Project Similarity Bias and Variance Neglect in Forecast Metric Evaluation

Shir Dekel[1], Micah Goldwater[1], Dan Lovallo[2], and Bruce Burns[1]

[1]The University of Sydney, School of Psychology

[2]The University of Sydney, Business School

Author Note

Shir Dekel https://orcid.org/0000-0003-1773-2446.

Portions of this work comprised Shir Dekel's doctoral dissertation.

Correspondence concerning this article should be addressed to Shir Dekel, Brennan MacCallum Building (A18) Camperdown, NSW 2006, Australia. E-mail: shir.dekel@sydney.edu.au

Abstract

Business executives often have to allocate resources across very dissimilar projects. They use financial measures, such as Net Present Value (NPV) that simplify this difficult comparison because they aim to be equally applicable to any kind of project, but these measures vary in their reliability. Psychological research suggests that comparing alignable objects will be easier than comparing non-alignable objects (Markman & Gentner, 1993; Markman & Medin, 1995). However, it is unclear how alignment might moderate people's use of financial measure such as NPV. We found that laypeople accommodate their use of a financial measure (NPV) based on its reliability (as explicitly described in the introduction to the task) when allocating resources to a set of alignable projects, but use it regardless of reliability when allocating to a set of non-alignable projects. However, when NPV reliability information was presented numerically using ranges, participants' allocation did not depend on the ranges—participants used NPV even when they had an opportunity to use the intrinsic features of the project. Overall, however, participants relied on NPV more when projects were low in alignment than when they were high in alignment. The result with numerical reliability was replicated with Masters of Management students. Our results demonstrate that considering dissimilar choices may hinder people's ability to evaluate their importance, and that people might not be using useful variance information in their decisions.

Keywords: structural alignment, similarity, variance neglect, Net Present Value, capital allocation

Word count: 7040

Project Similarity Bias and Variance Neglect in Forecast Metric Evaluation

One of the most important tasks faced by executives is the allocation of capital within their companies. This requires the ranking of projects by importance and predicted success, and allocating the limited capital accordingly (not unlike a scientific funding agency). Ranking of projects necessitates comparing them across a number of dimensions. For example, the executive of an oil company may have received multiple oil exploration proposals. Determining what makes one oil exploration project better than another is relatively simple. However, consider a different scenario in which the executive must allocate capital between an oil exploration project and an oil refinery project. The dimensions of oil refinery projects that distinguish superior from inferior projects may be totally different from those of oil exploration projects. Consider a funding agency having to decide between two cognitive scientists or between a cognitive scientist and a physicist in awarding a fellowship. What makes a physics proposal better for the field of physics than a cognitive science proposal for cognitive science?

Structure-mapping theory (Gentner, 1983; Gentner & Markman, 1997) provides a model of comparison that psychologically distinguishes these two kinds of allocation tasks. This framework models comparison as a process of mapping and alignment of the shared dimensions of two conceptual structures. This mapping process reveals the shared dimensions of the two structures as well as the differences in those shared dimensions (known as *alignable differences*). For example, when comparing two oil exploration projects, the process for measuring the quantity of hydrocarbons in a prospective oil field may be identical, but the specific quantities measured will differ. This is known as an alignable difference; that is, the difference constrained within the same dimension. However, when comparing an oil field and a refinery, there will be a significantly higher number of *non-alignable differences*, because the two domains do not share component dimensions. That is, the dimensional structure of processes in the exploration project will be substantially different from that of processes in the refinery project, making it difficult to find meaningful alignments. With a higher number of non-alignable differences, there are fewer opportunities to make meaningful comparisons, leading to greater difficulty in predicting project success and ranking projects. The present study experimentally examined project comparisons and how such comparisons may affect capital allocation decisions. The working hypothesis is that projects that have a higher number of alignable differences will lead to more precise and informed project predictions and rankings compared with projects with non-alignable differences.

However, what happens when a task demands that two domains be aligned but they are too disparate to align? Experimentally, this territory is somewhat uncharted. It is expected that, when required, people will grasp at any piece of information available and attempt to abstract and infer that which is reasonable to ease the alignment. This occurs frequently in business settings. Because corporate enterprises continue to embrace diversification strategies in their investments, they must constantly make capital allocation decisions involving highly disparate domains. To overcome these difficult comparisons, executives rely on various financial measures that, in theory, may be applied to any project or business proposal. These financial measures work well to ease the burden of difficult comparisons because they ignore the complexities of individual projects and focus solely on financial information such as total cost and projected profits. Therefore, projects that are difficult to compare may be evaluated more easily by comparing individual numerical measures.

The most common financial measure that is used by executives in order to value business project proposals is NPV (Graham et al., 2015; Graham & Harvey, 2001; Remer et al., 1993). NPV is the difference between the forecasted revenue of a project and the initial investment in its development (accounting for the time value of money), as shown in Equation (1):

$$\mathrm{NPV} = \sum_{t=0}^{n} \frac{R_t}{(1+i)^t}, \tag{1}$$

where t is the time of the cash flow, i is the discount rate, R_t is the net cash flow, and n is the total number of periods. NPV is commonly used in decisions about capital allocation and investment. The basic rule is that if a project has a positive NPV, it is financially viable, and if it has a negative NPV, it is not. However, the use of NPV has been criticised, by both academics and practitioners (Fox, 2008; Willigers et al., 2017). The main criticism is that there can be underlying sources of variance in NPV that are not reflected in the final measure, which is expressed as a single numerical value. For instance, NPV is dependent on the projected cash inflows for each year of the project. However, financial forecasting is frequently inaccurate and prone to optimism bias (Lovallo & Kahneman, 2003; Puri & Robinson, 2007). Therefore, there is bound to be variation in the reliability of NPV measures as a function of the forecasting error in the cash flow calculations. Project duration and the discount rate are further sources of variance that may be hidden by the single numerical value of NPV.

The secondary goal of this research is to investigate the extent to which people are sensitive to variance information (from financial forecasting) when making capital allocation decisions. This consideration is especially important in the capital allocation situations illustrated above, when executives need to compare projects with disparate domains and must, therefore, rely on NPV. This matters because the NPV of different domains may have different underlying forecasting error, potentially compromising the utility of using NPV as the basis of comparison. Do executives sufficiently account for the inherent sources of variance in the measure on which they rely so heavily? Research shows that people are effective at extracting variance information when exposed to numerical sequences (Rosenbaum et al., 2020). However, they struggle to use variance information when it is represented numerically (Batteux et al., 2020; Galesic & Garcia-Retamero, 2010; Konold et al., 1993; Vivalt & Coville, 2021).

Experiment Summary

Experiment 1 investigated the effect of project alignment on the decision-making of naive participants asked to allocate capital to a set of fictional projects. Naive participants were assumed to have no requisite knowledge about NPV reliability; thus, NPV reliability level was manipulated by directly telling participants whether or not the given NPV was reliable. For this experiment, it was predicted that when projects are alignable, participants who are told NPV is reliable would use it in their decision-making, while participants who are told that NPV was unreliable would not use it in their decision-making. However, when projects are not alignable, it was predicted that participants would use NPV, regardless of the stated NPV reliability level.

Experiment 2 investigated the decision-making of management students in a similar situation to Experiment 1. The main difference was that instead of telling participants whether or not the NPV was reliable, the level of *numerical* NPV reliability—that is, the width of the numerical range around the average NPV—was manipulated. Similar to Experiment 1, it was predicted that participants would rely more on NPV in non-alignable projects than in alignable projects. However, it was predicted that numerical reliability level would have no effect because there is little evidence that people are sensitive to variance information when it is shown numerically.

Experiment 3 also tested the effects of project alignment and reliability level in a non-business population but manipulated both verbal and numerical reliability to enable a direct comparison. The term *reliability level* is used to describe the manipulation of whether NPV was expressed as a reliable measure or not, while *reliability type* is used to describe the manipulation of whether reliability was expressed verbally or numerically. Experiment 3 predicted a reliability level effect for the verbal reliability condition but not the numerical reliability condition. Further, this experiment used project descriptions with clearer profitability indicators and added a larger selection of business industries.

Experiment 1

Experiment 1 investigated the effects of project alignment and explicit NPV reliability information on capital allocation decisions. The structural alignment literature suggests that people place more weight on alignable differences than they do on non-alignable differences. It was expected that participants would rely more on NPV than on other product attributes in their decision-making because NPV may be applied to every product. However, this effect should vary with participants' perceived NPV reliability level. That is, if other project dimensions are alignable, the use of NPV may depend on its reliability. However, it was predicted that in projects with low alignment, there will be a greater reliance on NPV as the sole alignable difference, regardless of its stated reliability. These effects were measured by considering the linear relationship between NPV and the money allocated to each project. Critically, the NPV and intrinsic features of each project shown to participants were inversely related. Therefore, a positive NPV trend will indicate a heavier reliance on NPV, whereas a negative trend will indicate a heavier reliance on the intrinsic project features. First, Experiment 1 tested the following omnibus hypothesis:

Hypothesis 1 (overall effect)**.** The alignment × reliability level × NPV interaction is significant.

Initially, specific effects were tested by excluding the no NPV condition (in which participants were not given NPV information). Given the difficulty of comparing dissimilar projects, participants were expected to rely more heavily on NPV when project attributes are not alignable compared with when they are alignable. Therefore, Experiment 1 tested the following hypothesis:

Hypothesis 2 (alignment effect)**.** The linear NPV trend will be stronger for projects with low alignment than for projects with high alignment.

Participants' budget allocations were expected to depend on the provided NPV reliability information. However, this is more likely when there are multiple aligned metrics from which to choose compared with when NPV only is alignable. Therefore, Experiment 1 tested the following hypothesis:

Hypothesis 3 (the NPV reliability level effect depends on alignment)**.** The NPV × reliability level interaction will be stronger in the high alignment than in low alignment.

Specifically, when projects are similar, it is expected that participants will rely more on NPV if they are told that NPV is reliable (leading to a positive NPV trend) but more on the intrinsic features of projects if they are told that NPV is unreliable (leading to a negative NPV trend). However, when projects are dissimilar, it is expected that participants will rely solely on NPV, regardless of what they are told about its reliability. Therefore, Experiment 1 tested the following hypotheses:

Hypothesis 4 (NPV reliability level in high alignment)**.** When projects have high alignment, the NPV trend will be stronger when NPV reliability is high compared with when NPV reliability is low.

Hypothesis 5 (NPV reliability level in low alignment)**.** When projects have low alignment, the NPV trend will not differ significantly between the two reliability level conditions.

A no NPV condition was used to gain a better understanding of participants' baseline response to materials when they had no information about NPV. The extent of participants' reliance on NPV was determined by comparing this no NPV condition to the conditions in which NPV was present. When projects are similar, this condition was expected to be equivalent to the low NPV reliability condition because in this condition participants should disregard NPV. When projects are dissimilar, this condition was expected to show the average participant value judgements of the project descriptions, because they only had the intrinsic project features for their evaluations. This was expected to result in a flat NPV trend. Therefore, Experiment 1 tested the following hypotheses:

Hypothesis 6 (effect of NPV information for projects with high alignment)**.** For projects with high alignment, the positive NPV trend will be stronger for projects with high NPV reliability compared with projects with no NPV information.

Hypothesis 7 (effect of NPV information for projects with low alignment)**.** For projects with low alignment, the positive NPV trend will be stronger for projects with both low and high NPV reliability compared with projects with no NPV information.

Method

Participants

One hundred and eighteen participants (55 female) were recruited from the online recruitment platform Prolific. Participants were compensated at a rate of £5 an hour (Prolific is based in the UK). The average age was 29.42 years (*SD* = 9.25, *min.* = 18, *max.* = 73). Table 1 shows the allocation of participants to the different conditions. NPV was varied within subjects.

Materials

Instructions. Participants, who did not necessarily have business experience, were first shown an instructions page with information about the task and NPV. These instructions also informed participants about whether NPV as a financial measure was reliable or unreliable for the specific project. Participants in the low NPV reliability level conditions were told that NPV was an unreliable metric, while those in the high NPV reliability level conditions were told that NPV was a reliable metric. Instructions provided to participants in the no NPV condition did not include an explanation of NPV or its reliability.

Table 1

Experiment 1 group allocation.

Project alignment	Reliability level of net present value (NPV)	N
High	High	26
High	Low	17
High	No NPV	17
Low	High	21
Low	Low	16
Low	No NPV	21
Total		118

Critically, participants were asked to invest in products with a high objective value (because a higher-quality product is not always better in the consumer goods market). Given that participants may not use this instruction when directly viewing the projects, Experiment 3 used projects whose attributes inherently expressed their quality. Appendix A shows the instructions used in Experiment 1.

Project Display. Participants were provided with a set of fictional business projects to which they were asked to allocate capital. Alignment manipulation was reinforced through visual presentation. Projects with high alignment were displayed in a table listing their various attributes (see Figure 1). In this group, each project involved the same product type with consistent concrete attributes. The table format was more appropriate for the high alignment condition because all dimensions were shared. In contrast, projects with low alignment were presented as paragraphs describing their relevant attributes (see Figure 2). In this group, each project was a different product with concrete attributes specific to that product. In both alignment conditions, each project description included an NPV. Critically, the values of the concrete attributes were always in conflict with the NPV. For instance, Project 4 always had the lowest value for each concrete attribute but always had the highest NPV. This meant that participants' allocations could be used as a proxy for their degree of dependence on NPV.

Presentation style was potentially a confounding factor. This was addressed in Experiment 3 by using the table format for both alignment conditions.

	Project 1	Project 2	Project 3	Project 4	Project 5
Product	Laptop	Laptop	Laptop	Laptop	Laptop
RAM (GB)	4	8	32	2	16
Hard drive (GB)	500	750	2000	250	1000
Resolution (px)	900	1080	1440	768	1200
Processor (Ghz)	2.4	3.2	3.8	1.6	3.6
NPV ($)	663	495	70	887	252

Figure 1

An example of a high alignment display in Experiment 1.

Allocation. Participants completed a capital allocation task (see Figure 3) adapted from Bardolet et al. (2011) in which they were asked to allocate a hypothetical yearly budget across the given five projects.

Additional Measures. Other measures apart from allocation were included. The stimuli for and analyses of these measures are reported in Appendix A. Specifically, participants were asked to forecast the future returns of the projects (see Figure A4), rank the projects (see Figure A5), indicate their confidence in their decisions (see Figure A6), and justify their decisions (see Figure A7).

Procedure

After reading the relevant instruction page, participants allocated to the low alignment conditions completed the forecasting task directly beneath each project display.

PROJECT 1

This project is about developing a new shampoo. It will have a 400 mL capacity, and the patented Dandruff Reduction Factor was 17 at testing. The fragrance was optimally effective for 3 metres, and the Safety Authority gave it a 81% safety rating. The NPV is estimated to be $685.

PROJECT 2

This project is about developing a new laundry machine. The machine will have a 12-star energy rating and an 8L capacity. The maximum speed rate is 900 rpm and it will have six different cycle programs. The NPV is estimated to be $500.

PROJECT 3

This project is about developing a new mountain bike. It will have a tensile strength of 910 megapascals, and a suspension for travel of 200mm. It will have a 12-speed cassette and is guaranteed for at least three tours. The NPV is estimated to be $81.

PROJECT 4

This project is about developing a new laptop computer. It will have 2GB of RAM and a hard drive with 250GB capacity. The resolution will be 768px, and the processor speed will be 1.6 Ghz. The NPV is estimated to be $894.

