- Two boxing is better independent of the value of $P(\texttt{Full})$.

Short

Bonus: Choose, but before opening, bet about the total value for a bonus.

- Your choice may indicate what you were inclined to choose in the past.
- And your past inclinations indicate the prediction made about you.
- Bet `Empty` if and only if you chose `TwoBox`.
- You might feel that you have been punished for your rationality.
- So it goes in mischievous though experiments!

Prisoners' Dilemma

PARADOX AND INFINITY
Benjamin Brast-McKie
April 29, 2024

Two Prisoners

Setup: Two separated prisoners are each offered \$1,000. They will be given an additional \$1,000,000 *iff* the other prisoner does not take the \$1,000.

- The prisoners' choices are causally independent.
- $P(\texttt{Take}_A \mathrel{\square\!\!\rightarrow} \texttt{Take}_B) = P(\neg\texttt{Take}_A \mathrel{\square\!\!\rightarrow} \texttt{Take}_B) = P(\texttt{Take}_B)$.
- $P(\texttt{Take}_A \mathrel{\square\!\!\rightarrow} \neg\texttt{Take}_B) = P(\neg\texttt{Take}_A \mathrel{\square\!\!\rightarrow} \neg\texttt{Take}_B) = P(\neg\texttt{Take}_B)$.
- We know that $P(\neg\texttt{Take}_B) = 1 - P(\texttt{Take}_B)$, but don't know $P(\texttt{Take}_B)$.
- Something similar may be said swapping 'A' and 'B' above.
- The prisoner's know everything except for the other's choice.
- What is it rational for prisoner A (similarly B) to do?

Dominant: Taking the \$1,000 is a *dominant strategy* for prisoner A (similarly B).

- Whether $\texttt{Take}_B$ or not, $v(\texttt{Take}_A) > v(\neg\texttt{Take}_A)$ for prisoner A.
- We get the following alternatives:

	$\texttt{Take}_B$	$\neg\texttt{Take}_B$
$\texttt{Take}_A$	$(A, B : \$1,000)$	$(A : \$1,001,000)$, $(B : \$0)$
$\neg\texttt{Take}_A$	$(A : \$0)$, $(B : \$1,001,000)$	$(A, B : \$1,000,000)$

- The setup assumes that neither prisoner cares about the other.
- If the prisoners cared about each other, that would be a different case.

Predictor: Given the circumstances, each prisoner is a good predictor of the other.

- $\texttt{Take}_A$ predicts that $\texttt{Take}_B$, i.e., $P(\texttt{Take}_B \mid \texttt{Take}_A)$ is high.
- Thus $P(\neg\texttt{Rich}_A \mid \texttt{Take}_A)$ is high since $\texttt{Take}_B$ *iff* $\neg\texttt{Rich}_A$.
- So if $\texttt{Take}_A$, then prisoner A has good reason to bet $\neg\texttt{Rich}_A$.
- We don't know what the probabilities $P(\texttt{Take}_A)$ or $P(\texttt{Take}_B)$.
- Does $\neg\texttt{Take}_A$ change the probability $P(\neg\texttt{Take}_B) = P(\texttt{Rich}_A)$?

Newcomb: $\texttt{Rich}_A$ *iff* it is predicted that $\neg\texttt{Take}_A$ (by $\neg\texttt{Take}_B$).

- $\neg\texttt{Take}_B$ is a *prediction instance* (a way of predicting $\neg\texttt{Take}_A$).
- The predication amounts to probabilistic dependence (not causal).
- When the prediction happens does not matter to the case.
- Is the prisoners' dilemma a Newcomp problem?

Dominance Calculations

Expected Causal Utility: Recall that: (a) $\texttt{Rich}_A$ *iff* $\neg\texttt{Take}_B$; and (b) $\texttt{Rich}_B$ *iff* $\neg\texttt{Take}_A$.

