Title
Why and How to Measure the Association between Choice Options

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 35(35)

ISSN
1069-7977

Authors
Andraszewicz, Sandra
Rieskamp, Jorg

Publication Date
2013

Peer reviewed

Why and How to Measure the Association between Choice Options

Sandra Andraszewicz (s.andraszewicz@unibas.ch)
University of Basel
Department of Psychology, Missionsstrasse 62A
Basel, 4055 Switzerland

Jörg Rieskamp (joerg.rieskamp@unibas.ch)
University of Basel
Department of Psychology, Missionsstrasse 62A
Basel, 4055 Switzerland

Abstract

Prominent theories of decision making, such as *proportional difference model*, *priority heuristics*, *decision field theory* and *regret theory* assume that people do not evaluate options independently of each other. Instead, these theories predict that people compare the options' outcomes with each other. Therefore the theories' predictions strongly depend on the association between outcomes. In the present work, we examine how the association between options can be best described. For options with two outcomes the standard correlation measure between option's outcomes does not provide a meaningful interpretation. Therefore, we propose the standardized covariance between options A and B, denoted as σ^*_{AB}. We describe the properties and interpretation of this measurement and show its similarities and differences with the correlation measurement. Finally, we show how the predictions of different models of decision making vary depending on the value of the standardized covariance.

Keywords: decision making models; covariance; gambles; two-outcome; risky choice

Introduction

Standard economic models of decision making like expected utility theory assume that people evaluate choice options independently of each other (Neumann & Morgenstern, 1944). However, contrary to this basic independence assumption a vast amount of evidence has shown that people evaluate choice option depending on the set of alternative choice options (see Rieskamp, Busemeyer, & Mellers, 2006 for more details).

For example, choice of between two health insurance offers is a choice between risky options with payouts depending on the occurence of an illness. An illness can occur with a certain probability that can be estimated based on teh patient's age and health history. When deciding between the two options, one would probably compare the insurance coverage in case of specific illnesses of both offers with each other, rather than first evaluate one offer and then another.

Many cognitive models of decision making assume that when people make choices between options they compare the options' outcomes with each other. For instance, the *priority heuristics* (Brandstätter, Giegerenzer, & Hertwig, 2006) assumes that people first compare all options with respect to their minimum outcomes. If these outcomes do not allow to discriminate the options, the options are compared with respect to the probability of the minimum outcomes, and so forth. *Regret theory* (Loomes & Sugden, 1982), *proportional difference model* (González-Vallejo, 2002), and *decision field theory* (Busemeyer & Townsend, 1993) are three other prominent models of decision making assuming that people compare options with respect to their outcomes. These comparisons are then accumulated to form an overall preference. Therefore the predictions of all these models depend on the associations between the outcomes of options.

How can these associations be best characterized? Using the covariance of the options' outcomes would be a sound solution, especially because many of the studies of decision making use monetary choices which are an analogy to investments. Indeed, in portfolio management covariance plays an important role for selection of assets (i.e. Pafka & Kondor, 2003; Disatnik & Benninga, 2007).

Also, many studies investigating decision making focus on comparing various models. In such case, the selection of the choice options, which are usually presented as gambles, is very important. As highlighted in the work on optimal experimental design, selecting gambles for discriminating between the models, is an essential issue that determines the effectiveness of the experiment (see Cavagnaro, Gonzalez, Myung, & Pitt, 2013; Myung & Pitt, 2009; Zhang & Lee, 2010) The main problem with using covariance is that its value depends on the range of the outcomes' values which makes it hard to interpret.

As an alternative measurement one could use the correlation between the outcomes. However, a large part of the research is done with two-outcome choice problems (e.g. González-Vallejo, 2002; Birnbaum, 2008), for which correlation is either 1 or -1 (see Rodgers & Nicewander, 1988), so that the correlation measurement does not provide a meaningful interpretation.

