Acting Irrationally to Improve Performance in Stochastic Worlds

Roman V. Belavkin (r.belavkin@mdx.ac.uk)

School of Computing Science,
Middlesex University, London NW4 4BT, UK

14 December 2005

OVERVIEW

1. Motivation of this work
2. Critique of the rational d.m. theory
3. Irrational decision–making
4. The architecture
5. Experimental results
6. Discussion

MOTIVATION

- Humans and animals always express some degree of randomness in their choice behaviour (Myers, Fort, Katz, & Suydam, 1963)
- Cognitive architectures add noise into utility to model the ‘irrational’ component (e.g. ACT–R, Anderson & Lebiere, 1998)
- Noise seems to play an important role optimising the behaviour (Belavkin & Ritter, 2003)
- The expected utility theory leads to many unexplained paradoxes.

EXPECTED UTILITY THEORY

- The classical decision–making theory is due to Bernoulli (1738/1954), von Neumann and Morgenstern (1944), Savage (1954) and Anscombe and Aumann (1963).
- The central idea is to represent preferences by some *utility* function $u : X \to \mathbb{R}$

$$x \succ y \iff u(x) > u(y)\,,$$

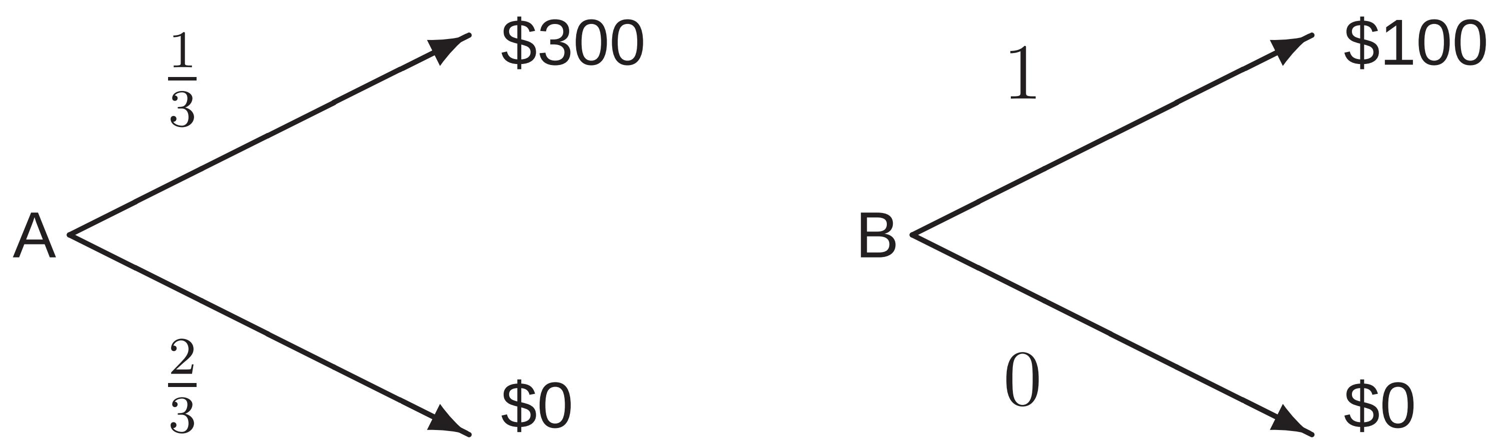
- Under uncertainty, the *expected utilities* (EUs) are considered

$$p \succ q \iff \sum_{z \in Z} p(z)\,u(z) > \sum_{z \in Z} q(z)\,u(z)\,,$$

where Z is a set of prizes, P a set of probability measures.

THE ALLAIS PARADOX

Due to Allais (1953). Also studied by Tversky and Kahneman (1974) in many interpretations. Consider two lotteries A and B



$$\frac{1}{3} \cdot \$300 + \frac{2}{3} \cdot \$0 = \$100 \qquad 1 \cdot \$100 + 0 \cdot \$0 = \$100$$

About 80% of subjects express preference $A \prec B$.

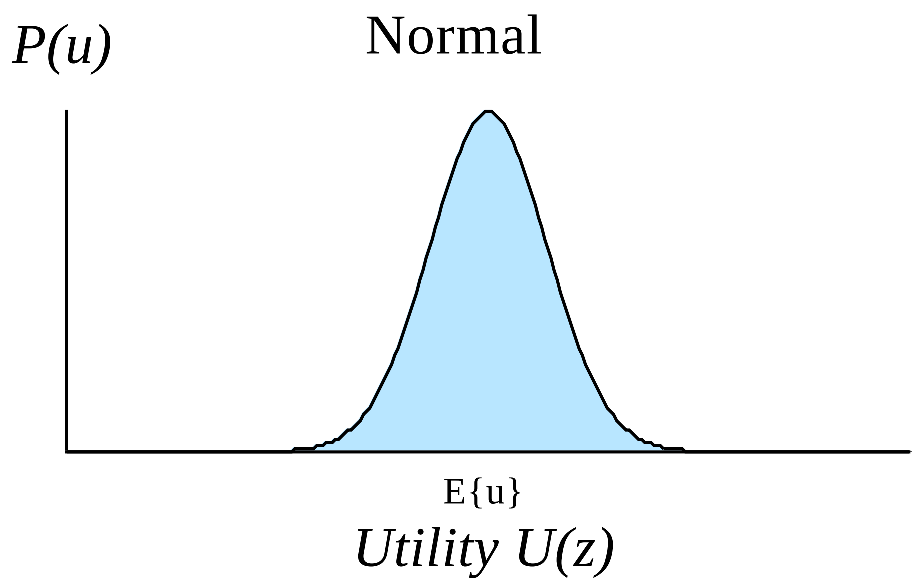
Professional traders behave this way too (List & Haigh, 2005).