PROJECT 5

This project is about developing a new backpack. It will have eight separate compartments for different types of storage, and a total capacity of 30L. The company will offer a four-year warranty, and the material is an 800-denier nylon fabric. The NPV is estimated to be $251.

Figure 2

An example of a low alignment display in Experiment 1.

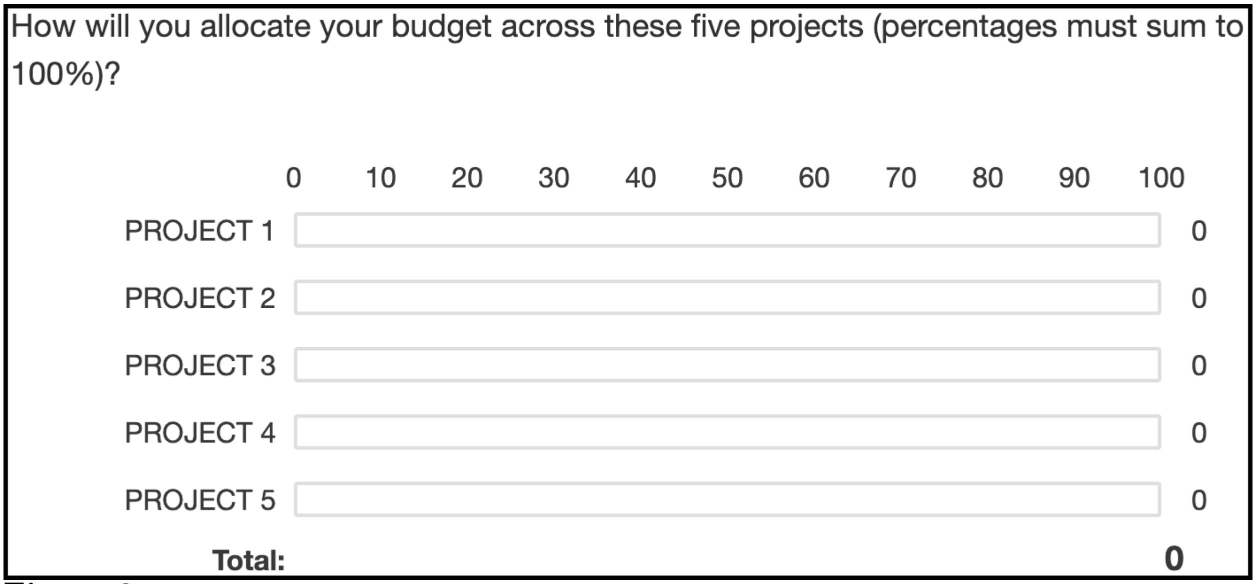


Figure 3

The allocation task.

For the high alignment conditions, this was done directly beneath all projects. Participants were then asked to rank the projects and subsequently answer the allocation, confidence, and justification questions.

Results

A mixed factorial ANOVA was conducted to investigate the effects of project alignment and NPV reliability level on participants' budget allocations. As shown in Figure 4, the alignment × NPV reliability level × NPV interaction was significant, $F(6.57, 367.76) = 2.18$, $p = .039$, $\hat{\eta}_p^2 = .038$. The analyses excluding the no NPV condition showed the expected results. The NPV trend averaged across both reliability level conditions was stronger for the low alignment conditions than for the high alignment conditions, $\Delta M = 61.70$, 95% CI $[33.02, 90.37]$, $t(76) = 4.29$, $p < .001$. This shows that people relied more on NPV when projects were dissimilar than when they were similar.

Further, the NPV × NPV reliability level interaction was stronger in the high alignment conditions than in the low alignment conditions, $\Delta M = 67.81$, 95% CI

$[10.47, 125.16]$, $t(76) = 2.36$, $p = .021$. Specifically, in the high alignment conditions, the NPV trend was stronger in the high NPV reliability condition than in the low NPV reliability condition, $\Delta M = -63.47$, 95% CI $[-100.00, -26.94]$, $t(112) = -3.44$, $p = .001$. In the low alignment conditions, there was no significant difference between the two reliability conditions, $\Delta M = 4.35$, 95% CI $[-34.52, 43.21]$, $t(112) = 0.22$, $p = .825$. This shows that participants only used the NPV reliability information in their allocation decisions when projects were similar, not when they were dissimilar.

The comparison with the no NPV condition revealed the expected pattern. For the high alignment group, the linear NPV trend was significantly weaker in the no NPV condition than in the high NPV reliability condition, $\Delta M = 75.70$, 95% CI $[39.17, 112.24]$, $t(112) = 4.11$, $p < .001$, but not the low NPV reliability condition, $\Delta M = 12.24$, 95% CI $[-27.94, 52.41]$, $t(112) = 0.60$, $p = .547$. However, in the low alignment group, the linear NPV trend was significantly weaker for the no NPV condition compared with both the low NPV reliability condition, $\Delta M = 64.63$, 95% CI $[25.76, 103.50]$, $t(112) = 3.29$, $p = .001$, and the high NPV reliability condition, $\Delta M = 60.29$, 95% CI $[24.14, 96.43]$, $t(112) = 3.30$, $p = .001$.

The mean ranking, confidence, and forecast data were all largely congruent with the allocation findings (see Appendix A). The results also show that the forecasts of those in the low alignment condition had higher standard deviations than those in the high alignment condition (see Appendix A). However, this was not replicated in follow-up experiments (Dekel, 2021, Appendix B.5, 2021, Appendix B.6).

Discussion

Experiment 1 found evidence for the effect of project alignment on laypeople's decision-making in capital allocation scenarios. Specifically, when projects were comparable, participants used NPV when they were told that it was reliable, but did not when they were

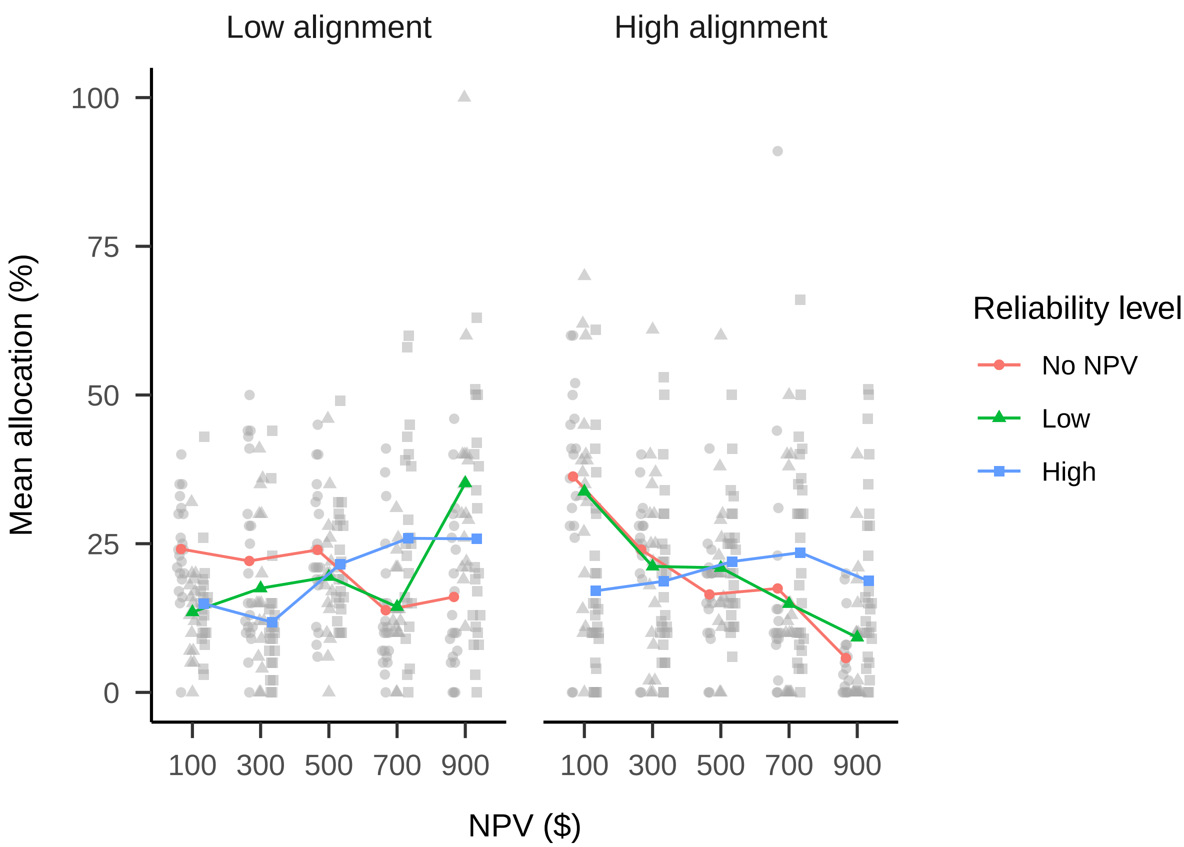


Figure 4

Mean allocation across NPV, by project alignment and reliability level conditions. In mixed factorial designs, error bars cannot be used to make inferences by "eye" across all conditions. Therefore, error bars are not included. Raw data are plotted in the background. When interpreting this figure, consider the linear trends in NPV.

told that it was unreliable. However, they used NPV regardless of its reliability when it was the only shared dimension across products.

Experiment 1 manipulated *verbal* NPV reliability. That is, participants were explicitly told whether NPV was considered to be a reliable metric or not. However, in the real-world the reliability of a metric is more commonly expressed in numerical form, such as a range around an estimate. Experiment 2 attempted to replicate the alignment effects, while manipulating the *numerical* NPV reliability associated with each project, rather than the verbal reliability as used in Experiment 1. Further, people with sufficient experience with financial theory and analysis may be able to successfully draw inferences from such information. Therefore, Experiment 2 used a sample of students enrolled in a Master of Management degree, instead of the laypeople used in Experiment 1.

Experiment 2

Experiment 2 investigated the effects of project alignment and numerically-expressed NPV reliability information on capital allocation decisions. In Experiment 1, the information about NPV reliability level was communicated explicitly (e.g., "NPV is unreliable"). However, in Experiment 2, only the actual NPV information itself was communicated without the conclusion about its reliability. Specifically, participants were given a range of predicted values (akin to a confidence interval). Therefore, while Experiment 1 manipulated *verbal* NPV reliability, Experiment 2 manipulated *numerical* NPV reliability. Further, Experiment 2 included participants with more business experience. This experiment tested whether the previous findings of an alignment effect would be replicated using participants with more business experience. The experiment also tested whether this population is sensitive to variance in forecasts.

Hypothesis 2 was again tested to investigate the alignment effect in the new sample. However, the other hypotheses tested in Experiment 1 were not retested because Experiment 2 manipulated numerical reliability and did not include a no NPV condition. Research has shown that people are poor at reasoning with numerical variance information (Batteux et al., 2020; Galesic & Garcia-Retamero, 2010; Konold et al., 1993; Vivalt & Coville, 2021). Therefore, Experiment 2 tested the following hypothesis:

Hypothesis 8 (no effect of numerical reliability)**.** The NPV $\times$ reliability level interaction is not significant in either alignment condition.

Experiment 2 also investigated the ability to quickly change participants' understanding, if they did not initially use numerical NPV reliability in their allocations.

Therefore, participants were presented with a short lecture on the importance of paying attention to variance in financial decision-making. However, this lecture was not sufficient to inform participants' use of numerical reliability (see Appendix B). Further, Experiment 2 investigated whether participants would be over-confident in their understanding of NPV (as in Long et al., 2018). These results are also reported in Appendix B because they are not highly relevant to the present study.

Method

Participants

Fifty-four participants (28 female) were recruited from a Master of Management degree at an Australian university. Age information was not recorded. Both the reliability level (low and high) and project alignment (low and high) conditions were presented to subjects, and the order of presentation was counterbalanced.

Materials

Instructions. Participants were shown similar instructions to those used in Experiment 1. However, they were given more NPV information (including discount rate and initial investment). Appendix B shows the full instructions.

NPV Test. Participants were asked to complete a short, simple test to check their understanding of NPV (see Appendix B).

Project Display. As shown in Figures 5 and 6, projects were displayed as they were in Experiment 1. However, a second set of projects with different product types and descriptions was added to enable within-subjects manipulation. Along with the single numerical NPV, participants were provided with the forecasted cash flow ranges used to calculate the NPV. In the low NPV reliability condition, ranges were ±85% around the mean (e.g., $150–$1,850 if forecast mean was $1,000); while in the high NPV reliability condition, ranges were ±5% around the mean (e.g., $950–$1,050 if the forecast mean was

$1,000). A wide range indicated that the measure had low reliability, while a narrow range indicated that the measure had high reliability. Participants were told to treat each set of projects independently.

PROJECT 1

This project is about developing a new pair of wireless earphones. They will have a frequency response of 16-40,000Hz and a sensitivity of 90 db/mW. The battery life is 8 hours and the pair will isolate noise up to 35 dB. The range of the cash inflow for the first year is $861-$10,619. The NPV is $227.27.

PROJECT 2

This project is about developing a new wrist watch. It will be water resistant up to 50m and will have one extra timing feature. The hardness of the glass face is rated 4 on the Moh scale and the strap is 10% leather. The range of the cash inflow for the first year is $966-$11,914. The NPV is $881.82.

PROJECT 3

This project is about developing a new treadmill. It will 12 training programs for different interests and abilities and 10 speed levels. It will also have two small compartments for storage and three adjustable inclination levels. The range of the cash inflow for the first year is $832.50-$10,267.50. The NPV is $81.82.

PROJECT 4

This project is about developing a new couch. It will have a guarantee for 10 years and a lightfastness level of 5. The cover's ability to resist abrasion has been tested to handle 50,000 cycles and it has a softness rating of 70%. The range of the cash inflow for the first year is $906-$11,174. The NPV is $490.91.

PROJECT 5

This project is about developing a new bottle of perfume. It will have a 100 mL capacity, and the scent will be able to last for 12 hours. The fragrance concentration will be 30%, and there will be two layers of notes to the scent. The range of the cash inflow for the first year is $925.50-$11,414.50. The NPV is $654.55.

Figure 5

An example of a low alignment, low reliability display in Experiment 2.

NPV Knowledge Ratings. Participants were asked to rate their confidence in knowledge of NPV at multiple points in the experiment. Appendix B shows an example of

	Project 1	Project 2	Project 3	Project 4	Project 5
Product	Laptop	Laptop	Laptop	Laptop	Laptop
RAM (GB)	4	8	32	2	16
Hard drive (GB)	500	750	2000	250	1000
Resolution (px)	900	1080	1440	768	1200
Processor (Ghz)	2.4	3.2	3.8	1.6	3.6
Cash inflow range for Year 1 ($)	$5,890-$6,510	$5,738-$6,342	$5,244-$5,796	$6,137-$6,783	$5,538.50-$6,121.50
NPV ($)	$636.36	$490.91	$18.18	$872.73	$300.00

Figure 6

An example of a high alignment, high reliability display in Experiment 2.

this display.

Variance Lecture. Participants were given a short lecture on the importance of paying attention to variance information in an attempt to increase their use of numerical reliability information in their allocations (see Appendix B for more details and the lecture slides).

Procedure

Participants were provided with the instructions and an explanation of NPV before completing a simple test to demonstrate their understanding of NPV. They then completed four counterbalanced capital allocation trials (one for each condition combination) before viewing a brief presentation on the importance of paying attention to variance in financial decision-making. Participants then repeated two of the trials that they had completed earlier. They were shown the allocation values they had provided earlier and were given the opportunity to change them. Participants rated their knowledge of NPV before and after completing the NPV test and then rated it again after completing the four project displays. They were then asked to rate their knowledge of NPV before and after the variance presentation.