Therefore, we propose an alternative measurement, *standardized covariance*, as a measure of the strength of the association between the options' outcomes, which can be easily interpreted. In the following sections we will first explain how the standardized covariance is determined and how it should be understood. Next, we will present relations between the covariance, variances and expected value of two-outcome gambles, which will clarify the construction of the standardized covariance. Finally, we will show how decision making models make different predictions depending on the strength of the standardized covariance.

Table 1: Twelve examples of choice options with different standardized covariance. In each example, the top row indicates the probability of the occurrence of two outcomes. Two consecutive rows display the possible outcomes of Option A and Option B.

$\sigma^*_{AB} = 1$			$\sigma^*_{AB} = -1$		
Example 1	60%	40%	Example 2	60%	40%
A	80	55	A	80	55
B	80	55	B	55	80
$\sigma^*_{AB} = .80$			$\sigma^*_{AB} = -.80$		
Example 3	60%	40%	Example 4	60%	40%
A	80	55	A	80	55
B	80	30	B	30	80
$\sigma^*_{AB} = .80$			$\sigma^*_{AB} = 1$		
Example 5	40%	60%	Example 6	40%	60%
A	80	55	A	80	55
B	80	30	B	70	45
$\sigma^*_{AB} = .32$			$\sigma^*_{AB} = .05$		
Example 7	60%	40%	Example 8	60%	40%
A	80	20	A	42	40
B	50	40	B	80	6

Standardized Covariance

We denote standardized covariance of a pair of dependent choice options A and B, with each having two possible outcomes, as σ^*_{AB}, where the non-standardized covariance is denoted as σ_{AB}. Standardized covariance is equal to twice the covariance divided by the sum of the variances σ_A and σ_B of each of the options (see Equation 1).

$$\sigma^*_{AB} = \frac{2\sigma_{AB}}{\sigma^2_A + \sigma^2_B} \tag{1}$$

For stochastically non-dominant options, σ^*_{AB} is a continuous variable ranging from -1 (strong negative association) to 1 (strong positive association). When $\sigma^*_{AB} = 0$ either the options are completely unrelated (e.g. they are statistically independent, where two options do not depend on one external event) or the covariance between the options' outcomes is equal to 0. The second case occurs, when one of the options is a sure thing. When σ^*_{AB} approaches 0, the variances of both options are low, and so is the association between the options.

Properties of σ^*_{AB}

When σ^*_{AB} reaches its maximum at 1 then the sum of the variances equals twice the covariance:

$$\sigma^*_{AB} = 1 \iff 2\sigma_{AB} = \sigma^2_A + \sigma^2_B. \tag{2}$$

Analogically, for negatively related gambles the relation is:

$$\sigma^*_{AB} = -1 \iff -2\sigma_{AB} = \sigma^2_A + \sigma^2_B. \tag{3}$$

Situation from Equation 2 occurs when both options are the same (Example 1 in Table 1) and, analogically, by interchanging the outcomes of Option B we can obtain choice options for which $\sigma^*_{AB} = -1$ (Example 2), which shows the symmetricity of options with positive and negative σ^*_{AB}. Further, as shown in Examples 3 and 4, by altering one outcome from Option B so that the options are not the same any more, we obtain options with slightly lower σ^*_{AB}. Interrestingly, the probabilities of the outcomes do not influence σ^*_{AB} (compare Examples 3 and 5). Also, the "perfect association" does not occur only when the options are identical, but also, when the difference between outcomes of option A and B is the same and this difference is the difference between expected values (Example 6 with the difference in expected values of 10 points). By making the outcomes corresponding to the same probabilities more dissimilar, one can decrease σ^*_{AB} (compare Examples 3 and 7). Finally, as shown in Example 8, when outcomes of one option are almost the same (almost a sure thing), while outcomes of the other option are dissimilar, σ^*_{AB} is almost 0.

For stochastically non-dominant options, σ^*_{AB} is not higher than 1 or lower than -1 because it is not true that $2\sigma_{AB} > (\sigma^2_A + \sigma^2_B)$. Below, we provide the mathematical proof.

