Asymmetric (Libertarian) Paternalism: Explanation Based on Decisions Under Interval Uncertainty, and Possible Applications to Education

Olga Kosheleva[1] and François Modave[2]
[1]Department of Teacher Education
[2]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968
olgak@utep.edu, fmodave@utep.edu

1. Outline

- In general, human being are rational decision makers.
- However, in many situations, they exhibit unexplained "inertia", reluctance to switch to a better decision.
- We show that this seemingly irrational behavior can be explained if we take uncertainty into account.
- We also explain how this phenomenon can be utilized in education.

2. Traditional Approach to Human Decision Making: A Brief Reminder

- *Situation:* we have alternatives $A_1, \ldots, A_n$.
- *Idea:* alternatives are characterized by their "utility values" $u(A_1), \ldots, u(A_n)$.
- *Preference:* A_i is preferable to A_j if and only if

$$u(A_i) > u(A_j).$$

- *Empirical testing:* we need to compare
  - empirically "testable" behavior (such as preferring one alternative A_i to another alternative A_j) and
  - difficult-to-test comparison between the (usually unknown) utility values.
- *Conclusion:* empirical testing is difficult.

3. A Testable Consequence of the Traditional Approach to Decision Making

- *Fact:* for every two alternatives A_i and A_j:
  - either $u(A_i) > u(A_j)$, i.e., the alternative A_i is better,
  - or $u(A_j) > u(A_i)$, i.e., the alternative A_j is better.
- *Comment:* exact equality of $u(A_i)$ and $u(A_j)$ is highly improbable.
- In the first case $u(A_i) > u(A_j)$,
  - if we originally only had A_i, and then we add A_j, then we stick with A_i;
  - on the other hand, if we originally only had A_j, and then we add A_i, then we switch our choice to A_i.
- Similarly, in the second case $u(A_j) > u(A_i)$.

4. The Above Testable Consequence is in Perfect Agreement with Common Sense

- *Claim:* the above behavior is in perfect agreement with common sense.
- *Case 1:* the alternative A_i is preferable to the alternative A_j.
- *Expected behavior:* choose A_i irrespective of whether we started with only A_i or only A_j.
- *Case 2:* the alternative A_j is preferable to the alternative A_i.
- *Expected behavior:* choose A_j irrespective of whether we started with only A_i or only A_j.

5. For Close Alternatives, Decision Makers Do Not Behave in This Rational Fashion

- *Empirical result:* when the alternatives are close in value, decision maker exhibit "inertia".
- *Example:* selecting between two similar retirement plans A_i and A_j.
- *Case 1:* we start with the plan A_i and then add A_j.
- *Typical behavior:* stick to A_i.
- *Case 2:* we start with the plan A_j and then add A_i.
- *Typical behavior:* stick to A_j.
- *Why this is counter-intuitive:*
  - if A_i is better, then in Case 2, people should switch to A_i;
  - if A_j is better, then in Case 1, people should switch to A_j.

6. Maybe Human Behavior Is Irrational?

- How can we explain this seemingly irrational behavior?
- One possible explanation is that many people do often make bad (irrational) decisions:
  - waste money on gambling,
  - waste one's health or alcohol and drugs, etc.
- However, the above inertial behavior occurs among the most successful (otherwise rational) people.
- It is therefore reasonable to look for an explanation of this seemingly irrational behavior.
- It turns out that
  - we can come up with such an explanation
  - if we take into account uncertainty related to decision making.

7. How to Take Into Account Uncertainty in Decision Making Situations

- In practice, we can predict the consequences of alternatives only approximately, with some accuracy ε.
- So, instead of the exact values $u(A_i)$ and $u(A_j)$, we only know approximate values $\widetilde{u}_i$ and $\widetilde{u}_j$.
- The actual utility values can be within intervals $\mathbf{u}_i = [\widetilde{u}_i - \varepsilon, \widetilde{u}_i + \varepsilon]$ and $\mathbf{u}_j = [\widetilde{u}_j - \varepsilon, \widetilde{u}_j + \varepsilon]$.
- If the estimates are close, i.e., if $|\widetilde{u}_i - \widetilde{u}_j| < 2\varepsilon$, then
  - there exist values $u_i \in \mathbf{u}_i$ and $u_j \in \mathbf{u}_j$ s.t. $u_i < u_j$; and
  - there exist values $u_i \in \mathbf{u}_i$ and $u_j \in \mathbf{u}_j$ s.t. $u_i > u_j$.
- Thus, switching may decrease utility.
- So, it is prudent not to switch (especially since often switching comes with a penalty).

8. Another Case when Inertia is Beneficial: Control of a Mobile Robot

- We change direction based on the moment-by-moment measurements of the robot's location and/or velocity.
- Measurements are never 100% accurate.
- The resulting measurement noise leads to random deviations – shaking and "wobbling".
- Each change in direction requires that energy from the robot's battery go to the robot's motor.
- So, this wobbling drains the batteries and slows down the robot's motion.
- *Natural idea:* only change if it's clear (beyond uncertainty) that this will improve the performance.
- *Result:* UTEP robot's 1st place at 1997 AAAI competition.

9. Asymmetric Paternalism: Practical Application of Present-Biased Preferences

- *Fact:* the decision-making inertia is used in practice, to encourage desirable behavior.
- *Example:* a kid can drink either a healthy fruit juice or a soda drink which has no health value.
- *Traditional paternalism:* prohibit undesirable choices.
- *Problem:* this enforcement rarely works.
- *More efficient idea:*
  - at first provide only the desired alternative,
  - and then introduce all the other alternatives.
- *Example:* have only healthy drinks for the first few weeks of school, but then allow all the choices.
- *Result:* due to inertia, kids tend to stick to their original healthier choice.

10. How Does Our Explanation Help?

- *Fact:* asymmetric paternalism works.
- *Natural question:* do we need any explanation to make it work?
- *Problem:* sometimes this approach works, and sometimes it does not.
- *Additional problem:* it is now known how to predict when it will work.
- *Our solution:* this approach works when $|\widetilde{u}_i - \widetilde{u}_j| < 2\varepsilon$.

11. Potential Applications to Education

- *Current applications:* in economy and in health.
- *Our idea:* use it in education.
- *Example:*
  - when the students just come to class from recess or from home, it is difficult to get their attention;
  - once they get engaged in the class, it is difficult for them to stop when the bell rings.
- *Objective:* prevent students from switching to a passive state A_j.
- *How to use this phenomenon:*
  - to start a class with engaging fun material, to get them into the studying state A_i;
  - they will (hopefully) remain in A_i even when a somewhat less fun necessary material is presented.