Reexamining Sequential Rationality in Games with Imperfect Recall

Nicholas Butler

Princeton University

October 7, 2015
Motivation

- Imperfect Recall is a salient attribute of real decision makers.
- Models which incorporate imperfect recall remain rare.
  1. Non-existence of equilibrium in behavioral strategies
  2. Issues surrounding the notion of sequential rationality
     - Paradox of the Absentminded Driver
- Try to address the second problem
Literature: Paradox of the Absentminded Driver

- Single Self Approach
  - Piccione and Rubinstein (1997)
  - Halpern and Grove (1997)

- Multi-Self Approach
  - Gilboa (1997)
Goal

- Outline the most well known formulations and their criticisms
  - Single Self (PR)
  - Multi-Self (AHP)
- Argue that existing resolutions are not entirely satisfactory
- Propose a notion of beliefs and sequential rationality satisfying
  - Equivalence between Ex-Ante Optimality and Sequential Rationality
  - Reduces to the standard notion of sequential rationality in decision problems with perfect recall.
Decision problem $G = \langle A, T, \rho, b_0, H, u \rangle$

- $A$ - set of actions
- $A^*$ - set of all ordered tuples of elements of $A$
- $T \subset A^*$ is a tree satisfying
  1. $(a_1, \ldots, a_n) \in T \Rightarrow (a_1, \ldots, a_{n-1}) \in T$
  2. $\phi \in T$

- $A(w) = \{a \in A | (w, a) \in T\}$ - actions available at $w \in T$
- $Z$ - set of terminal nodes
- $\rho : T \rightarrow \{0, 1\}$ - the player function ($0$ denotes nature)
- $u : Z \rightarrow \mathbb{R}$ - utility index
Information Partition

- $H$ is a partition of $T_1$
  - $w, w' \in h \in H \Rightarrow A(w) = A(w')$
- $h(w)$ is the unique information set containing $w \in T_1$
- $P(w)$ is the set of all predecessors of $w \in T_1$
- $a(\tilde{w}, w)$ is the action taken at $\tilde{w} \in P(w)$ on the way to $w \in T_1$

Perfect Recall

$H$ satisfies perfect recall if for all $w \in T_1$

1. $w' \in h(w) \Rightarrow w' \notin P(w) \land w \notin P(w')$
2. $w' \in h(w) \land \tilde{w} \in P(w) \Rightarrow \exists \tilde{w}' \in P(w') \cap h(\tilde{w}) \land a(\tilde{w}', w') = a(\tilde{w}, w)$
Behavioral Strategies

The space of behavioral strategies of the player is given by

\[ B_1 = \prod_{h \in H} \Delta(A(h)) \]

The behavioral strategy of nature is given by \( b_0 \). The probability of reaching a node \( w \in T \) under \( b \in B \) is

\[ p(w|b) = \prod_{\tilde{w} \in P(z)} b(a(\tilde{w}, w)|\tilde{w}) \]

The probability of reaching \( w \in T \) from \( w' \in P(w) \) is

\[ p(w|b, w') = \prod_{\tilde{w} \in P(w)/P(w')} b(a(\tilde{w}, w)|\tilde{w}) \]
Ex-Ante Optimality

A strategy $b$ is ex-ante optimal if for all $b' \in B$

$$\sum_{z \in Z} p(z|b)u(z) \geq \sum_{z \in Z} p(z|b')u(z)$$

- Ex-ante Optimality (PR) $\leftrightarrow$ Planning Optimality (AHP)
  - Strategy delivers the highest utility as measured at the outset.
AHP and PR use essentially the same notion of consistent beliefs

- Beliefs are fixed when players consider deviating
- Long run proportion of times visiting the node
- Not a consequence of Baye’s Rule

- Interpreted as the probability of being at $x$ conditional on being in the information set $X$.
- In principle locational beliefs are unconstrained if there is absentmindedness.

Beliefs $\mu$ are consistent with $b$ if for any $X \in H$ such that $p(X|b) > 0$

$$\mu(x|X) = \frac{p(x|b)}{\sum_{x \in X} p(x|b)}$$
Sequential Rationality

A strategy $b$ is sequentially rational if for any $X$ and $\mu$ consistent with $b$, such that $p(X|b) > 0$

$$\sum_{x \in X} \mu(x|X) \sum_{z \in Z} p(z|b, x) u(z) \geq \sum_{x \in X} \mu(x|X) \sum_{z \in Z} p(z|b', x) u(z)$$

for all $b' \in B$.

