People Prefer More Information About Uncertainty, But Perform Worse When Given This Information: An Explanation of the Paradoxical Phenomenon

Jieqiong Zhao[1], Olga Kosheleva[2], and Vladik Kreinovich[3]
School of Computing and Augmented Intelligence, Arizona State University
699 S Mill Ave, Tempe, AZ 85281, USA, jieqiong.zhao@asu.edu
[2,3]Departments of [2]Teacher Education, [3]Computer Science
University of Texas at El Paso, El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu

1. When making a decision, it is desirable to have as much information as possible

- To make a decision, it is desirable to have as much information about the decision making situation as possible.
- Indeed, additional information can help make a better decision.
- This desirability has been confirmed by many polls.

2. This includes the need for more information about uncertainty

- In complex situations, to make decisions, we usually use computers.
- In general, computers process numbers.
- Thus, the information about a situation usually consists of several numbers, e.g., the values of the corresponding physical quantities.
- These values usually come from measurements.
- It could be direct measurement or so-called indirect measurement, i.e., processing of measurement results.
- Measurements are never absolutely accurate.
- There is always a non-zero difference $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$ between:
  - the measurement result $\widetilde{x}$ and
  - the actual (unknown) value x of the corresponding quantity.
- This difference is known as the *measurement error*.

3. This includes the need for more information about uncertainty (cont-d)

- It is desirable to get as much information as possible.
- So, it is desirable to provide the users not only the measurement results, but also with the information about the measurement error.
- The desirability of this information was indeed confirmed by the polls.

4. What do we know about the measurement errors: probabilistic and interval uncertainty

- Ideally, we should know:
  - what are the possible values of the measurement error, and
  - what is the frequency (probability) with which each of these values will appear in the actual measurements.
- In other words, ideally, we should know the probability distribution of the measurement errors.
- To determine this probability distribution, we need to *calibrate* the measuring instrument, i.e., to compare:
  - the values $\widetilde{x}_k$ measured by this instrument and
  - the values x'_k measured (for the same actual quantity) by a much more accurate measuring instrument (known as *standard*).

5. What do we know about the measurement errors: probabilistic and interval uncertainty (cont-d)

- The standard measuring instrument (SMI) is so much more accurate that:
  - in comparison with the measurement errors of the tested measuring instrument (MI),
  - we can safely ignore the measurement errors of the SMI.
- Thus, we can assume that SMI's measurement results x'_k are equal to the corresponding actual values x_k.
- So, the differences $\widetilde{x}_k - x'_k$ can be safely assumed to equal to the values of the measurement errors $\widetilde{x}_k - x_k$.
- Hence, from the resulting sample of these differences, we can determine the desired probability distribution.
- This procedure is rather time-consuming and expensive – since the use of the complex standard measuring instruments is not cheap.

6. What do we know about the measurement errors: probabilistic and interval uncertainty (cont-d)

- As a result, in practice, such a detailed calibration is often not performed.
- Instead, the manufacturer of the measuring instruments provides a guaranteed upper bound Δ on $|\Delta x|$: $|\Delta x| \le \Delta$.
- Under this information:
  - after we perform the measurement and get the measurement result $\widetilde{x}$,
  - the only information that we have about the actual (unknown) value x of the measured quantity is that $x \in [\underline{x}, \overline{x}] \stackrel{\text{def}}{=} [\widetilde{x}-\Delta, \widetilde{x}+\Delta]$.
- This uncertainty is known as *interval uncertainty*.

7. Probability distribution of the measurement error is often Gaussian

- In many practical cases, the measurement error is caused by the large number of small independent factors.
- It is known that the probability distribution of the joint effect of many independent small factors is close to Gaussian.
- The corresponding mathematical result is known as the Central Limit Theorem.
- Thus, we can safely assume that the measurement error is distributed according to the normal (Gaussian) distribution.
- This assumption is in good accordance with the empirical data, according to which:
  - in the majority of the cases,
  - the distribution is indeed Gaussian.

8. Probability distribution of the measurement error is often Gaussian (cont-d)

- The Gaussian distribution is uniquely determined by two parameters: mean μ and standard deviation σ.
- By comparing the measurement results with the results of a standard measuring instrument:
  - we can determine the bias
  - as the arithmetic average of the observed differences between the measurements by two instruments: $\mu \approx \frac{1}{n} \cdot \sum_{k=1}^{n} (\widetilde{x}_k - x'_k)$.
- Once we know μ, we can recalibrate the scale – i.e., subtract this mean value μ from each measurement result.
- In other words, we replace each value $\widetilde{x}$ with $\widetilde{x}' \stackrel{\text{def}}{=} \widetilde{x} - \mu$.
- After that, the mean will be 0.

9. Probability distribution of the measurement error is often Gaussian (cont-d)

- Also, we can determine the standard deviation as the mean squared value of the (re-calibrated) difference

$$\sigma \approx \sqrt{\frac{1}{n} \cdot \sum_{k=1}^{n} (\widetilde{x}'_k - x'_k)^2}.$$

- So, it makes sense to assume that the measurement error is normally distributed with mean 0 and the known standard deviation σ.