Results

A within-subjects factorial ANOVA was conducted to investigate the effects of NPV, project alignment, and numerical NPV reliability on participants' project allocations (see Figure 7). The alignment × NPV reliability level × NPV interaction was significant, $F(2.81, 148.75) = 3.95$, $p = .011$, $\hat{\eta}_p^2 = .069$. However, this appeared to be driven by the difference between alignment conditions in the interaction between the quadratic NPV trend and NPV reliability level, $\Delta M = -42.28$, 95% $\text{CI}_{\text{Sidak(5)}}$ $[-76.96, -7.59]$, $t(53) = -3.14$, $p_{\text{Sidak(5)}} = .011$, even after applying a Šidák correction. The same interaction but using a the linear NPV trend was not significant, $\Delta M = -6.13$, 95% $\text{CI}_{\text{Sidak(5)}}$ $[-31.50, 19.25]$, $t(53) = -0.62$, $p_{\text{Sidak(5)}} = .954$. Further, the linear NPV trend did not differ between the reliability level conditions in either the low alignment condition, $\Delta M = -3.19$, 95% CI $[-18.77, 12.40]$, $t(53) = -0.41$, $p = .683$ or the high alignment condition, $\Delta M = 2.94$, 95% CI $[-12.63, 18.52]$, $t(53) = 0.38$, $p = .706$. However, averaged across reliability level, the linear NPV trend was stronger in the low alignment condition than in the high alignment condition, $\Delta M = 28.19$, 95% CI $[5.57, 50.81]$, $t(53) = 2.50$, $p = .016$. This suggests that participants relied more on NPV when projects were dissimilar compared with when they were similar.

The ranking data were congruent with these results, while the confidence data were less so. Further, the findings on over-confidence from Long et al. (2018 Study 1) were not replicated with NPV knowledge, and the variance lecture did not facilitate participants' use of numerical reliability information. These analyses are reported in Appendix B.

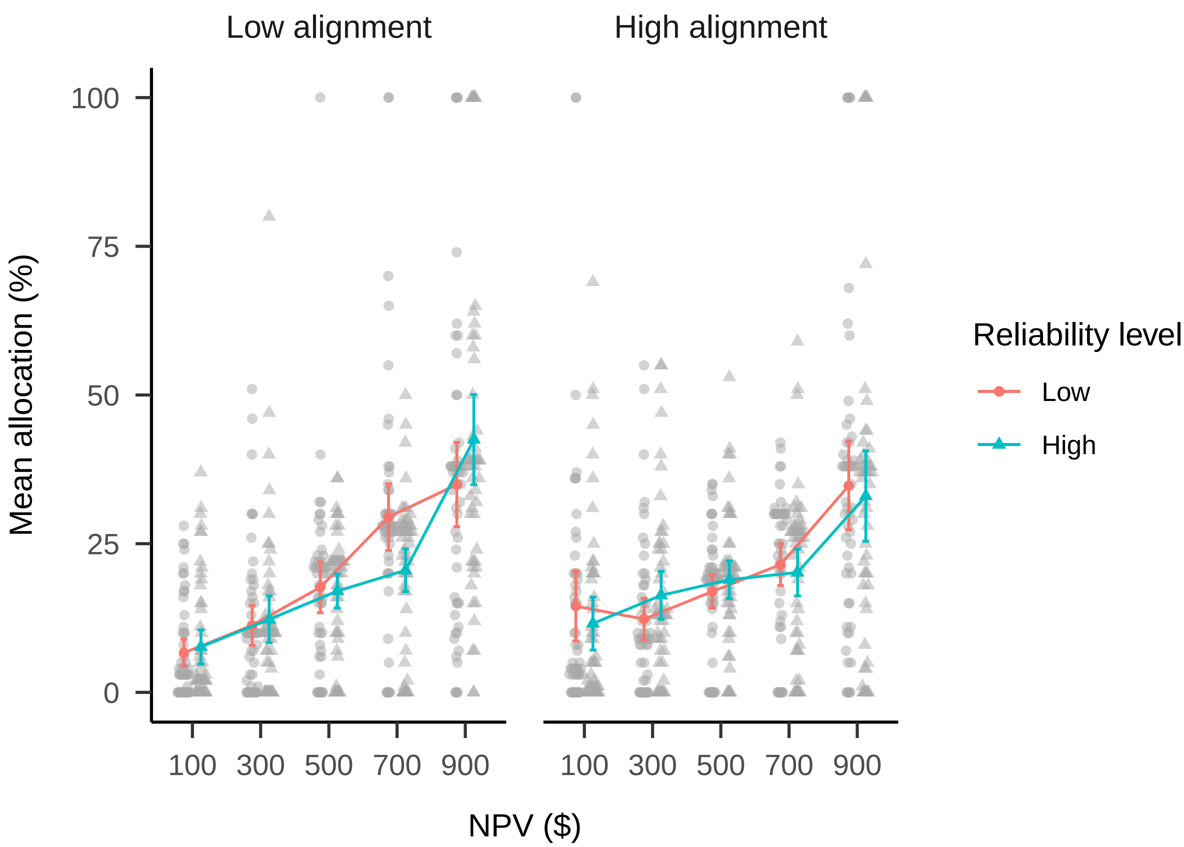


Figure 7

Mean allocation across NPV, by project alignment and reliability level conditions. Error bars represent 95% confidence intervals, calculated from the within-subjects standard errors using the method from Cousineau and O'Brien (2014). Raw data are plotted in the background.

Discussion

Based on participants with real-world business experience, Experiment 2 replicated the alignment effect found in Experiment 1. That is, participants relied more on NPV when faced with a set of dissimilar projects than when faced with similar projects, supporting Hypothesis 2. Experiment 2 also found evidence for Hypothesis 8, with no significant differences between the numerical reliability conditions. While Experiment 2 did not replicate the interaction found in Experiment 1, it should be emphasised that these are two different effects. In Experiment 1, participants were explicitly told whether the NPV measure was reliable, while in Experiment 2, they were provided with variance information that merely implied NPV reliability. Thus, the results of Experiment 2 show that business students were affected by the comparability of projects but not by numerical NPV reliability information. Specifically, participants appeared to focus only on the NPV itself for a specific project, not on the underlying noisiness of the measure.

The participants in Experiment 2 seemed to rely on NPV more than those in Experiment 1. This was seen by the steeper linear trends in Experiment 2. This discrepancy may be due to the difference in domain experience and exposure to financial metrics in formal study. However, the extra explanation and testing of NPV for the management students may have also increased its salience. In sum, the Experiment 2 sample showed clearer trends of NPV reliance, but importantly was still affected by similarity even when it was manipulated within-subjects.

Experiment 1 tested NPV reliability expressed verbally, while Experiment 2 tested NPV reliability expressed numerically. However, the difference in findings was confounded by the different populations that were sampled. Further, in both experiments, the business projects consisted of a limited number of domains. It is unclear to what extent these specific domains influenced the results. These projects were centred around consumer products, which were chosen to be more easily accessible to participants without business experience. However, the individual features of a project do not necessarily indicate its profitability. For instance, a laptop with a low storage capacity may be more profitable than one with a high storage capacity because of consumer goods markets. Experiment 3 addressed these limitations.

Another limitation of Experiments 1 and 2 was the potential confounding effect of presentation style. The two alignment conditions differed in the number of alignable differences, but also in the way that the information was presented. The information in the low alignment condition was presented as paragraphs, while the information in the high alignment condition was presented as a table. While it is likely that these data types would be presented in this way in the business setting, it is important to rule out that this difference did not unnecessarily increase task difficulty. Therefore, Experiment 3 attempted to replicate this effect while controlling for presentation style

Experiment 3

Experiment 3 investigated the effects of project alignment, NPV, NPV reliability type and NPV reliability level on participants' budget allocations. Experiment 1 manipulated NPV reliability level using verbal prompts. That is, participants were explicitly told whether or not NPV was reliable for a certain project. Experiment 2 investigated whether people were able to extract the same reliability information using numerical prompts. That is, participants were provided with NPVs with either wide or narrow ranges, indicating either low or high reliability, respectively. Moreover, given that laypeople were sampled for Experiment 1, and Master of Management students were sampled for Experiment 2, it was not possible to compare the two reliability types (verbal and numerical) without ruling out the potential confounding effect of population type. Thus, similar to Experiments 1 and 2, Experiment 3 manipulated project alignment, NPV and NPV reliability level but also added reliability type. Further, presentation style was a possible confounding factor in the previous experiments. That is, projects in the high alignment condition were always displayed in a table, while projects in the low alignment condition were always displayed as paragraphs. This possible confounder was excluded in Experiment 3 by using the same presentation style for both alignment conditions.

In Experiment 3, the expected results for the verbal reliability condition replicated those of Experiment 1. The numerical reliability condition may replicate the findings of Experiment 2. However, a pilot experiment (Dekel, 2021, Appendix B.8) found no significant differences between numerical reliability conditions. Appendix C shows a simulation of the hypothesised effects, with the numerical reliability effects based on the findings of the pilot experiment. Therefore, Experiment 3 retested Hypotheses 1, 2, 3, 4, and 5 for the verbal

Table 2

Experiment 3 group allocation.

Project alignment	Reliability type	N
High	Explicit	112
High	Implicit	112
Low	Explicit	112
Low	Implicit	112
Total		448

reliability condition, but was agnostic between whether the numerical reliability condition will look more like the pattern found in the pilot experiment or the pattern found in Experiment 2.

Method

Participants

Four hundred and forty-eight participants (176 female) were recruited from the online recruitment platform Prolific. Participants were compensated at a rate of £5 an hour (Prolific is based in the UK). The average age was 41.65 years (*SD* = 10.3, *min.* = 29, *max.* = 78). Participants reported an average of 6.94 years (*SD* = 8.23, *min.* = 0, *max.* = 43) working in a business setting, and an average of 3.73 years (*SD* = 6.27, *min.* = 0, *max.* = 45) of business education. The mean completion time of the task was 11.35 min (*SD* = 10.79, *min.* = 1.92, *max.* = 183.7). Table 2 shows the allocation of participants to the different conditions. The two reliability level conditions (low and high) were presented within subjects and the order of their presentation was randomised. Similar to the previous experiments, NPV varied within subjects. Therefore, each participant saw two separate project displays. Appendix C describes the power analysis conducted to arrive at the sample size.

Materials

Instructions. Participants were given instructions similar to those in the previous experiments, with an added explanation about the NPV reliability information for each reliability type (see Appendix C). Further, they completed a test of basic NPV understanding. Further, they completed a test on basic NPV understanding, which also functioned as an attention check.

Project Display. The project displays were similar to those used in the previous experiments. However, participants were given the same presentation style for both alignment conditions. Each display had a table describing the projects in the set, including ranking and allocation inputs. Project details were presented as bullet points within the relevant cells of the table. Figure 8 shows an example of a low alignment, low verbal reliability display; and Figure 9 shows an example of a high alignment, high numerical reliability display.

Three elements were counterbalanced: (a) the association between reliability level and project set (two variations), (b) the association between business name and NPV (five latin square variations), and (c) project variation (five variations per alignment condition). When counterbalancing for the high alignment group, projects varied by project type (e.g., whether the five projects all described oil wells or microchips). When counterbalancing for the low alignment group, projects varied by their intrinsic features (e.g., whether the oil well project in the set indicated a probability of finding oil of 96% or 90%). Table column order and project display order were both randomised.

Interstitial Page. Prior to each project being displayed, participants were shown an interstitial page, which was used to (a) introduce the next display, and (b) check the participant's attention (given that no input was required, participants could easily skip the page without reading the text). See Appendix C for an example.

Carefully read through the project descriptions below and then do the following: 1. Rank the projects between 1 (highest) and 5 (lowest) in order of investment priority in the relevant "Project Ranking" row input; and 2. Allocate each project a percentage (a number between 0 and 100) of the total budget in the relevant "Project Allocation" row input.

Relevant information	Project 1	Project 2	Project 3	Project 4	Project 5
Project ranking	Ranking:	Ranking:	Ranking:	Ranking:	Ranking:
Project allocation (%)	Allocation:	Allocation:	Allocation:	Allocation:	Allocation:
Business name	Pressbloom	Cweb	Pharmacore	Erectic	Railmont
Project type	national newspaper	software	pharmaceutical	high-rise construction	railway
Predicted project features	• Newspapers printed: 50,000 a day • Number of weekly advertisers: 80 • Ink that is not discarded due to impurities: 5,000L a day	• Code written: 1,000 lines a day • Security rating: 60% • Number of potential customers in first year: 3 million	• Pills pressed: 300,000 an hour • Shelf life: 20 months • Probability of symptom reduction after a week: 90%	• High-rises built: 8 a year • Probability that the builders complete construction within a month of the due date: 70% • Number of tenant expressions of interest: 100	• Railway lines built: 5 a decade • Number of seats filled by paying customers at peak hour: 2,000 • Time before the train carriages will need to be serviced: 12 years
NPV ($)	501 million. (In this industry, NPV is an **unreliable** predictor of a project's profits.)	611 million. (In this industry, NPV is an **unreliable** predictor of a project's profits.)	722 million. (In this industry, NPV is an **unreliable** predictor of a project's profits.)	806 million. (In this industry, NPV is an **unreliable** predictor of a project's profits.)	416 million. (In this industry, NPV is an **unreliable** predictor of a project's profits.)

Continue

Figure 8

An example of a low alignment, low verbal reliability display in Experiment 3.

Carefully read through the project descriptions below and then do the following: 1. Rank the projects between 1 (highest) and 5 (lowest) in order of investment priority in the relevant "Project Ranking" row input; and 2. Allocate each project a percentage (a number between 0 and 100) of the total budget in the relevant "Project Allocation" row input.

Relevant information	**Project 1**	**Project 2**	**Project 3**	**Project 4**	**Project 5**
Project ranking	Ranking:	Ranking:	Ranking:	Ranking:	Ranking:
Project allocation (%)	Allocation:	Allocation:	Allocation:	Allocation:	Allocation:
Business name	Liquid Pipeline	Enfuel	Petroyield	Refinera	Oilpier
Project type	oil well	oil well	oil well	oil well	oil well
Predicted project features	• Oil extracted: 3,400L an hour • Time the machinery lasts before requiring maintenance: 11 years • Probability of finding oil: 96%	• Oil extracted: 2,000L an hour • Time the machinery lasts before requiring maintenance: 7 years • Probability of finding oil: 90%	• Oil extracted: 3,870L an hour • Time the machinery lasts before requiring maintenance: 13 years • Probability of finding oil: 99%	• Oil extracted: 2,470L an hour • Time the machinery lasts before requiring maintenance: 8 years • Probability of finding oil: 92%	• Oil extracted: 2,940L an hour • Time the machinery lasts before requiring maintenance: 10 years • Probability of finding oil: 94%
NPV ($)	494-546 million. (Midpoint: 520.)	792-876 million. (Midpoint: 834.)	409-453 million. (Midpoint: 431.)	697-771 million. (Midpoint: 734.)	598-662 million. (Midpoint: 630.)

Continue

Figure 9

An example of a high alignment, high numerical reliability display in Experiment 3.