- Sequential Rationality (PR) $\leftrightarrow$ Action Optimality (AHP)
  - Strategy achieves highest utility as measured at each information set.

Nicholas Butler (Princeton University)
Paradox of the Absentminded Driver

Nicholas Butler (Princeton University)

Sequential Rationality
Let \( p \) be the probability of continuing.

The ex-ante utility of the player is

\[
U(p) = 0(1 - p) + 4p(1 - p) + p^2 = 4p - 3p^2
\]

\[
\frac{dU}{dp} = 4 - 6p \Rightarrow p^* = \frac{2}{3}
\]
Single Self Paradox

The players’ utility upon arriving at the information set is

\[ U(p, q, \alpha) = \alpha[(1 - p)0 + p(1 - q)4 + pq] + (1 - \alpha)[(1 - p)4 + p] \]

Continuing with probability \( p \) is sequentially rational if

\[ p \in \arg \max_{\tilde{p} \in [0,1]} U(\tilde{p}, \tilde{p}, \frac{1}{1 + p}) \]

\[ U(\tilde{p}, \tilde{p}, \frac{1}{1 + p}) = \frac{1}{1 + p}[4\tilde{p}(1 - \tilde{p}) + \tilde{p}^2] + \frac{p}{1 + p}[4(1 - \tilde{p}) + \tilde{p}] \]

\[ \frac{d}{d\tilde{p}} U(\tilde{p}, \tilde{p}, \frac{1}{1 + p}) = \frac{4 - 6\tilde{p}}{1 + p} - \frac{3p}{1 + p} \Rightarrow p^* = \frac{4}{9} \]
Implications: Single Self

1. There are ex-ante optimal strategies that are not sequentially rational.

2. There are sequentially rational strategies that are not ex-ante optimal.

3. These results are not sensitive to the notion of consistent beliefs adopted, since the planning stage and action stage F.O.Cs differ unless $\alpha = 0$.

4. These are undesirable properties.
Criticisms

- **AHP**
  - Nodes not information sets are the point of decision making.
    - Maximizing over $p$ and $q$ subject to $p = q$ implies that the decision maker controls his actions at other nodes.
    - Should adopt a fixed point approach.
  - If you can control your action at the other nodes in the information set then $\alpha$ should depend on $p$.

- **Halpern and Grove**
  - The distribution over terminal nodes implied by the ex-ante and continuation utilities are different.
Recall that the utility function takes the form

\[ U(p, q, \alpha) = \alpha[(1 - p)0 + p(1 - q)4 + pq] + (1 - \alpha)[(1 - p)4 + p] \]

Planning Stage: (Objectively Correct)

\[ P_E(0|p) = (1 - p) \]
\[ P_E(4|p) = (1 - p)p \]
\[ P_E(1|p) = p^2 \]

Action Stage: assuming \( p = q \)

\[ P_A(0|p) = \alpha(1 - p) \leq P_E(0|p) \]
\[ P_A(4|p) = \alpha p(1 - p) + (1 - \alpha)(1 - p) \geq P_E(4|p) \]
\[ P_A(1|p) = \alpha p^2 + (1 - \alpha)p \geq P_E(1|p) \]
The existence of a “paradox” is not surprising.

Planning stage and Action stage reasoning are not equally valid

- Planning stage probabilities coincide with the objectively correct long run frequencies of following the strategy $p$.
- Action stage beliefs coincide with counterfactual long run frequencies.
- The player does objectively worse by utilizing the incorrect action stage probabilities.

If during the planning stage the player knows he will be wrong in the future, why not choose not to think in the future?
Multi-Self Partial Resolution

The utility of the player upon arriving at the information set is

$$U(p, q, \frac{1}{1+q}) = \frac{1}{1+q} [4p(1-q) + pq] + \frac{q}{1+q} [4(1-p) + p]$$

Continuing with probability $p$ is “action optimal” if

$$p \in \arg \max_{\tilde{p} \in [0,1]} U(\tilde{p}, p, \frac{1}{1+p})$$

$$\frac{d}{d\tilde{p}} U(\tilde{p}, p, \frac{1}{1+p}) = \frac{4-3p}{1+p} + \frac{-3p}{1+p} \Rightarrow p^* = \frac{2}{3}$$

In general, under the multi-self approach, any ex-ante optimal strategy is also sequentially rational.
(Multi)-Self Defeating Sequential Rationality
The ex-ante utility is given by

\[ U(p) = 1 - p + 2p^2 \]