THE MAX EU PRINCIPLE

- Many paradoxes occur when we try to apply

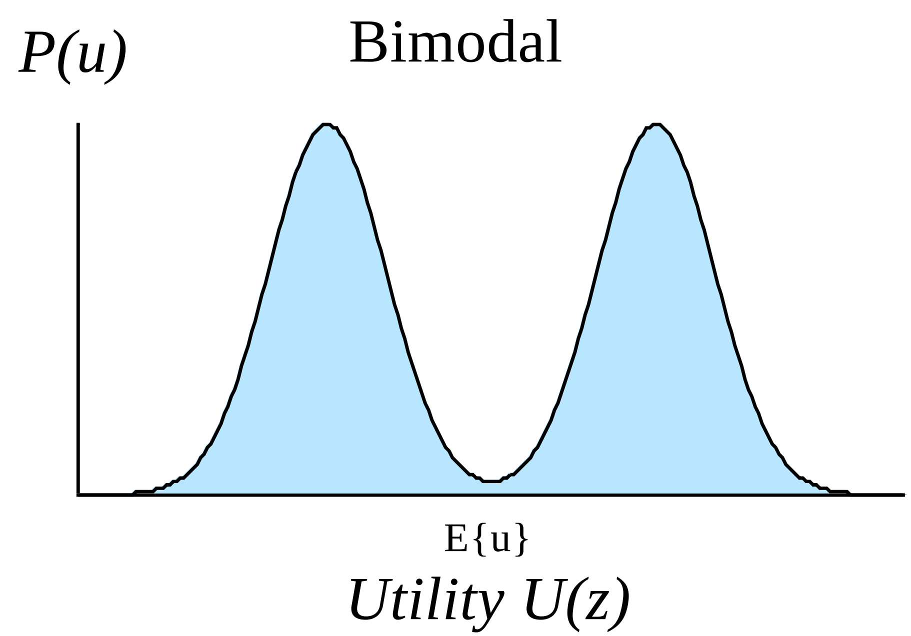
$$x = \arg\max_{x \in X} \sum_{z \in Z} p(z \mid x) u(z)$$

- For Gaussian distributions, $E\{u\}$ corresponds to $\max P(u)$



NON–GAUSSIAN DISTRIBUTIONS

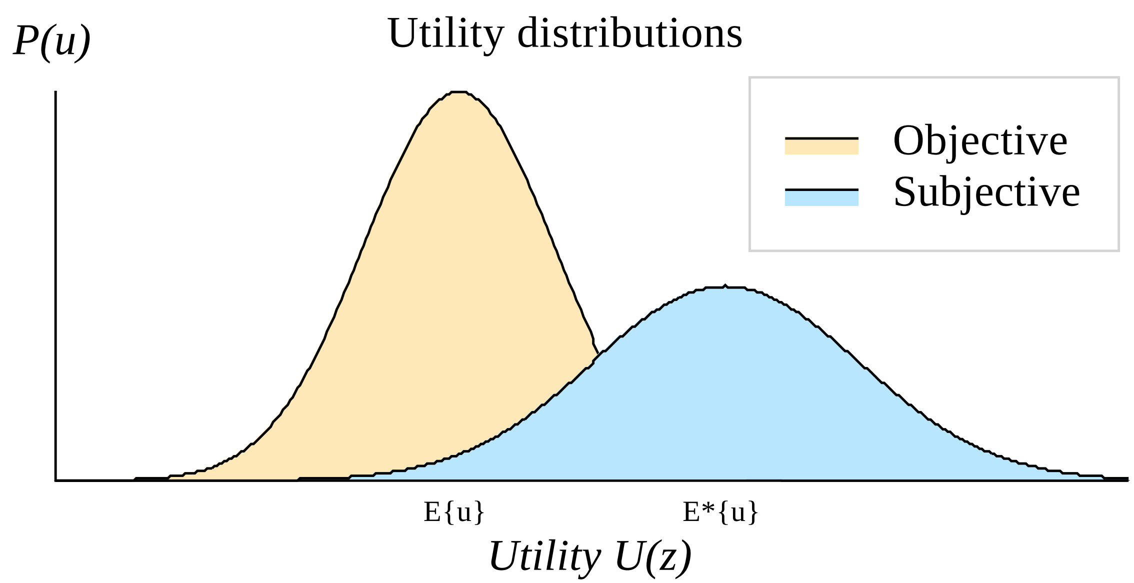
- In general, $E\{u\} \neq \arg\max P(u)$



- Often, $E\{x\} \notin X$, such as $\$100 \notin \{\$0, \$300\}$

EXPLORATION vs EXPLOITATION

- Distributions P^* an agent uses may be only approximations of the objective probabilities P, and $E^*\{u\} \neq E\{u\}$.



- Is $\max E\{u\}$ a good sampling strategy?
- Distributions $P(z)$ may depend on the agent's actions.

RANDOM DECISION–MAKING

- Instead of $\sum_z p(z \mid x)u(z)$, we can use $p(z \mid x)$ to draw z randomly (i.e. Monte–Carlo). The utility of this outcome is called *random utility* $u(z)$, and it can be used to choose x

$$x = \arg\max_{x \in X} u(z) \,, \text{ where } z \leftarrow p(z \mid x)$$

- Sampling can be implemented using the inverse PDF method

$$\text{Outcome} = F^{-1}(p) \,, \quad \text{where } p \in (0,1)$$

- On average, $z = \arg\max p(z)$

THE AGENT ARCHITECTURE

- The following decision–theoretic agents architecture is used

$$
\begin{aligned}
X &= \{x_1, \dots, x_m\} && \text{percepts} \\
Y &= \{y_1, \dots, y_n\} && \text{preferences (e.g. } Y = \{\text{success, failure}\}) \\
Z &= \{z_1, \dots, z_k\} && \text{actions}
\end{aligned}
$$

- Transitions (x_i, y_j, z_k) are recorded in M_{ij}^k.
- Normalised M_{ij}^k is used as a Markov transition model $P(X, Y, Z) = (p_{ij}^k)$, where $p_{ij}^k = p(x_i, y_j, z_k)$.
- We can use Bayesian inference

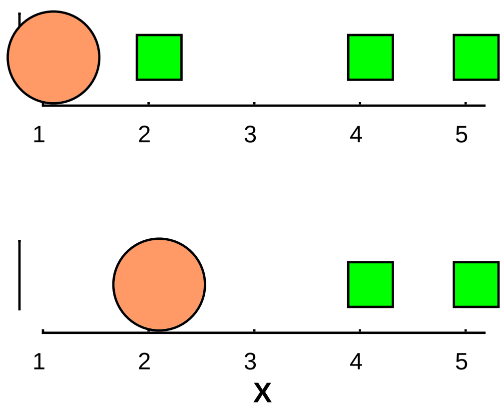
$$P(Y \mid X, Z) = \alpha P(X, Y, Z)\,, \quad \text{where } \alpha = \frac{1}{\|P(Y|X,Z)\|}$$

LEARNING

- The memory is initialised to $M_{ij}^k = I = (1)$. Normalised, it represents uniform distribution that yields maximum entropy $\max H(X, Y, Z) = \ln(m \times n \times k)$ (i.e. no information).
- The preference relations influence the action selection mechanism, which makes X, Y and Z statistically dependent.
- Mutual information can measure this dependence

$$I(X, Y, Z) = H(X) + H(Y) + H(Z) - H(X, Y, Z)$$

THE EXPERIMENT



Percepts: $X = \{x_1, \ldots, x_5\}$

Preferences: $Y = \{\text{success}, \text{failure}\}$

Actions: $Z = \{\text{left}, \text{stay}, \text{right}\}$

- The rewards appear randomly according to some distribution law and at different rates.
- The performance of agents can be measured and compared by the number of rewards they manage to collect.