Proof. $2\sigma_{AB} > (\sigma^2_A + \sigma^2_B)$ is false

In stochastically non-dominant options the variances of options' outcomes are unequal, thus $\sigma^2_A < \sigma^2_B \vee \sigma^2_A > \sigma^2_B$.

$\sigma^2_A < \sigma^2_B \iff \sigma^2_B = \sigma^2_A + s \wedge s \in \mathbb{R}^+$

Then,

$2\sigma_{AB} > \sigma^2_A + \sigma^2_B$

$2\sigma_{AB} > 2\sigma^2_A + s$

$0 > \sigma^2_A + \frac{s}{2} - \sigma_{AB}$

$0 > E[a^2] + \frac{E[b^2]-E[a^2]}{2} - E[ab]$

$E[a^2] < E[b^2] \iff b = a + g \wedge g \in \mathbb{R}^+ \wedge g = const.$

By expanding the inequality we get

$0 > \frac{g^2}{2}$

Since $g^2 > 0$ the inequality is false.

□

Standardized Covariance vs. Correlation

Examples presented in Table 1 indicate that there are similarities between correlation and standardized covariance and the "perfect" correlation overlaps with the "perfect" standardized covariance (e.g. Examples 1, 2 and 3 in Table 1). Correlation coefficient r equals to

$$r = \frac{\sigma_{AB}}{\sigma_A \sigma_B}. \tag{4}$$

The relation between correlation coefficient and standardized covariance is as follows:

$$\sigma^*_{AB} = \frac{2r\sigma_A\sigma_B}{\sigma^2_A + \sigma^2_B} \tag{5}$$

and correlation is equal to standardized covariance when

$$2\sigma_A\sigma_B = \sigma_A^2 + \sigma_B^2. \quad (6)$$

Because for stochastically non-dominant options with two outcomes correlation is always either -1 or 1, $\sigma_{AB} = \sigma_A\sigma_B$. Therefore, the relation between the sum of variances and the product of the standard deviation is the same as the relation between the sum of variances and the covariance. Thus, standardized covariance could also be written as

$$\sigma_{AB}^* = \frac{2\sigma_A\sigma_B}{\sigma_A^2 + \sigma_B^2} \quad (7)$$

when the association between the options' outcomes is positive. When the association between the options' outcomes is negative, $\sigma_{AB} = -2\sigma_A\sigma_B$.

Variances and Covariance of Two-Outcome Options

In stochastically non-dominant pairs of options which are not identical, one option has higher variance than the other (compare range of outcomes of options A and B in Table 1 in Examples 3, 4, 5, 7 and 8, to Example 6 which contains a stochastically dominant pair of options). Therefore the sum of variances is composed of a smaller and larger variance.

The relation between the smaller variance and covariance is close to linear and is symmetric with respect to the x-axis. In contrast, the relation of the larger variance to covariance takes the shape of a triangular area and is also symmetric with respect to the x-axis, as shown in Figure 1. This figure shows a very interesting pattern in which the graph on the right side fits into the graph on the left side like "key and lock". The data points in Figure 1 were obtained from 100000 randomly generated two-outcome choice options, with various differences between expected values. The outcomes' values ranged between 1 and 100, and probabilities of their occurrence ranged between 1 and 99%.

The relation between the sum of variances and covariance is symmetric with respect to the x-axis (Figure 2). The gray data points, which lay on the diagonal of the graph in Figure 2, are the only ones, for which standardized covariance is equal to correlation. This is the key property of standardized covariance, as the gray points correspond to the "perfect correlation" between the options, for which equation 6 holds. In contrast, all black points represent the pairs of options whose relation varies between -1 and 1 (not perfect correlation).