- What are the *expected causal utilities* of $\texttt{Take}_A$ and $\neg\texttt{Take}_A$?
- $ECU(\neg\texttt{Take}_A) = \$1,000,000 \times P(\neg\texttt{Take}_A \mathbin{\square\!\!\rightarrow} \texttt{Rich}_A) + \$0 \times P(\neg\texttt{Take}_A \mathbin{\square\!\!\rightarrow} \neg\texttt{Rich}_A)$
  $= \$1,000,000 \times P(\neg\texttt{Take}_B)$ by (a).
- $ECU(\texttt{Take}_A) = \$1,001,000 \times P(\texttt{Take}_A \mathbin{\square\!\!\rightarrow} \texttt{Rich}_A) + \$1,000 \times P(\texttt{Take}_A \mathbin{\square\!\!\rightarrow} \neg\texttt{Rich}_A)$.
  $= \$1,001,000 \times P(\neg\texttt{Take}_B) + \$1,000 \times P(\texttt{Take}_B)$ by (a).
  $= \$1,001,000 \times P(\neg\texttt{Take}_B) + \$1,000 \times (1 - P(\neg\texttt{Take}_B))$.
  $= \$1,000,000 \times P(\neg\texttt{Take}_B) + \$1,000$.
  $= ECU(\neg\texttt{Take}_A) + \$1,000$.
- Taking the money is better for prisoner A (and similarly for B).

Accuracy

Clash: The predication does not have to be very accurate for the expected utility calculation to clash with causal expected utility (i.e. $> .5005$).

- Suppose $P(\texttt{Take}_B \mid \texttt{Take}_A) = P(\neg\texttt{Take}_B \mid \neg\texttt{Take}_A) = .5006$.
- So $P(\texttt{Rich}_A \mid \neg\texttt{Take}_A) = P(\neg\texttt{Take}_B \mid \neg\texttt{Take}_A) = .5006$.
- And $P(\texttt{Rich}_A \mid \texttt{Take}_A) = P(\neg\texttt{Take}_B \mid \texttt{Take}_A) = 1 - P(\texttt{Take}_B \mid \texttt{Take}_A) = .4994$.
- $EV(\neg\texttt{Take}_A) = \$1,000,000 \times P(\texttt{Rich}_A \mid \neg\texttt{Take}_A) + \$0 \times P(\neg\texttt{Rich}_A \mid \neg\texttt{Take}_A)$
  $= \$1,000,000 \times P(\texttt{Rich}_A \mid \neg\texttt{Take}_A)$
  $= \$500,600$.
- $EV(\texttt{Take}_A) = \$1,001,000 \times P(\texttt{Rich}_A \mid \texttt{Take}_A) + \$1,000 \times P(\neg\texttt{Rich}_A \mid \texttt{Take}_A)$
  $= \$1,000,000 \times P(\texttt{Rich}_A \mid \texttt{Take}_A) + \$1,000$
  $= \$500,400$.
- Even if prisoner A is an inaccurate predictor of prisoner B, the expected utility and expected causal utility calculations are bound to come apart.

Upshot

Common: Newcomb's problem is fanciful, but prisoners' dilemmas are common.

- Prisoners' dilemmas support *causal decision theory* on their own.
- No need to appeal to Newcomb cases to motivate CDT.

Comparison: Should a oneboxer also avoid taking the money?

- Does comparing the cases put any pressure on the oneboxer to twobox?

Prisoners' Dilemma

PARADOX AND INFINITY
Benjamin Brast-McKie
May 1, 2024

Instance Thesis

Argument: Lewis argues that (P) is an instance of (N):

(P) $\texttt{Rich}_A$ *iff* $\neg\texttt{Take}_B$.

(N) $\texttt{Rich}_A$ *iff* it is predicted that $\neg\texttt{Take}_A$.

Inessentials: Lewis claims (N) is equivalent to the following:

$(N)'$ $\texttt{Rich}_A$ *iff* a certain potentially predictive process (which may go on before, during, or after my choice) yields an outcome which could warrant a prediction that I do not take my $1,000.

- Lewis claims that $(N)'$ eliminates inessentials from (N).
- Focusing on (N), we may take $(N)'$ to elaborate what (N) intends.

Instance: Is (P) an instance of (N)?

- Does $\neg\texttt{Take}_B$ predict that $\neg\texttt{Take}_A$?
- Lewis says 'yes' when prisoners A and B are sufficiently similar.
- What is sufficient for $\neg\texttt{Take}_B$ to predict that $\neg\texttt{Take}_A$?

Prediction and Probability

Motivation: Lewis appeals to the prisoner's dilemma to motivate CDT.

- All that matters is that $\neg\texttt{Take}_B$ raises the likelihood of $\neg\texttt{Take}_A$ enough.
- Letting $r = \frac{\$1{,}000}{\$1{,}000{,}000}$, the probability must greater than $\frac{1+r}{2} = .5005$.
- $\neg\texttt{Take}_B$ predicts $\neg\texttt{Take}_A$ *if* $P(\neg\texttt{Take}_A \mid \neg\texttt{Take}_B) > .5005$.