ACTION SELECTION

Three agents were used in tests. The only difference was how actions were selected from Z

max EU

$$z = \arg\max_{z \in Z} \sum_{y \in Y} p(y \mid x, z) u(y)$$

rand Act

$$z \leftarrow p(z \mid x, \arg\max_{y \in Y} u(y))$$

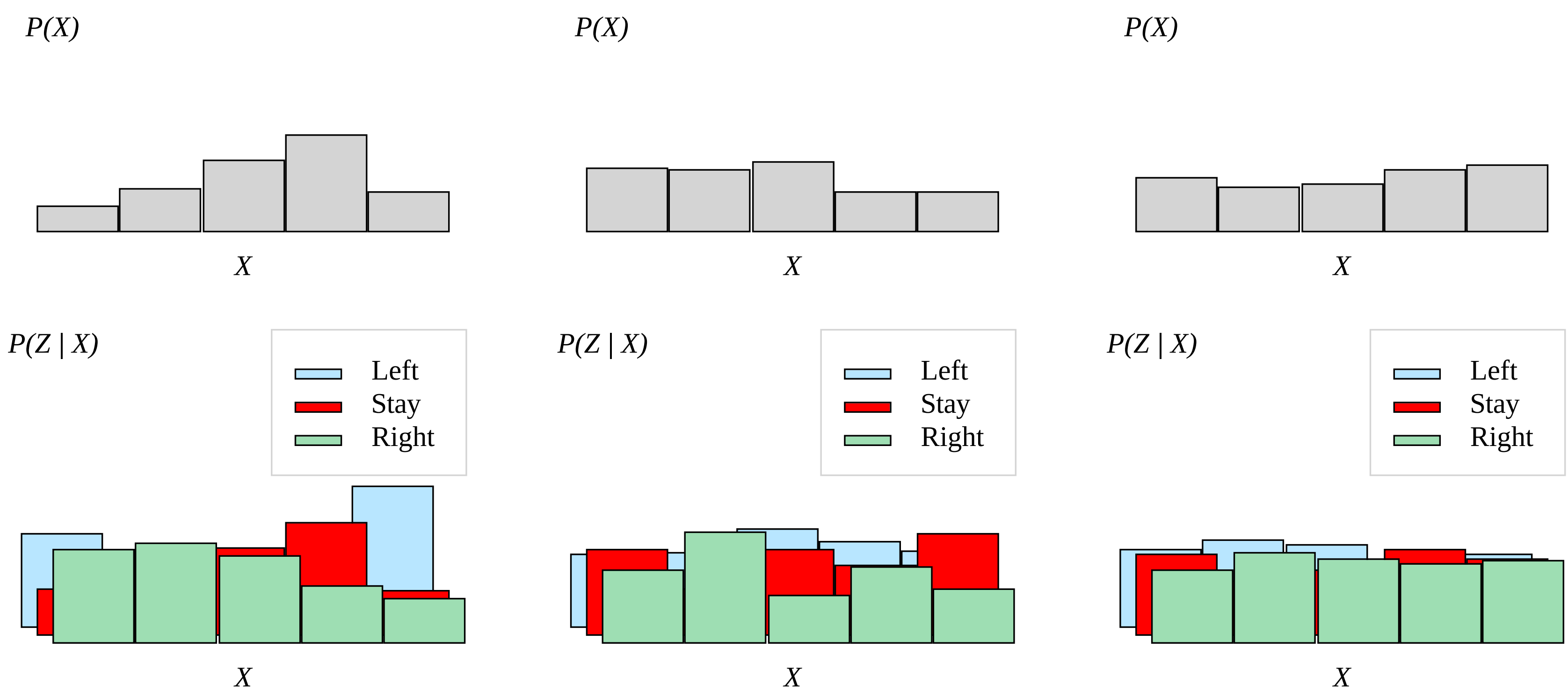
rand U

$$z = \arg\max_{z \in Z} u(y) \,, \quad \text{where } y \leftarrow p(y \mid x, z)$$