Options with Negative and Mixed Outcomes

Until now, we have discussed the properties of standardized covariance, covariance and variances of options which generate only positive outcomes. However, some experiments might include choice options which generate only losses or might generate both, gains and losses. As a consequence, we randomly generated stochastically non-dominant options with only negative outcomes ($N = 100000$), to which we will refer as *negative options*, and options that have one positive

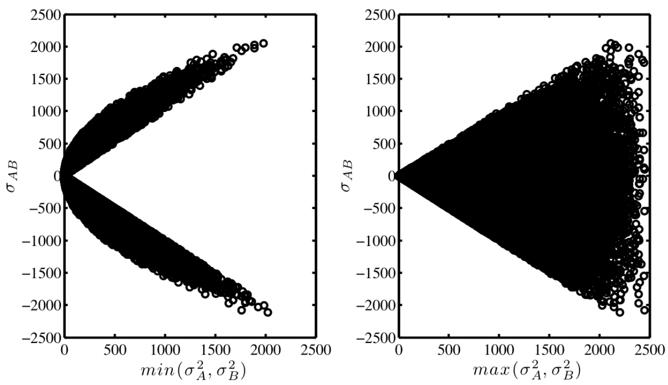

Figure 1: Left: relation between the lower variance (more secure option) and covariance between two-outcome options, Right: relation between the higher variance (more risky option) and covariance between two-outcome options.

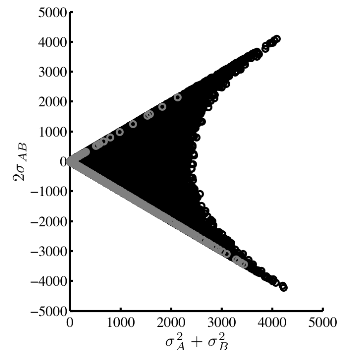

Figure 2: Relation of the sum of variances to twice the covariance. Gray points indicate the cases for which standardized covariance overlaps with correlation, such that $\sigma_{AB}^* = r = 1$ or $\sigma_{AB}^* = r = -1$.

and one negative outcome ($N = 100000$), which we call *mixed options*. The outcomes of *negative options* varied between -100 and -1 points, while the outcomes of *mixed options* were in range $[-100, 1]$ and $[1, 100]$ points. The probabilities of these outcomes ranged from 1% to 99%. We repeated the analysis of the relation between variances and covariance, as well as the sum of variances and the product of standard deviations of the options for the two new types of options.

The obtained results for the *negative* and *mixed options* were the same as for the *positive options*. Thus, the properties of the choice options regarding their variances, covariance and the standardized covariance, depicted in Figures 1 and 2, apply to various kinds of choice options. The only difference is that the range of the values of variances and covariance of *mixed options* is much greater (range: $[-10000, 10000]$). This is due to the fact that the range of the possible values is twice as big compared to the *positive* and *negative options*. In sum, standardized covariance is a stable measure of association between options.

Table 2: Ranges of values of standardized covariance, ratio of the smaller to the larger variance and the amount of pairs of options generated for each of the five differences between expected values between the options.

ΔEV	σ^*_{AB}	$\frac{min(\sigma^2_A,\sigma^2_B)}{min(\sigma^2_A,\sigma^2_B)}$	N
10	.02-.99	.00-.72	975021
15	.02-.94	.00-.49	377750
20	.02-.94	.00-.49	418251
25	.02-.94	.00-.49	244734
30	.02-.89	.00-.36	119241

Standardized Covariance vs. Expected Value

Standardized covariance is sensitive to the differences between expected values of two choice options. We generated all possible pairs of options with outcomes and probabilities as previously, for which the difference between the expected values was either 10, 15, 20, 25 or 30. As shown in Table 2, the greater the difference between expected values, the more narrow the range of possible values of σ^*_{AB}. Thus, when manipulating the difference between the expected value of pairs of options, our analysis shows that this manipulation will most likely also change the covariance of the options outcomes. Thus, when not controlling for this aspect, then variations of the expected value differences will often be confounded with variations of covariance differences. Therefore, the experimenters should keep in mind that the strength of the association between the gambles that they present to the participants may depend on the differences between expected values (ΔEV).