Coin: Does getting heads $7/10$ times *predict* heads is more likely?

- If the coin is fair, heads is just as likely as tails.
- The fairness of the coin justifies the prediction that heads is .5 likely.

Similarity: What could justify that $P(\neg\texttt{Take}_A \mid \neg\texttt{Take}_B) > .5005$?

- Lewis claims that simulation is a predictive process *par excellence*.
- "To predict whether I will take my thousand, make a replica of me, put my replica in a replica of my predicament, and see whether my replica takes his thousand." —Lewis (1979, p. 237)
- Is prisoner B a good enough replica of prisoner A?

Conclusion: If so, then (P) is an instance of (N) as Lewis claims.

Optimal Rationality

Collaboration: Suppose that both prisoners are (optimally) rationally.

- Suppose they know that they are each optimally rational.
- Suppose they have all the same information and values.
- Does this mean that they act in the same way?
- One sort of answers claims 'yes': optimal rationality is unique.
- Thus there are only two possible outcomes: $\texttt{Take}_{AB}$ or $\neg\texttt{Take}_{AB}$.
- Moreover, $v(\neg\texttt{Take}_{AB}) \gg v(\texttt{Take}_{AB})$.
- Can we conclude that optimally rational prisoners will collaborate?

Theory: Is optimal rationality unique?

- Is there just one rational action for each agent in each case?
- Does optimal rationality require knowing whether optimal rationality is unique?
- One needn't know a final linguistic theory to be fluent in English.
- Nor does one need to know physics in order to hit a baseball.
- Being rational doesn't require knowing what rationality is.
- In particular, one needn't know if optimal rationality is unique.

Uniqueness: Can the prisoners assume that they will act in the same way?

- Even if optimal rationality is unique, can't assume they know this.
- Thus they can't conclude they will act in the same way.
- So the prisoner's can't run the reasoning above to $\neg\texttt{Take}_{AB}$.
- This reasoning also fails if rationality is not unique.

Modulo Theory

Rationality: What is it to be rational?

- Takes an epistemic state and values as input and action choice as output.
- There are the various ways people act given their values and info.
- Holding the inputs fixed, can the outputs be totally ordered?
- If totally ordered, must there be a maximally rational output?

Theory: Is it the task of a theory of rationality to provide a total ordering?

- For instance, EDT and CDT recommend opposing choices.
- Should we assume the same theory will be universally applicable?
- If not, how are we to decide which theory to choose when?
- For instance, we saw before that a twoboxer might use CDT to choose, but then use EDT to bet against themselves.

Surprise Exam Paradox

PARADOX AND INFINITY
Benjamin Brast-McKie
May 6, 2024

The Exam

Setup: A single surprise exam is announced for next week (9am m, w, or f).

- Let 'E_i' read 'The exam occurs on i' where $i \in \{m, w, f\}$.
- Let '$\mathcal{B}_i(A)$' read 'The students believe A at 9am on i'.
- E_i is a *surprise iff* $E_i \wedge \neg\mathcal{B}_i(E_i)$.
- Let $S = (E_m \wedge \neg\mathcal{B}_m(E_m)) \vee (E_w \wedge \neg\mathcal{B}_w(E_w)) \vee (E_f \wedge \neg\mathcal{B}_f(E_f))$.
- The students believe this announcement S throughout the week.

Closure: If $\mathcal{B}_i(A)$ for all $A \in \Gamma$ and $\Gamma \vdash B$, then $\mathcal{B}_i(B)$.

- We only need limited instances of *Closure* to hold.

Informed: The students learn each day if there is an exam, forming a true belief.

Memory: The students maintain their beliefs from the previous days.