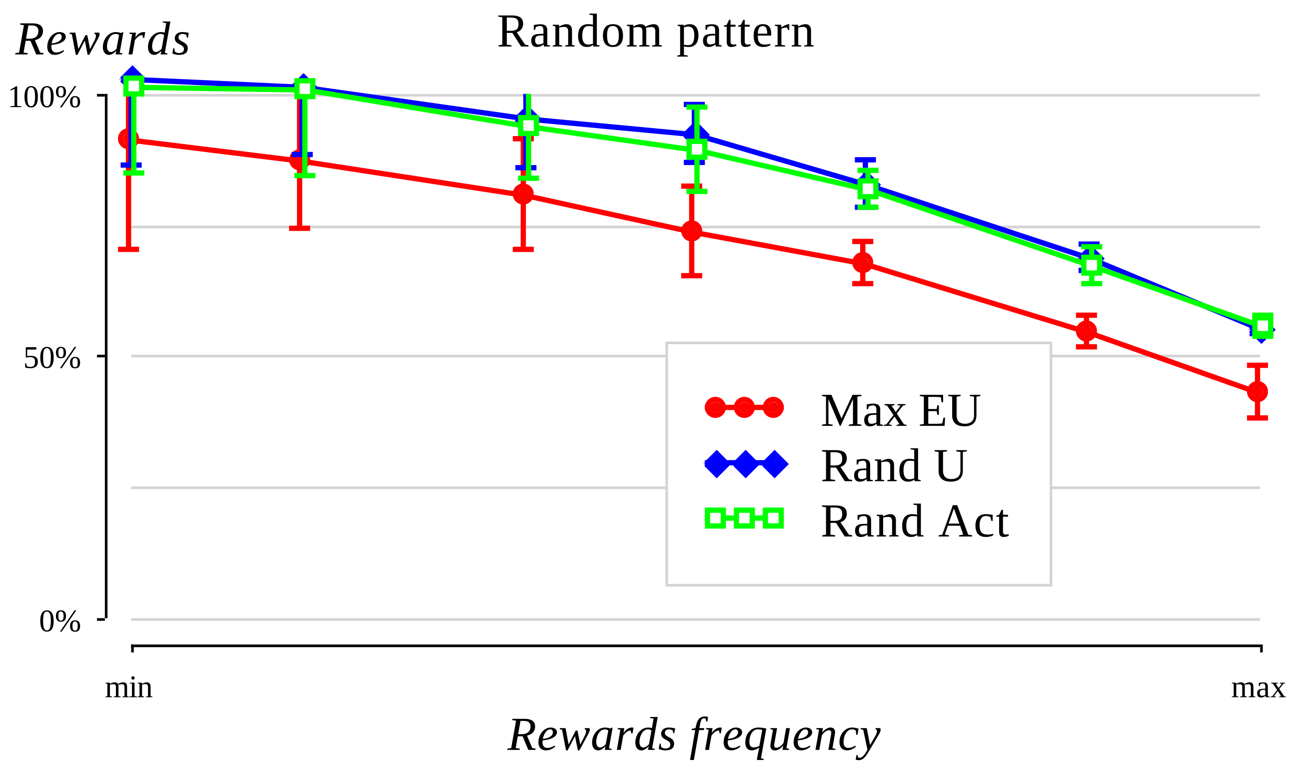
COMPLETELY RANDOM WORLD

Rewards appear without any pattern (i.e. uniformly).

$\max EU$ | Rand Act | Rand U

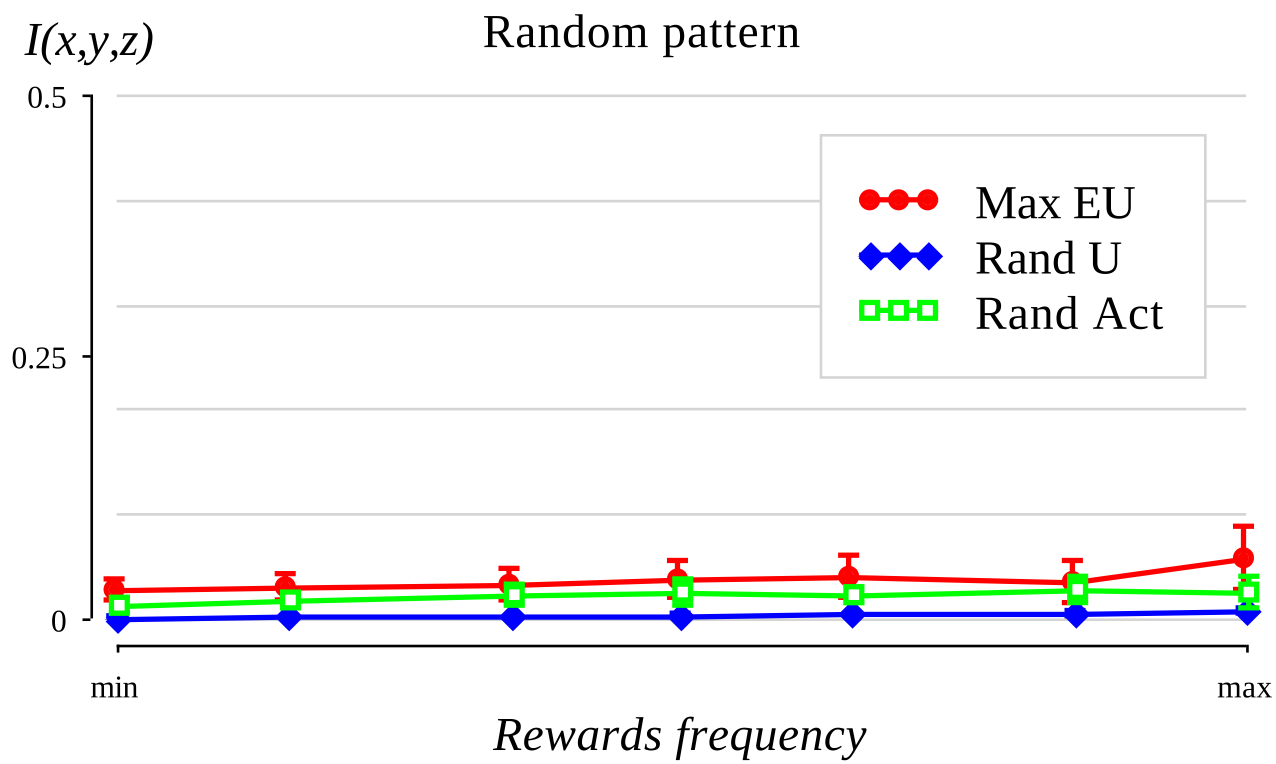


PERFORMANCE (NO PATTERN)