Results

A mixed factorial ANOVA was conducted to investigate the effects of NPV, project alignment, NPV reliability level, and NPV reliability type on participants' project allocations (see Figure 10 for the main results and Appendix C for the remainder of the hypothesised allocation effects). The four-way interaction (alignment × reliability level × NPV × reliability type) was not significant, $F(3.20, 1, 420.19) = 0.71$, $p = .555$, $\hat{\eta}_p^2 = .002$. Regardless, the primary hypotheses were supported.

Verbal Reliability

The three-way interaction (alignment × reliability level × NPV amount) in the verbal reliability condition was not significant, $\Delta M = 13.42$, 95% CI $[-1.27, 28.11]$, $t(444) = 1.80$, $p = .073$. This is because NPV reliability level interacted with NPV in both alignment conditions. This is a different pattern from Experiment 1 where there was no effect of NPV reliability level in the low alignment condition. In the high alignment condition, the interaction between the linear NPV trend and NPV reliability level was significant, $\Delta M = -36.63$, 95% CI $[-47.02, -26.25]$, $t(444) = -6.93$, $p < .001$. Specifically, the trend was stronger for the high reliability condition, $\Delta M = 27.26$, 95% CI $[17.69, 36.83]$, $t(444) = 5.60$, $p < .001$, compared with the low reliability condition, $\Delta M = -9.38$, 95% CI $[-18.86, 0.11]$, $t(444) = -1.94$, $p = .053$. This shows that, similar to Experiment 1, participants' allocations depended on verbally expressed NPV reliability. In low alignment, there was also an interaction between the linear NPV trend and NPV reliability level, $\Delta M = -23.21$, 95% CI $[-33.60, -12.83]$, $t(444) = -4.39$, $p < .001$. This suggests that allocations also depended on verbal reliability in the low alignment condition.

However, another aspect of the data suggests a greater use of NPV in the low alignment condition. The linear NPV trend was stronger in the low alignment condition than in the high alignment condition when averaged over reliability level, $\Delta M = 28.97$, 95% CI $[17.68, 40.26]$, $t(444) = 5.04$, $p < .001$. This suggests that when NPV reliability was expressed verbally, similar to Experiment 1, participants relied more on NPV when projects were dissimilar than when they were similar.

Overall, participants used NPV less when it was described as less reliable in both high and low alignment conditions, and further, used NPV more when projects were less alignable regardless of how reliable NPV was described to be.

Numerical Reliability

The numerical reliability data were analysed differently to the verbal reliability data because the effects of interest here were the alignment and reliability level effects. The linear NPV trend was stronger in the low alignment condition, averaged over reliability level (with Bonferroni adjustment), $\Delta M = 15.19$, 95% $\mathrm{CI}_{\mathrm{Bonferroni(5)}}$ $[0.78, 29.60]$, $t(444) = 2.64$, $p_{\mathrm{Bonferroni(5)}} = .034$. This pattern was the same as that found for the verbal reliability condition above and in Experiment 2. Further, the linear NPV trend was not significantly different between the reliability level conditions for both the low alignment condition, $\Delta M = 1.64$, 95% $\mathrm{CI}_{\mathrm{Bonferroni(5)}}$ $[-11.61, 14.90]$, $t(444) = 0.31$, $p_{\mathrm{Bonferroni(5)}} > .999$, and high alignment condition, $\Delta M = -1.21$, 95% $\mathrm{CI}_{\mathrm{Bonferroni(5)}}$ $[-14.46, 12.05]$, $t(444) = -0.23$, $p_{\mathrm{Bonferroni(5)}} > .999$. This indicates that participants did not use numerical NPV reliability to inform their allocations.

Similar to the verbal reliability condition, the use of NPV was stronger in the low alignment condition than it was in the high alignment condition. However, unlike the verbal reliability condition, allocations did not depend on numerical reliability in either the low or the high alignment condition. In the verbal reliability condition, allocations depended on NPV reliability level in both alignment conditions.

Discussion

Hypotheses 1, 2, 3, and 4 were supported in the verbal reliability condition. This shows that, while overall participants preferred to use NPV as a proxy for project quality in

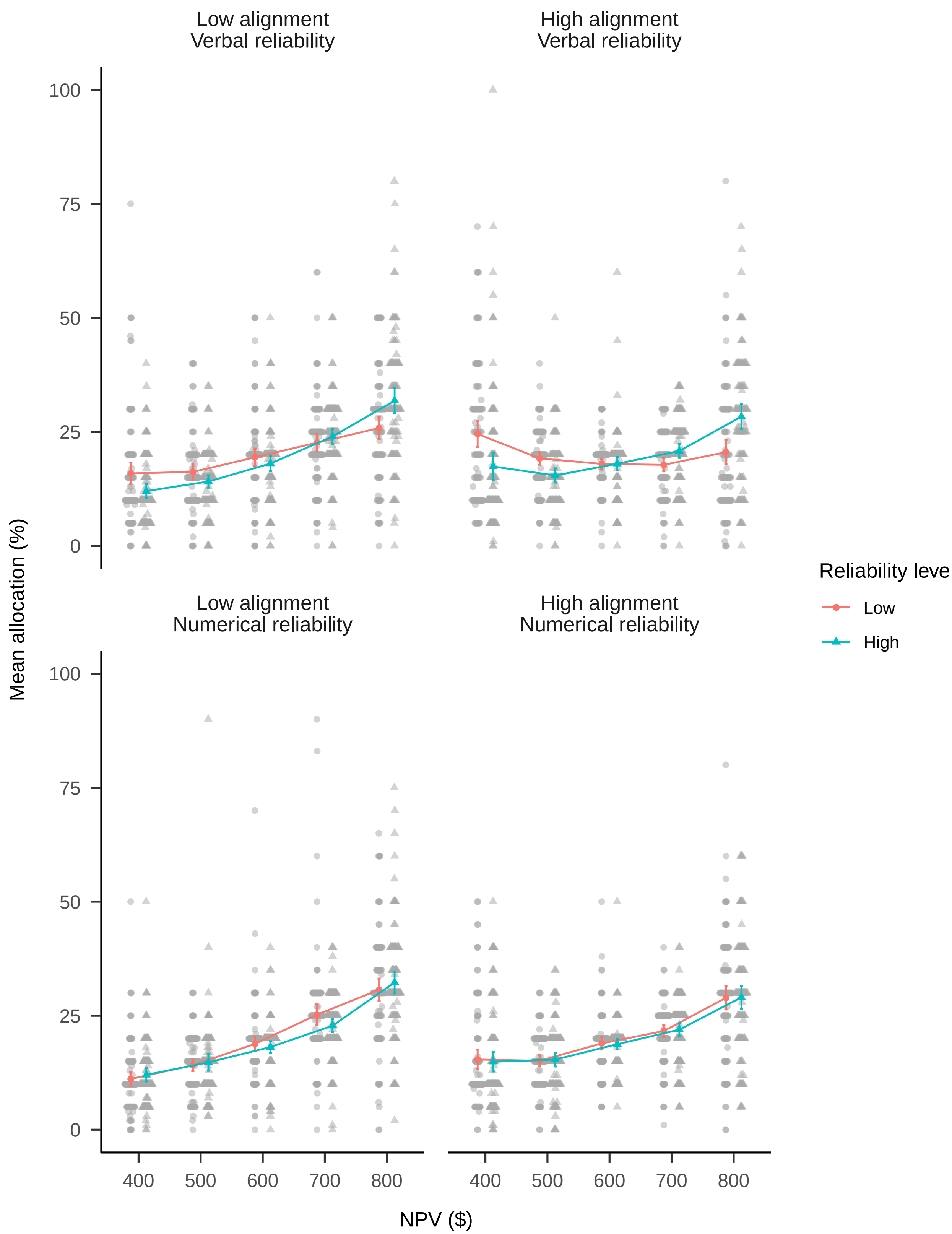


Figure 10

Mean allocation across NPV, by alignment, reliability level, and reliability type conditions. Error bars represent 95% confidence intervals, calculated from the within-subjects standard errors using the method from Cousineau and O'Brien (2014). Raw data are plotted in the background.

their allocations, they still used verbal reliability information. Specifically, when projects were similar, participants used NPV when they were told that it was reliable, and used alternative metrics when told that it was not reliable. However, in Experiment 3, no support was found for Hypothesis 5. It was expected that participants in the low alignment condition would use NPV regardless of the reliability level conditions, as in Experiment 1. Rather, they used NPV less when told that it was unreliable. However, they primarily used NPV overall, as shown by the positive NPV trend in both reliability level conditions.

Further, Experiment 3 replicated the finding of Experiment 2 for the numerical reliability condition. Specifically, participants relied more on NPV when projects were dissimilar but, critically, did not use numerical range information to influence their allocations. A pilot study (Dekel, 2021, Appendix B.8) replicated the results of Experiment 1 in the verbal reliability condition, but did not replicate the results of Experiment 2 in the numerical reliability condition. That is, when faced with numerical ranges as the NPV reliability information, participants did not even use the midpoint in their decisions. The results of Experiment 3 suggest that the finding in the pilot experiment may have been spurious or due to an unexplored component of the experimental design, but this can only be deteremined with future research.

General Discussion

Across three experiments there were two main findings: (a) NPV is used more when options are difficult to compare in the low alignment conditions; and (b) people do not consider numerical variance information, despite this being important to the reliability of the NPV forecasts. This pattern with numerical reliability information contrasted with the frequent use of verbal indicators of reliability level. This numerical variance neglect is surprising, since other work showed that people can readily extract variance information when experiencing numerical sequences (Rosenbaum et al., 2020). Both the verbal and numerical effects were consistent for both naive and experienced participants, indicating their persistence. People make use of metrics with alignable differences when required to compare disparate options. However, they do not always use alternative metrics, even when they are available.

Experiment 1 found that participants did not use NPV in their allocation decisions when they were told that it was unreliable but did use it when told it was reliable. Experiment 2 found that participants with some business experience relied more on NPV for capital allocation when the rest of the information was non-alignable compared with when it was alignable. However, they did not take into account numerical reliability information when making these decisions. Experiment 3 found further evidence of these effects within one experimental design.

Alignable differences have been shown to be important into decision-making in many settings (Markman & Loewenstein, 2010; Markman & Medin, 1995). The experiments presented in the present study are novel in terms of the effects of project alignment on capital allocation. Further, these experiments considered the extent to which the reliability of an alignable measure (NPV) affects the way in which it is used. This depended on the availability of other alignable differences in the set of choices. If other alignable differences were available, then participants were willing to reduce their use of a reportedly unreliable alignable measure (or use it when told that it was reliable). However, when no other alignable differences were available, then the alignable, albeit unreliable, measure was more likely to be used. This was found in both Experiments 1 and 3, as well as in a pilot study to a lesser extent (reported in Dekel, 2021, Appendix B.4).

Financial measures such as NPV are useful because of their alignability. That is, they may serve as an alignable difference, regardless of the inherent similarities between a set of projects. Psychologically, these measures are useful because they allow for relevant inferences (Lassaline, 1996) and because they offer an abstraction of concrete details (Doumas & Hummel, 2013). However, the structural alignment account does not directly speak to real-world implications when there is a need for non-alignable comparisons. NPV is a type of abstraction that facilitates the comparison of different aspects of a company. For instance, the use of NPV may facilitate the comparison of an oil field project with a refinery project. However, this increased alignment could actually hide important information because it does not consider the finer complexities inherent in each business unit. The forecasts used to calculate NPV for each business unit are based on different indicators, and there are likely to be differences between each unit's estimates. Thus, one can imagine a continuum of comparisons in which the usefulness of comparison increases with the level of alignability but depends on the level of abstraction that is required to achieve the alignment.

The finding that participants, even those with some business experience, did not sufficiently consider variance information is surprising but understandable. It is surprising because financial decision-making largely depends on the consideration of different sources of variance (e.g., risk, volatility, and uncertainty). At the same time, it is understandable because research from psychology and statistics education shows that statistics students and people in general have a poor ability to draw statistical inferences (e.g., Galesic & Garcia-Retamero, 2010; Konold et al., 1993). Future research should investigate the conditions under which individuals' sensitivity to variance information may be facilitated. For instance, it is unclear whether it is merely salience that is lacking, meaning that visual aids could be useful, or whether a further explicit explanation of statistical inference is necessary. The findings of a pilot experiment suggest that participants struggle to use numerical NPV reliability information, even when given explicit instructions (Dekel, 2021, Appendix B.7).

A possible limitation of these experiments is the use of NPV as the only financial metric. In the business world, there are many metrics that serve similar functions and are used as tools to deal with non-alignable options. Therefore, future research should attempt to replicate the current findings using different financial measures.

Future research should also investigate the boundary conditions of the reliability type effect. That is, people appear to respond to explicit reliability information but not to variance information that only implies reliability. Future research should attempt to identify the minimal variance information that participants need to understand the relevant implications for reliability. Participants may simply not notice the variance information or assume that it is irrelevant. For instance, future research could test participants in a condition in which the variance information is more salient.

References

Bardolet, D., Fox, C. R., & Lovallo, D. (2011). Corporate capital allocation: A behavioral perspective. *Strategic Management Journal*, *32*(13), 1465–1483. https://doi.org/10/cn6xsb

Batteux, E., Bilovich, A., Johnson, S., & Tuckett, D. (2020). *Impressed by Numbers: The Extent to Which Novice Investors Favor Precise Numerical Information in a Context of Uncertainty* (SSRN Scholarly Paper ID 3595409). Social Science Research Network. https://doi.org/10.2139/ssrn.3595409

Cousineau, D., & O'Brien, F. (2014). Error bars in within-subject designs: A comment on Baguley (2012). *Behavior Research Methods*, *46*(4), 1149–1151. https://doi.org/10/f6vdsw

Dekel, S. (2021). *The Psychology of Managerial Capital Allocation* [Doctoral dissertation, The University of Sydney]. https://thesis.shirdekel.com/

Doumas, L. A. A., & Hummel, J. E. (2013). Comparison and Mapping Facilitate Relation Discovery and Predication. *PLoS ONE*, *8*(6), 1–8. https://doi.org/10/gjscsn

Fox, R. (2008). A brief critical history of NPV. *BAA Conference*, 16. http://usir.salford.ac.uk/id/eprint/9291/2/NPV_paper.pdf

Galesic, M., & Garcia-Retamero, R. (2010). Statistical Numeracy for Health: A Cross-cultural Comparison With Probabilistic National Samples. *Arch Intern Med*, *170*(5), 462–468. https://doi.org/10/fmj7q3

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, *7*(2), 155–170. https://doi.org/10/dw52z8

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*(1), 45–56. https://doi.org/10/fm4rrb

Graham, J. R., & Harvey, C. R. (2001). The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics*, *60*(2), 187–243. https://doi.org/10/fpdzrj

Graham, J. R., Harvey, C. R., & Puri, M. (2015). Capital allocation and delegation of decision-making authority within firms. *Journal of Financial Economics*, *115*(3), 449–470. https://doi.org/10/gfvz8d

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in Students' Reasoning about Probability. *Journal for Research in Mathematics Education*, *24*(5), 392. https://doi.org/10/bq4hvm

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10/gdj7s9

Lassaline, M. E. (1996). Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 754–770. https://doi.org/10/fq9fww

Long, A. R., Fernbach, P. M., & De Langhe, B. (2018). Circle of Incompetence: Sense of Understanding as an Improper Guide to Investment Risk. *Journal of Marketing Research*, *55*(4), 474–488. https://doi.org/10/gjscr7

Lovallo, D., & Kahneman, D. (2003). Delusions of Success: How Optimism Undermines Executives' Decisions. *Harvard Business Review*, *81*(7).

