San Skulrattanakulchai
Gustavus Adolphus College

Rice's Theorem

Sipser Problem 5.28

Theorem (Rice's Theorem)**.** *Let P be a language consisting of Turing machine descriptions where P fulfills two conditions. First, P is nontrivial—it contains some, but not all, TM descriptions. Second, P is a property of the TM's language—whenever $L(M_1) = L(M_2)$, we have $\langle M_1 \rangle \in P$ if and only if $\langle M_2 \rangle \in P$. Here, M_1 and M_2 are any TMs. Prove that P is an undecidable language.*

Proof. (Sipser reduces from A_{TM} but we'll reduce from $HALT_{\text{TM}}$ instead.) Suppose P is decidable and let R be a TM that decides it. We'll construct a TM S to decide

$$HALT_{\text{TM}} = \{ \langle M, w \rangle : M \text{ is a TM and } M \text{ halts on input string } w \},$$

which will give the desired contradiction since $HALT_{\text{TM}}$ is known to be undecidable. First assume that $\langle T_\emptyset \rangle \notin P$, where $T_\emptyset$ is the TM that rejects all input, i.e., $L(T_\emptyset) = \emptyset$. Since P is nontrivial, there exists some TM T with $\langle T \rangle \in P$. Our TM S works as follows. S="On input $\langle M, w \rangle$, where M is a TM and w is a string:

1. Construct an encoding $\langle M_w \rangle$ of a TM M_w that works as follows.

   M_w='On input string x:

   (a) Run M on w.

   (b) If M halts, run T on x. If it accepts, *accept*. If it rejects, *reject*.'

2. Run R on $\langle M_w \rangle$.

3. If R accepts, *accept*. If R rejects, *reject*."

We now prove that S decides $HALT_{\text{TM}}$. So suppose $\langle M, w\rangle$ is an input to S.

First suppose M halts on w. Then running M_w on any input x is exactly like running T on the same input x, i.e., $L(M_w) = L(T)$. We have picked T so that $\langle T\rangle \in P$. So by the second condition assumption we have $\langle M_w\rangle \in P$. Hence, R accepts $\langle M_w\rangle$. Thus, S accepts $\langle M, w\rangle$.

Next suppose M loops on w. Then M_w loops on all input; therefore, $L(M_w) = \emptyset$, i.e., $L(M_w) = L(T_\emptyset)$. We have $\langle T_\emptyset\rangle \notin P$ by supposition. So by the second condition assumption we have $\langle M_w\rangle \notin P$. Hence, R rejects $\langle M_w\rangle$. Thus, S rejects $\langle M, w\rangle$.

Therefore, S decides $HALT_{\text{TM}}$.

Next assume that $\langle T_\emptyset\rangle \in P$. Then $\langle T_\emptyset\rangle \notin \overline{P}$. It's easy to check that, just like P, the language $\overline{P}$ also fulfills the two conditions stated in the statement of Rice's Theorem. So, replacing P by $\overline{P}$ everywhere in the above proof shows that $\overline{P}$ is undecidable. Therefore, P is undecidable since decidable languages are closed under complementation. $\square$