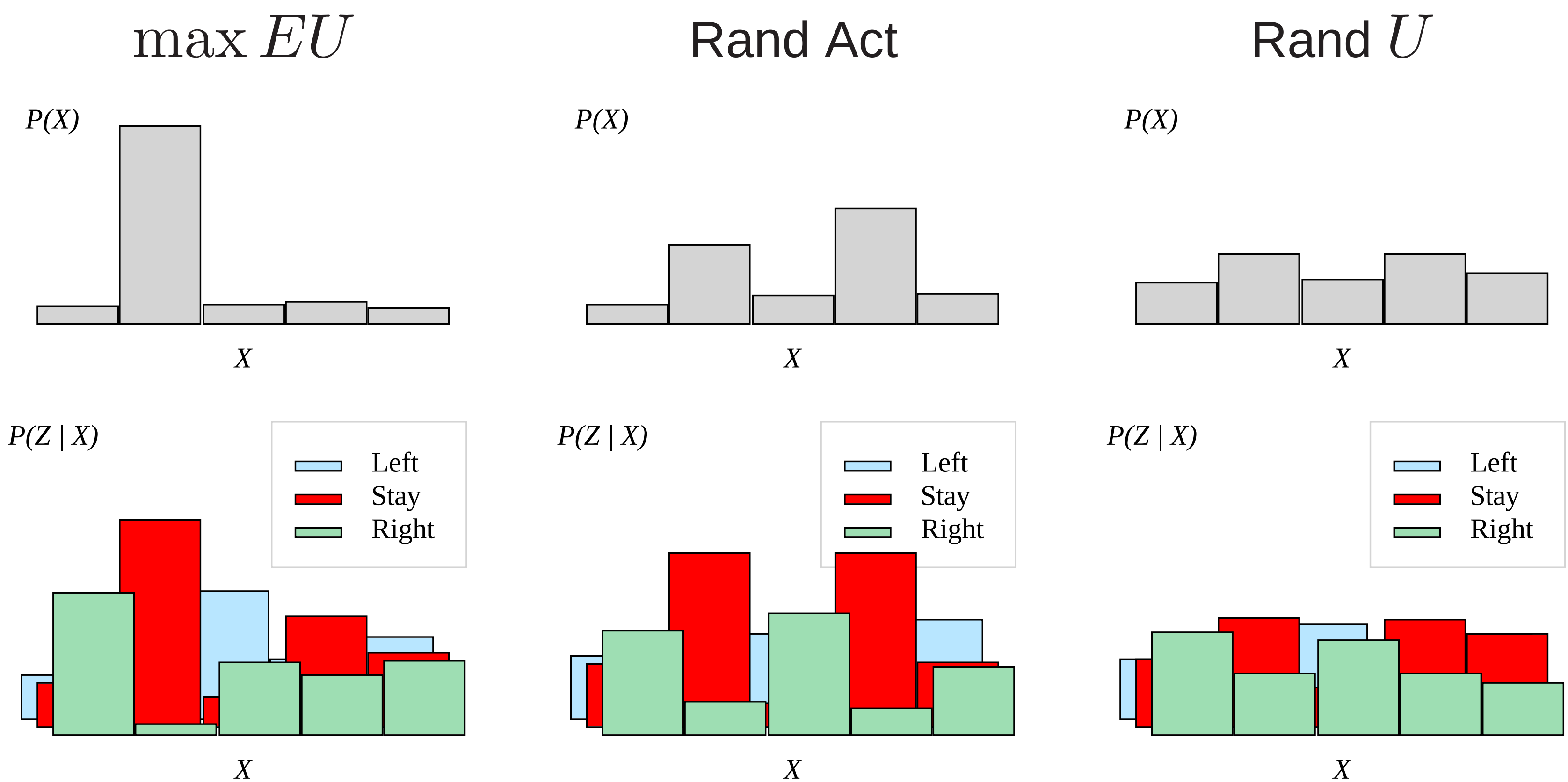
The random agents are not doing too bad!

MUTUAL INFORMATION (NO PATTERN)

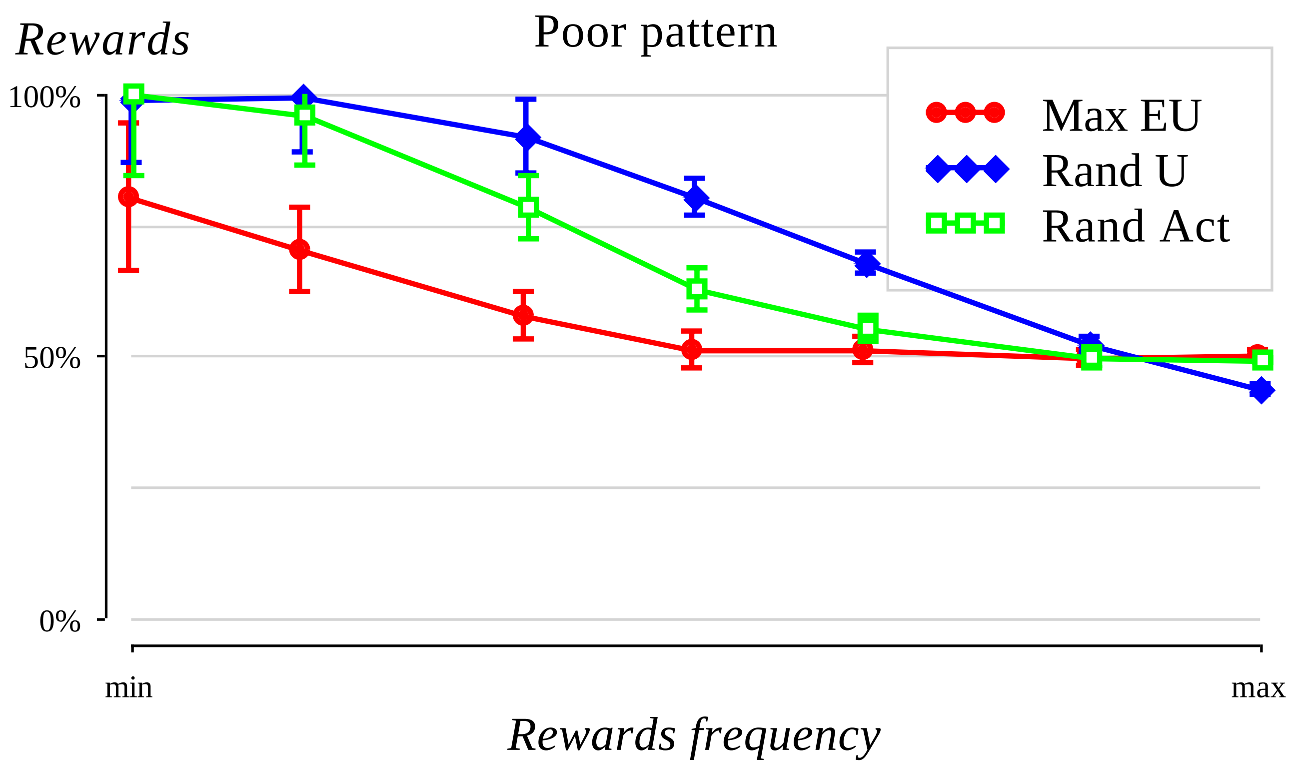


REGULAR PATTERN (POOR)

