Why Under Stress Positive Reinforcement Is More Effective? Why Optimists Study Better? Why People Become Restless? Simple Utility-Based Explanations

Francisco Zapata[1], Olga Kosheleva[2], and Vladik Kreinovich[3]
[1]Department of Industrial, Manufacturing, and Systems Engineering
[2]Department of Teacher Education
[3]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 7998, USA
fazg74@gmail.com, olgak@utep.edu, vladik@utep.edu

Abstract

In this paper, we use the utility-based approach to decision making to provide simple answers to the following three questions: Why under stress positive reinforcement is more effective? Why optimists study better? Why people become restless?

1 Why Under Stress Positive Reinforcement Is More Effective?

Phenomenon. To encourage a person to do something, we can use both positive reinforcement – when we reward a person for doing this, and negative reinforcement – when we penalize a person for not doing the task. Both approaches have their strengths and limitations.

It has been observed that in stress situations, when a person's mood is negative, the relative strength of positive reinforcement increases; see, e.g., [11, 21]. Why?

Let us formulate this problem in precise terms. In traditional decision theory, human preferences are described by *utilities*; see, e.g., see, e.g., [2, 3, 14, 17, 19]. A utility is defined as follows: we select a very good situation A_1 and a very bad situation A_0 and then compare each situation A with the lottery $L(p)$ in which:

- we get A_1 with probability p and
- we get A_0 with the remaining probability $1 - p$.

For small p, $L(p)$ is close to A_0 and is, thus, much worse than A: $L(p) < A$. For p close to 1, $L(p)$ is close to A_1 and is, thus, mush better than A: $A < L(p)$

There is therefore a threshold value p_0 such that:

- for $p > p_0$, we have $A < L(p)$, while
- for $p < p_0$, we have $l(p) < A$.

This threshold value – for which A is (in this sense) equivalent to $A(p_0)$ – is called the *utility* $u(a)$ of the alternative A.

If $p < p'$, then, of course, $L(p')$ is better than $L(p)$. Thus, among several alternatives, we should select a one for which the utility $u(a)$ is the largest.

It is known that the utility of monetary rewards or losses is approximately proportional to the square root of the amount m of money:

- $u(m) = a_+ \cdot \sqrt{m}$ for $m \geq 0$ and
- $u(m) = -a_- \cdot \sqrt{|m|}$ for $m < 0$,

for some values a_+ and a_-; see, e.g., [6, 12, 13].

We can measure the relative strength of positive and negative reinforcement by comparing the changes in utility if we add or subtract a certain amount of money m.

If we start with a neutral situation, in which we have no money, then the original utility value is 0. Then, after adding the amount m we get the utility $a_+ \cdot \sqrt{m}$, while after subtracting amount m, we lose the utility amount $a_- \cdot \sqrt{m}$. In this case, the ration of positive-to-negative reinforcement effects is

$$\frac{a_+ \cdot \sqrt{m}}{a_- \cdot \sqrt{m}} = \frac{a_+}{a_-}. \quad (1)$$

What if we start with a stressful situation, in which the initial amount of money is small but negative: $-m_0 < 0$? In this case, the initial value of the utility is $a_- \cdot \sqrt{m_0}$. After adding m, we get $m - m_0$, with the utility $a_+ \cdot \sqrt{m - m_0}$. Thus, the utility gain is $a_+ \cdot \sqrt{m - m_0} + a_- \cdot \sqrt{m_0}$.

If we subtract the money amount m, the we end up with the negative amount $-(m + m_0)$, whose utility is $-a_- \cdot \sqrt{m + m_0}$. Thus, the loss of utility is the difference $a_- \cdot \sqrt{m + m_0} - a_- \cdot \sqrt{m_0}$.

Thus, the ratio describing the relative strength of possible reinforcement takes the form

$$\frac{a_+ \cdot \sqrt{m - m_0} + a_- \cdot \sqrt{m_0}}{a_- \cdot \sqrt{m + m_0} - a_- \cdot \sqrt{m_0}}. \quad (2)$$

Our explanation. We will show that the ratio (2) is larger than the ratio (1). This explains the empirical fact that under stress, positive reinforcement is more efficient.