Friday: On Monday 8am, the students reason as follows:

- If E_f, then $\neg E_m$ and $\neg E_w$, so $\mathcal{B}_m(\neg E_m)$ and $\mathcal{B}_w(\neg E_w)$ by *Informed*.
- So $\mathcal{B}_f(\neg E_m)$ and $\mathcal{B}_f(\neg E_w)$ by *Memory*, where $\mathcal{B}_f(S)$ is a premise.
- But $S, \neg E_m, \neg E_w \vdash E_f$, and so $\mathcal{B}_f(E_f)$ by *Closure*.
- Thus E_f is not a surprise, i.e., $\neg\neg\mathcal{B}_f(E_f)$, and so $\neg(E_f \wedge \neg\mathcal{B}_f(E_f))$.
- In this way, the students come to $\mathcal{B}_m(\neg(E_f \wedge \neg\mathcal{B}_f(E_f)))$.
- However, $S, \neg(E_f \wedge \neg\mathcal{B}_f(E_f)) \vdash (E_m \wedge \neg\mathcal{B}_m(E_m)) \vee (E_w \wedge \neg\mathcal{B}_w(E_w))$.
- By *Closure*, $\mathcal{B}_m(S')$ where $S' = (E_m \wedge \neg\mathcal{B}_m(E_m)) \vee (E_w \wedge \neg\mathcal{B}_w(E_w))$.

Wednesday: The students (on Monday 8:05am) turn to reason about Wednesday:

- If E_w, then $\neg E_m$, so $\mathcal{B}_m(\neg E_m)$ by *Informed* and $\mathcal{B}_w(\neg E_m)$ by *Memory*.
- However, $\mathcal{B}_m(S')$ by *Friday*, and so $\mathcal{B}_w(S')$ by *Memory*.
- Since $S', \neg E_m \vdash E_w$, it follows by *Closure* that $\mathcal{B}_w(E_w)$.
- Thus if E_w, then E_w is not a surprise, and so $\neg(E_w \wedge \neg\mathcal{B}_w(E_w))$.
- In this way, the students come to $\mathcal{B}_m(\neg(E_w \wedge \neg\mathcal{B}_w(E_w)))$.
- However, $S', \neg(E_w \wedge \neg\mathcal{B}_w(E_w)) \vdash E_m \wedge \neg\mathcal{B}_m(E_m)$.
- By *Closure*, $\mathcal{B}_m(S'')$ where $S'' = E_m \wedge \neg\mathcal{B}_m(E_m)$.

Monday The students now turn to consider Monday (still on Monday 8:10am):

- $\mathcal{B}_m(S'')$ entails $\mathcal{B}_m(E_m)$ and $\mathcal{B}_m(\neg\mathcal{B}_m(E_m))$.
- $\mathcal{B}_m(E_m) \vdash \mathcal{B}_m(\mathcal{B}_m(E_m))$ leads to believing a contradiction.
- So the students would seem to have reason to reject $\mathcal{B}_m(S'')$.

Moore's Problem

Rain: It is raining (R) but I do not believe that it is raining $\neg\mathcal{B}(R)$.

- Can be true, but can't be asserted (normally).
- Can *assert* either R or $\neg\mathcal{B}(R)$, but not both.
- OK to assert: It is raining but *you* do not believe that it is raining.

Belief Norm: Don't assert what you don't yourself believe (in normal circumstances).

- One could appeal to this norm to infer $\mathcal{B}(R)$ from an assertion of R.
- Similarly, $\mathcal{B}(\neg\mathcal{B}(R))$ can be inferred from an assertion of $\neg\mathcal{B}(R)$.
- Can *believe* R or $\neg\mathcal{B}(R)$, but not both?

Introspection: The following introspection principles have many true instances.

(*Positive*) $\mathcal{B}(A) \vdash \mathcal{B}(\mathcal{B}(A))$. (*Negative*) $\neg\mathcal{B}(A) \vdash \mathcal{B}(\neg\mathcal{B}(A))$.

- Nothing seems to block introspection for $A = E_m$.
- As above, $\mathcal{B}_m(S)$ entails $\mathcal{B}_m(E_m)$ and $\mathcal{B}_m(\neg\mathcal{B}_m(E_m))$.
- So $\mathcal{B}_m(\mathcal{B}_m(E_m))$ follows by *Positive Introspection*.
- Moreover $\mathcal{B}_m(E_m), \neg\mathcal{B}_m(E_m) \vdash \mathcal{B}_m(E_m) \wedge \neg\mathcal{B}_m(E_m)$.
- Hence $\mathcal{B}_m(\mathcal{B}_m(E_m) \wedge \neg\mathcal{B}_m(E_m))$ follows by *Closure*.
- But $\mathcal{B}_m(E_m) \wedge \neg\mathcal{B}_m(E_m)$ is a contradiction.

Contradiction: Don't believe contradictions (revise your beliefs accordingly).