Rewards appeared according to $\{0, 1, 0, 1, 0\}$

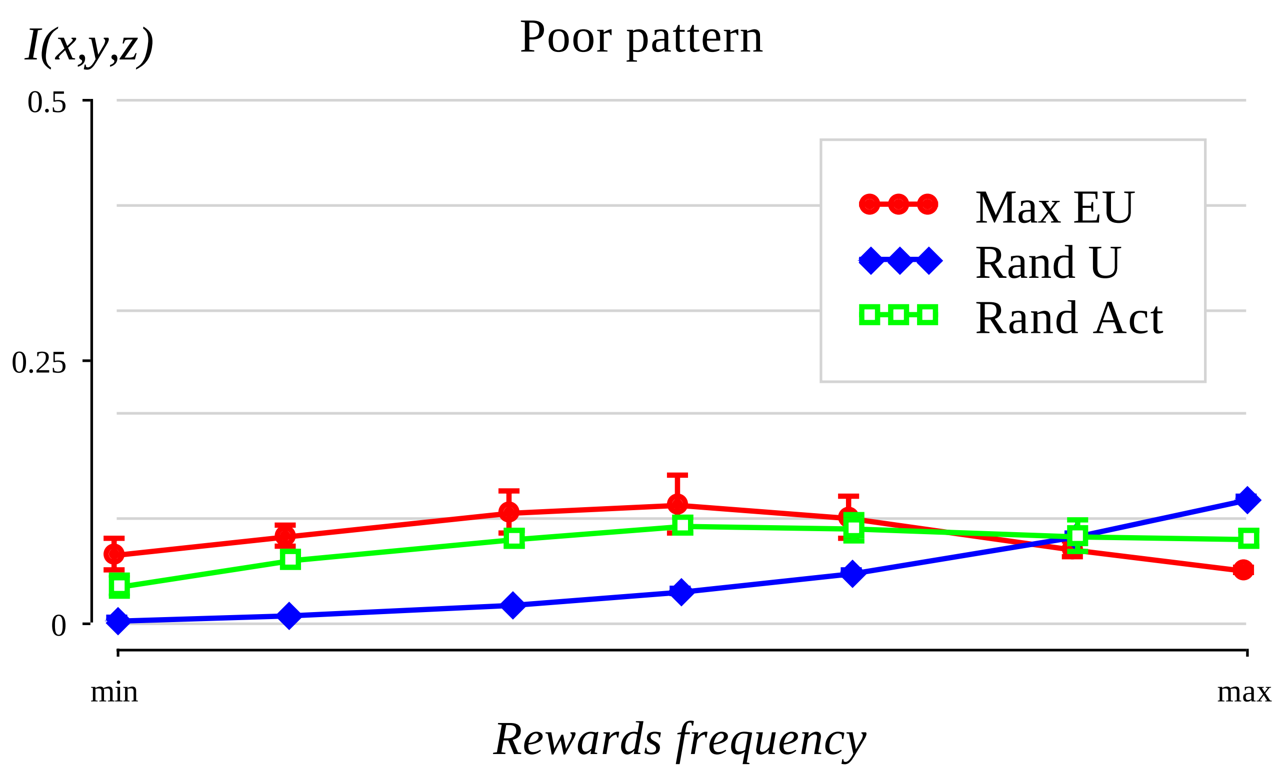


PERFORMANCE (REGULAR 1)



The random agents outperform $\max EU$ almost 2:1.

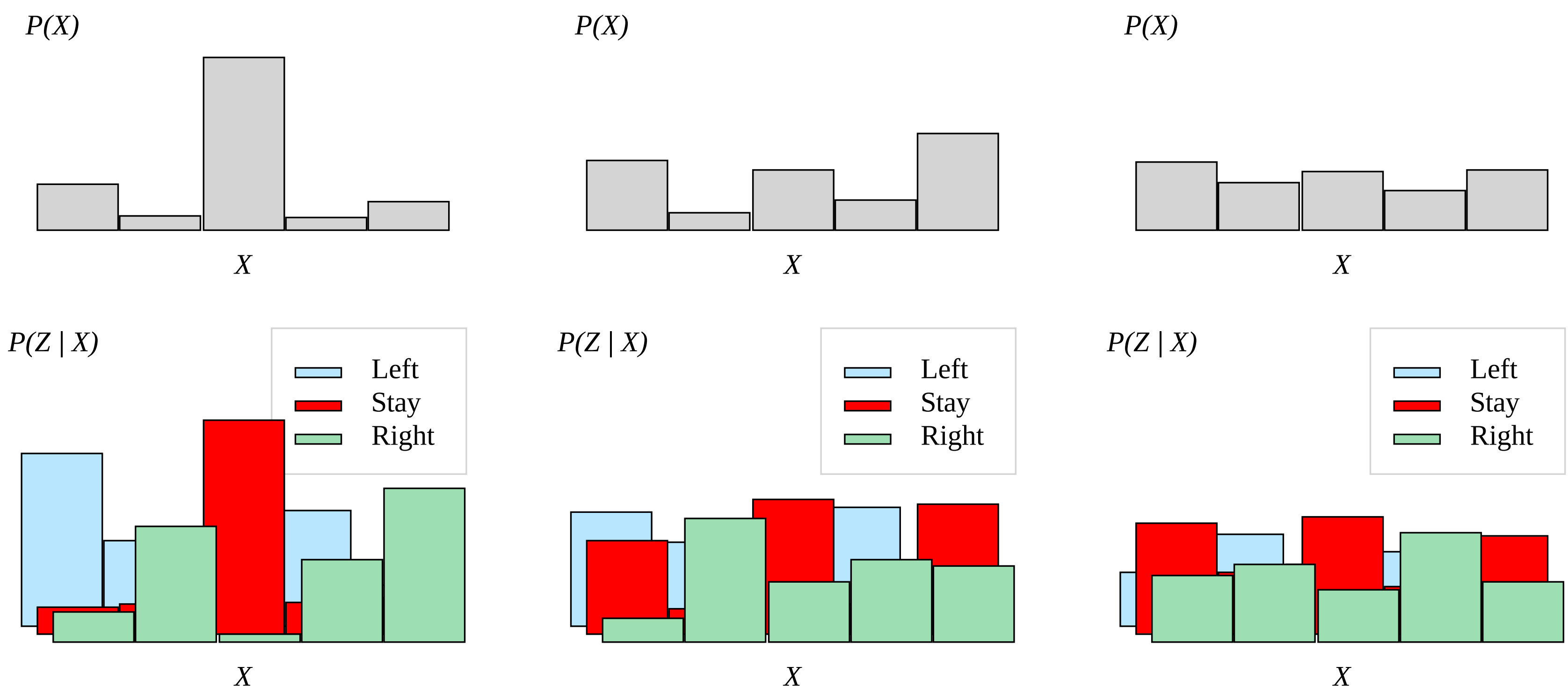
MUTUAL INFORMATION (REGULAR 1)