Markman, A. B., & Gentner, D. (1993). Structural Alignment during Similarity

Comparisons. *Cognitive Psychology*, *25*(4), 431–467. https://doi.org/10/cqtx7q

Markman, A. B., & Loewenstein, J. (2010). Structural comparison and consumer choice. *Journal of Consumer Psychology*, *20*(2), 126–137. https://doi.org/10/d7b49c

Markman, A. B., & Medin, D. L. (1995). Similarity and Alignment in Choice. *Organizational Behavior and Human Decision Processes*, *63*(2), 117–130. https://doi.org/10/c8z7r9

Puri, M., & Robinson, D. T. (2007). Optimism and economic choice. *Journal of Financial Economics*, *86*(1), 71–99. https://doi.org/10/c9839j

Remer, D. S., Stokdyk, S. B., & Van Driel, M. (1993). Survey of project evaluation techniques currently used in industry. *International Journal of Production Economics*, *32*(1), 103–115. https://doi.org/10/bsc6bs

Rosenbaum, D. M., Glickman, M., & Usher, M. (2020). *Extracting summary statistics of rapid numerical sequences* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/6scav

Vivalt, E., & Coville, A. (2021). *How Do Policy-Makers Update Their Beliefs?* (p. 51). http://evavivalt.com/wp-content/uploads/How-Do-Policymakers-Update.pdf

Willigers, B. J. A., Jones, B., & Bratvold, R. B. (2017). The Net-Present-Value Paradox: Criticized by Many, Applied by All. *SPE Economics & Management*, *9*(04), 090–102. https://doi.org/10/gjscsx

Appendix A

Experiment 1

Instructions

Figures A1, A2, and A3 show the instructions given to those in the low NPV reliability, high NPV reliability, and no NPV condition, respectively.

Forecasting

Participants were asked to respond to a forecasting task (adapted from Long et al., 2018), seen in Figure A4. Participants were asked to predict each project's rate of return after one month. This allowed to calculate each participant's forecasting mean and standard deviation (the latter as inversely proportional to forecasting precision).

Ranking

As shown in Figure A5, participants were asked to rank the projects in order of investment priority.

Confidence

As Figure A6 shows, participants were asked to indicate how confident they were about each of their allocation decisions on a scale from 0 ("Not confident at all") to 100 ("Extremely confident").

Justification

As Figure A7 shows, participants were asked to justify their allocation decision in a free-response text-box.

You will be shown information about a number of projects that a consumer products firm is considering to invest in. Some specific information about the product itself is provided. In addition to those numbers, you will find each project's net present value (NPV), which is the company's estimation of the future returns of the project. An NPV that is greater than 0 (zero) indicates that there is an expectation of profit. **The higher the NPV, the better the expectations for each project.** However, it is important to note that NPV is a very noisy measure relative to the other more specific measures because it relies on future forecasting. As such, **NPV is very unreliable and should be relied upon only as a last result; the specific project's measures should be used instead.**

We would like you to take the role of the manager in charge of capital allocation for the firm. This firm is specifically interested in investing in the development of high-end goods, so your valuations should reflect this. That is, even though there might be a market for the lower-end products in the descriptions that you will see, **you should be aiming to invest in the products with the highest objective value**. The features of the products that are listed matter because they reflect the direct value of the product, whereas financial measures such as NPV may reflect other factors, thus making it noisier, as mentioned above.

You will see a set of five different projects for which you must predict the investment returns after one month. For example, how likely is it that the project will return more than 9% after one month, how likely is it that the project will return 7% to 9%, etc.

You will also decide how to rank the projects in order of investment priority, and decide how to allocate the capital available for investment this year among the different projects. Note that this is not the operational budget (advertising, etc.), but rather the funds to be used for investment in developing the new products. You will do this by selecting a percentage value for each project, such that the budget is allocated completely among each set of projects.

Figure A1

Experiment 1 low reliability instructions.

You will be shown information about a number of projects that a consumer products firm is considering to invest in. Some specific information about the product itself is provided. In addition to those numbers, you will find each project's net present value (NPV), which is the company's estimation of the future returns of the project. An NPV that is greater than 0 (zero) indicates that there is an expectation of profit. **The higher the NPV, the better the expectations for each project.** However, it is important to note that NPV is a very noisy measure relative to the other more specific measures because it relies on future forecasting. As such, NPV is very unreliable and should be relied upon only as a last result; the specific project's measures should be used instead.NPV is a very useful measure relative to the other more specific measures because it can be calculated regardless of the type of product. As such, **NPV is very reliable in most cases.**

We would like you to take the role of the manager in charge of capital allocation for the firm. This firm is specifically interested in investing in the development of high-end goods, so your valuations should reflect this. That is, even though there might be a market for the lower-end products in the descriptions that you will see, **you should be aiming to invest in the products with the highest objective value.**

You will see a set of five different projects for which you must predict the investment returns after one month. For example, how likely is it that the project will return more than 9% after one month, how likely is it that the project will return 7% to 9%, etc.

You will also decide how to rank the projects in order of investment priority, and decide how to allocate the capital available for investment this year among the different projects. Note that this is not the operational budget (advertising, etc.), but rather the funds to be used for investment in developing the new products. You will do this by selecting a percentage value for each project, such that the budget is allocated completely among each set of projects.

Figure A2

Experiment 1 high reliability instructions.

You will be shown information about a number of projects that a consumer products firm is considering to invest in. Some specific information about the product itself is provided.

We would like you to take the role of the manager in charge of capital allocation for the firm. This firm is specifically interested in investing in the development of high-end goods, so your valuations should reflect this. That is, even though there might be a market for the lower-end products in the descriptions that you will see, **you should be aiming to invest in the products with the highest objective value**. The features of the products that are listed matter because they reflect the direct value of the product, whereas financial measures may reflect other factors.

You will see a set of five different projects for which you must decide how to rank in order of investment priority, and decide how to allocate the capital available for investment this year among the different projects. Note that this is not the operational budget (advertising, etc.), but rather the funds to be used for investment in developing the new products. You will do this by selecting a percentage value for each project, such that the budget is allocated completely among each set of projects.

Figure A3

The instructions for the no NPV condition in Experiment 1.

Additional Analyses

Ranking

A mixed factorial ANOVA was conducted to investigate the effects of alignment and verbally-instructed NPV reliability on participants' rankings of the target project. As shown in Figure A8, the alignment × reliability level × NPV interaction was significant, $F(6.62, 370.54) = 2.70$, $p = .011$, $\hat{\eta}^2_p = .046$. This effect seems to be driven by the differences between the no NPV condition and the conditions with NPV across the two alignment conditions. Specifically, in the low alignment condition, the linear NPV trend was significantly lower in the no NPV condition than both the low reliability condition, $\Delta M = -6.56$, 95% CI $[-10.26, -2.85]$, $t(112) = -3.50$, $p = .001$, and the high reliability condition, $\Delta M = -7.38$, 95% CI $[-10.83, -3.93]$, $t(112) = -4.24$, $p < .001$. However, in the

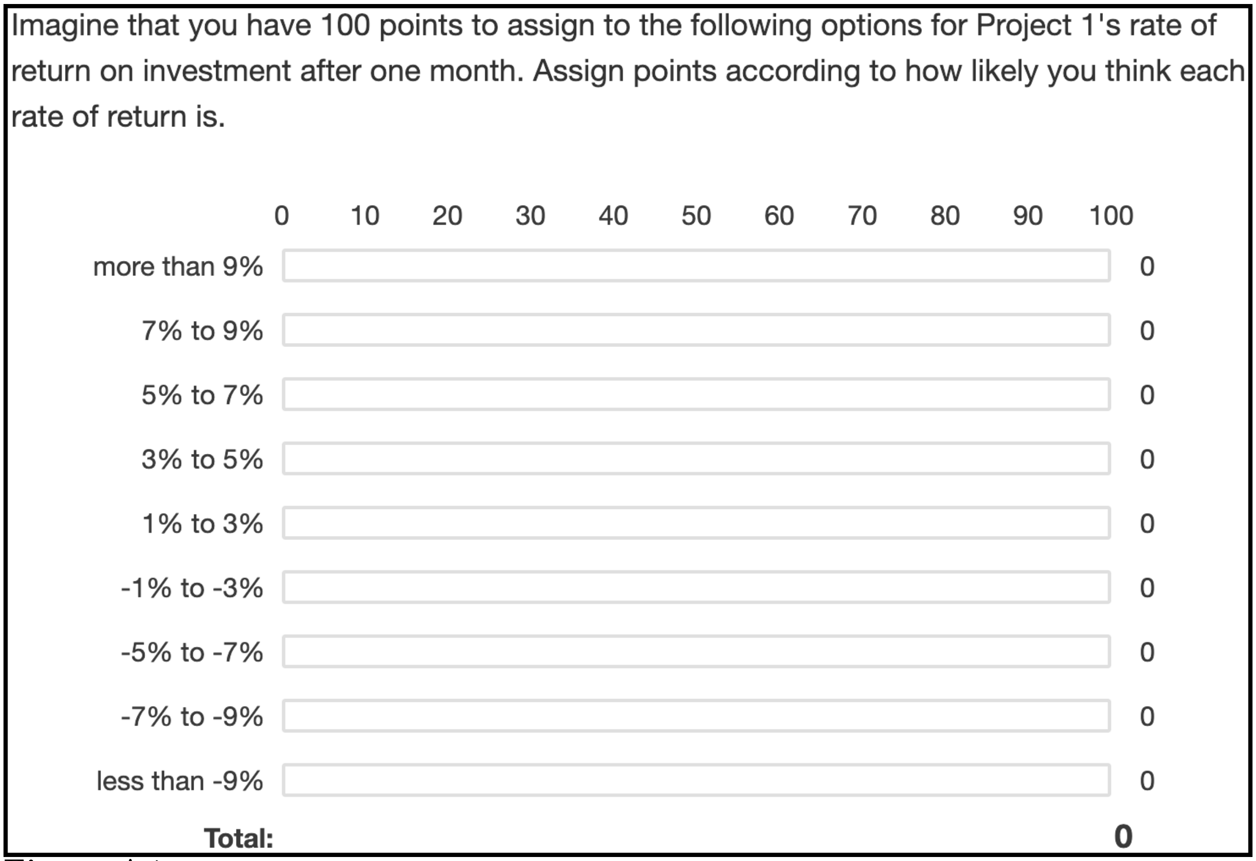


Figure A4

The forecasting task.

high alignment condition, the linear NPV trend was only significantly lower in the no NPV condition than the high reliability condition, $\Delta M = -8.37$, 95% CI $[-11.85, -4.88]$, $t(112) = -4.76$, $p < .001$, and not the low reliability condition, $\Delta M = -1.71$, 95% CI $[-5.54, 2.13]$, $t(112) = -0.88$, $p = .380$.

Confidence

A mixed factorial ANOVA was conducted to investigate the effects of alignment and verbally-instructed NPV reliability on participants' confidence rating of their decisions. As shown in Figure A9, the alignment × reliability level × NPV interaction was not significant, $F(7.47, 418.08) = 1.26$, $p = .267$, $\hat{\eta}_p^2 = .022$. Contrary to the allocation and ranking data, in the low alignment condition, there were no significant differences in the linear NPV trend

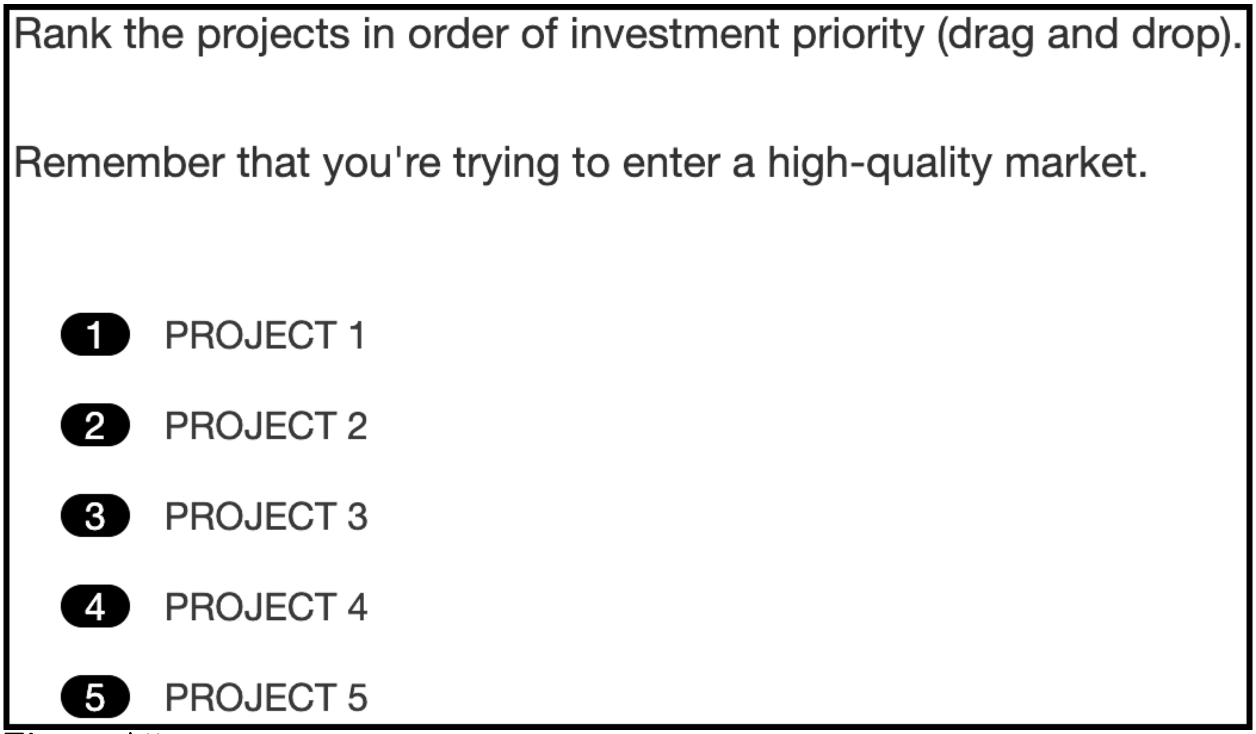


Figure A5

The ranking task.

between the no NPV condition and low reliability condition, $\Delta M = 10.73$, 95% CI $[-30.15, 51.61]$, $t(112) = 0.52$, $p = .604$, nor the high reliability condition, $\Delta M = 13.05$, 95% CI $[-24.97, 51.07]$, $t(112) = 0.68$, $p = .498$. However, as above, in the high alignment condition, the linear NPV trend was significantly lower in the no NPV condition than the high reliability condition, $\Delta M = 65.14$, 95% CI $[26.72, 103.57]$, $t(112) = 3.36$, $p = .001$, and not the low reliability condition, $\Delta M = 31.88$, 95% CI $[-10.38, 74.14]$, $t(112) = 1.49$, $p = .138$.