- Since $\mathcal{B}_m(S) \vdash \mathcal{B}_m(\mathcal{B}_m(E_m) \wedge \neg\mathcal{B}_m(E_m))$, we get $\neg\mathcal{B}_m(S)$.
- But the students are able to believe that there will be a surprise exam.

Blindspot: Is the paradox solved by claiming that it is impossible for $\mathcal{B}_m(S)$?

- Is it still possible for S to be true?

Surprise Exam Paradox

PARADOX AND INFINITY
Benjamin Brast-McKie
May 8, 2024

Disbelief

Knowledge: Sometimes the argument is developed in terms of knowledge.

- Were going to stick with belief.

Belief: Can the students believe the instructor?

- Yes, easily so long as they don't do too much reasoning (bad answer).
- Let $S = (E_m \wedge \neg \mathcal{B}_m(E_m)) \vee (E_w \wedge \neg \mathcal{B}_w(E_w)) \vee (E_f \wedge \neg \mathcal{B}_f(E_f))$.
- The interesting question is whether $\mathcal{B}_m(S)$ given *Closure*, etc.

Logic: Can logically omniscient students believe S on Monday?

- It might seem that the arguments show that the answer is 'No'.
- But it seems like there can be surprises, and so S could be true.
- So are the logically omniscient students missing out on a true belief?

Repost: Perhaps the good reasons for belief are overturned by the argument.

- Even the most expert testimonies can be overturned, why not this?
- Remains to accommodate the possibility of a surprise exam.
- But we also want to maintain reasonably strong epistemic principles.

Doubts

Setup: Why can't the students believe S?

- One explanation claims that S happens to be false.
- But surely S is possible, and if so, assume such a case.
- Another strategy looks to spot the mistake in our reasoning before.

One Day: Can there be an announced surprise exam on just one day?

- Announcement: "There is a surprise exam on Monday."
- Seems that the announcement ensures that it is false.

Two Days: Can there be an announced surprise exam on one of two days?

- Suppose the exam is held on Monday (as opposed to Wednesday).
- Would it come as a surprise to the students?
- On Monday, how could they be sure that it wasn't on Wednesday?
- Because if it was on Wednesday, it wouldn't be a surprise *then*.
- But we might be surprised *today* to find out that it is on Wednesday.

Surprise

Timing: It would come as a surprise on Monday that it is/isn't on Monday.

- It wouldn't be a surprise on Wednesday if it didn't happen Monday.
- Do we need to maintain that it is a surprise on the day of?
- Why not take something to be a surprise by referencing the day before?

Analysis: Let E_i be a *surprise iff* $E_i \wedge \neg\mathcal{B}_{i-1}(E_i)$.

- Assume $m - 1 = f'$ (on the week before), $w - 1 = m$, and $f - 1 = w$.
- If E_f, then since $\neg\mathcal{B}_{f-1}(E_f) = \neg\mathcal{B}_w(E_f)$, so it is a surprise.
- Really the surprise takes place on Wednesday.
- On Wednesday, the surprise is about whether E_w or E_f.

Surprise: Does this new analysis capture a natural notion of surprise?

- No less reasonable than the first analysis, and blocks the argument.
- So there can be surprise exams, just not of the first kind of surprise.
- Is this adequate?

Learning: Compare learning something new: you go from $\neg\mathcal{B}(X)$ to $\mathcal{B}(X)$.

- Suppose that you learn something now about something in the future.
- Suppose Ali will go on a walk tomorrow.
- Learning this today, must we be surprised?
- You might say, "I'm not surprised," since Ali often goes on walks.
- But this has the same form as before: $\texttt{Walk}_i \wedge \neg\mathcal{B}_{i-1}(\texttt{Walk}_i)$.

Belief: Could weaken our analysis to a mere necessary condition.

- Partial analysis risks being fairly weak, though still true.
- Consider the exclamations: "I don't believe it!", "I am in disbelief!".
- We say these things when we believe something that surprises us.
- It's not just that we learn something new, it has to be anticipated.

Credences: But couldn't we anticipate Ali's walk without being surprised?

- Merely contemplating a future event is not enough to anticipate it.
- Instead of changing our beliefs, consider updating our credences.
- The bigger the jump in credences, the more surprising.
- When no test is given Wednesday, we go from $\frac{1}{2}$ to 1 that it is on Friday.
- Couldn't our expectation that Ali goes on a walk be similar?
- What makes the exam a surprise and Ali's walk anything but?