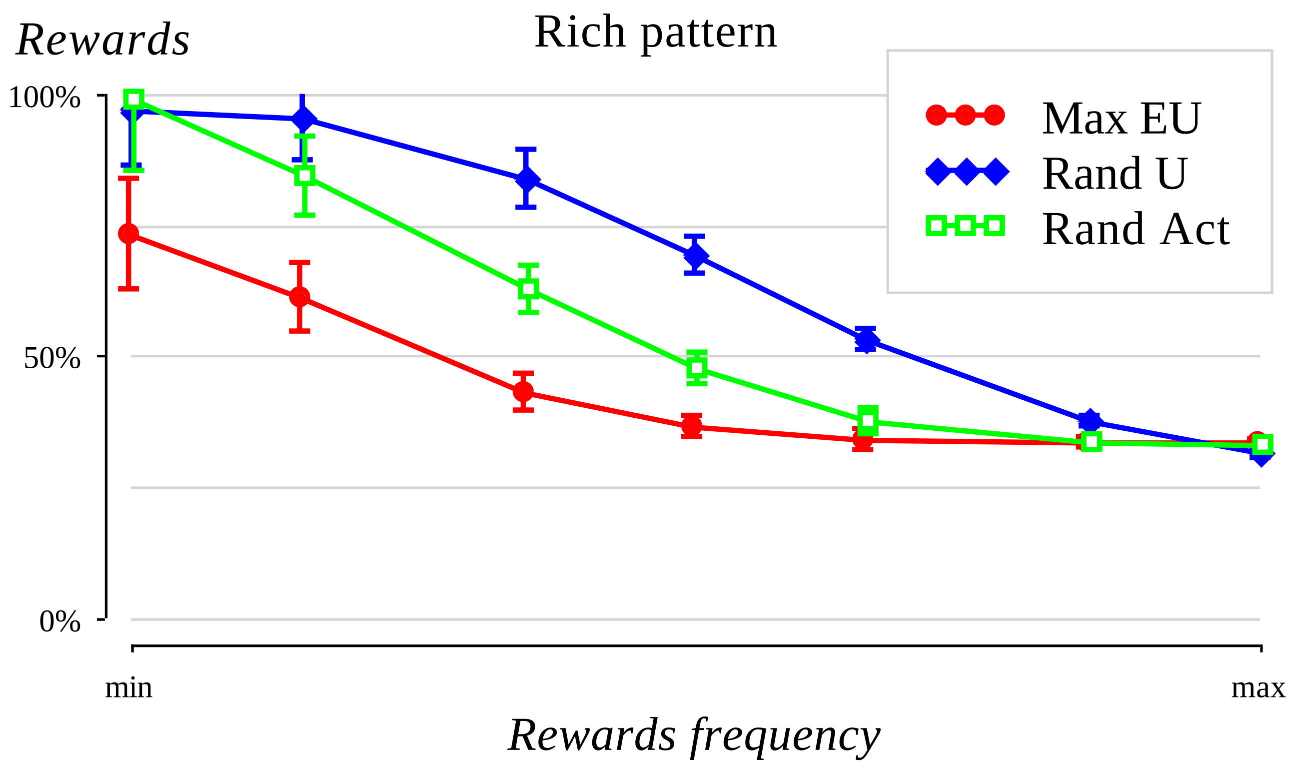
REGULAR PATTERN (RICH)

Rewards appeared according to $\{1, 0, 1, 0, 1\}$

$\max EU$ | Rand Act | Rand U

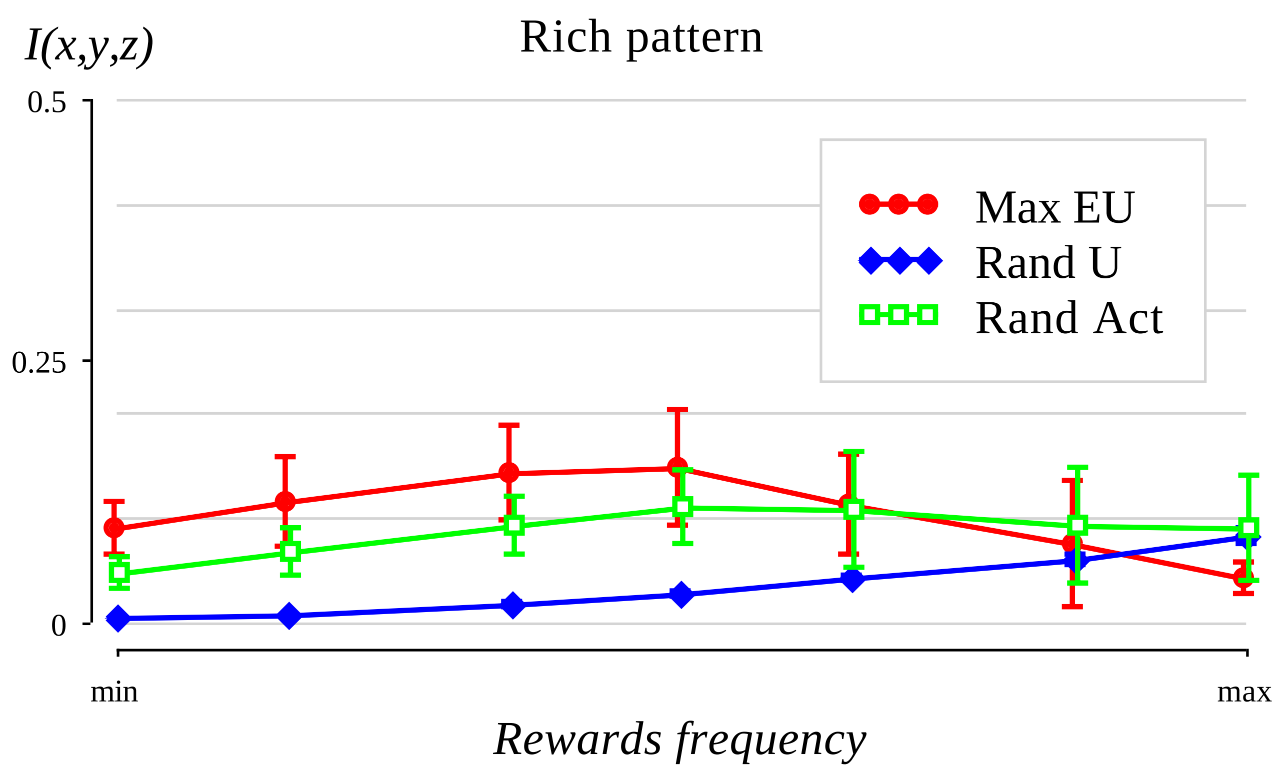


PERFORMANCE (REGULAR 2)