Forecast Mean

A mixed factorial ANOVA was conducted to investigate the effects of alignment and verbally-instructed NPV reliability on participants' forecast means. As seen in Figure A10, the alignment × reliability level × NPV interaction was not significant, $F(5.26, 142.10) = 1.89$, $p = .095$, $\hat{\eta}_p^2 = .066$. However, the alignment × NPV interaction was

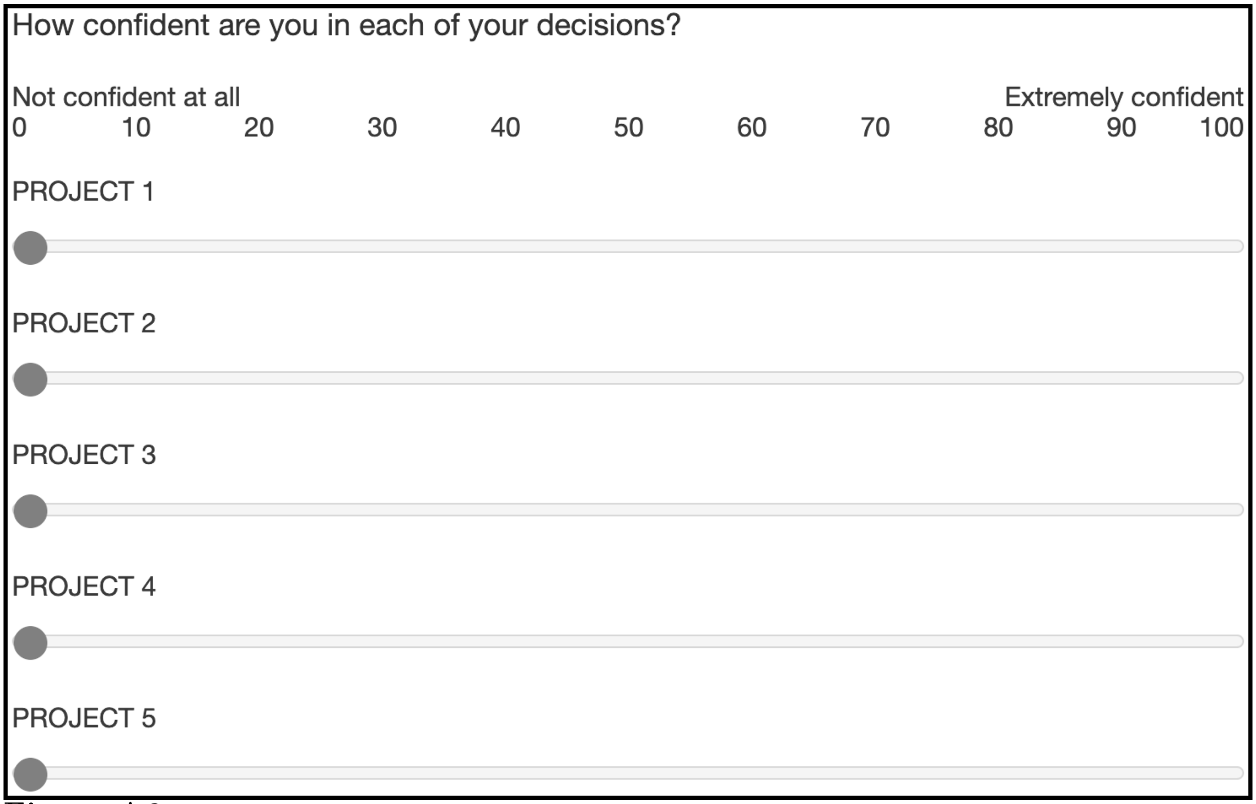


Figure A6

The confidence task.

Figure A7

The justification task.

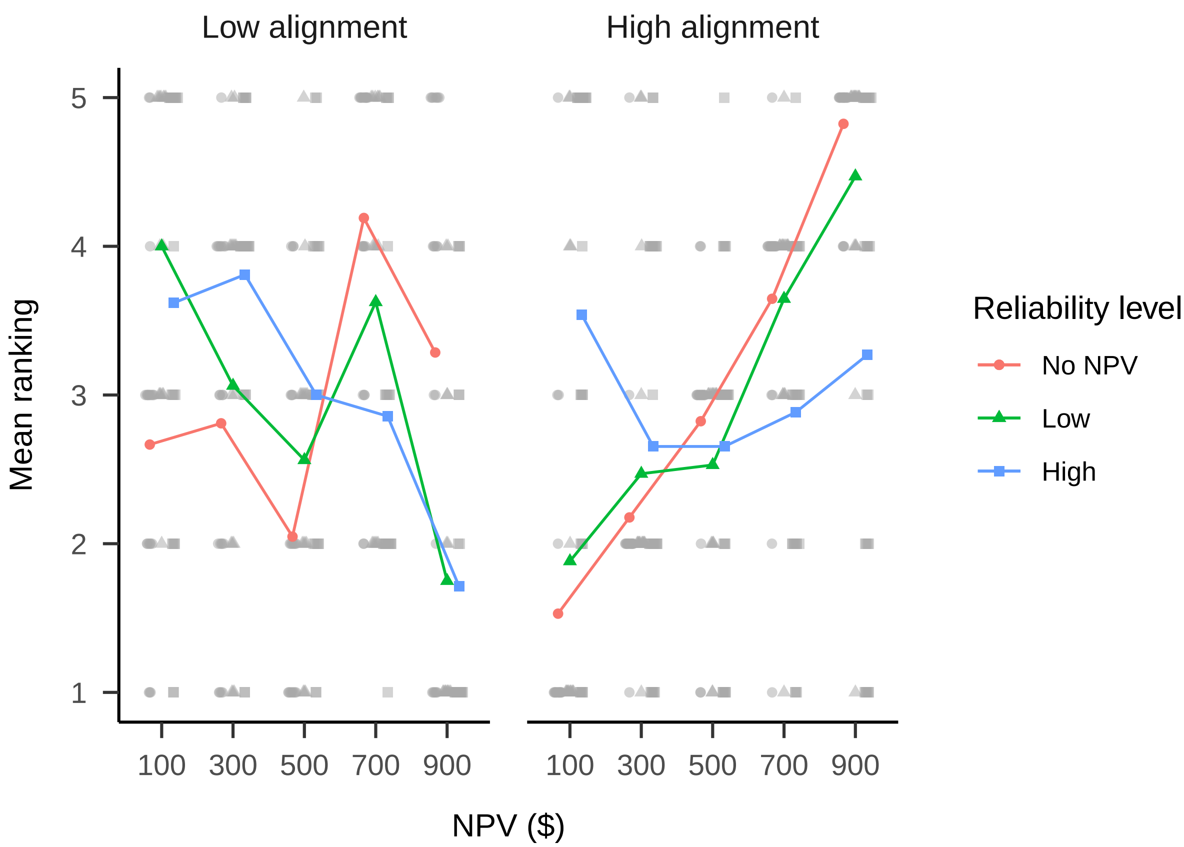


Figure A8

Mean ranking.

significant, $F(2.63, 142.10) = 2.89$, $p = .044$, $\hat{\eta}_p^2 = .051$; as well as the reliability level × NPV interaction, $F(5.26, 142.10) = 7.91$, $p < .001$, $\hat{\eta}_p^2 = .227$. The simple effects appear to be as above. Specifically, in the low alignment condition, the linear NPV trend was significantly lower in the no NPV condition than both the low reliability condition, $\Delta M = 0.19$, 95% CI $[0.09, 0.30]$, $t(54) = 3.63$, $p = .001$, and the high reliability condition, $\Delta M = 0.16$, 95% CI $[0.04, 0.28]$, $t(54) = 2.75$, $p = .008$. However, in the high alignment condition, the linear NPV trend was only significantly lower in the no NPV condition than the high reliability condition, $\Delta M = 0.22$, 95% CI $[0.11, 0.32]$, $t(54) = 4.04$, $p < .001$, and not the low reliability condition, $\Delta M = 0.08$, 95% CI $[-0.04, 0.21]$, $t(54) = 1.30$, $p = .198$.

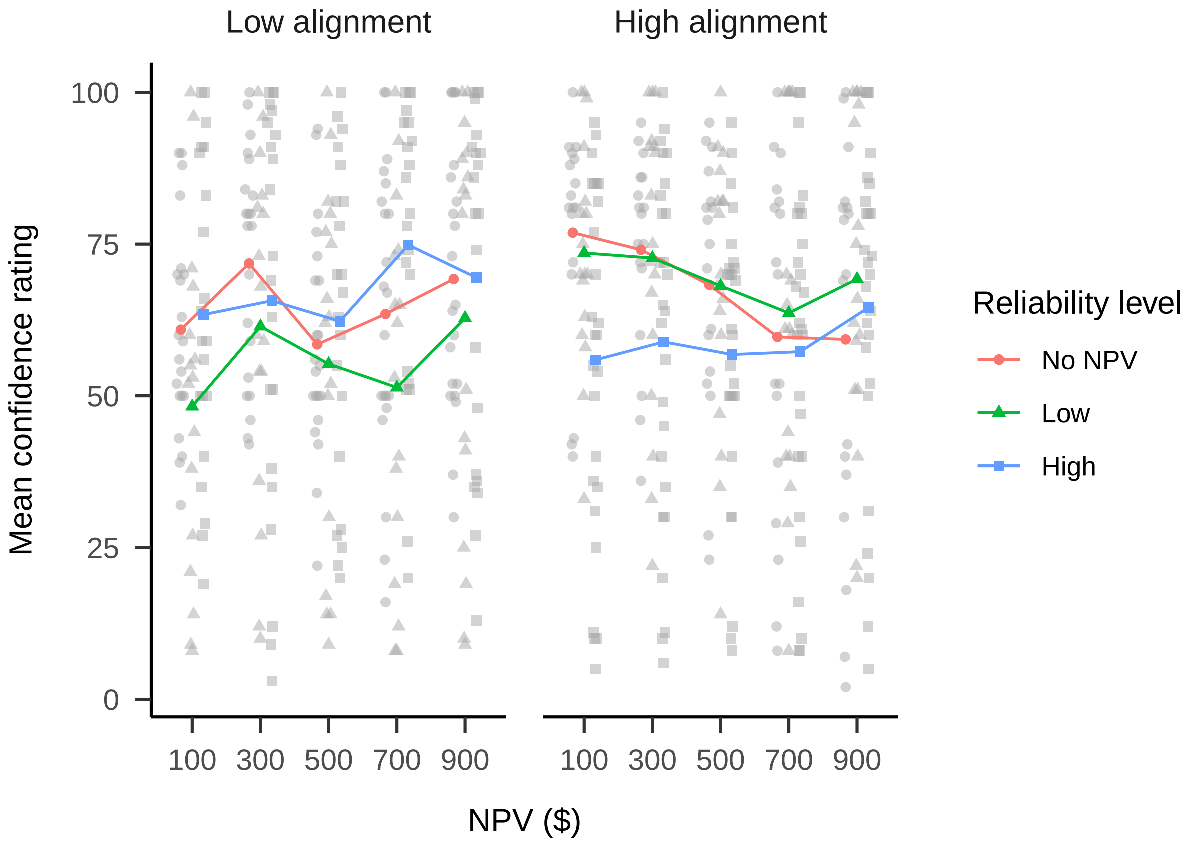


Figure A9

Mean confidence.

Forecast SD

A mixed factorial ANOVA was conducted to investigate the effects of alignment and verbally-instructed NPV reliability on participants' forecast SDs. As seen in Figure A11, the alignment × reliability level × NPV interaction was significant, $F(6.87, 185.42) = 2.91$, $p = .007$, $\hat{\eta}_p^2 = .097$. However, none of the linear NPV trends were significantly different from each other as above. Of relevance, the low alignment condition on average had higher SDs than those in the high alignment condition, $F(1, 54) = 5.77$, $p = .020$, $\hat{\eta}_p^2 = .097$.

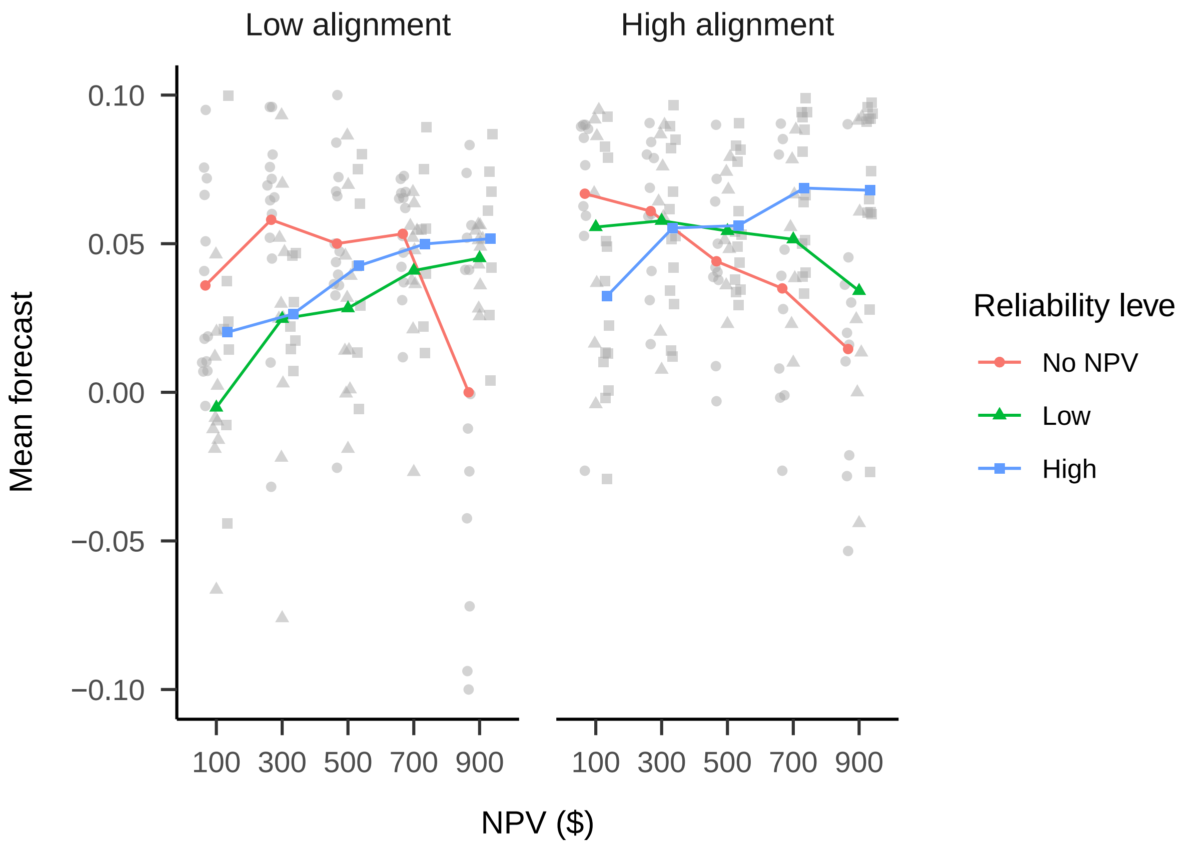


Figure A10

Mean forecasts.

