Notes on Gibbard's Theorem

Branden Fitelson
September 18, 2008

Let $\mathscr{L}$ be a sentential (object) language containing atoms 'A', 'B', …, and two *logical* connectives '&' and '→'. In addition to these two *logical* connectives, $\mathscr{L}$ will also contain another binary connective '$\rightsquigarrow$', which is intended to be interpreted as the English indicative. In the meta-language for $\mathscr{L}$, we will have two meta-linguistic operations: '$\Vdash$' and '$\vdash$'. '$\Vdash$' is a binary relation between individual sentences in $\mathscr{L}$. It will be interpreted as "single premise entailment" (or "single premise deducibility in $\mathscr{L}$"). '$\vdash$' is a monadic predicate on sentences of $\mathscr{L}$. It will be interpreted as "logical truth of the logic of $\mathscr{L}$" (or "theorem of the logic of $\mathscr{L}$"). We will not presuppose anything about the relationship between '$\Vdash$' and '$\vdash$'. Rather, we will state explicitly all assumptions about these meta-theoretic relations that will be required for Gibbard's Theorem.

Below, I will report both "weak" and "strong" versions of Gibbard's Theorem (the "weak" one is the one typically seen in the literature — I think the "strong" one is new). I'll do the "weak" Theorem first. Here are 10 (independent) assumptions that are sufficient to entail the "weak" version of Gibbard's Theorem. First, two preliminary remarks: (a) the "if…then" and "and" I'm using in the meta-meta-language of $\mathscr{L}$ to state the assumptions of the theorem are assumed to be classical, and (b) these assumptions are all *schematic* (*i.e.*, they are to be interpreted as allowing *any instances* that can be formed from sentences of $\mathscr{L}$).

1. If $\vdash p \to q$ and $\vdash p$, then $\vdash q$.

    - What (1) says is that *modus ponens* for the *logical* conditional '→' preserves *logical* truth (or *theoremhood* in $\mathscr{L}$). [The distinction between *modus ponens* as a truth-preserving *vs* validity preserving rule will be important. See my final remarks at the end of this note, for more on this.]

2. $\vdash p \to (q \to p)$

    - What (2) says is that "Weakening" is a theorem for '→' in $\mathscr{L}$. Weakening is a theorem of classical logic, as well as intuitionistic logic (and some sub-logics of intuitionistic logic). It is not a theorem of most relevance logics (which is why Gibbard's Theorem is inapplicable to relevant logics).

3. $\vdash (p \to q) \to ((p \to r) \to (p \to (q \,\&\, r)))$

    - (3) is a form of "conjunction introduction" axiom, which is valid classically and intuitionistically.

4. $\vdash (p \to (q \to r)) \to ((p \,\&\, q) \to r)$

    - (4) is the importation axiom for $\langle \to, \& \rangle$. Again, this is valid classically and intuitionistically.

5. $(p \,\&\, q) \rightsquigarrow r \Vdash p \rightsquigarrow (q \rightsquigarrow r)$

    - (5) is a '$\Vdash$' exportation *rule* for $\langle \rightsquigarrow, \& \rangle$. This says that the indicative satisfies the "rule-form" of exportation, with respect to our "entailment" relation. Note: so far, we have said nothing about the relationship between '$\Vdash$' and '$\vdash$'. So, it would be a mistake to think that anything like an exportation *axiom* is required to hold for $\langle \rightsquigarrow, \& \rangle$. At this point, that does *not* follow. I'll comment further on this issue, below, when we have all assumptions in place (including those assumptions which explicitly *relate* '$\Vdash$' and '$\vdash$'). People are rather sloppy about this in the literature.

6. $p \rightsquigarrow q \Vdash p \to q$

    - (6) says that the indicative conditional is "at least as strong as" the logical conditional — from the point of view of our "entailment" relation '$\Vdash$'. Again, this does not (yet) imply anything about *axioms* (or theorems) relating the two connectives.

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

    - (7) is a conjunction-elimination *axiom* for $\langle \rightsquigarrow, \& \rangle$. This holds for every conditional I can think of.