Easy to see that \( p^* = 1 \) is the ex-ante optimal strategy
The continuation utility is given by,

\[ U(p, q, \alpha) = \alpha[(1 - p)1 + p(1 - q)0 + 2pq] + (1 - \alpha)[(1 - p)0 + 2p] \]

- \( p = 1 \) is sequentially rational since it achieves the highest possible utility.
- \( p = 0 \) is sequentially rational since if \( \alpha = 1 \) and \( q = 0 \)

\[ U(p, 0, 1) = (1 - p) \]
We seek a fixed point of the form

\[ p \in \text{arg max} \ U(\tilde{p}, p, \frac{1}{1 + p}) \]

\[ \tilde{p} \in [0, 1] \]

\[ U(p, q, \frac{1}{1 + q}) = \frac{1}{1 + q}[(1 - p) + 2pq] + \frac{q}{1 + q}2p = \frac{1 - p + 4pq}{1 + q} \]

\[ \frac{d}{d\tilde{p}} U(\tilde{p}, p, \frac{1}{1 + p}) = \frac{-1 + 4p}{1 + p} = 0 \Rightarrow p^* = \frac{1}{4} \]

\[ U(p, \frac{1}{4}, \frac{4}{5}) = \frac{1 - p + 4\left(\frac{1}{4}\right)p}{1 + \frac{1}{4}} = \frac{4}{5} \]

So \( p = \frac{1}{4} \) is also sequentially rational
Recall the ex-ante utility is given by

\[ U(p) = 1 - p + 2p^2 \]

The F.O.C. for an extremum is

\[ \frac{d}{dp} U(p) = -1 + 4p = 0 \implies p^* = \frac{1}{4} \]

The S.O.C. is

\[ \frac{d^2}{dp^2} U(p) = 4 > 0 \]

So \( p^* = \frac{1}{4} \) is the minimum achievable ex-ante utility.
Implications: Multi-Self

- There can be sequentially rational strategies that are not ex-ante optimal.

- It is possible for a sequentially rational strategy to be the worst possible choice ex-ante.

- Since ex-ante utility is identical to the objective expected utility, a planning optimal strategy need not be “good” in any objective sense.
  
  - Ok if you believe in a “planning stage” where the selves can coordinate on the good equilibrium.
  - Problematic if you don’t believe in such a stage.
First: The driver makes a decision at each intersection through which he passes. Moreover, when at one intersection, he can determine the action only there, and not at the other intersection - where he isn’t Second: Since he is in completely indistinguishable situations at the two intersection, whatever reasoning obtains at one must also obtain at the other, and he is aware of this.
PR seek a fixed point of the form

\[ p \in \arg \max_{\tilde{p} \in [0,1]} U(\tilde{p}, \tilde{p}, \frac{1}{1 + p}) \]

- PR’s formulation appears to contradict AHP’s second principle
- Since decisions are made at nodes and not information sets, AHP assert that this requires a “mysterious psychic process”

Can one interpret PR’s formulation without simultaneous control?
Yes it is possible! If we interpret information sets carefully

- I know that whatever reasoning process I am currently going through, I will also go through at the other node.
- When considering a deviation, I can conjecture that whatever deviation I choose as a result of my reasoning process, I will also deviate to (or will have already deviated to) at the other nodes since my reasoning at the other nodes will be (or has been) identical.
- This conjecture is confirmed by the observable play.
Information sets describe indistinguishable histories and “states of mind”

- Imagine the DM is a computer whose internal state is reset every time he reaches a node in the same information set.
- His specific reasoning process (programming) does not matter.
- Matters that he knows he will reason in exactly the same way at the other nodes in the same information set.

I agree with AHP’s first principle, so simultaneous control is problematic.

However, PR’s formulation can be interpreted without simultaneous control.
Interpreting Deviations

The Multi-Self Story:
- Coordination between the selves is an equilibrium condition that is not internalized.
- Selves are not aware that their reasoning is identical to that of the other selves.
- Allowed to consider the possibility of a deviation leading to counterfactual miscoordination.

The Single Self + Conjecture Story:
- Doesn’t entertain the possibility of miscoordination.
- DM makes separate choices at each information set; however, he understands that his reasoning (and therefore his choice) will be identical.
- Knows he is unable to “outthink” himself.
Under the Single Self + Conjecture approach, what happens when a DM decides to deviate?