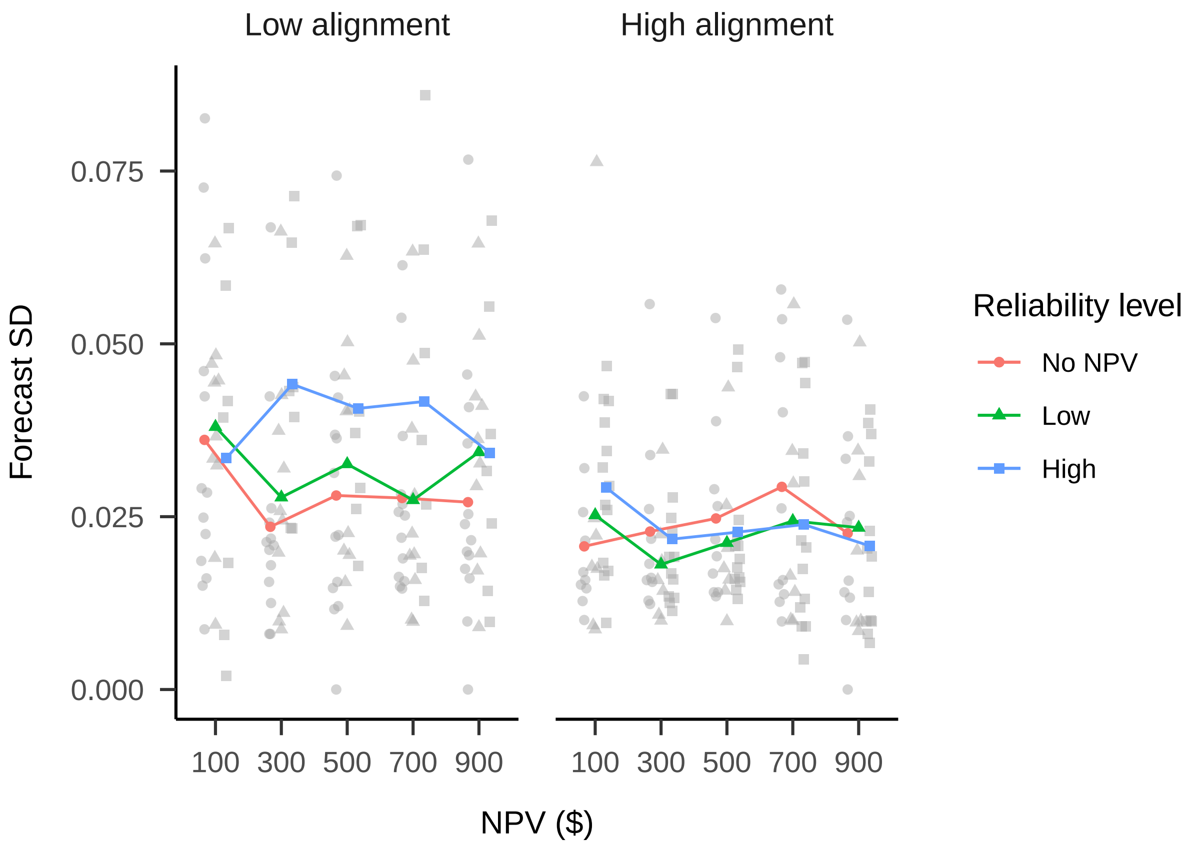


Figure A11

Mean forecast SD.

Appendix B

Experiment 2

Instructions

Figure B1 shows the instructions.

Investment task

You will be shown information about a number of projects that a consumer products firm is considering to invest in. Some specific information about the product itself is provided. In addition to those numbers, you will find each project's projected cash inflow for each year, and the net present value (NPV) that was calculated using those figures. The discount rate will always be 10% and the initial investment will always be $5000. These are taken into account in the NPV calculations.

We would like you to take the role of the manager in charge of capital allocation for the firm. This firm is specifically interested in investing in the development of high-end goods, so your valuations should reflect this. That is, even though there might be a market for the lower-end products in the descriptions that you will see, **you should be aiming to invest in the products with the highest intrinsic quality.**

You will decide how to rank the projects in order of investment priority, and decide how to allocate the capital available for investment this year among the different projects. Note that this is not the operational budget (advertising, etc.), but rather the funds to be used for investment in developing the new products. You will do this by selecting a percentage value for each project, such that the budget is allocated completely among each set of projects.

Importantly, each page's set of five projects should be treated independently of the other pages' project sets.

Figure B1

Experiment 2 instructions.

NPV Test

Participants were given more extensive information about NPV than in the previous experiment and were tested on their ability to calculate simple averages from given numerical ranges, as shown in Figures B2 and B3.

NPV Knowledge Ratings

A similar design to Long et al. (2018 Study 1) was used to test whether this sample may be overconfident in their understanding on NPV. Therefore, participants were asked to rate their knowledge of NPV in various points in the study. Figure B4 shows an example of one such display.

Understanding NPV

Net Present Value (NPV) is used as a measure of a project's potential profitability. A positive value indicates that the project is profitable, while a negative value indicates that a project is not profitable.

When calculating NPV, the potential future cash inflows are converted to their "present values". This is important, because we know that an amount of money is more valuable in the present than it is in the future. The time value of money is accounted for by dividing each year's cash inflow by the discount rate. Finally, the sum of all the present values is deducted from the value of the initial investment.

To calculate the NPV you need the following components:

1. The cash inflow for each year of the project
2. The discount rate
4. The initial investment

Below is the generic formula for calculating NPV:

$$NPV = \frac{Cash\ inflow\ for\ year\ 1}{(1 + discount\ rate)^1} + \frac{Cash\ inflow\ for\ year\ 2}{(1 + discount\ rate)^2} + \frac{Cash\ inflow\ for\ year\ 3}{(1 + discount\ rate)^3} \ldots - Initial\ investment$$

Some of the time, it might be unclear exactly what the future cash inflow is, so it might be given as a range of possible values.

Below is an example of an NPV calculation with the discount rate calculations and initial investment already filled in. Notice that instead of a single cash inflow value, a range is provided (assume that the distribution is uniform). The value that should be used as the cash inflow is the mid point of these values. This is done by calculating the average of the two values.

For this session, you will get some practise in calculating NPV. However, we will give you the value that is in the denominator (the discount rate calculation) and the initial investment. All you need to do is calculate each year's cash inflow and enter them into the formula.

Example 1

$$NPV = \frac{[range: 1500 - 2500]}{1.1} + \frac{[range: 750 - 1250]}{1.21} + \frac{[range: 1875 - 3125]}{1.331} - 3000$$

Please calculate the mid-points for these ranges and type them in below:

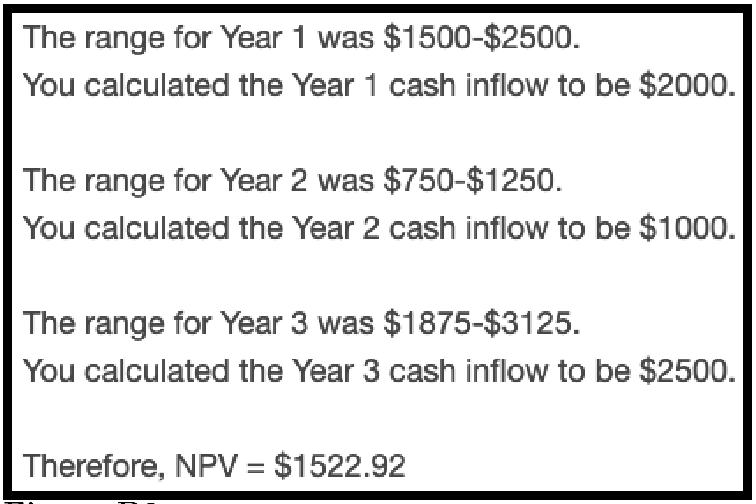
Year 1 cash inflow

Year 2 cash inflow

Year 3 cash inflow

Figure B2

Experiment 2 NPV test.


The range for Year 1 was $1500-$2500.
You calculated the Year 1 cash inflow to be $2000.

The range for Year 2 was $750-$1250.
You calculated the Year 2 cash inflow to be $1000.

The range for Year 3 was $1875-$3125.
You calculated the Year 3 cash inflow to be $2500.

Therefore, NPV = $1522.92

Figure B3

Experiment 2 NPV test answers.

Variance Lecture

See below the slides for the variance lecture.

Additional Analyses

Ranking

A mixed factorial ANOVA was conducted to investigate the effects of NPV, alignment, and numerical NPV reliability on participants' project rankings. Figure B19 shows these data. The alignment × reliability level × NPV interaction was not significant, $F(3.00, 159.10) = 2.44$, $p = .066$, $\hat{\eta}_p^2 = .044$. However, the alignment × NPV interaction was significant, $F(1.99, 105.72) = 16.97$, $p < .001$, $\hat{\eta}_p^2 = .243$; as well as the reliability amount × NPV interaction, $F(3.28, 173.93) = 2.67$, $p = .044$, $\hat{\eta}_p^2 = .048$. As in the allocation data, the linear NPV trend did not differ between reliability level condition in neither the low alignment condition, $\Delta M = 0.43$, 95% CI $[-0.77, 1.63]$, $t(53) = 0.71$, $p = .480$, nor the high alignment condition, $\Delta M = 0.46$, 95% CI $[-0.92, 1.84]$, $t(53) = 0.67$, $p = .504$. However, averaging over reliability level, the linear NPV trend was higher in the low alignment condition than in the high alignment condition, $\Delta M = -4.54$, 95% CI $[-6.39, -2.68]$,

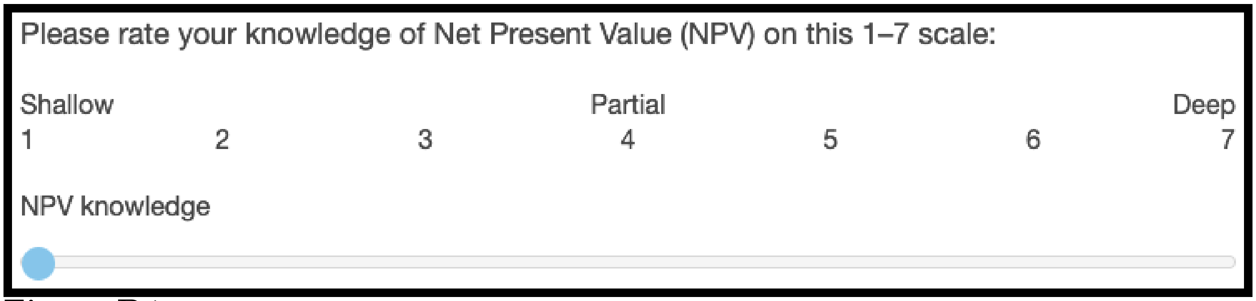


Figure B4

Experiment 2 NPV knowledge rating task.

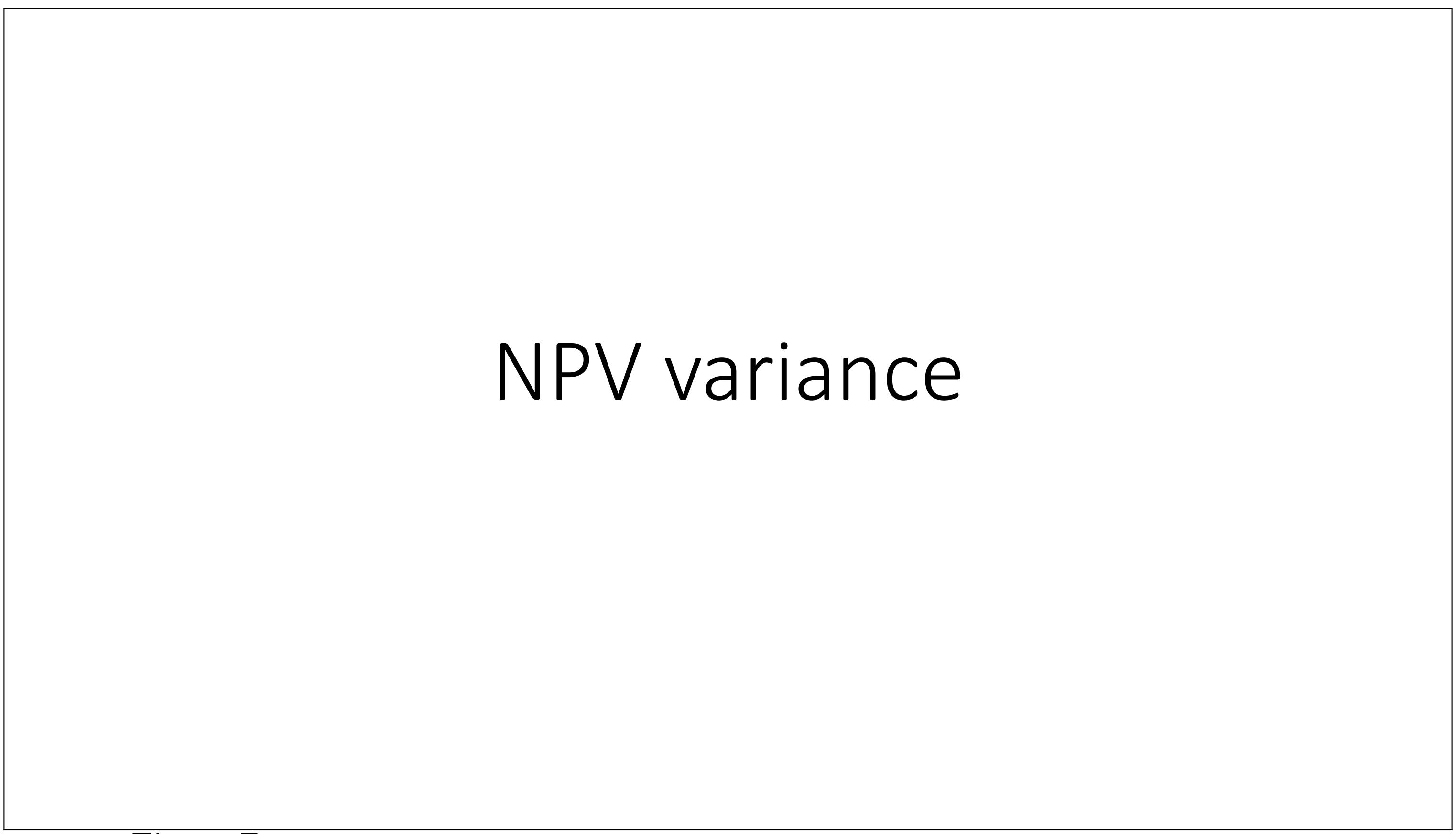


Figure B5

Variance lecture slide 1.