Indeed, for small x_0, by taking the first two terms of the corresponding Taylor series, we get

$$\sqrt{m - m_0} = \sqrt{m} - \frac{1}{2\sqrt{m}} \cdot m_0 + o(m_0).$$

For small m_0, we have $\sqrt{m_0} \gg m_0$, thus in the ifrst approximation, we can ignore the terms proportional to m_0 and only consider terms proportional to $\sqrt{m_0}$. So, the numerator of the ratio (2) takes the form

$$a_+ \cdot \sqrt{m - x_0} + a_- \cdot \sqrt{m_0} \approx a_+ \cdot \sqrt{m} + a_- \cdot \sqrt{m_0}.$$

Similarly, we have

$$\sqrt{m + m_0} = \sqrt{m} + \frac{1}{2\sqrt{m}} \cdot m_0 + o(m_0)$$

and thus, in the first approximation, the denominator of the formula (2) takes the form

$$a_- \cdot \sqrt{m + m_0} - a_- \cdot \sqrt{m_0} \approx a_- \cdot \sqrt{m} - a_- \cdot \sqrt{m_0}.$$

Thus. in the first approximation, the ratio (2) has the form

$$\frac{a_+ \cdot \sqrt{m} + a_- \cdot \sqrt{m_0}}{a_- \cdot \sqrt{m} - a_- \cdot \sqrt{m_0}}. \tag{3}$$

We can see that, in comparison to the ratio (1), we increased the numerator and decreased the denominator – as a result, the ratio increases.

This is exactly what we wanted to explain.

Auxiliary analysis: beyond explanation. A natural question is: what if instead of considering stress, we consider euphoria, i.e., we consider situations in which we have a positive initial amount of money m_0. How will this affect the relative strength of positive and negative reinforcements?

In this case, we start with the utility $a_+ \cdot \sqrt{m_0}$. When we add the amount m, we get the utility $a_+ \cdot \sqrt{m + m_0}$, so the increase in utility is equal to

$$a_+ \cdot \sqrt{m + m_0} - a_+ \cdot \sqrt{m_0}.$$

Vice versa, if we take away the amount m, we get the new utility $-a_- \cdot \sqrt{m - m_0}$, so the loss in utility is equal to

$$a_- \cdot \sqrt{m - m_0} - a_+ \cdot \sqrt{m_0}.$$

In this situation, the ratio describing the relative strength of positive and negative reinforcements takes the form

$$\frac{a_+ \cdot \sqrt{m + m_0} - a_+ \cdot \sqrt{m_0}}{a_- \cdot \sqrt{m - m_0} - a_+ \cdot \sqrt{m_0}}. \tag{4}$$

Similarly to the stress case, in the first approximation, the numerator is approximately equal to $a_+ \cdot \sqrt{m} - a_+ \cdot \sqrt{m_0}$, while the denominator is approximately equal to $a_- \cdot \sqrt{m} - a_+ \cdot \sqrt{m_0}$. Thus, in the first approximation, the ratio (4) takes the form

$$\frac{a_+ \cdot \sqrt{m} - a_+ \cdot \sqrt{m_0}}{a_- \cdot \sqrt{m} - a_+ \cdot \sqrt{m_0}}.$$

We can see that, in comparison to the ratio (1), we decreased the numerator and increased the denominator – as a result, the ratio decreases.

Thus, we conclude that in happy situations, negative reinforcements are more efficient that the positive ones (but do not tell that to your bosses :-).

2 Why Optimists Study Better?

Empirical fact. It is a known fact that optimists study better; see, e.g., [22] and references therein.

Let us describe this situation in precise terms. What does optimism mean in precise terms?

According to the traditional decision theory, if for each possible alternative a, we know the probabilities $p_i(a)$ and utilities $u_i(a)$ of different outcomes i, then a rational person should select an alternative a for which the value $u(a) \stackrel{\text{def}}{=} \sum_i p_i(a) \cdot u_i(a)$ is the largest [2, 3, 14, 17, 19]. In such situations, there is only one rational choice, there is no possibility to show optimism or pessimism.

In practice, however, we rarely know the exact probability and the exact utility of different outcomes. Usually, we only know the bounds $\underline{p}_i(a)$, $\overline{p}_i(a)$, $\underline{u}_i(a)$, and $\overline{u}_i(a)$ on possible values of $p_i(a)$ and $u_i(a)$: $\underline{p}_i(a) \le p_i(a) \le \overline{p}_i(a)$ and $\underline{u}_i(a) \le u_i(a) \le \overline{u}_i(a)$. For different values of $p_i(a)$ and $u_i(a)$ from the corresponding intervals, we get different values of the overall utility $u(a)$. Thus, instead of a single value $u(a)$, we have an *interval* $[\underline{u}(a), \overline{u}(a)]$ of possible values. How should we make decisions if for each alternative a, we know such an interval $[\underline{u}(a), \overline{u}(a)]$?