8. If $\vdash p \to q$, then $p \Vdash q$.

- (8) is one direction of a "deduction theorem" relating *theorems* of the underlying $\langle \to, \& \rangle$-logic and single premise "*entailments*" under '$\Vdash$'. This, to my mind, is the key assumption in all of these arguments. This is where most of the real work is done. By forging a connection between the logical conditional and the "entailment" relation, one can illicitly smuggle-in properties of the underlying $\langle \to, \& \rangle$-logic, which can trickle back down (*via* $\Vdash$) to the logic of the indicative. Everyone seems to be sloppy about this too. It is here where we start to enforce *strong connections* between the logics of '$\to$' and '$\rightsquigarrow$'. [Note, also, that we're presupposing a *common conjunction sign* for both the logical and the indicative conditional. That's also doing some work here.]

9. If $p \approx q$ and $\vdash p$, then $\vdash q$. [Here, $\ulcorner p \approx q \urcorner$ is an abbreviation for $\ulcorner p \Vdash q$ and $q \Vdash p \urcorner$.]

    - (9) is another crucial assumption *connecting* '$\vdash$' and '$\Vdash$'. It's presupposed by everyone I've seen. It says that if p and q "entail each other" (in the '$\Vdash$' sense of entailment), then if p is a theorem/logical truth, then so is q. It seems to me that (8) and (9) are where all the action is here.

10. If $\vdash p \to q$ and $\vdash q \to p$, then $p = q$.

    - (10) says that if $\ulcorner p \to q \urcorner$ and $\ulcorner q \to p \urcorner$ are both logical truths/theorems of $\mathscr{L}$, then p and q can be inter-substituted for each other in all sentences of $\mathscr{L}$ (*i.e.*, they can be treated as *the same sentence* of $\mathscr{L}$). Note that this assumption only involves '$\to$' and '$\vdash$'. Note, also, that this inter-substitutivity assumption is satisfied by both classical $\langle \to, \& \rangle$-logic and intuitionistic $\langle \to, \& \rangle$-logic. [I think this holds generally for *positive* sub-logics of intuitionism, but I need to look into that.]

OK, now we're ready for the first "weak" version of Gibbard's Theorem:

> **Theorem 1** (Gibbard). If $\mathscr{L}$ and its meta-theoretic operations '$\Vdash$' and '$\vdash$' satisfy assumptions (1)–(10) above, then the following must also be the case:
>
> (†) $$p \to q \Vdash p \rightsquigarrow q$$

Hence, by (†) and (6), we have $p \to q \approx p \rightsquigarrow q$, and therefore the "equivalence" of '$\to$' and '$\rightsquigarrow$', from the point of view of our "entailment" relation '$\Vdash$'. This is the "weak" version of Gibbard's Theorem (and it is the version that is most often discussed in the literature). I will also present a "strong" version (which I haven't seen elsewhere) shortly. But, first, two key remarks concerning the "weak" Gibbard theorem are in order.

- Assumptions (1)–(10) do *not* imply that either '$\to$' or '$\rightsquigarrow$' is classical. This can be shown rigorously, as I can report models of (1)–(10) that are compatible with both of the following:

  $\nVdash ((p \to q) \to p) \to p$
  $\nVdash ((p \rightsquigarrow q) \rightsquigarrow p) \rightsquigarrow p$

  In other words, assumptions (1)–(10) do *not* entail that Peirce's Law is a theorem for '$\to$' or for '$\rightsquigarrow$'.

- Moreover, assumptions (1)–(10) do not even imply that either '$\to$' or '$\rightsquigarrow$' is intuitionistic. This can be shown rigorously, as I can report models of (1)–(10) that are compatible with both of the following:

  $\nVdash (p \to (q \to r)) \to ((p \to q) \to (p \to r))$
  $\nVdash (p \rightsquigarrow (q \rightsquigarrow r)) \rightsquigarrow ((p \rightsquigarrow q) \rightsquigarrow (p \rightsquigarrow r))$

  This form of transitivity is a theorem of intuitionism, but it does not follow from (1)–(10). Thus, the class of underlying $\langle \to, \& \rangle$-logics for which the weak version of Gibbard's Theorem holds includes classical logic, intuitionistic logic, and other logics besides, which are sub-logics of intuitionism. I don't have a more precise characterization at this point, but I think this is enough to give you a taste.