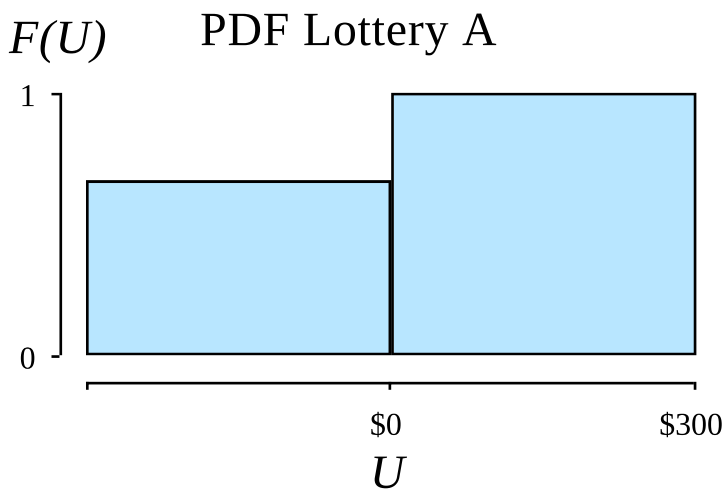


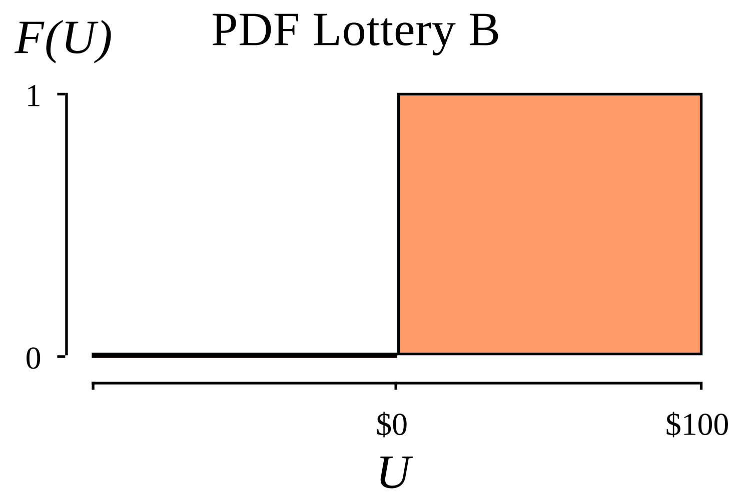
Again, random outperform $\max EU$ as much as 2:1.

MUTUAL INFORMATION (REGULAR 2)



INVERSE PDF (A and B)

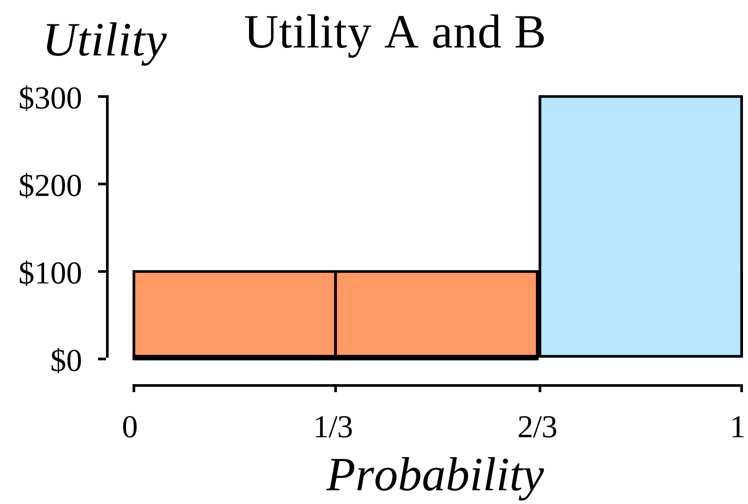




Utility $= F^{-1}(P)$

$RU_A < RU_B$ 2 out of 3 times, which supports experimental evidence

$$A \prec B$$



CONCLUSIONS

- Here, the $\max EU$ method turned to be inferior to the Monte–Carlo methods.
- The random agents not only maximise the utility, but also sample the distributions (exploration vs exploitation).
- Monte–Carlo methods use all statistics (not just E).
- Random d.m. accommodates better human choice behaviour.
- The inverse PDF provides some clues for the Allais paradox.
- Should we go back to the game theory? (e.g. the *Prisoners' dilemma*)

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'École americaine. *Econometrica*, *21*, 503–546.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum.

Anscombe, F. J., & Aumann, R. J. (1963). A definition of subjective probability. *Annals of Mathematical Statistics*, *34*, 199–205.

Belavkin, R. V., & Ritter, F. E. (2003, April). The use of entropy for analysis and control of cognitive models. In F. Detje, D. Dörner, & H. Schaub (Eds.), *Proceedings of the Fifth International Conference on Cognitive Modelling* (pp. 21–26). Bamberg, Germany: Universitäts–Verlag Bamberg. (ISBN 3-933463-15-7)

Bernoulli, D. (1954). Commentarii acad [English translation]. *Econometrica*, *22*, 23–36. (Reprinted from *Scientiarum Imperialis Petropolitanae*, 1738, *5*, 175–192)

List, J. A., & Haigh, M. S. (2005). A simple test of expected utility theory using professional traders. *PNAS*, *102*(3), 945–948.

Myers, J. L., Fort, J. G., Katz, L., & Suydam, M. M. (1963). Differential monetary gains and losses and event probability in a two–choice situation. *Journal of Experimental Psychology*, *77*, 453–359.

Neumann, J. von, & Morgenstern, O. (1944). *Theory of games and economic behavior* (first ed.). Princeton, NJ: Princeton University Press.

Savage, L. (1954). *The foundations of statistics*. New York: John

Wiley & Sons.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.