MAD (mutually assured destruction)
- If player 1 does nothing, then everybody lives happily ever after.
- If player 1 destroys the country of player 2, then player 2 has a threat deterrance.
- However, no soldier has any incentive to do so after his country is destroyed.

\[
\text{Movie: Dr. Strangelove}
\]

Sussman-prefered Equilibrium

Let \( \Gamma = (\mathcal{N}, A, H, P, (v_i)) \) be an E(WP).

Def (Sussman)

The Sussman of \( \Gamma \) rooted at history \( h \) is the E(WP)
\[ \Gamma(h) = (\mathcal{N}, A, H_h, P_h, (v_i)) \]
where
- \( H_h = \{ h' \mid (h, h') \in H \} \)
- \( P_h(h') = P((h, h')) \)
- \( v_i(h) = v_i((h, h')) \) for all \((h, h') \in \mathcal{H_h} \)

Strategies relativized to histories.
For each strategy \( s_i \) in \( \Gamma \), let \( s_i(h) \) denote the outcome function of \( \Gamma(h) \) determined by \( s_i \).
Def (SPE)
A subgame-perfect equilibrium (SPE) of a
Extended Game \( G \) is a strategy profile \( \sigma^* \) s.t.

such that for each history \( h \) of \( G \):

\[ \sigma^*_h \in \sigma^* \]

is an NE of \( G(h) \).

Example (Shale-Jamar)
A hex for play 1: (2,0), (L,A), (0,2)
A hex for play 2: Y agree to SPE, d
- nobody gets anything.

Questions
- Does all SPE always exist?
- Under which condition?
- How to compute it?
- When do the come truly?

We show:
- It is easy to verify if a profile is an SPE
  \( \Rightarrow \) "one deviation property" (for finite two-person games)

  For finite games we can easily compute the SPE by "backwards induction" (Kuhn's Theorem).
Notation: If $\Gamma$ is an FPGP, then $L(\Gamma)$ denotes the length of the longest history in $\Gamma$.

Lemma (omission deviation property)

Let $\Gamma = \langle h, a, H, F, v >$ be a FPGP having finite Fews. Then a strategy profile $s^*$ is an SPE of $\Gamma$ if and only if for every player $i \in H$ and every history $h$ for which $P(h,i)$, we have:

$u^*_i (\alpha_i (s^*_i, s^*_i)) \geq u^*_i (\alpha_i (s^*_i, s_i))$

for every strategy $s_i$ of player $i$ in the subgame $P(h,i)$ that differs from $s^*_i$ only in the action after the initial history of $P(h,i)$.

Without the uninformed parts, it is just the def of SPE.

Proof:

$\Rightarrow$: obvious.

$\Leftarrow$: By contradiction.

Suppose that $s^*$ is not an SPE. Then there is a history $h$ and a player $i$ such that $s_i$ is a profitable deviation for player $i$ in the subgame $P(h)$. We see the number of histories $h'$ with $s_i(h') \neq s^*_i(h')$ is at most $E(\Gamma(h))$ and hence finite (finite history assumption). Since deviations not on the subgame outcome path are irrelevant,
Choose profitable deviation $s_i \in \Gamma(h)$ with minimal number of deviation points.

Let $h^*$ be longest history in $\Gamma(h)$ with $s_i(h^*) = s^*_1(h^*)$, i.e., "deepest" deviation point for $s_i$.

Then in $\Gamma((h, h^*))$, $s_i(h^*)$ differs from $s^*_1(h^*)$ only in the initial history. Moreover, $s_i(h^*)$ is a profitable deviation in $\Gamma((h, h^*))$, since $h^*$ is the longest history in $\Gamma(h)$ with $s_i(h^*) = s^*_1(h^*)$.

So $\Gamma((h, h^*))$ is the desired subgame where a one-step deviation is sufficient to improve the utility.

The corresponding proposition for infinite horizon games does not hold.

Consider: example

\[
\begin{array}{cccccccc}
D & D & D & D & D & D & \cdots & D \\
D & 0 & 0 & 0 & 0 & 0 & & \\
\end{array}
\]

Strategy $s$ with $s(h) = 0$ for all $h \in H \setminus \emptyset$ satisfies "one deviation property," but

- is not a SPE, since it is dominated by $s^*$ with $s^*(h) = A$ for all $h \in H \setminus \emptyset$.