The propositional calculus is complete*

Hans Halvorson

We already know that the proof rules are *safe*, or *not too strong*: if there is a correctly written proof $(A_1, \ldots, A_n \vdash B)$ then $A_1, \ldots, A_n$ semantically imply B. We wish now to show that the proof rules are *strong enough*, i.e. if $A_1, \ldots, A_n$ semantically imply B, then there is a correctly written proof $(A_1, \ldots, A_n \vdash B)$. We state this fact as a theorem:

Completeness Theorem. *For any sentences $A_1, \ldots, A_n, B$, if $A_1, \ldots, A_n \models B$ then $A_1, \ldots, A_n \vdash B$.*

We claim that the following Lemma is sufficient to prove completeness. In other words, if we can prove this Lemma, then completeness will be a straightforward consequence.

Main Lemma. *Let A be an arbitrary sentence, and let P be an atomic sentence. If $A \nvdash P \wedge \neg P$, then A is consistent. In other words, if A is an inconsistency, then $A \vdash P \wedge \neg P$.*

Let's see why the Main Lemma implies completeness. Suppose that $A_1, \ldots, A_n \nvdash B$. Then

$$A_1 \wedge \cdots \wedge A_1 \wedge \neg B \ \nvdash \ P \wedge \neg P,$$

because if there were a proof of a contradiction from $A_1 \wedge \cdots \wedge A_n \wedge \neg B$, then starting with $A_1, \ldots, A_n, \neg B$, we could use $\wedge$-Introduction to derive $P \wedge \neg P$; so using RAA we get $\neg\neg B$, and using DN we get B. Hence the Main Lemma shows that the sentence $A = A_1 \wedge \cdots \wedge A_n \wedge \neg B$ is consistent. Therefore, $A_1, \ldots, A_n \not\models B$.

So how do we prove the Main Lemma? Here is the idea: We show first that for any sentence A there is a sentence A^d in disjunctive normal form such that $\vdash A \leftrightarrow A^d$. (This fact is proved as Lemma 4 below. The definition of *disjunctive normal form* is in the Appendix.) So, if $A \nvdash P \wedge \neg P$ then also $A^d \nvdash P \wedge \neg P$. Since A^d is in disjunctive normal form, it can be written as:

$$A^d = \bigvee_{i=1}^{n} S_i,$$

where the S_i are conjunctions of literals. We claim that there is an i such that for all sentences X, $S_i \nvdash X \wedge \neg X$. Indeed, suppose for reductio ad absurdum that for all $i = 1, \ldots, n$, there is a sentence X_i such that $S_i \vdash X_i \wedge \neg X_i$. Clearly, if $S_i \vdash X_i \wedge \neg X_i$ then $S_i \vdash P \wedge \neg P$. So it would follow that $S_i \vdash P \wedge \neg P$ for all i. But then $\vee$E would yield $A^d \vdash P \wedge \neg P$, which we supposed is not true. Hence there is an i such that for all sentences X, $S_i \nvdash X \wedge \neg X$. Now

*Version 0.1, revised April 24, 2012. Please send suggestions and corrections to: hhalvors@princeton.edu.

we define a truth-valuation v on atomic sentences $\{P_1, P_2, \dots\}$ by setting $v(P_j) = T$ if P_j occurs in S_i, and $v(P_j) = F$ otherwise. We claim that $v(A) = T$. Indeed, $v(S_i) = T$ since for any literal X occuring in the conjunction S_i, $v(X) = T$; and since $S_i \models A^d$, it follows that $v(A) = T$. Hence A is consistent.

So now we only need to show that each sentence A is provably equivalent to a sentence A^d in disjunctive normal form. For this, we need a few preliminary lemmas.

Definition. Let A and B be sentences. We say that A is *provably equivalent* to B, and write $A \equiv B$, just in case $\vdash A \leftrightarrow B$.

