Russellian Descriptions[1] & Gibbardian Indicatives
(Two Case Studies Involving Automated Reasoning)

Branden Fitelson

Philosophy Department
Rutgers University

&

Center for Advanced Studies
Ludwig-Maximilians-Universität München

`branden@fitelson.org`

---

[1] The Russellian descriptions part is joint work with Paul Oppenheimer.

William Walker McCune (1953 – 2011)

- Bill McCune was one of the true pioneers of automated reasoning. The systems OTTER [6], MACE, [7] and PROVER9 [4] were masterfully implemented (almost entirely) by Bill.
- Bill's program EQP [5] (a predecessor of PROVER9) discovered the first proof of the Robbins Conjecture [8]. He was awarded the 2000 Hebrand Award for his AR research.
- Bill was a kind and gentle soul. He will be greatly missed.

- IMHO, the most challenging aspect of applying automated reasoning tools to philosophical problems is the (so-called) *knowledge representation problem* (KRP) [11, Chapter 6].
- Automated reasoning tools come in many flavors. There are theorem provers, model finders, decision procedures, verification systems, symbolic algebra systems, *etc.*
  - Also: special purpose *vs* general purpose, higher-order *vs* first-order, deterministic *vs* non-deterministic, *etc.*
- For those of you who are interested in using these sorts of tools, I highly recommend Harrison's encyclopedic [2].
- Today, I'll be talking about attacking two types of problems (problems arising in some branches of formal philosophy) using *first-order theorem provers and model finders.*
  - There lots of terrific online resources for such systems. I'd start with Geoff Sutcliffe's TPTP [12] & CASC [13] websites.
- The best ATP is Vampire [14]; the best AMF is Paradox [3].

- First-order theorem provers and model finders can only work with assumptions that are *expressible in FOL* (by which I mean classical FOL with identity and function symbols).
- So, in order to apply these AR-tools to arguments involving "non-classical stuff", one needs to find a way to represent such arguments in FOL — in a way that *preserves validity*.
- *E.g.*, representing arguments in Ed's *theory of abstract objects* introduces various KRPs (some of which we've seen).
- In Ed's work with Paul on the ontological argument, we find various arguments involving definite descriptions.
- That is, we have arguments involving an ι operator, which is treated as a — possibly *non-denoting*! — singular term.
- This is tricky, from an FOL point of view, since (primitive) singular terms are assumed to have denotations.
- Our approach is to use a (broadly) "Russellian" strategy for representing these arguments in FOL. This is a nice KRP...

- Before explaining our (ultimate) Russellian technique, I will briefly discuss two *failed* attempts to represent ι in FOL.
- **Attempt #1.** Render $\ulcorner G\iota xFx \urcorner$ ($\ulcorner$the F is $G\urcorner$) as:

$$\ulcorner G\iota xFx \urcorner \stackrel{\text{def}}{=} \ulcorner \mathtt{x} = \mathtt{the}(\mathtt{F})\,\&\,\mathtt{Gx} \urcorner,$$

  where $\mathtt{the}(\cdot)$ is a function from properties to objects.
- Menzel pointed out that (in our multi-sorted representation of object theory) this representation leads to *triviality*.
  (1) `all x (Object(x) -> -Property(x)).`
  (2) `all x all F (x = The(F) -> (Object(x) & Property(F))).`
- Typing conditions (1) and (2) *entail a contradiction* in FOL!
- **Attempt #2.** Render $\ulcorner G\iota xFx \urcorner$ ($\ulcorner$the F is $G\urcorner$) as:

$$\ulcorner G\iota xFx \urcorner \stackrel{\text{def}}{=} \ulcorner \mathtt{IsThe}(\mathtt{x},\mathtt{F})\,\&\,\mathtt{Gx} \urcorner,$$

  where $\mathtt{IsThe}(\cdot,\cdot)$ *relates* objects and properties.
- This avoids Menzel's typing triviality, but it is not suitable for application to problems like the ontological argument.

- The reason Attempt #2 is unsuitable in the context of the ontological argument is that it ends-up forcing some of the (FOL) premises to *beg questions* about denotation/existence.
- In the ontological argument, we have a "definition of God" [$g \stackrel{\text{def}}{=} \iota xFx$], which will now have to be rendered as follows:

$$\mathtt{IsThe}(\mathtt{g},\mathtt{F})$$

  where $\mathtt{g}$ is a constant symbol in FOL and $\mathtt{F}$ is a property.
- In FOL, this ends-up entailing the existence of some object $\mathtt{g}$ that has property $\mathtt{F}$ — that is to say, it ends-up entailing that the corresponding definite description *denotes*.
- This isn't Kosher — it *begs the (existence) question*.
- So, we'll need another approach for dealing with ι in FOL.
- Paul Oppenheimer has come up with an algorithm for *eliminating* all ι expressions ("Russell-style"), which *preserves truth*. This algorithm is inspired by Rosser's [10].