Interestingly, the greater the difference between the expected values, the fewer choice options could be obtained (see Table 2). Also, the greater the difference between expected values, the more narrow the ranges of possible values of standardized covariance and ratio between the lower and the higher variance within the pair of options (see Table 2). Therefore, in experiments that control for the expected value difference, it might be the standardized covariance between the options that influences people's choice, rather than the expected value.

Further, we selected a group of options for which $\Delta EV = 15$. For these options with fixed difference between expected values, we tested the relation between the variances of both options. As shown in Figure 3, the data points create a pattern that is symmetric with respect to the diagonal of the graph. In other words, when ΔEV is fixed, the variances of both options are related to each other with respect to a certain ratio, whose ranges we listed in Table 2.

Options with More than Two Outcomes

In order to analyze in more detail the relation between the correlation measure and the standardized covariance, one would have to extend the problem to choices with more than two possible outcomes. Therefore, we generated 10000 pairs of stochastically non-dominant options with four outcomes. The outcomes varied between 1 and 100 points, and probabilities varied betweeen 1% and 40%.

Firstly, we investigated the relations between the variances and the covariance. As shown in Figure 4, the relations between both variances and covariance do not display the "key-lock" pattern as in Figure 1, and the patterns are not symmetrical. Analogically, the relation between the sum of variances and twice the covariance is not symmetric with respect to the x-axis.

Secondly, we looked at the relation between correlation and standardized covariance. Figure 5 shows a strong relation between the correlation measure and standardized covariance. Pearson correlation between these two measures is very strong, $r = .98, p < .001$. In the current sample of generated pairs of options, covariance and correlation are equal to each other for 24% of the cases. Also, the slope of the regression line is high and the intercept very small (see caption of Figure 5). Thus, standardized covariance is a similar measure as correlation, but it has the advantage that it can be

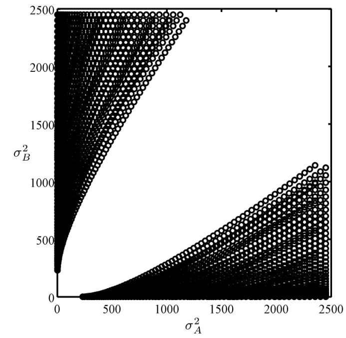

Figure 3: Relation between variances of two options with two outcomes.

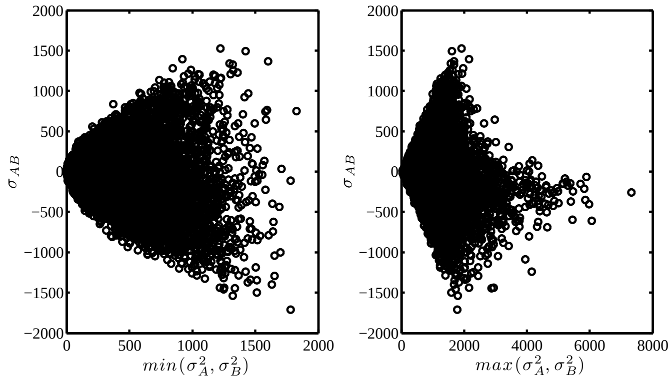

Figure 4: Left: relation between the smaller variance and the covariance between four-outcome options. Right: relation between the larger variance and the covariance between four-outcome options.

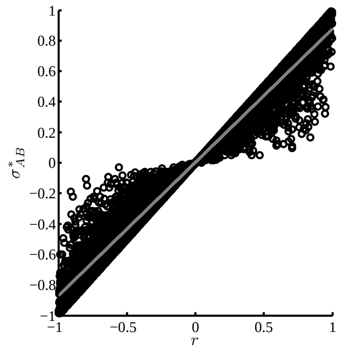

Figure 5: Relation between the correlation coefficient and standardized covariance of options with four outcomes. Gray line indicates the regression line, with the slope of .87 and intercept .0027.

applied to both two-outcome choice options and options with several outcomes.