10. Confidence intervals

- It is known that with high probability, all the values of the normally distributed random variable are located within the interval

$$[\mu - k_o \cdot \sigma, \mu + k_0 \cdot \sigma] \text{ for an appropriate } k_0 :$$

  - for $k_0 = 2$, this is true with probability 95%,
  - for $k_0 = 3$, this is true with probability 99.9%, and
  - for $k_0 = 6$, this is true with probability $1 - 10^{-8}$.

- These intervals are particular cases of *confidence intervals*.
- These intervals is what is often supplied to the users as a partial information about the probability distributions.

11. What happened when decision maker received confidence intervals: paradoxical situation

- The decisions makers expressed the desire to receive information about uncertainty.
- So, they were supplied, in addition to measurement results, with the corresponding confidence intervals.
- And here, a strange thing happened.
- One would expect that this additional information would:
  - help the decision makers make better decisions,
  - or at least did not degrade the quality of their decisions.

12. What happened when decision maker received confidence intervals: paradoxical situation (cont-d)

- However, in reality:
  - the decisions of those users who got this additional information were worse than
  - the decisions make by users from the control group – that did not receive this information.
- How can we explain this paradoxical phenomenon?

13. Preliminaries

- To understand the above paradoxical phenomenon, let us recall how rational people make decisions.
- Let us start with the case of full information, when we know all the probabilities.
- According to decision theory, preferences of each rational person can be described by assigning:
  - to each alternative x,
  - a numerical value $u(x)$ called its *utility*.
- To assign these numerical values, one need to select two alternatives:
  - a very good alternative A_+ which is better than anything that the decision maker will actually encounter, and
  - a very bad alternative A_- which is worse than anything that the decision maker will actually encounter.

14. Preliminaries (cont-d)

- Once these two alternatives are selected, we can form, for each value p from the interval $[0, 1]$, a lottery $L(p)$ in which:
  - we get A_+ with probability p, and
  - we get A_- with the remaining probability $1 - p$.
- For any actual alternative x, when p is close to 0, the lottery $L(p)$ is close to A_- and is, thus, worse than x: $A_- < x$.
- When p is close to 1, the lottery $L(p)$ is close to A_+ and is, thus, better than x: $x < A_+$.
- When we move from 0 to 1, there is a threshold value at which the relation $L(p) < x$ is replaced with $x < L(p)$.
- This threshold value $u(x)$ is what is called the utility of x.

15. Preliminaries (cont-d)

- By definition of utility, each alternative x is equivalent, to the decision maker, to a lottery $L(u(x))$ in which:
  - we get A_+ with probability $u(x)$, and
  - we get A_- with the remaining probability $1 - u(x)$.
- In general, the larger the probability p of getting the very good alternative A_+, the better the lottery.
- Thus, if we need to select between several lotteries of this type, we should select the lottery with the largest values of the probability p.
- Each alternative x is equal to the lottery $L(p)$ with probability $p = u(x)$.
- This means that we should always select the alternative with the largest value of utility.

16. Preliminaries (cont-d)

- What is the utility of an action in which we get n possible outcomes, for each of which we know:
  - its probability p_i and
  - its utility u_i?
- By definition of utility, each outcome is equivalent to a lottery $L(u_i)$ in which:
  - we get A_+ with probability u_i, and
  - we get A_- with probability $1 - u_i$.
- Thus, to the user, the action is equivalent to a 2-stage lottery in which:
  - we first select each i with probability p_i and
  - then, depending on what i we selected on the first stage, select A_+ with probability u_i and A_- with probability $1 - u_i$.
- As a result of this two-stage lottery, we get either A_+ or A_-.

17. Preliminaries (cont-d)

- The probability u of getting A_+ is determined by the formula of full probability: $u = p_1 \cdot u_1 + \ldots + p_n \cdot u_n$.
- Thus, by definition of utility, this value u is the utility of the action under consideration.
- The right-hand side of the formula for u is the expected value of the utility; so, we can conclude that:
  - the utility of an action with random consequences
  - is equal to the expected value of the utility of different consequences.

18. Comment

- The numerical value of the utility depends on the selection of the alternatives A_+ and A_-.
- If we select a different pair A'_+ and A'_-, then we will have different numerical values $u'(x)$.
- It turns out that for every two pairs of alternatives, there exist real numbers $a > 0$ and b for which, for each x, we have $u'(x) = a \cdot u(x) + b$.
- In other words, utility is defined modulo a linear transformation.
- This is similar to the fact that, e.g., the numerical value of the moment of time also depends:
  - on what starting point we use to measure time, and on what measuring unit we use.
- For time, all the scales are related to each other by an appropriate linear transformation $t' = a \cdot t + b$ for some $a > 0$ and b.