- I won't go through the algorithm in detail. I will merely illustrate it using the (simplified) ontological argument. But, first, I will explain why a "naïve" elimination of ι *fails*.
- Let p be a statement containing some expression(s) of form:

$$\psi_v^{\iota x\phi}$$

  where ψ is some atomic predication (or identity) formula in which exactly one variable v occurs free.
- Naïvely, one might think that one can simply "replace each such expression" $\psi_v^{\iota x\phi}$ in p with its "Russellian correlate":

$$\exists y(\phi_x^y \,\&\, \forall z(\phi_x^z \to z = y) \,\&\, \psi_v^y)$$

  thereby yielding an *equivalent* statement p' [of course, this would require changing bound variables if necessary to avoid incorrect capture of variables, *etc.*].
- And, *something like this* is exactly what Paul's algorithm does. But, one must be very careful about the *scope* of definite descriptions — *they must be given narrowest scope.*

- Consider this simplified OA (where $F \stackrel{\text{def}}{=} E!$, and the universe of discourse is restricted to to the *conceivable* objects):
  1. $\neg F\iota x\neg\exists wGwx \to \exists uGu\iota x\neg\exists wGwx$.

  $\therefore$ 2. $F\iota x\neg\exists wGwx$.
- *Naïvely* eliminating $\psi_v^{\iota x\phi}$ expressions from (1) & (2), yields:

$$(1') \quad \exists x(\neg\exists yGyx \,\&\, \forall z(\neg\exists y\, Gyx \to z = x) \,\&\, \neg Fx) \to$$
$$\exists x(\neg\exists y\, Gyx \,\&\, \forall z(\neg\exists y\, Gyx \to z = x) \,\&\, \exists y\, Gyx)$$

$$(2') \quad \exists x(\neg\exists yGyx \,\&\, \forall z(\neg\exists yGyz \to z = x) \,\&\, Fx)$$

- And, the argument from (1′) to (2′) is *invalid* in FOL!
  - There is a simple countermodel $\mathcal{M} = \langle D, I\rangle$ in which $D = \{0\}$, $I(F) = \varnothing$, and $I(G) = \{\langle 0,0\rangle\}$.
- Paul's algorithm (always) yields an *equivalent* argument (*i.e.*, an argument in pure FOL that is equivalent to the original argument — *in the richer language, containing ι and the Russell axiom*). The key is to eliminate $\psi_v^{\iota x\phi}$ expressions in such a way that *the ι-operator always takes narrowest scope.*

- Re-consider the simplified OA ($F \overset{\text{def}}{=} E!$, and the universe of discourse is restricted to to the *conceivable* objects):

  1. $\neg F\iota x\neg\exists w Gwx \rightarrow \exists u Gu\iota x\neg\exists w Gwx$.

  ∴ 2. $F\iota x\neg\exists w Gwx$.

- Using Paul's algorithm to eliminate ι from (1) & (2), yields:

$$(1^\star)\quad \neg\exists x_3(\neg\exists w Gwx_3 \,\&\, \forall x_4(\neg\exists w Gwx_4 \rightarrow x_4 = x_3) \,\&\, Fx_3) \rightarrow \exists u\exists x_1(\neg\exists w Gwx_1 \,\&\, \forall x_2(\neg\exists w Gwx_2 \rightarrow x_2 = x_1) \,\&\, Gux_1)$$

$$(2^\star)\quad \exists x(\neg\exists y Gyx \,\&\, \forall z(\neg\exists y Gyz \rightarrow z = x) \,\&\, Fx)$$

- The argument from $(1^\star)$ to $(2^\star)$ is *valid* in FOL, as it must be, since (a) the original argument is valid in the richer language containing ι and the Russell axiom, and (b) Paul's algorithm generates logically equivalent statements.
- This is a nice case study in the perils (and thrils!) of the FOL–KRP in the context of computational metaphysics.

- Gibbard [1] argues that indicatives ($\rightsquigarrow$) satisfying certain constraints must *collapse* to material conditionals ($\supset$).
- One of Gibbard's key assumptions is *import-export* for $\rightsquigarrow$.

  (IE) $B \rightsquigarrow (A \rightsquigarrow C)$ is *logically equivalent* to $(A \,\&\, B) \rightsquigarrow C$.

- Suppose $\rightsquigarrow$ satisfies (IE). Then, (1) is equivalent to (2).

  (1) $(A \supset C) \rightsquigarrow (A \rightsquigarrow C)$.

