1 The universal language is undecidable

In class I was too hasty with the proof, and neglected to account for inputs that aren't valid encodings of TMs, as well as for TMs that don't halt. The proof below is more careful about both points, and uses a reduction from the halting problem.

Theorem 1. *The language* $U = \{\langle M\rangle, x : M(x) = 1\}$ *is undecidable.*

Proof. Assuming that U is decidable by a TM A, we will construct another TM B that decides $HALT$. That will be a contradiction.

Given input x, if x isn't a valid encoding $\langle M\rangle$ of a TM M then B rejects. Otherwise, B constructs a machine M' that acts as follows:

> When given input y, the TM $M'(y)$ simulates $M(y)$, and outputs 1 if $M(y)$ halts, and doesn't halt if $M(y)$ doesn't halt.[1]

Then B computes $A(\langle M'\rangle, \langle M\rangle)$ and outputs whatever A outputted.

We claim that B decides $HALT$.

- If $x \in HALT$ then $x = \langle M\rangle$ for some TM M and $M(\langle M\rangle)$ halts, hence $M'(\langle M\rangle) = 1$. In this case $\langle M'\rangle, \langle M\rangle \in U$ and A will accept, hence $B(x) = 1$.
- Otherwise, $x \notin HALT$. If x isn't a valid encoding of a TM then $B(x) = 0$. And if $x = \langle M\rangle$ for some TM M, then $M'(\langle M\rangle)$ is undefined (because $M(\langle M\rangle)$ doesn't halt), hence $\langle M'\rangle, \langle M\rangle \notin U$, so A will reject and $B(x) = 0$.

Thus, B decides $HALT$, a contradiction. ■

[1] Convince yourself that given a description of M, we can compute a description of M'.