Model Predictions Depending on the Standardized Covariance

In the previous sections, we have described features of stochastically non-dominant options with two outcomes and how one could measure the association between the options with the use of standardized covariance. In this section, we will show why the association between two options is important, based on two prominent theories of decision making, *regret theory* and *decision field theory*. Models of decision making generate different predictions for options with various differences between expected values. Thus, we focused on the choice options for which $\Delta EV = 15$.

For all of these options we generated predictions of the two models. Following Pathan, Bonsall, and Jong (2011), we defined the regret function of choosing option A over option B with outcomes x_i, $i \in \{1,2\}$ as

$$R_{iA} = ln(1 + exp(\beta(x_i - max(x_{iA}, x_{iB})))). \quad (8)$$

The total regret from choosing option A equals to

$$R_A = \Sigma_{i=1}^{2} R_{iA}. \quad (9)$$

Further, the probability of choosing option A over option B is estimated using softmax rule

$$Pr(A|A,B) = \frac{1}{1 + exp(\theta(R_B - R_A))}. \quad (10)$$

β and θ are free parameters of the model. More details regarding *regret theory* is provided in Loomes and Sugden (1982). A parsimonious version of *decision field theory* was used, as described by Busemeyer and Townsend (1993).

The models' predictions are expressed as probabilities of choosing option A over B. We converted these results to the prediction that the option with the higher expected value

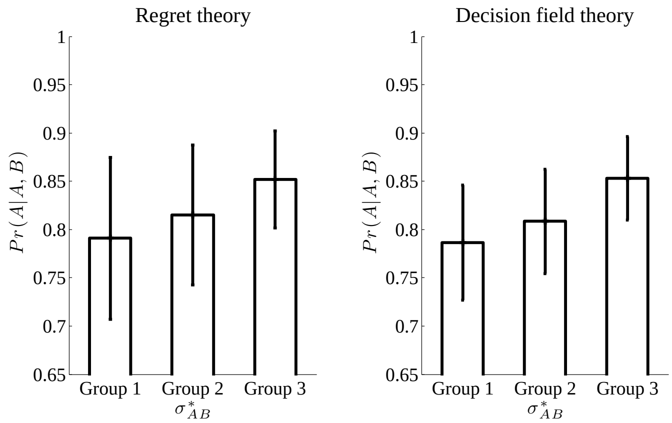

Figure 6: Average predictions of *regret theory* and *decision field theory*. To generate predictions the following parameters were used: *regret theory* $\beta = .05$, $\theta = 4.6$, *decision field theory* $\theta = 1.19$. The parameter of *decision field theory* was based on Rieskamp (2008) and the parameters of *regret theory* were adjusted so that the predictions of both models are at the same level. Error bars indicate standard deviations.

would be chosen, which resulted in all predictions ranging between 0.5 and 1. Next, we grouped the options depending on their standardized covariance, such that group 1: $\sigma^*_{AB} < 0.2$ (21.2%), group 2: $0.2 < \sigma^*_{AB} \leq 0.5$ (34.7%), group 3: $0.5 < \sigma^*_{AB}$ (44.2%). For each of the three groups we calculated the mean prediction and its standard deviation.

As shown in Figure 6, the models' predictions differ among the three groups. This constitutes evidence that some theories of decision making not only assume on a theoretical level that the relation between the options' outcomes play an important role in decision making, but also provide quantitative evidence. Therefore, one should control for the association between the options. This is a crucial property of the standardized covariance, because in experiments in which the association between options was not examined, the results might depend on the selected set of choice options.

Furthermore, the models' predictions for choices with the same level of association could differ depending on the difference between expected values. From each group of options with difference between expected values of 10, 15, 20, 25 and 30 points, we picked all options for which $\sigma^*_{AB} = .3$ and we generated models' predictions using the same set of parameters as previously. As shown in Figure 7, *decision field theory* makes very systematic predictions in which the higher the expected value difference, the higher the probability of choosing the option with the higher expected value. In contrast, *regret theory* indicates some differences but no trend can be observed.