  (2) $((A \supset C) \,\&\, A) \rightsquigarrow C$.

- Substitutivity implies that (2) [and ∴ (1)] is equivalent to

  (3) $(A \,\&\, C) \rightsquigarrow C$.

- If (3) is a logical truth (as Gibbard assumes), so are (1) & (2).
- Finally, suppose $P \rightsquigarrow Q$ *entails* $P \supset Q$. Then, (1) entails

  (4) $(A \supset C) \supset (A \rightsquigarrow C)$.

- So (4) is a logical truth. Thus, $A \supset C$ entails $A \rightsquigarrow C$.
- Therefore, $P \rightsquigarrow Q$ entails $P \supset Q$ and $P \supset Q$ entails $P \rightsquigarrow Q$.
- That is, in general, $\rightsquigarrow$ and $\supset$ are logically equivalent. *QED*.

- Gibbard's argument is informal, and presupposes lots of things. I will now work up to a *purely formal* version of it.
- I will assume a propositional object language $\mathcal{L}$, containing:
  - Atomic sentences: 'A', 'B', 'C', *etc.*
  - An "indicative" conditional connective '$\rightsquigarrow$'.
  - A "logical" conditional connective '$\rightarrow$'.
  - A conjunction connective '&'.
- And, I will assume a metalanguage £ for $\mathcal{L}$, containing:
  - A binary "entailment" or "logical consequence" relation '$\Vdash$', which is a relation between pairs of sentences in $\mathcal{L}$.
    - That is, I will explicate $\ulcorner p$ entails $q\urcorner$ as $\ulcorner p \Vdash q\urcorner$.
  - A unary "theoremhood" or "logical truth" relation '$\vdash$', which is a property of individual sentences of $\mathcal{L}$.
  - £ will also contain (English language) logical connectives, which will be assumed to be *classical* (Boolean) operations.
    - *e.g.*, $\ulcorner p$ is logically equivalent to $q\urcorner$ gets explicated as $\ulcorner p \Vdash q$ and $q \Vdash p\urcorner$, where 'and' is assumed to be *classical*.

- Next, I will give *axioms* for $\rightsquigarrow$, $\rightarrow$, &, $\Vdash$, and $\vdash$, which will lead to a *precise formalization* of "Gibbard's Theorem".

  (1) $\vdash (p \,\&\, q) \rightarrow p$.

  - (1) is a "conjunction-elimination" *axiom* for $\langle\rightarrow, \&\rangle$.

  (2) $\vdash p \rightarrow (q \rightarrow r)$ if and only if $\vdash (p \,\&\, q) \rightarrow r$.

  - (2) is an *import-export* rule for $\langle\rightarrow, \&\rangle$-*theorems* of $\mathcal{L}$ (*i.e.*, that import-export for the *logical* conditional $\rightarrow$ preserves *validity* in $\mathcal{L}$). This is kosher — classically & intuitionistically.

  (3) $\vdash (p \,\&\, q) \rightsquigarrow r$ if and only if $\vdash p \rightsquigarrow (q \rightsquigarrow r)$.

  - (3) is an *import-export* rule for $\langle\rightsquigarrow, \&\rangle$-*theorems* of $\mathcal{L}$ (it says that import-export for the *indicative* $\rightsquigarrow$ preserves *validity*).

  (4) If $\vdash p \rightsquigarrow q$, then $\vdash p \rightarrow q$.

  - (4) says that the indicative conditional is "at least as strong as" the logical conditional — but *only* in the sense that if an indicative conditional is a *theorem* of $\mathcal{L}$, then the corresponding logical conditional is *also* a *theorem* of $\mathcal{L}$.

(5) $\vdash (p \,\&\, q) \rightsquigarrow q$.

- (5) is a "conjunction-elimination" *axiom* for $\langle \rightsquigarrow, \& \rangle$.

(6) If $p \Vdash q$ and $p \Vdash r$, then $p \Vdash q \,\&\, r$.

- (6) is a form of the *conjunction introduction rule*. This is kosher (for just about any entailment relation I can think of).

(7) If $\vdash p \to q$, then $p \Vdash q$.

- (7) says (informally) that if a logical conditional is a logical truth (theorem), then its antecedent entails its consequent.

(8) If $p \Vdash q$ and $q \Vdash p$, then $p = q$.

- (8) says (informally) that if p and q are logically equivalent, then they are *inter-substitutable salva veritate* (*i.e.*, it is kosher to treat them as *one and the same formula*).