Reasonable requirements on rationality of a decision maker lead to the following solution (first proposed by the future Nobel Prize winner Leo Hurwicz): we should select a number $\alpha \in [0, 1]$ and select an alternative for which the combination $\alpha \cdot \overline{u}(a) + (1 - \alpha) \cdot \underline{u}(a)$ is the largest possible; see, e.g., [5, 10, 14].

When $\alpha = 1$, this means that when making a decision, we only take into account the most favorable situation, when the utility $u(a)$ attains its largest possible value $\overline{u}(a)$. This is clearly the case of extreme optimism.

When $\alpha = 0$, this means that when making a decision, we only take into account the least favorable situation, when the utility $u(a)$ attains its smallest possible value $\underline{u}(a)$. This is clearly the case of extreme pessimism.

Values α intermediate between 0 and 1 describe realistic decision makers. The larger α, the more the decision maker takes into account the most optimistic scenario and the less he/she takes into account the most pessimistic scenarios.

Thus, the value α can serve as a quantitative measure of the decision maker's optimism: the larger α, the more optimistic the decision maker.

This explains why optimistic study better. In education, we invest some efforts now and get rewards in the future. Time for studying is taken from time of having fun: we have less chances to go to a movie, to watch TV. etc. So, in comparison with not studying, this part of the learning process brings less positive utility.

We do study, because we know that there will be a future reward: better knowledge, better job, etc. So, when deciding how much time we dedicate to studying (or whether to study at all), we take into account both the utility decrease now and the potential utility increase in the future.

The decrease now is clear, we thus know the value $u_d < 0$. About the future rewards, we are not 100% certain: now there is a demand for tour major, who knows what will happen four years from now, when we graduate with a degree? Thus, for future rewards, instead of the exact value u_r, we only know the interval of possible values $u_r \in [\underline{u}_r, \overline{u}_r]$. The overall utility therefore takes all possible values form $\underline{u} = u_d + \underline{u}_r$ to $\overline{u} = u_d + \overline{u}_r$. A person with an optimism value α selects to study if the Hurwicz combination $\alpha \cdot u + (1 - \alpha) \cdot \underline{u}$ is larger than the value 0 corresponding to not studying.

Here, as one can easily check, the Hurwicz combination is equal to

$$u_d + \alpha \cdot \overline{u}_r + (1 - \alpha) \cdot \underline{u}_r = u_d + \underline{u}_r + \alpha \cdot (\overline{u}_r - \underline{u}_r).$$

This value increases with α. If this value was larger that 0 for some α, it will be still larger than 0 for $\alpha' > \alpha$ – and for $\alpha' > \alpha$, in some situations when the Hurwicz value was negative, it may becomes positive.

Thus, the larger the level α of a person's optimism, the more there are situations in which this person will start studying. This explains why optimists are better students.

3 Why People Become Restless?

Phenomenon. When a person's salary is increased, this person becomes happy. If a few years pass and the salary remains the same, then, while objectively, the person has the same good life as before, he or she becomes restless, unhappy. Why?

The situation is the same as in the past years, so why is not level of happiness the same?

Towards an explanation. It is known that our utility depends not only on what we have now, it also depends on what we expect in the future: otherwise, we would act without thinking of possible consequences. The expected future values of utility come with some discounting, usually, the exponential discounting, when – just like when you invest money in a bank – the utility T moments in the future gets multiply by β^T for some $\beta < 1$; see, e.g., [1, 4, 7, 8, 9, 15, 16, 18, 20]. (For the bank, β is 1 minus interest; e.g., if the interest rate is 3%, $\beta = 0.97$.) As a result, if o_t is the utility caused by the current situation at moment t, the actual utility u_t at moment t is equal to

$$u_t = o_t + \alpha \cdot o_{t+1} + \alpha^2 \cdot o_{t+2} + \ldots$$

We do not know the future values, we get them by extrapolation, based on the previous several values o_t, o_{t-1}, ...

The simplest possible extrapolation is linear extrapolation which is based on the last two values o_t and o_{t-1}. Here, $o_{t+j} = o_t + j \cdot (o_t - o_{t-1})$. In the year t in which a salary got increased, the difference is positive, so $o_{t+1} > o_t$, $o_{t+2} > o_t$, etc., hence

$$u_t = o_t + \alpha \cdot o_{t+1} + \alpha^2 \cdot o_{t+2} + \ldots > o_t + \alpha \cdot o_t + \alpha^2 \cdot o_t + \ldots = o_t \cdot (1 + \alpha + \alpha^2 + \ldots).$$

A few years later, when $o_t = o_{t-1}$, all extrapolated values are the same: $o_{t+1} = o_{t+2} = \ldots = o_t$, thus

$$u_t = o_t + \alpha \cdot o_t + \alpha^2 \cdot o_t + \ldots = o_t \cdot (1 + \alpha + \alpha^2 + \ldots).$$

We see that the utility in the first year is indeed larger than the utility a few years after – this is exactly what we observe.

