Peak-End Rule: A Utility-Based Explanation

Olga Kosheleva, Martine Ceberio, and Vladik Kreinovich

University of Texas at El Paso
El Paso, Texas 79968, USA
olgak@utep.edu, mceberio@utep.edu
vladik@utep.edu
1. Peak-End Rule: Description and Need for an Explanation

• Often, people judge their overall experience by the peak and end pleasantness or unpleasantness.

• In other words, they use only the maximum (minimum) and the last value.

• This is how we judge pleasantness of a medical procedure, quality of the cell phone perception, etc.

• There is a lot of empirical evidence supporting the peak-end rule, but not much of an understanding.

• At first glance, the rule appears counter-intuitive: why only peak and last value? why not average?

• In this talk, we provide such an explanation based on the traditional decision making theory.
2. Towards an Explanation

• Our objective is to describe the peak-end rule in terms of the traditional decision making theory.

• According to decision theory, preferences of rational agents can be described in terms of utility.

• A rational agent selects an action with the largest value of expected utility.

• Utility is usually defined modulo a linear transformation.

• In the above experiments, we usually have a fixed status quo level which can be taken as 0.

• Once we fix this value at 0, the only remaining non-uniqueness in describing utility is scaling $u \rightarrow k \cdot u$.

• We want to describe the “average” utility corresponding to a sequence of different experiences.
3. Need for a Utility-Averaging Operation

- We assume that we know the utility corresponding to each moment of time.

- To get an overall utility value, we need to combine these momentous utilities into a single average. Hence:
  - if we have already found the average utility corresponding to two consequent sub-intervals of time,
  - we then need to combine these two averages into a single average corresponding to the whole interval.

- In other words, we need an operation \( a \ast b \) that:
  - given the average utilities \( a \) and \( b \) corresponding to two consequent time intervals,
  - generates the average utility of the combined two-stage experience.
4. Natural Properties of the Utility-Averaging Operation

- If two stages have the same average utility \( a = b \), then two-stage average should be the same: \( a * a = a \).

- In mathematical terms, this means that the utility-averaging operation \( * \) should be idempotent.

- If we make one of the stages better, then the resulting average utility should increase (or at least not decrease).

- In other words, the utility-averaging operation \( * \) should be monotonic: if \( a \leq a' \) and \( b \leq b' \) then \( a * b \leq a' * b' \).

- Small changes in one of the stages should lead to small changes in the overall average utility.

- In precise terms, this means that the function \( a*b \) must be continuous.
5. Properties of Utility Averaging (cont-d)

- For a three-stage situation, with average utilities $a$, $b$, and $c$:
  - we can first combine $a$ and $b$ into $a \ast b$, and then combine this with $c$, resulting in $(a \ast b) \ast c$;
  - we can also combine $b$ and $c$, and then combine with $a$, resulting in $a \ast (b \ast c)$.

- The resulting three-stage average should not depend on the order: $(a \ast b) \ast c = a \ast (b \ast c)$.

- In mathematical terms, the operation $a \ast b$ must be associative.

- Finally, since utility is defined modulo scaling $u \rightarrow k \cdot u$, the utility-averaging does not change with scaling:
  $$(k \cdot a) \ast (k \cdot b) = k \cdot (a \ast b).$$
6. Main Result

Let \( a \ast b \) be a binary operation on the set of all non-negative numbers which satisfies the following properties:

1) it is idempotent, i.e., \( a \ast a = a \) for all \( a \);
2) it is monotonic: \( a \leq a' \) and \( b \leq b' \) imply \( a \ast b \leq a' \ast b' \);
3) it is continuous as a function of \( a \) and \( b \);
4) it is associative, i.e., \( (a \ast b) \ast c = a \ast (b \ast c) \);
5) it is scale-invariant, i.e., \( (k \cdot a) \ast (k \cdot b) = k \cdot (a \ast b) \) for all \( k \), \( a \) and \( b \).

Then, \( \ast \) coincides with one of the following four operations:

- \( a_1 \ast \ldots \ast a_n = \min(a_1, \ldots, a_n) \);
- \( a_1 \ast \ldots \ast a_n = \max(a_1, \ldots, a_n) \);
- \( a_1 \ast \ldots \ast a_n = a_1 \);
- \( a_1 \ast \ldots \ast a_n = a_n \).
7. Discussion

- Every utility-averaging operation which satisfies the above reasonable properties means that we select:
  - either the worst
  - or the best
  - or the first
  - or the last utility.
- This (almost) justifies the peak-end phenomenon.
- The only exception that in addition to peak and end, we also have the start as one of the options:
  \[ a_1 \ast \ldots \ast a_n = a_1. \]
- A similar result can be proven if we take negative \( a_i \).
8. First Open Problem

- Following the psychological experiments, we only considered:
  - the case when all experiences are positive and
  - the case when all experiences are negative.

- What happens in the general case?

- If we impose an additional requirement of shift-invariance, then we can get a result similar to the above:
  \[(a + u_0) \ast (b + u_0) = a \ast b + u_0.\]

- But what if we do not impose this additional requirement?
9. Second Open Problem

- Are all five conditions necessary? Some are necessary:
  1) $a \ast b = a + b$ satisfies all the conditions except for idempotence;
  4) $a \ast b = \frac{a + b}{2}$ satisfies all the conditions except for associativity;
  5) the closest-to-1 value from $[\min(a, b), \max(a, b)]$ satisfies all the conditions except for scale invariance.