There is also a "strong" version of Gibbard's Theorem, which adds the following additional assumption:

11. $p \rightsquigarrow (q \rightsquigarrow r) \Vdash (p \,\&\, q) \rightsquigarrow r$

    - (11) is a '$\Vdash$' *importation rule* for $\langle \rightsquigarrow, \& \rangle$. This says that the indicative satisfies the "rule-form" of importation, with respect to our "entailment" relation. [This is the other direction of (5).]

If we add the full *import-export* rule for $\langle \rightsquigarrow, \& \rangle$ with respect to '$\Vdash$', then we can prove the following "strong" version of Gibbard's Theorem.

> **Theorem 2** (Me?). If $\mathscr{L}$ and its meta-theoretic operations '$\Vdash$' and '$\vdash$' satisfy assumptions (1)–(11) above, then the following must also be the case:
>
> (‡) $$p \to q = p \rightsquigarrow q$$

In other words, if (1)–(11) are assumed, then it follows that $\ulcorner p \to q \urcorner$ and $\ulcorner p \rightsquigarrow q \urcorner$ are *inter-substitutible* — that they can be treated as if they are *the same sentence* of $\mathscr{L}$. This is stronger than Theorem 1 (because assumptions 1–10 entail reflexivity of '$\Vdash$': $\ulcorner p \Vdash p \urcorner$). And, as far as I know, nobody has proven this stronger result before. A few final remarks about this "strong" version of Gibbard's Theorem are in order.

- Assumptions (1)–(11) do *not* imply that '$\to$' (which now must be *identical* to '$\rightsquigarrow$') is classical. This can be shown rigorously, as I can report models of (1)–(11) that are compatible with the following:

  $\nvdash ((p \to q) \to p) \to p$

  In other words, assumptions (1)–(11) do *not* entail that Peirce's Law is a theorem for '$\to$'/'$\rightsquigarrow$'.

- Assumptions (1)–(11) *do* imply that '$\to$' (which now must be *identical* to '$\rightsquigarrow$') is intuitionistic. This can be shown rigorously, as I have a (very non-trivial!) proof from (1)–(11) of the following:

  $\vdash (p \to (q \to r)) \to ((p \to q) \to (p \to r))$

  And, together with our other assumptions, this entails that '$\to$' (which now must be *identical* to '$\rightsquigarrow$') is intuitionistic. This is interesting, but it still does *not* get us collapse to a *classical* conditional logic.

Further Remarks on Some *Non*-Consequences of Our Assumptions. It is important that certain things do *not* follow from our assumptions. Here are a list of things that are *not* implied by our assumptions, and so are *not essential* to Gibbard's Theorem. As such, Gibbard's Theorem *per se* doesn't mandate such revisions.

- *Modus Ponens* in its *axiom* form does *not* follow for the indicative conditional, from the "weak" set of assumptions (1)–(10). That is, assumptions (1)–(10) are consistent with:

  $\nvdash (p \,\&\, (p \rightsquigarrow q)) \rightsquigarrow q$

  As such, the traditional "weak" version of Gibbard's Theorem (which is similar to theorems proved by various others, such as McGee and Katz) only (essentially) presupposes that *modus ponens* preserves validity or logical truth or theoremhood (of the indicative), which seems more plausible.

- Given the "strong" set of assumptions (1)–(11), the axiom form of *modus ponens does* follow for '$\rightsquigarrow$'. As a result, the "strong" theorem does come along with a stronger commitment to *modus ponens.*

- *Modus ponens* in its *rule* form — with respect to '$\Vdash$' — does *not* follow here either, since '$\Vdash$' is only a *single premise* relation. Thus, we can't even *express modus ponens* in its *rule* form, with respect to '$\Vdash$'.

- Transitivity of '$\Vdash$' does *not* follow either — not even from our "strong" set of assumptions (1)–(11).

- Our assumptions only apply to a restricted language $\mathscr{L}$ containing $\langle \to, \&, \rightsquigarrow \rangle$. We make no claims about what happens to any of our assumptions if we enrich the language with additional connectives. [This is crucial, since, *e.g.*, if negation is added, then assumption (10) fails for intuitionism.] As such, *e.g.*, one needn't reject (10) because of Gibbard's Theorem *per se*, if the reason one is rejecting (10) is that it fails in a language *richer than* $\mathscr{L}$. That extra richness (just like the extra richness one would get if one added *modus ponens* as an axiom for '$\rightsquigarrow$' to 1–10) goes beyond what is *essential* to the Theorem.