MOTIVATING HILBERT SYSTEMS

Hilbert systems are formal systems that encode the notion of a mathematical proof, which are used in the branch of mathematics known as proof theory. There are other formal systems that proof theorists use, with their own advantages and disadvantages. The advantage of Hilbert systems is that they are simpler than the alternatives in terms of the number of primitive notions they involve.

Formal systems for proof theory work by deriving theorems from axioms using *rules of inference*. The distinctive characteristic of Hilbert systems is that they have very few primitive rules of inference—in fact a Hilbert system with just one primitive rule of inference, *modus ponens*, suffices to formalize proofs using first-order logic, and first-order logic is sufficient for all mathematical reasoning. *Modus ponens* is the rule of inference that says that if we have theorems of the forms $p \to q$ and p, where p and q are formulas, then χ is also a theorem. This makes sense given the interpretation of $p \to q$ as meaning "if p is true, then so is q".

The simplicity of inference in Hilbert systems is compensated for by a somewhat more complicated set of axioms. For minimal propositional logic, the two axiom schemes below suffice:

(1) For every pair of formulas p and q, the formula $p \to (q \to p)$ is an axiom.
(2) For every triple of formulas p, q and r, the formula $(p \to (q \to r)) \to ((p \to q) \to (p \to r))$ is an axiom.

One thing that had always bothered me when reading about Hilbert systems was that I couldn't see how people could come up with these axioms other than by a stroke of luck or genius. They are rather complicated, and even more so, the proofs that one generates using them directly are rather complicated. To illustrate, here's an example of a proof in a Hilbert system using these two axioms:

Theorem 1. *For every formula p, the formula $p \to p$ is a theorem.*

Proof.

(1) From axiom scheme 1 we have that $p \to (p \to p)$ is an axiom.
(2) From axiom scheme 1 we have that $p \to ((p \to p) \to p)$ is an axiom.
(3) From axiom scheme 2 we have that $(p \to ((p \to p) \to p)) \to ((p \to (p \to p)) \to (p \to p))$ is an axiom.
(4) Applying *modus ponens* to the theorems proved in steps 2 and 3, we see that $(p \to (p \to p)) \to (p \to p)$ is a theorem.
(5) Applying *modus ponens* to the theorems proved in steps 1 and 4, we see that $p \to p$ is a theorem.

□

This proof is rather difficult to come up with, because the reasoning in it is totally unlike how people naturally do mathematical reasoning. A more natural informal proof that $p \to p$ is a theorem would probably go something like this:

(1) Assume that p is true.
(2) Then (repeating ourselves) we have that p is true.

(3) It follows that $p \to p$ is a theorem by applying the rule that of inference saying that if we assume some formula q is true and then derive some formula r, then $q \to r$ is a theorem.

The rule of inference mentioned in step 3 is called *deduction*. The rules of inference of *modus ponens* and deduction together encapsulate the interpretation of the conditional connective $\to$ as meaning "if ... then". Whereas *modus ponens* tells us what we can prove *from* an "if ... then" statement, deduction tells us how we can get *to* an "if ... then" statement.

Although deduction isn't a primitive rule of inference in Hilbert systems, it can be proven to be a valid *derived* rule of inference. This requires a formalization of the notion of assumption; fortunately this doesn't require any extra machinery. The assumption of a formula p can be thought of as a movement from the original Hilbert system into another Hilbert system that has p as an extra axiom. Let us introduce some convenient notation: if we identify Hilbert systems with their sets of axioms, then given a Hilbert system Γ and a formula p, we can write the Hilbert system obtained by adding p as an axiom to Γ as $\Gamma \cup \{p\}$, using set notation. Then we can formally state the deduction theorem like so:

Theorem 2 (Deduction Theorem)**.** *For every Hilbert system Γ and every pair of formulas p and q such that q is a theorem of $\Gamma \cup \{p\}$, the formula $p \to q$ is a theorem of Γ.*

The proof of the deduction theorem is quite straightforward if we use an inductive technique. Since *modus ponens* is the only rule of inference in Hilbert systems, the set of the theorems of a Hilbert system Γ can be defined as the smallest set X with the two properties listed below:

(1) Every axiom of Γ is a member of X.
(2) For every pair of members of X of the forms $p \to q$ and p, where p and q are formulas, the formula q is also a member of X.

Therefore, in order to prove that a set X contains every theorem of Γ, it suffices to prove that X has these two properties.

For the deduction theorem, if we are given some fixed formula p, we may consider the set X of the theorems q of $\Gamma \cup \{p\}$ such that $p \to q$ is also a theorem of Γ. Then, to prove the first property, we have to prove that X contains every axiom of $\Gamma \cup \{p\}$. By Theorem 1 we have that $p \to p$ is a theorem of Γ, so X certainly contains Γ. As for the axioms of Γ, by axiom scheme 1 we have that for every formula q, including the axioms of Γ, the formula $q \to (p \to q)$ is a theorem of Γ. Applying *modus ponens*, it follows that $p \to q$ is a theorem of Γ for every axiom q of Γ.

The second property is even more straightforward to prove. Suppose q and r are formulas and $q \to r$ and q are members of X, so that $p \to (q \to r)$ and $p \to q$ are theorems of Γ. By axiom scheme 2, we have that $(p \to (q \to r)) \to ((p \to q) \to (p \to r))$ is a theorem of Γ. Applying *modus ponens* to this theorem and $p \to (q \to r)$, it follows that $(p \to q) \to (q \to r)$ is a theorem of Γ. Applying *modus ponens* again to this theorem and $p \to q$, it follows that $p \to r$ is a theorem of Γ and hence $r \in X$. This completes the proof of the deduction theorem.

Now, I realized just the other day that this proof can be used to explain where axiom schemes 1 and 2 come from. Suppose we were trying to prove the deduction theorem using *modus ponens* only without knowing what axiom schemes to start

with. The proof would proceed as usual except for the final steps in proving each of the two properties. For the first property, we would need to be able to prove that $p \to p$ is a theorem of Γ in one case, and in the other case we would get to a state where we would have that a formula q is a theorem of Γ, and we would need to be able to conclude that $p \to q$ is also a theorem of Γ. For the second property, we would get to a state where we would have that formulas of the forms $p \to (q \to r)$ and $p \to q$, where q and r are formulas, are theorems of Γ, and we would need to be able to conclude that $p \to r$ is also a theorem of Γ. This would then naturally motivate us to introduce the three axiom schemes listed below, which are exactly the theorems we need to have available in order to draw the required conclusions straightforwardly by repeatedly applying *modus ponens*.

(1) For every formula p, we have $p \to p$.
(2) For every pair of formulas p and q, we have $p \to (q \to p)$.
(3) For every triple of formulas p, q and r, we have $(p \to (q \to r)) \to ((p \to q) \to (q \to r))$.

At some point we would have to realize that the first axiom scheme was unnecessary, of course, and this would probably have to happen as a surprising discovery after playing around with the system.

(One would also naturally wonder if it would be possible to prove the second axiom from the first and third, or the third axiom from the first and second, instead of the first axiom from the second and third; or even the second axiom from the third alone, or the third axiom from the second alone. Presumably it is impossible in all cases, but I don't know how prove this.)

Anyway, I feel a lot more comfortable with Hilbert systems now that I can see how one might be directed towards their definition.