Lemma 1. *The relation $\equiv$ has the following properties:*

- *Reflexive: for all A, $A \equiv A$;*
- *Symmetric: for all A, B, if $A \equiv B$ then $B \equiv A$;*
- *Transitive: for all A, B, C, if $A \equiv B$ and $B \equiv C$ then $A \equiv C$.*

Proof. (Reflexive) By the Rule of Assumptions, $A \vdash A$, hence $\vdash A \to A$ by CP. Then $\wedge$I and the definition of $\leftrightarrow$ give $\vdash A \leftrightarrow A$.

(Symmetric) Suppose that $A \equiv B$, that is $\vdash A \leftrightarrow B$. By the definition of $\leftrightarrow$ and $\wedge$E, $\vdash A \to B$ and $\vdash B \to A$. Putting these back together again in the opposite order gives $\vdash B \leftrightarrow A$, and hence $B \equiv A$.

(Transitivity) Left as an exercise for the reader. ☐

Lemma 2. *For all sentences A, A', B, B' if $A \equiv A'$ and $B \equiv B'$ then $\neg A \equiv \neg A'$ and $A \star B \equiv A' \star B'$, where $\star$ stands for any of $\wedge, \vee, \to$.*

Proof. Suppose that $A \equiv A'$, that is $\vdash A \leftrightarrow A'$. We first show that $\vdash \neg A \to \neg A'$. By the Rule of Assumptions $\neg A \vdash \neg A$. Applying $\wedge$-elimination to $\vdash A \leftrightarrow A'$ gives $\vdash A' \to A$. Then applying MTT to $\neg A \vdash \neg A$ and $\vdash A' \to A$ gives $\neg A \vdash \neg A'$. Finally, CP gives $\vdash \neg A \to \neg A'$. A similar argument shows that $\vdash \neg A' \to \neg A$, hence $\wedge$I gives $\vdash \neg A \leftrightarrow \neg A'$, that is $\neg A \equiv \neg A'$.

Now suppose that both $A \equiv A'$ and $B \equiv B'$. We show that $A \vee B \equiv A' \vee B'$. Let's show first that $\vdash (A \vee B) \to (A' \vee B')$. By assumption, $\vdash A \leftrightarrow A'$ and $\vdash B \leftrightarrow B'$, hence $\vdash A \to A'$ and $\vdash B \to B'$ by $\wedge$E. By the Rule of Assumptions, $A \vee B \vdash A \vee B$. Furthermore, $\vdash A \to A' \vee B'$ follows from $\vdash A \to A'$ by MPP and $\vee$I, and similarly $B \vdash A' \vee B'$ follows from $\vdash B \to B'$. Collecting these facts:

$$A \vee B \vdash A \vee B, \qquad A \vdash A' \vee B', \qquad B \vdash A' \vee B'.$$

Thus $\vee$E gives $A \vee B \vdash A' \vee B'$, and CP gives $\vdash (A \vee B) \to (A' \vee B')$. A similar argument shows that $\vdash (A' \vee B') \to (A \vee B)$. Therefore $A \vee B \equiv A' \vee B'$.

We leave the cases of $A \wedge B \equiv A' \wedge B'$ and $A \to B \equiv A' \to B'$ as exercises for the reader. ☐

The previous lemma has the following immediate generalization.

Lemma 3. *Let $A_1, \dots, A_n$ and $A'_1, \dots, A'_n$ be sentences such that $A_i \equiv A'_i$ for $i = 1, \dots, n$. Then $(\bigvee_{i=1}^n A_i) \equiv (\bigvee_{i=1}^n A'_i)$.*

We omit the proof of this lemma. You should try to prove it yourself using induction on the number n of disjuncts in $\bigvee_{i=1}^n A_i$.

Lemma 4. *For any sentence A, there is a sentence A^d in disjunctive normal form such that $A \equiv A^d$*

Proof. It is actually easier to prove the following, apparently stronger claim: for any sentence A, there is a sentence A^c in conjunctive normal form, and a sentence A^d in disjunctive normal form, such that $A \equiv A^c \equiv A^d$. We prove this latter claim by induction on the complexity of formulas.

