COMP 455, Models of Languages and Computation, Spring 2011
Generating a Theorem that is True but Unprovable
NOT REQUIRED

Suppose L is a sound system of logic. Suppose L is powerful enough to prove all true statements of the form "Turing machine M halts on input x." (This can be done just by simulating the Turing machine until it halts.) Let T_j be the Turing machine which, on input i, halts if the statement "T_i fails to halt on input i" is provable in L, and loops otherwise.

Theorem The statement "T_j fails to halt on input j" is true but not provable in L.

Proof Suppose T_j halts on input j. By definition of T_j, this means that in L one can prove that T_j does not halt on input j. Because L is sound, this means that T_j does not halt on input j. Thus there is a contradiction. Therefore T_j does not halt on input j. By definition of T_j, this means that in L it is not provable that T_j does not halt on input j. **End of proof**

This result can also be formalized in the *encode* notation as follows:

Suppose L is a sound system of logic. Suppose L is powerful enough to prove all true statements of the form "Turing machine M halts on input x." (This can be done just by simulating the Turing machine until it halts.) Let T be the Turing machine which, on input $encode(M)$, halts if the statement "M fails to halt on input $encode(M)$" is provable in L, and loops otherwise.

Theorem The statement "T fails to halt on input $encode(T)$" is true but not provable in L.

Proof Suppose T halts on input $encode(T)$. By definition of T, this means that in L one can prove that T does not halt on input $encode(T)$. Because L is sound, this means that T does not halt on input $encode(T)$. Thus there is a contradiction. Therefore T does not halt on input $encode(T)$. By definition of T, this means that in L it is not provable that T does not halt on input $encode(T)$. **End of proof**

Letting X_L be the statement "T fails to halt on input $encode(T)$" where T is defined as above from L, then for any sound system L of logic that can simulate Turing computations, X_L is true but not provable in L. Thus humans have the ability to get "outside" of any fixed logical system L and generate the statement X_L that is true but not provable in L. This seems to indicate that humans do not reason within any fixed logical system.