Recap of semi-honest protocol:
- Alice garbles $f$, sends $(F,d)$ to Bob
- Alice sends her garbled input
- Using 1-out-of-2 OT, Bob picks up his garbled input
- Bob evaluates and sends result to Alice

Could go wrong if Alice/Bob malicious?
- Alice sends invalid $F,y,d$
- Alice sends invalid garbled input (including OTs)
- Bob reports wrong result

Cut-and-choose [LP07]
- I lay down $n$ pieces of paper
  - □ yes □ --- □ yes □
  - For each piece of paper, you ask me to turn it over with probability $\frac{1}{2}$
  - Suppose every piece of paper that I turned over says "yes"

Q: Could 1 of the face-down pages say "no"? w/ prob $\frac{1}{2}$ you won't detect this page
Q: Could majority of the face-down pages say "no"?
  - Ex.: $\frac{3}{4}$ say "yes", $\frac{1}{4}$ say "no"
  - $Pr[\text{turn over only "yes" pages}] = \left(\frac{1}{2}\right)^{\frac{n}{4}} \approx \text{negligible in } n$
open each item w/ prob $\frac{1}{2}$
if all opened items are "good"
then majority of unopened ones are "good"
except w/ negligible probability

Alice can generate n garbled circuits
Bob asks for random subset to be "opened"
(both wire labels on each wire) "check circuits"
(or, can give seed to RNG:
rest of garbling is deterministic,
Bob can check)

if all circuits check out, so
majority of unopened circuits are likely valid
Bob evaluates unopened circuits, report the
majority value "evaluation circuits"

Try lots of things that don't work
Protocol #1: Alice garbles $s$ circuits $(F,d)$, sends them. Bob asks to open random subset. If any circuit invalid, Bob aborts.

For each evaluation circuit:
- Alice sends her input wire labels.
- For each input wire $i$ belonging to Bob:
  - Alice sends $\langle 0 \rangle$ wire labels for input wire $i$, all eval circ.
  - $\langle 1 \rangle$ wire labels for input wire $i$, all eval circ.

As inputs to OT (Bob gets only 1 of them).

Bob evaluates each eval-circuit.
Bob outputs majority value.

Problems: no validation of Alice's OT inputs.

Ex: instead of sending correct wire labels $(A_0, A_1)$ to the OT, Alice sends $(A_0, junk)$ (in all evaluation circuits).

If Bob's input bit is 0, then evaluation works.
If 1, then evaluation fails & Bob aborts.

Selective failure attack!
Bob's failure probability depends on his input.

[LP07] approach:
evaluate \( \hat{f} \) instead of \( f \)

Bob encodes his input randomly

input bit to \( f \) was \( b \) = \( s \) input bits for \( \hat{f} \)
whose XOR is \( b \)