- Simultaneously alters his conjecture regarding his actions at other nodes.
- Simultaneously modifies his beliefs to be consistent with the deviation and altered conjecture.
- Consistent with AHP’s criticism of PR
  - Maximizing over p and q simultaneously implies $\alpha$ should depend on $p$. 

Nicholas Butler (Princeton University)
Lessons from previous models

- Address the “Probability Paradox”
  - Make sure action stage probabilities agree with the objective conditional probabilities.
- Don’t get stuck in bad equilibria.
  - Avoid the multi-self approach.
- The Single Self + Conjecture story is at least as plausible as the Multi-Self story.
  - If we adopt the single self approach should have self locating beliefs.
Let $X \subset H$ be an information set

**Definition: Covering**

$C_X \subset X$ is a *covering* of $X$ if for any $z \in Z$ such that there exists a $x \in X$ such that $x \in P(z)$, there exists a $c \in C_X$ such that $c \in P(z)$.

For an information set $X$ with covering $C_X$, define $c : Z \to C_X$ as follows.

1. $c(z) \in P(z) \cap C_X$
2. $x \in P(z) \cap C_X \Rightarrow c(z) \notin P(x)$

$C : H \to 2^T$ is composed of coverings if for every $X \in H$, $C(X)$ is a covering of $X$. 
Two coverings $C_1 = \{x_1\}$, $C_2 = \{x_1, x_2\}$

- $C_1 \Rightarrow c(0) = c(4) = c(1) = x_1$
- $C_2 \Rightarrow c(0) = x_1$ and $c(4) = c(1) = x_2$
For $c \in C(X)$, beliefs are of the form

$$\mu(c|X, b) = \frac{p(c|b)}{p(X|b)}$$

Interpreted as the probability of reaching $c$ under strategy $b$ conditional on reaching the information set $X$.

**C-Sequential Rationality**

A strategy $b$ is $C$-sequentially rational if for every $X \in H$ such that $p(X|b) > 0$ there is no $b' \in B$ such that

$$\sum_{z \in Z} \mu(c(z)|X, b)p(z|b, c(z))u(z) \geq \sum_{z \in Z} \mu(c(z)|X, b')p(z|b', c(z))u(z)$$
Sequential Rationality

Why might this work?

\[
\sum_{z \in Z} \mu(c(z) | X, b) p(z | b, c(z)) u(z) \\
= \sum_{z \in Z} \frac{p(c(z) | b)}{p(X | b)} p(z | b, c(z)) u(z) \\
= \sum_{z \in Z} p(c(z) | X, b) p(z | b, c(z)) u(z) \\
= \sum_{z \in Z} p(c(z) | X, b) p(z | b, c(z), X) u(z) \\
= \sum_{z \in Z} p(c(z) \land z | X, b) u(z) \\
= \sum_{z \in Z} p(z | X, b) u(z) = U(b | X)
\]
Absentminded Driver Example

- Using the covering $C_1 = \{x_1\}$

$$U(p|X) = \sum_{z \in Z} \mu(c(z)|X, p)p(z|p, c(z))u(z)$$

$$= 1(1 - p)0 + 1(p(1 - p))4 + 1(p^2)1$$

$$= p^2 + 4p(1 - p) = U(p)$$

- Using the covering $C_2 = \{x_1, x_2\}$

$$U(p|X) = \sum_{z \in Z} \mu(c(z)|X, p)p(z|p, c(z))u(z)$$

$$= 1(1 - p)0 + p(1 - p)4 + p(p)1$$

$$= p^2 + 4p(1 - p) = U(p)$$
Conjectures

Conjecture 1:
If $C$ and $C'$ are composed of coverings, then a strategy $b$ is $C$–sequentially rational if and only if $b$ is $C'$-sequentially rational.

So sequential rationality is independent of the choice of $C$, as long as it is composed of coverings.

Conjecture 2:
For any $C$ composed of coverings, a strategy $b$ is ex-ante optimal if and only if it is $C$-sequentially rational.
Wrapping Up

- No “probability paradox”
  - Action stage probabilities implied by beliefs agree with objective conditional probabilities.

- No Absentmindedness $\Rightarrow$ only one covering, the entire information set.
  - Reduces to the standard notion of sequential rationality in games with perfect recall.

- With absentmindedness there are multiple ways players can form beliefs that work.

- Requires beliefs to be self locating, and non-locational in general

- Easily extended to extensive form games.