Acknowledgments

This work was supported in part by the National Science Foundation grant HRD-1242122 (Cyber-ShARE Center of Excellence).

References

[1] T. S. Critchfield and S. H. Kollins, "Temporal discounting: basic research and the analysis of socially important behavior", *Journal of Applied Bheavior Analysis*, 2001, Vol. 34, pp. 101–122.

[2] P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.

[3] P. C. Fishburn, *Nonlinear Preference and Utility Theory*, The John Hopkins Press, Baltimore, Maryland, 1988.

[4] S. Frederick, G. Loewenstein, and T. O'Donoghue, "Time discounting: a critical review", *Journal of Economic Literature*, 2002, Vol. 40, pp. 351–401.

[5] L. Hurwicz, *Optimality Criteria for Decision Making Under Ignorance*, Cowles Commission Discussion Paper, Statistics, No. 370, 1951.

[6] D. Kahneman, *THinking Fast and Slow*, Farrar, Straus, and Girous, New York, 2011.

[7] G. R. King. A. W. Logue, and D. Gleiser, "Probability and delay in reinforcement: an examination of Mazur's equivalence rule", *Behavioural Processes*, 1992, Vol. 27, pp. 125–138.

[8] K. N. Kirby, "Bidding on the future: evidence against normative discounting of delayed rewards", *Journal of Experimental Psychology (General)*, 1997, Vol. 126, pp. 54–70.

[9] C. J. Konig and M. Kleinmann, "Deadline rush: a time management phenomenon and its mathematical description", *The Journal of Psychology*, 2005, Vol. 139, No. 1, pp. 33–45.

[10] V. Kreinovich, "Decision making under interval uncertainty (and beyond)", In: P. Guo and W. Pedrycz (eds.), *Human-Centric Decision-Making Models for Social Sciences*, Springer Verlag, 2014, pp. 163–193.

[11] N. R. Lighthall, M. A. Gorlick, A. Schoeke, M. J. Frank, and M. Mather, "Stress modulates reinforcement learning in younger and older adults", *Psychology of Aging*, 2013, Vol. 28, No. 1, pp. 35–46.

[12] J. Lorkowski and V. Kreinovich, "Granularity helps explain seemingly irreation features of human decision making", In: W. Pedrycz and S.-M. Chen (eds.), *Granular Computing and Decision-Making: Interactive and Iterative Approaches*, Springer Verlag, Cham, Switzerland, 2015, pp. 1–31.

[13] J. Lorkowski and V. Kreinovich, *Bounded Rationality in Decision Making Under Uncertainty: Towards Optimal Granularity*, Springer Verlag, Cham, Switzerland, 2018.

[14] R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

[15] J. E. Mazur, "An adjustment procedure for studying delayed reinforcement", In: M. L. Commons, J. E. Mazur, J. A. Nevin, and H. Rachlin (eds.), *Quantitative Analyses of Behavior. Vol. 5, The Effect of Delay and Intervening Events*, Erlbaum, Hillsdale, 1987.

[16] J. E. Mazur, "Choice, delay, probability, and conditional reinforcement", *Animal Learning Behavior*, 1997, Vol. 25, pp. 131–147.

[17] H. T. Nguyen, O. Kosheleva, and V. Kreinovich, "Decision making beyond Arrows 'impossibility theorem', with the analysis of effects of collusion and mutual attraction", *International Journal of Intelligent Systems*, 2009, Vol. 24, No. 1, pp. 27–47.

[18] H. Rachlin, A. Raineri, and D. Cross, "Subjective probability and delay", *Journal of the Experimental Analysis of Behavior*, 1991, Vol. 55, pp. 233–244.

[19] H. Raiffa, *Decision Analysis*, Addison-Wesley, Reading, Massachusetts, 1970.

[20] F. Zapata, O. Kosheleva, V. Kreinovich, and T. Dumrongpokaphan, "Do it today or do it tomorrow: empirical non-exponential discounting explained by symmetry ideas", In: V.-N. Huynh, M. Inuiguchi, D.-H. Tran, and T. Denoeux (eds.), *Proceedings of the International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making IUKM'2018*, Hanoi, Vietnam, March 13–15, 2018.

[21] K. Zimmer, 'Tough choices", *The Scientist*, 2018, Vol. 32, No. 3, pp. 20–21.

[22] M. Seligman, *The Optimistic Child: A Proven Program to Safeguard Children Against Depression and Build Lifelong Resilience*, Martiner Books, new York, 2007.