- Axioms (1)–(8) are *independent*. And, they *suffice* to ensure that the indicative conditional collapses to the logical conditional. That is to say, we have the following theorem:

Theorem. Assuming (1)–(8), the following must hold

$$p \to q \Vdash p \rightsquigarrow q \quad \text{and} \quad p \rightsquigarrow q \Vdash p \to q.$$

- The *precise* underlying logic of $\langle \to, \& \rangle$ is *inessential* to *our* "Gibbardian Collapse Theorem". That is to say, our premises (1)–(8) do *not* ensure "collapse" to *classical* $\supset$.
- Here is an interesting fact about the $\vdash$ of "theory" (1)–(8):

(9) $\nvdash ((p \rightsquigarrow q) \rightsquigarrow p) \rightsquigarrow p$.

- *Peirce's Law* need not hold for the "collapsed" $\rightsquigarrow$.

- So, (1)–(8) do *not* imply collapse to $\supset$. But, we *do* have:

(10) $\vdash (p \rightsquigarrow (q \rightsquigarrow r)) \rightsquigarrow ((p \rightsquigarrow q) \rightsquigarrow (p \rightsquigarrow r))$.

- (10) is valid intuitionistically and classically.

(11) $\vdash p \rightsquigarrow (q \rightsquigarrow p)$.

- (10) and (11) jointly entail that the indicative $\rightsquigarrow$ must be *at least as strong as the intuituionistic conditional.*

- Q: what's the "weakest" ("most interesting") assumption X involving $\rightsquigarrow$, $\to$, &, $\Vdash$, $\vdash$ s.t. (1) – (9) + X collapses $\rightsquigarrow$ to $\supset$?

- There are other "triviality arguments" involving $\rightsquigarrow$ and "import-export" principles. Here's one [9]. Suppose:

1. For all t, S's degree of belief structure at time t $\langle \mathcal{B}_t, \Pr_t \rangle$ is a *probability model*. Thus, the space of propositions $\mathcal{B}_t$ over which they assign credences (at t) is a Boolean algebra.
2. S's Boolean algebra of propositions $\mathcal{B}_t$ (at t) is *closed* under indicative implication: $\forall a, c \in \mathcal{B}_t$, $a \rightsquigarrow c \in \mathcal{B}_t$.
3. S's credence function (at t) $\Pr_t$ satisfies The Equation:
$$\Pr_t(a \rightsquigarrow c) = \Pr_t(c \mid a).$$
4. S's credence function (at t) $\Pr_t$ satisfies the following (synchronic!) "*probabilistic* import-export" law:
$$\Pr_t(b \rightsquigarrow (a \rightsquigarrow c)) = \Pr_t((a \,\&\, b) \rightsquigarrow c).$$

Theorem. (1)–(4) imply that S's credence function (at t) $\Pr_t$ is such that, for all *contingent* [$\Pr_t \in (0,1)$] $a, c \in \mathcal{B}_t$:

$$\Pr_t(a \rightsquigarrow c) = \Pr_t(a \supset c).$$

- This is similar to Gibbard's (desired) "$\supset$-collapse" result.

---

[1] A. Gibbard, *Two recent theories of conditionals*, in W.L. Harper, R. Stalnaker & G. Pearce (*eds.*), *Ifs: Conditionals, Belief, Decision, Chance and Time*, Reidel, 1981.

[2] J. Harrison, *Handbook of Practical Logic and Automated Reasoning*, CUP, 2009.

[3] K. Claessen and N. Sörensson, *New Techniques that Improve MACE-style Finite Model Finding*, *Proceedings of the CADE-19 Workshop*, `http://fitelson.org/paradox.pdf`.

[4] W. McCune, Prover9 *Manual (Version 2009-11A)*, November 2009, URL: `http://www.cs.unm.edu/~mccune/prover9/manual/2009-11A/`.

[5] ———, Eqp*: Equational Prover*, `http://www.cs.unm.edu/~mccune/eqp/`.

[6] ———, *Otter 3.3 Reference Manual*, Tech. Memo ANL/MCS-TM-263, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, August 2003.

[7] ———, *Mace4 Reference Manual and Guide*, Tech. Memo ANL/MCS-TM-264, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, August 2003.

[8] ———, *Solution of the Robbins problem*, *Journal of Automated Reasoning*, December 1997.

[9] Milne, P., *The simplest Lewis-style triviality proof yet?*, *Analysis*, 2003.

[10] J.B. Rosser, *On consistency of Quine's New Foundations for Mathematical Logic*, *JSL*, 1939.

[11] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, 1995.

[12] G. Sutcliffe, *The TPTP Problem Library for ATP*, `http://www.cs.miami.edu/~tptp/`.

[13] ———, *The CADE ATP System Competition*, `http://www.cs.miami.edu/~tptp/CASC/`.

[14] A. Voronkov, *Vampire Home Page*, `http://www.vprover.org/`.