19. How rational people make decisions under interval uncertainty

- As we have mentioned:
  - in many practical situations,
  - we only know the values of the quantities (that describe the state of the world) with interval uncertainty.
- In this case, we can describe the consequences – and their utility – also under interval uncertainty.
- In other words, for each possible decision x:
  - instead of the exact value $u(x)$ of the corresponding utility,
  - we only know the interval $[\underline{u}(x), \overline{u}(x)]$ of possible utility values.
- How can we make a decision in this case?
- To make a decision, we need to assign, to each interval $[\underline{u}, \overline{u}]$, an equivalent numerical value $u(\underline{u}, \overline{u})$.

20. How rational people make decisions under interval uncertainty (cont-d)

- As we have mentioned, utility is defined modular a linear transformation.
- There is no fixed selection of the alternatives A_+ and A_-.
- So it makes sense to require that the function $u(\underline{u}, \overline{u})$ remains the same for all the scales, i.e., that:
  - if $u = u(\underline{u}, \overline{u})$, then
  - for all $a > 0$ and b, we should have $u' = u(\underline{u}', \overline{u}')$, where $u' = a \cdot u + b$, $\underline{u}' = a \cdot \underline{u} + b$, and $\overline{u}' = a \cdot \overline{u} + b$.
- Let us denote $\alpha \stackrel{\text{def}}{=} u(0, 1)$.
- If we know that the utility is between 0 and 1, then the situation is:
  - clearly better (or at least as good) than when utility is 0, and
  - worse (or at least as good) then when utility is 1.

21. How rational people make decisions under interval uncertainty (cont-d)

- Thus, we must have $0 \le \alpha \le 1$.
- Every interval $[\underline{u}, \overline{u}]$ can be obtained from the interval $[0, 1]$ by a linear transformation $u \mapsto a \cdot t + b$ for $a = \overline{u} - \underline{u}$ and $b = \underline{u}$.
- Thus, due to invariance, from $\alpha = u(0, 1)$, we can conclude that

$$u(\underline{u}, \overline{u}) = a \cdot \alpha + b = \alpha \cdot (\overline{u} - \underline{u}) + \underline{u} = \alpha \cdot \overline{u} + (1 - \alpha) \cdot \underline{u}.$$

- This formula was first proposed by the Nobelist Leo Hurwicz and is thus known as *Hurwicz optimism-pessimism criterion.*

22. How rational people make decisions under interval uncertainty (cont-d)

- The name comes from the fact that:
  - for $\alpha = 1$, this means only taking into account the best-case value $\overline{u}$ – the case of extreme optimism, while
  - for $\alpha = 0$, this means only taking into account the worst-case value $\underline{u}$ – the case of extreme pessimism.
- The value α is different for different decision makers, depending on their level of optimism and pessimism.

23. Finally, an explanation

- Now we are ready to produce the desired explanation.
- Let us consider the simplest possible setting, when the decision maker is provided with the information about each possible decision x:
  - the estimate $\widetilde{u}(x)$ of the utility $u(x)$, and
  - the information that the difference $\Delta u(x) \stackrel{\text{def}}{=} \widetilde{u}(x) - u(x)$ is normally distributed with 0 mean and standard deviation σ.
- In this *ideal* case:
  - the decision maker's equivalent utility of each possible decision x
  - is equal to the expected value of the random utility value $u(x) = \widetilde{u}(x) - \Delta u(x)$, i.e., to the value $\widetilde{u}(x)$.
- This was the ideal case.
- In our experiment, we never report the whole distribution to the decision maker.

24. Finally, an explanation (cont-d)

- Instead, we report either a single value $\widetilde{u}(x)$ or the confidence interval
$$[\widetilde{u}(x) - k_0 \cdot \sigma, \widetilde{u}(x) + k_0 \cdot \sigma].$$
- In the first case, when we supply no information about uncertainty, the decision maker uses the provided value $\widetilde{u}(x)$ in his/her decisions.
- It so happens that this value is exactly what we would get if we knew the exact distributions.
- In other words, in this case, the decision maker makes an optimal decision.
- On the other hand:
  - if we provide the decision maker with the confidence interval,
  - then the decision maker – using Hurwicz criterion – will assign, to each possible decision x, an equivalent value
$$\alpha \cdot (\widetilde{u}(x) + k_0 \cdot \sigma) + (1 - \alpha) \cdot (\widetilde{u}(x) - k_0 \cdot \sigma) = \widetilde{u}(x) + (2\alpha - 1) \cdot k_0 \cdot \sigma.$$

25. Finally, an explanation (cont-d)

- Thus:
  - for almost all possible values α from the interval $[0, 1]$ – with the only exception of the value $\alpha = 0.5$,
  - this value will be different from the optimal value (corresponding to the full information).
- So, the decision based on such values will be not as good as the optimal decision.
- This is exactly what we observe in the above-described seemingly paradoxical experiment.
- Note that the worsening of the decision happens when we provide the decision make with *partial* information about uncertainty.
- If we provide the decision maker with full information, the decision will, of course, be optimal.

26. Acknowledgments

This work was supported in part by:

- National Science Foundation grants 1623190, HRD-1834620, HRD-2034030, and EAR-2225395;
- AT&T Fellowship in Information Technology;
- program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and
- a grant from the Hungarian National Research, Development and Innovation Office (NRDI).