Peak-End Rule: A Utility-Based Explanation

Olga Kosheleva, Martine Ceberio, and Vladik Kreinovich

University of Texas at El Paso, El Paso, TX 79968, USA
{olgak,mceberio,vladik}@utep.edu

Abstract. In many practical situations, people judge their overall experience by only taking into account the peak and the last levels of pleasantness or unpleasantness. While this peak-end rule is empirically supported by numerous psychological experiments, it seems to contradict our general theoretical ideas about people's preferences. In this paper, we show that, contrary to this impression, the end-peak rule can be justified based on the main ideas of the traditional utility-based decision theory.

1 Peak-End Rule: Description and Need for an Explanation

Peak-end rule: empirical fact. In many situations, people judge their overall experience by the peak and end pleasantness or unpleasantness, i.e., by using only the maximum (minimum) and the last value; see, e.g., [1, 4].

This is true for people's perception of the unpleasantness of a medical procedure, of the quality of the cell phone perception, etc.

Need for an explanation. There is a lot of empirical evidence supporting the peak-end rule, but not much of an understanding. However, at first glance, the rule appears somewhat counter-intuitive: why only peak and last value? why not some average? In this paper, we provide such an explanation based on the traditional decision making theory.

2 Towards an Explanation

Traditional decision making theory: a brief reminder of utility approach. 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* (see, e.g., [2, 3]): a rational agent selects an action with the largest value of expected utility.

Utility is not uniquely defined. 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 $u \to k \cdot u$.

Need for a utility-averaging operation. We want to describe the "average" utility corresponding to a sequence of different experiences. We assume that we know the utility corresponding to each moment of time. To get an average utility value, we need to combine these momentous utilities into a single average.

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 $a*b$ that, given the average utilities a and b corresponding to two consequent time intervals, generates the average utility of the combined two-stage experience.

Natural properties of the utility-averaging operation.

1) If we had the same average utility level $a = b$ on both stages, then this same value should be the two-stage average, i.e., we should have $a * a = a$. In mathematical terms, this means that the utility-averaging operation $*$ should be *idempotent.*

2) If we make one of the stages better, then the resulting average utility should increase (or at least not decrease) as well. In other words, the utility-averaging operation $*$ should be *monotonic* in the sense that if $a \le a'$ and $b \le b'$ then $a * b \le a' * b'$.

3) 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.*

4) For a three-stage situation, with average utilities a, b, and c corresponding to the three stages, we can compute the average utility in two different ways:

- we can first combine the utilities of the first two stages into an average value $a * b$, and then combine this average with c, resulting in $(a * b) * c$;
- alternatively, we can first combine the utilities b and c into $b * c$, and then combine a with $b * c$, resulting in $a * (b * c)$.

The resulting three-stage average should not depend on the order in which we combined the stages, so we should have $(a * b) * c = a * (b * c)$; in mathematical terms, the operation $a * b$ must be *associative.*

5) Finally, since utility is defined modulo scaling, it is reasonable to require that the utility-averaging operation does not change with scaling:

- In the original scale, we combine a and b and get $a * b$. In the new scale corresponding to a factor $k > 0$, this combined value has the form $k \cdot (a * b)$.
- After re-scaling, the original utilities get the new values $a' = k \cdot a$ and $b' = k \cdot b$. Averaging these two values leads to $a' * b' = (k \cdot a) * (k \cdot b)$ in the new scale.

The resulting average should not depend on how we deduced it, i.e., we should have $(k \cdot a) * (k \cdot b) = k \cdot (a * b)$ for all k, a and b.

What we plan to do. Let us show that the above reasonable requirements largely explain the peak-end phenomenon.

3 Main Result

Proposition 1. *Let $a * b$ be a binary operation on the set of all non-negative numbers which satisfies the following properties:*

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

Then, this operation coincides with one of the following four operations:

- $a_1 * \ldots * a_n = \min(a_1, \ldots, a_n)$;
- $a_1 * \ldots * a_n = \max(a_1, \ldots, a_n)$;
- $a_1 * \ldots * a_n = a_1$;
- $a_1 * \ldots * a_n = a_n$.

Comment. Thus, 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, with the only exception that in addition to peak and end, we also have the start $a_1 * \ldots * a_n = a_1$ as one of the options.

Proof.

1°. For every $a \ge 1$, let us denote $a * 1$ by $\varphi(a)$. For $a = 1$, due to the idempotence, $\varphi(1) = 1 * 1 = 1$. Due to monotonicity, $a \le a'$ implies that $\varphi(a) \le \varphi(a')$, i.e., that the function $\varphi(a)$ is (non-strictly) increasing.

2°. Due to associativity, for every a, we have $(a * 1) * 1 = a * (1 * 1)$. Due to idempotence, $1 * 1 = 1$, so the above equality takes the form $(a * 1) * 1 = a * 1$, i.e., the form $\varphi(\varphi(a)) = \varphi(a)$. Thus, for every value t from the range of the function $\varphi(a)$ for $a \ge 1$, we have $\varphi(t) = t$.

3°. Since the operation $a * b$ is continuous, the function $\varphi(a) = a * 1$ is also continuous. Thus, its range $S \stackrel{\text{def}}{=} \varphi([1, \infty))$ for $a \in [1, \infty)$ is a connected set, i.e., an interval (finite or infinite). Since the function $\varphi(a)$ is monotonic, and $\varphi(1) = 1$, this interval must start with 1. So, 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.

3.1°. In the first case, $\varphi(a) = a * 1 = 1$ for all a. From scale invariance, we can now conclude that for all $a \ge b$, we have $a * b = b \cdot \left(\dfrac{a}{b} * 1\right) = b \cdot 1 = b$.