In sum, predictions of models of decision making result from the interaction between the difference between expected values and the strength of the association between the choice options. This finding is very important, as in most studies

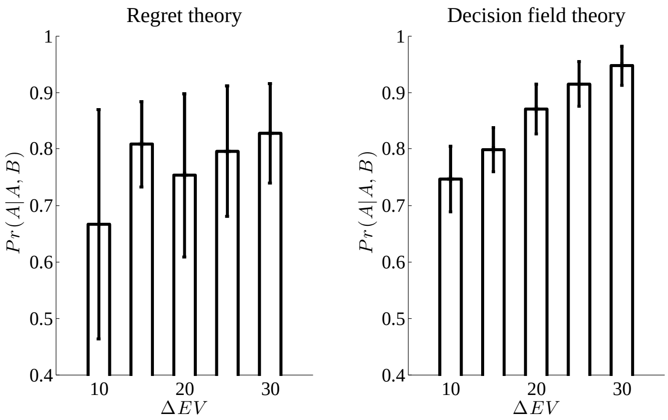

Figure 7: Average predictions of *regret theory* and *decision field theory* for choices with $\sigma^*_{AB} = .3$ and various expected values. Error bars indicate standard deviations.

the researchers do not even consider the covariance of the options' outcomes, but only report the differences in expected values.

Discussion

The association between two options' outcomes may play an important role in testing models of decision making. As we have shown, models can generate different predictions depending on the combination of the expected value difference and the association between the options. Experiments that control only for the expected value difference may obtain confounded results.

Therefore, a simple measure of the strength of this association is needed. For experiments that employ two-outcome choice options, we proposed the *standardized covariance*. Its values range between -1 and 1, where 1 indicates "perfect positive association" and -1 indicates "perfect negative association". When standardized covariance equals to 0, one of the options is a sure thing. When standardized covariance equals to -1, 0 and 1 it overlaps with correlation.

There is a strong association between correlation measure and standardized covariance. This constitutes solid evidence in favor of the reliability of the standardized covariance as a measure of the association between two choice options. Interestingly, there are very clear patterns of relations between variances and covariance of the two-outcome options. In contrast, these patterns are different when there are more outcomes. Therefore, future empirical research is needed to test the applicability of the standardized covariance and its perception by human decision makers. Also, as a future investigation, we suggest that one should test whether the predictions of the aforementioned models of decision making reflect the real human choice behavior.

In sum, this work was based on extensive simulations of random choice options and choice options with specific properties. We have shown that standardized covariance is a robust measure, with similar properties to the correlation. Finally, we showed that the covariance strongly influences the prediction of different cognitive models of decision making and should be given more attention in empirical work.

References

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review*, 463–501.

Brandstätter, E., Giegerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*(2), 409-432.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432-459.

Cavagnaro, D. R., Gonzalez, R., Myung, J. I., & Pitt, M. A. (2013). Optimal decision stimuli for risky choice experiments: An adaptive approach. *Management Science*, *59*(2), 358–375.

Disatnik, D., & Benninga, S. (2007). Shrinking the covariance matrix. *The Journal of Portfolio Management*, *33*(4), 55–63.

González-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review*, *109*(1), 137.

Loomes, G., & Sugden, R. (1982, December). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, *92*(368), 805-24.

Myung, J. I., & Pitt, M. A. (2009). Optimal experimental design for model discrimination. *Psychological review*, *116*(3), 499.

Neumann, J. V., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Pafka, S., & Kondor, I. (2003). Noisy covariance matrices and portfolio optimization ii. *Physica A: Statistical Mechanics and its Applications*, *319*, 487–494.

Pathan, A., Bonsall, P., & Jong, G. de. (2011, June). How the choice of mode is conditioned by information source. In *International choice modelling conference, leeds, uk.*

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1446.

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006, September). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, *44*(3), 631-661.

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, *42*, 59–66.

Zhang, S., & Lee, M. D. (2010). Optimal experimental design for a class of bandit problems. *Journal of Mathematical Psychology*, *54*(6), 499–508.