- However, it is not clear whether monotonicity and continuity are needed to prove our results.
10. Acknowledgments

This work was supported in part:

- by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721,
- by Grants 1 T36 GM078000-01 and 1R43TR000173-01 from the National Institutes of Health, and
- by a grant N62909-12-1-7039 from the Office of Naval Research.
11. Proof

- For every $a \geq 1$, let us denote $a \ast 1$ by $\varphi(a)$.
- For $a = 1$, due to the idempotence, $\varphi(1) = 1 \ast 1 = 1$.
- Due to monotonicity, $\varphi(a)$ is (non-strictly) increasing.
- Due to associativity, $(a \ast 1) \ast 1 = a \ast (1 \ast 1)$.
- Due to idempotence, $1 \ast 1 = 1$, so $(a \ast 1) \ast 1 = a \ast 1$, i.e., $\varphi(\varphi(a)) = \varphi(a)$.
- Thus, for every value $t$ from the range of the function $\varphi(a)$ for $a \geq 1$, we have $\varphi(t) = t$.
- Since $a \ast b$ is continuous, $\varphi(a) = a \ast 1$ is also continuous.
- Thus, the range of $\varphi(a)$ is an interval (finite or infinite).
- Since the function $\varphi(a)$ is monotonic, and $\varphi(1) = 1$, this interval $S$ must start with 1.
12. Proof (cont-d)

- Thus, we have three possible options:
  - $S = \{1\}$;
  - $S = [1, k]$ or $S = [1, k)$ for some $k \in (1, \infty)$;
  - $S = [1, \infty)$.

- Let us consider these three options one by one.

- When $S = \{1\}$, we have $\varphi(a) = a \ast 1 = 1$ for all $a$.

- From scale invariance, we can now conclude that for all $a \geq b$, we have $a \ast b = b \cdot \left(\frac{a}{b} \ast 1\right) = b \cdot 1 = b$.

- When $S = [1, k]$ or $S = [1, k)$, every value $t$ between 1 and $k$ is a possible value of $\varphi(a)$.

- Thus, $\varphi(t) = t \ast 1 = t$ for all such values $t$.

- In particular, for every $\varepsilon > 0$, for the value $t = k - \varepsilon$, we have $\varphi(k - \varepsilon) = k - \varepsilon$. 
13. Proof (cont-d)

- From $\varphi(k - \varepsilon) = k - \varepsilon$ and continuity, we get $\varphi(k) = k$.
- For $t \geq k$, due to monotonicity, we have $\varphi(t) \geq k$; since $\varphi(t) \in S \subseteq [1, k]$, we have $\varphi(t) \leq k$, so $\varphi(t) = k$.
- Due to associativity, we have $l = r$, where 
  
  \[ l = ((k - \varepsilon)^2 \ast (k - \varepsilon)) \ast 1 \; \text{and} \; r = (k - \varepsilon)^2 \ast ((k - \varepsilon) \ast 1) \; . \]

- Here, due to scale-invariance,
  
  \[ (k - \varepsilon)^2 \ast (k - \varepsilon) = (k - \varepsilon) \ast ((k - \varepsilon) \ast 1) = (k - \varepsilon) \ast \varphi(k - \varepsilon) = (k - \varepsilon) \ast (k - \varepsilon) = (k - \varepsilon)^2 \; . \]

- Thus, $((k - \varepsilon)^2 \ast (k - \varepsilon)) \ast 1 = (k - \varepsilon)^2 \ast 1 = \varphi((k - \varepsilon)^2)$.

- For $k > 1$, we have $k^2 > k$ and thus $(k - \varepsilon)^2 > k$ for sufficiently small $\varepsilon > 0$; so, $l = \varphi((k - \varepsilon)^2) = k$.

- Since $(k - \varepsilon) \ast 1 = k - \varepsilon$, we have $r = (k - \varepsilon)^2 \ast (k - \varepsilon) = (k - \varepsilon)^2 > k$; this contradicts to $r = l = k$. 
14. Proof (cont-d)

- The contradiction proves that the case $S = [1, k]$ or $S = [1, k)$ is impossible.

- When $S = [1, \infty)$, every value $t \geq 1$ is a possible value of $\varphi(a)$, thus $\varphi(t) = t \ast 1 = t$ for all values $t \geq 1$.

- Thus, for all $a \geq b$, we have $a \ast b = b \cdot \left(\frac{a}{b} \ast 1\right) = b \cdot \frac{a}{b} = a$.

- So, we have one of the following two cases:
  
  $\geq_1$: for all $a \geq b$, we have $a \ast b = b$;
  
  $\geq_2$: for all $a \geq b$, we have $a \ast b = a$.

- Similarly, by considering $a \leq b$, we conclude that in this case, we also have two possible cases:
  
  $\leq_1$: for all $a \leq b$, we have $a \ast b = b$;
  
  $\leq_2$: for all $a \leq b$, we have $a \ast b = a$. 
15. Proof (conclusion)

• By combining each of the $\geq$ cases with each of the $\leq$ cases, we get the following four combinations:

$\geq_1, \leq_1$: in this case, $a \ast b = b$ for all $a$ and $b$, and therefore, $a_1 \ast \ldots \ast a_n = a_n$;

$\geq_1, \leq_2$: in this case, $a \ast b = \min(a, b)$ for all $a$ and $b$, and therefore,

$$a_1 \ast \ldots \ast a_n = \min(a_1, \ldots, a_n);$$

$\geq_2, \leq_1$: in this case, $a \ast b = \max(a, b)$ for all $a$ and $b$, and therefore,

$$a_1 \ast \ldots \ast a_n = \max(a_1, \ldots, a_n);$$

$\geq_2, \leq_2$: in this case, $a \ast b = a$ for all $a$ and $b$, and therefore, $a_1 \ast \ldots \ast a_n = a_1$.

• The main result is proven.