(Base case) Atomic sentences are already in both CNF and DNF.

(Inductive step $\neg$) Suppose that the result is true for A. We show that the result is true for $\neg A$. Let A^c be a CNF sentence and A^d a DNF sentence such that $A \equiv A^c \equiv A^d$. Write out $A^c = \bigwedge_{i=1}^n D_i$, where D_i is a disjunction of literals, and $A^d = \bigvee_{j=1}^m C_j$, where C_j is a conjunction of literals. By DeMorgan's rules (which are provable in our system) and DN, $\neg D_i$ is provably equivalent to a sentence $\overline{D}_i$ that is a *conjunction* of literals. Then

$$\neg A^c \;=\; \neg(\bigwedge_{i=1}^n D_i) \;\equiv\; \bigvee_{i=1}^n \neg D_i \;\equiv\; \bigvee_{i=1}^n \overline{D}_i,$$

and the latter sentence is DNF. Here the first equivalence follows from DeMorgan's rules, and the second follows from Lemma 3. Since $\neg A \equiv \neg A^c$ (again by Lemma 3), it follows that $\neg A$ is provably equivalent to a DNF sentence.

(Inductive step $\wedge$) Suppose that the result is true for A and B. Thus, there are CNF sentences A^c, B^c and DNF sentences A^d, B^d such that $A \equiv A^c \equiv A^d$ and $B \equiv B^c \equiv B^d$. We show that the result is also true for $A \wedge B$. First we claim that $A^c \wedge B^c$ is a CNF sentence (that's obvious!), and $A \wedge B \equiv A^c \wedge B^c$. The latter claim follows from Lemma 3. Next we claim that $A^d \wedge B^d$ is equivalent to a DNF sentence. Write out $A^d = \bigvee_{i=1}^n S_i$ and $B^d = \bigvee_{j=1}^m T_j$, where each S_i and T_j is a conjunction of literals. Then

$$A^d \wedge B^d \;=\; (\bigvee_{i=1}^n S_i) \wedge (\bigvee_{j=1}^m T_j) \;\equiv\; \bigvee_{i=1}^n (S_i \wedge (\bigvee_{j=1}^m T_j)) \;\equiv\; \bigvee_{i=1}^n \bigvee_{j=1}^m (S_i \wedge T_j).$$

The two equivalences follow from the fact that $X \wedge (Y \vee Z) \equiv (X \wedge Y) \vee (X \wedge Z)$, which generalizes to the fact that

$$X \wedge (\bigvee_{i=1}^n Y_i) \;\equiv\; \bigvee_{i=1}^n (X \wedge Y_i).$$

(A rigorous proof of this fact would use induction on the number of disjuncts.) Now the sentence $\bigvee_{i=1}^n \bigvee_{j=1}^m (S_i \wedge T_j)$ is clearly DNF. Thus $A \wedge B$ is provably equivalent to $A^d \wedge B^d$, which is provably equivalent to a DNF sentence.

(Inductive steps $\vee$ and $\to$) One could give a proof similar to the one we gave for $\wedge$; or one could invoke the fact (also provable by induction!) that every sentence is provably equivalent to one containing only $\wedge$ and $\neg$. We leave the remaining steps as an exercise for the reader. $\square$

As discussed on the first page, the Main Lemma follows from the fact that each sentence is provably equivalent to a DNF sentence. Furthermore, the completeness of the propositional calculus follows from the Main Lemma. So, with the proof of Lemma 4 completed, we have established the completeness of the propositional calculus.

A Appendix: Definitions

Definition. A sentence X is said to be a *literal* just in case it is atomic or the negation of an atomic. A sentence X is said to be a *simple conjunction* just in case it is a conjunction of literals. Similarly, a sentence X is said to be a *simple disjunction* just in case it's a disjunction of literals. A sentence X is said to be in *disjunctive normal form* if it is a disjunction of simple conjunctions. A sentence X is said to be in *conjunctive normal form* if it is a conjunction of simple disjunctions.