3.2°. In the second case, every value t between 1 and k is a possible value of $\varphi(a)$, thus $\varphi(t) = t * 1 = t$ for all such values t. In particular, for every $\varepsilon > 0$, for the value $t = k - \varepsilon$, we have $\varphi(k - \varepsilon) = k - \varepsilon$. Due to monotonicity, the value

$\varphi(k)$ must be not smaller than all these values $k-\varepsilon$, hence not smaller than k. On the other hand, all the values $\varphi(a)$ are less than or equal than k, so we must have $\varphi(k)=k$ as well. Similarly, for values $t \ge k$, due to monotonicity, we have $\varphi(t) \ge k$ and since always $\varphi(t) \le k$, we conclude that $\varphi(t)=k$ for all $t \ge k$. Now, due to associativity, we have

$$((k-\varepsilon)^2 * (k-\varepsilon)) * 1 = (k-\varepsilon)^2 * ((k-\varepsilon) * 1). \quad (1)$$

Here, due to scale-invariance,

$$(k-\varepsilon)^2 * (k-\varepsilon) = (k-\varepsilon) \cdot ((k-\varepsilon) * 1) = (k-\varepsilon) \cdot \varphi(k-\varepsilon) =$$
$$(k-\varepsilon) \cdot (k-\varepsilon) = (k-\varepsilon)^2, \quad (2)$$

and therefore,

$$((k-\varepsilon)^2 * (k-\varepsilon)) * 1 = (k-\varepsilon)^2 * 1 = \varphi((k-\varepsilon)^2).$$

For $k > 1$, we have $k^2 > k$ and thus, for sufficiently small $\varepsilon > 0$, we have $(k-\varepsilon)^2 > k$. So, $\varphi((k-\varepsilon)^2) = k$, i.e., the left-hand side of the equality (1) is equal to k.

Let us now compute the right-hand side of the equality (1). Here, $(k-\varepsilon)*1 = k-\varepsilon$ and thus, the right-hand side has the form $(k-\varepsilon)^2 * (k-\varepsilon)$ which, as we already know (Equation (2)), is equal to $(k-\varepsilon)^2$. We already know that the left-hand side is equal to k, and that $(k-\varepsilon)^2 > k$. Thus, the equality (1) cannot be satisfied. This proves that the second case is impossible.

3.3°. In the third case, every value $t \ge 1$ is a possible value of $\varphi(a)$, thus

$$\varphi(t) = t * 1 = t$$

for all values $t \ge 1$. Thus, for all $a \ge b$, we have $a * b = b \cdot \left(\frac{a}{b} * 1\right) = b \cdot \frac{a}{b} = a$.

4°. Due to Part 3 of this proof, we have one of the following two cases:

$\ge_1$: for all $a \ge b$, we have $a * b = b$;
$\ge_2$: for all $a \ge b$, we have $a * b = a$.

Similarly, by considering $a \le b$, we conclude that in this case, we also have two possible cases:

$\le_1$: for all $a \le b$, we have $a * b = b$;
$\le_2$: for all $a \le b$, we have $a * b = a$.

By combining each of the $\ge$ cases with each of the $\le$ cases, we get the following four combinations:

$\ge_1, \le_1$: in this case, $a * b = b$ for all a and b, and therefore, $a_1 * \ldots * a_n = a_n$;
$\ge_1, \le_2$: in this case, $a * b = \min(a, b)$ for all a and b, and therefore,

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

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

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

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

The proposition is proven.

Case of negative utilities. The above formula shows how to combine positive experiences. A similar result can be proven for situations in which we need to combine unpleasant experiences, i.e., experience corresponding to negative utilities; the proof of this result is similar to the proof of Proposition 1.

Remaining open problems. 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 $(a + u_0) * (b + u_0) = a * b + u_0$, then we can get a result similar to Proposition 1 for this general case as well. But what if we do not impose this additional requirement?

Are all five conditions in Proposition 1 necessary? Some are necessary:

1) $a * b = a + b$ satisfies all the conditions except for idempotence;
4) $a * b = \dfrac{a + b}{2}$ satisfies all the conditions except for associativity;
5) the operation $a * b$ that returns the value from the interval $[\min(a, b), \max(a, b)]$ which is the closest to 1 satisfies all the conditions except for scale invariance.

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

Comment. In analyzing the need for these conditions, it may help to know that the set $\{z : z * 1 = z\}$ is a semigroup: indeed, if $z_1 * 1 = z_1$ and $z_2 * 1 = z_2$, then $(z_1 \cdot z_2) * (z_1 * 1) = (z_1 \cdot z_2) * z_1 = z_1 \cdot (z_2 * 1) = z_1 \cdot z_2$ and $((z_1 \cdot z_2) * z_1) * 1 = (z_1 \cdot z_2) * 1$, so associativity implies that $(z_1 \cdot z_2) * 1 = z_1 \cdot z_2$.

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.

References

1. Kahneman, D.: Objective happiness, in: Kahneman, D., Diener, E., and Schwarz, N. (eds.), Well-Being: The Foundations of Hedonistic Psychology, Russell Sage, New York (1999), pp. 3–25.
2. Luce, R. D., Raiffa, R.: Games and Decisions: Introduction and Critical Survey, Dover, New York (1989)
3. Raiffa, H.: Decision Analysis, Addison-Wesley, Reading, Massachusetts (1970)
4. McFadden, D. L.: The New Science of Pleasure, National Bureau of Economic Research, Working Paper 18687 (2013)
5. Bishop, E., Bridges, D. S.: Constructive Analysis, Springer, N.Y. (1985)