Stakes: One thought is that the *stakes* play a role.

- The higher the stakes, the more surprising something can be.

Surprise Exam Paradox

PARADOX AND INFINITY
Benjamin Brast-McKie
May 13, 2025

Disbelief

Knowledge: Sometimes the argument is developed in terms of knowledge.

- Were going to stick with belief.

Belief: Can the students believe the instructor?

- Yes, easily so long as they don't do too much reasoning (bad answer).
- Let $S = (E_m \wedge \neg\mathfrak{B}_m(E_m)) \vee (E_w \wedge \neg\mathfrak{B}_w(E_w)) \vee (E_f \wedge \neg\mathfrak{B}_f(E_f))$.
- The interesting question is whether $\mathfrak{B}_m(S)$ given *Closure*, etc.

Logic: Can logically omniscient students believe S on Monday?

- It might seem that the arguments show that the answer is 'No'.
- But it seems like there can be surprises, and so S could be true.
- So are the logically omniscient students missing out on a true belief?

Repost: Perhaps the good reasons for belief are overturned by the argument.

- Even the most expert testimonies can be overturned, why not this?
- Remains to accommodate the possibility of a surprise exam.
- But we also want to maintain reasonably strong epistemic principles.

Doubts

Setup: Why can't the students believe S?

- One explanation claims that S happens to be false.
- But surely S is possible, and if so, assume such a case.
- Another strategy looks to spot the mistake in our reasoning before.

One Day: Can there be an announced surprise exam on just one day?

- Announcement: "There is a surprise exam on Monday."
- Seems that the announcement ensures that it is false.

Two Days: Can there be an announced surprise exam on one of two days?

- Suppose the exam is held on Monday (as opposed to Wednesday).
- Would it come as a surprise to the students?
- On Monday, how could they be sure that it wasn't on Wednesday?
- Because if it was on Wednesday, it wouldn't be a surprise *then*.
- But we might be surprised *today* to find out that it is on Wednesday.

Surprise

Timing: It would come as a surprise on Monday that it is/isn't on Monday.

- It wouldn't be a surprise on Wednesday if it didn't happen Monday.
- Do we need to maintain that it is a surprise on the day of?
- Why not take something to be a surprise by referencing the day before?

Analysis: Let E_i be a *surprise iff* $E_i \wedge \neg\mathfrak{B}_{i-1}(E_i)$.

- Assume $m - 1 = f'$ (on the week before), $w - 1 = m$, and $f - 1 = w$.
- If E_f, then since $\neg\mathfrak{B}_{f-1}(E_f) = \neg\mathfrak{B}_w(E_f)$, so it is a surprise.
- Really the surprise takes place on Wednesday.
- On Wednesday, the surprise is about whether E_w or E_f.

Surprise: Does this new analysis capture a natural notion of surprise?

- No less reasonable than the first analysis, and blocks the argument.
- So there can be surprise exams, just not of the first kind of surprise.
- Is this adequate?

Learning: Compare learning something new: you go from $\neg\mathfrak{B}(X)$ to $\mathfrak{B}(X)$.

- Suppose that you learn something now about something in the future.
- Suppose Ali will go on a walk tomorrow.
- Learning this today, must we be surprised?
- You might say, "I'm not surprised," since Ali often goes on walks.
- But this has the same form as before: $\texttt{Walk}_i \wedge \neg\mathfrak{B}_{i-1}(\texttt{Walk}_i)$.

Belief: Could weaken our analysis to a mere necessary condition.

- Partial analysis risks being fairly weak, though still true.
- Consider the exclamations: "I don't believe it!", "I am in disbelief!".
- We say these things when we believe something that surprises us.
- It's not just that we learn something new, it has to be anticipated.

Credences: But couldn't we anticipate Ali's walk without being surprised?

- Merely contemplating a future event is not enough to anticipate it.
- Instead of changing our beliefs, consider updating our credences.
- The bigger the jump in credences, the more surprising.
- When no test is given Wednesday, we go from $\frac{1}{2}$ to 1 that it is on Friday.
- Couldn't our expectation that Ali goes on a walk be similar?
- What makes the exam a surprise and Ali's walk anything but?

Stakes: One thought is that the *stakes* play a role.

- The higher the stakes, the more surprising something can be.