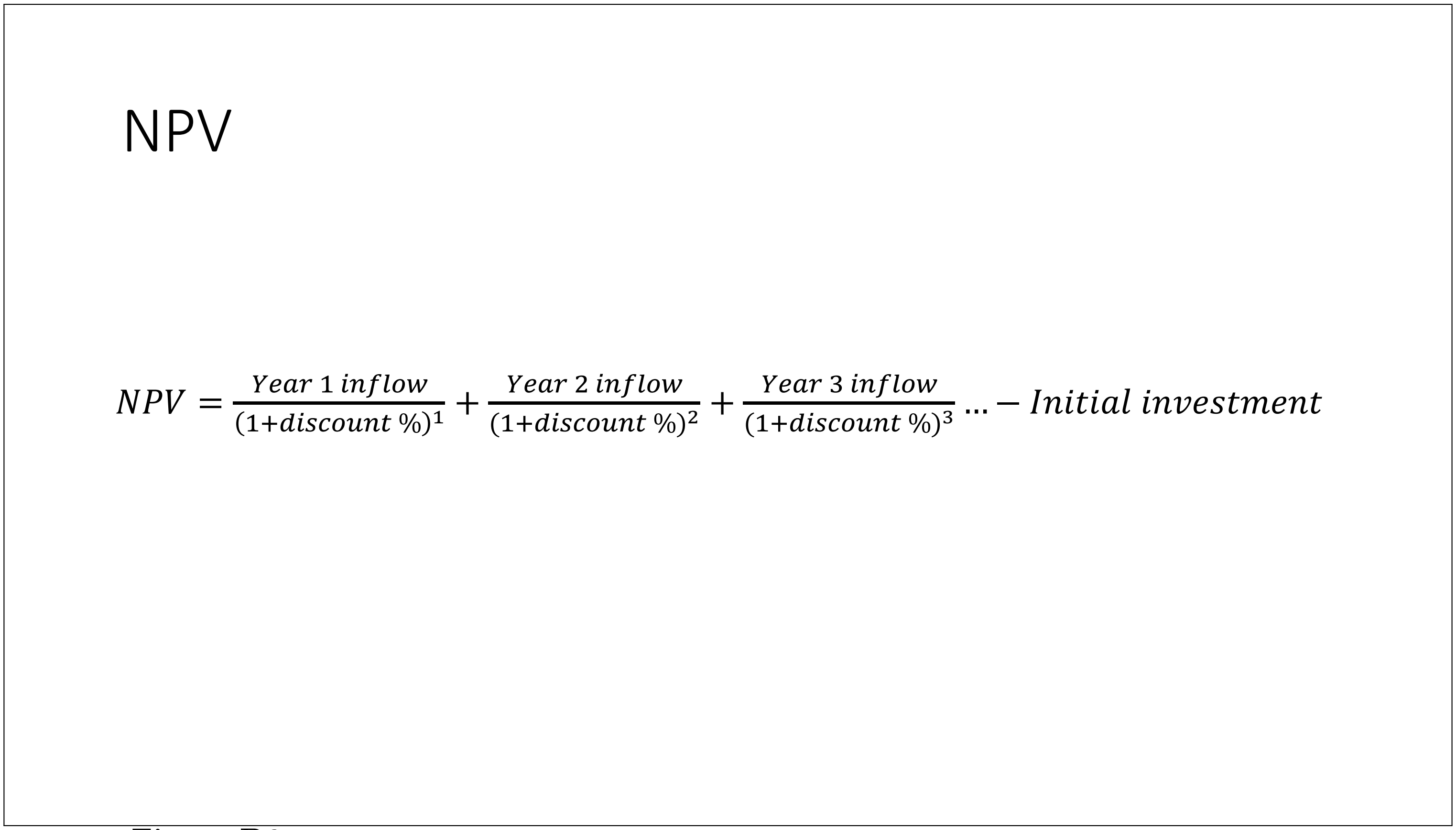


Figure B6

Variance lecture slide 2.

$t(53) = -4.91$, $p < .001$.

Confidence

A mixed factorial ANOVA was conducted to investigate the effects of NPV, alignment, and numerical NPV reliability on participants' confidence ratings. Figure B20 shows these data. Only the main effect of NPV was significant, $F(2.62, 139.08) = 2.97$, $p = .041$, $\hat{\eta}^2_p = .053$.

Variance Lecture

The allocation and ranking data show that participants were affected by the similarity of options, but were not affected by variance information. After the main task of

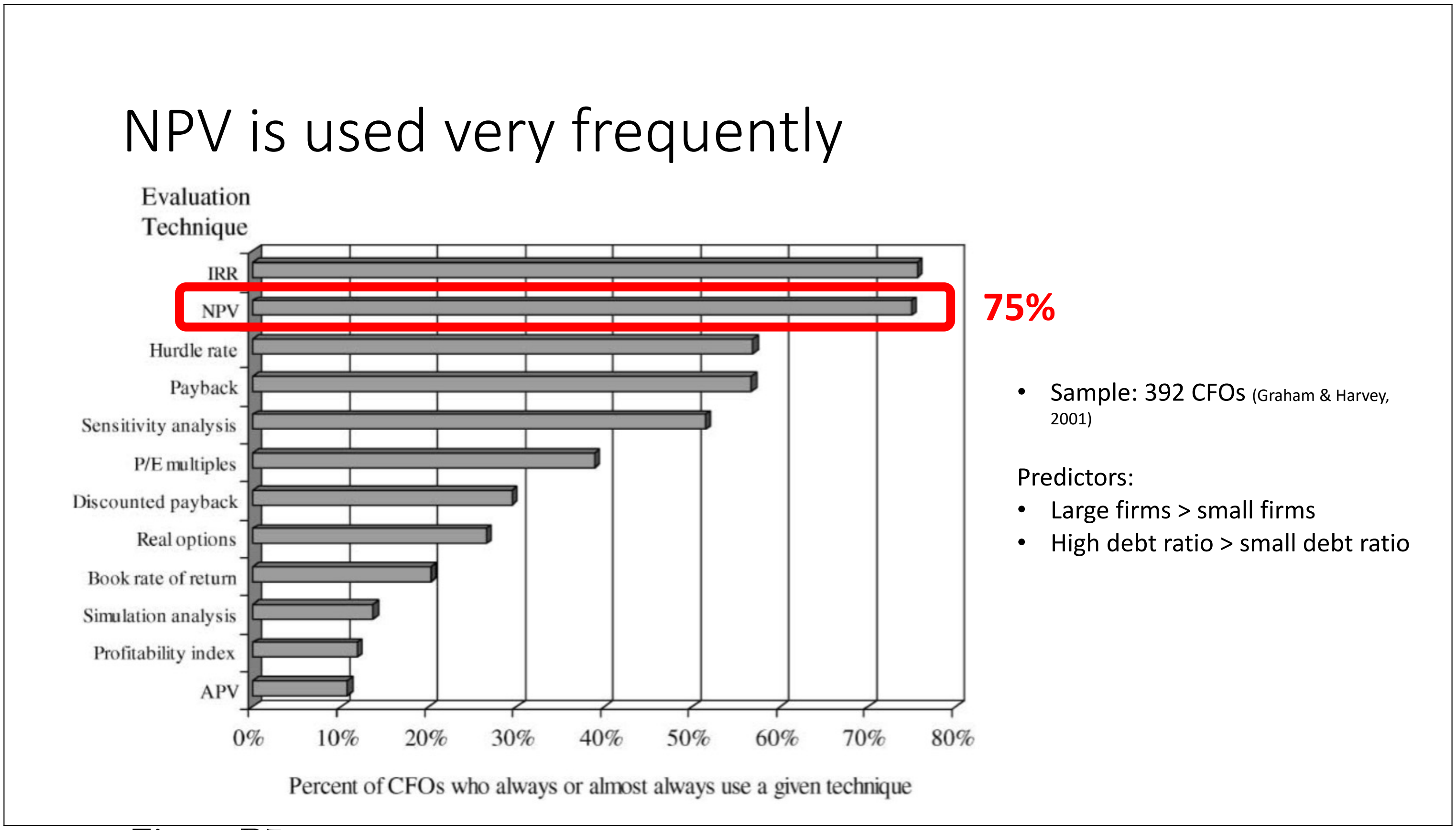


Figure B7

Variance lecture slide 3.

this experiment, participants were shown a short lecture about the importance of variance information when making allocation decisions. They were then presented with half of their previous allocations and gave them an opportunity to amend their allocations. It was hypothesised that participants will be more sensitive to variance after the educational intervention.

A mixed factorial ANOVA was conducted to investigate the effects of phase on participants' project allocations. As shown in Figure B21, the four-way interaction was not significant, $F(2.56, 133.09) = 1.74$, $p = .169$, $\hat{\eta}_p^2 = .032$. Further, the NPV × phase × reliability level interactions were not significant for either the low alignment condition, $\Delta M = 4.43$, 95% CI $[-23.71, 32.58]$, $t(52) = 0.32$, $p = .753$; or the high alignment conditions,

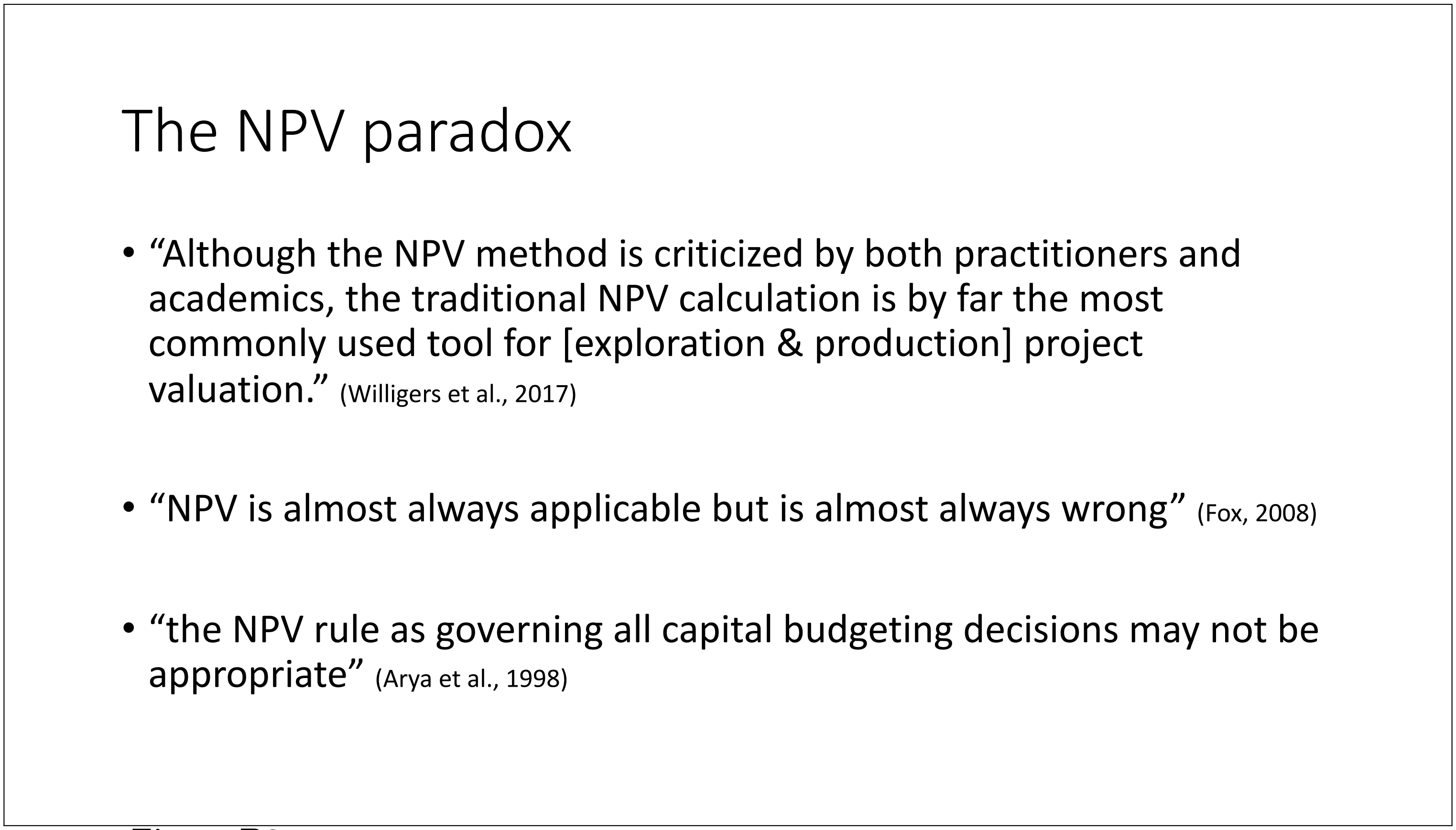


Figure B8

Variance lecture slide 4.

$\Delta M = -11.92$, 95% CI $[-43.39, 19.55]$, $t(52) = -0.76$, $p = .451$. In the low alignment condition, the linear NPV trend (averaged over reliability level) was significantly weaker after the lecture, compared with the linear NPV trend before the lecture, $\Delta M = -12.85$, 95% CI $[-24.08, -1.62]$, $t(52) = -2.30$, $p = .026$. However, this comparison was not significant in the high alignment condition, $\Delta M = -6.37$, 95% CI $[-18.93, 6.18]$, $t(52) = -1.02$, $p = .313$. These results suggest that participants did not become better informed by NPV numerical reliability after the variance lecture. There was, however, some reduction in reliance on NPV overall when projects were dissimilar.

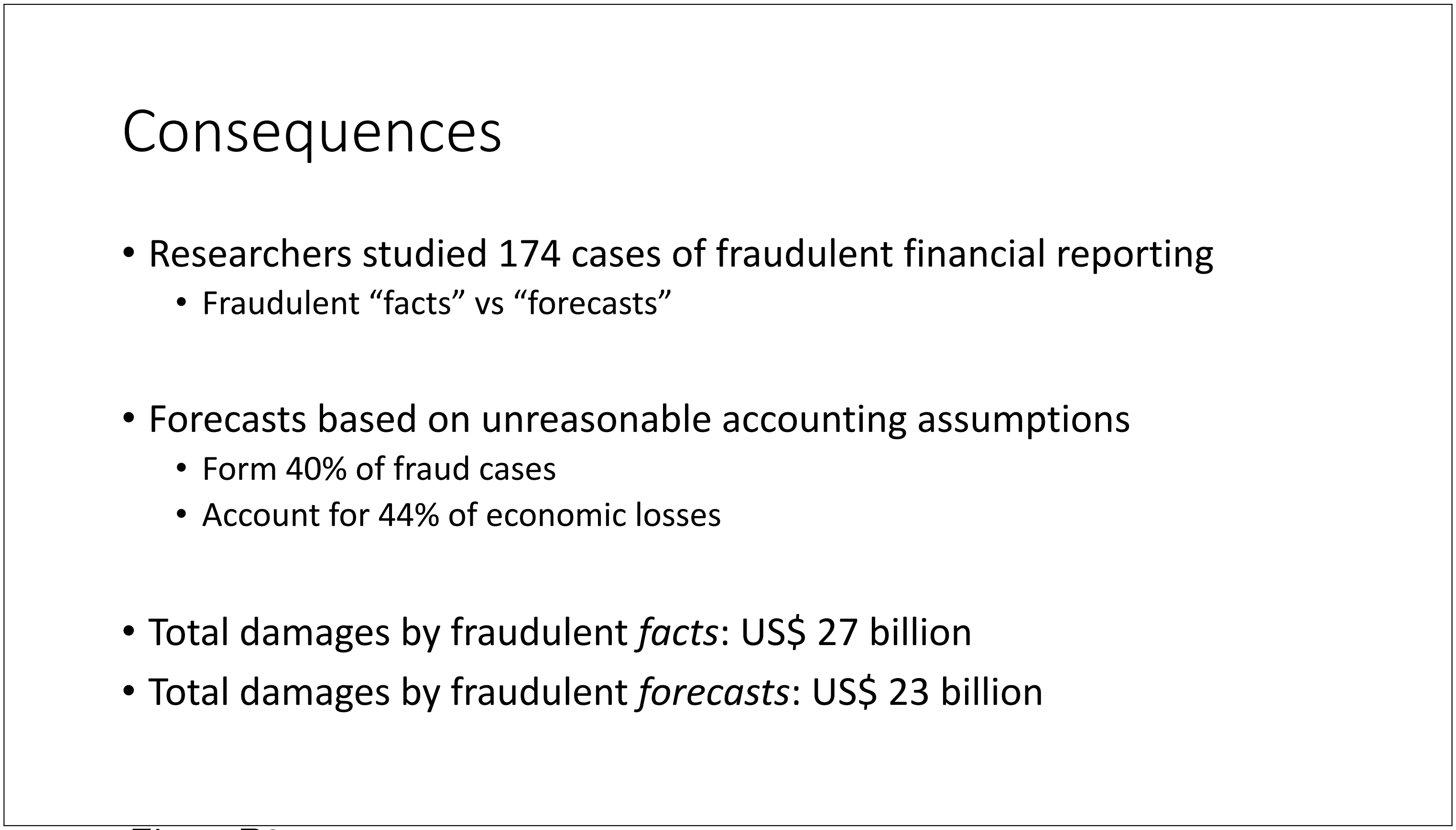


Figure B9

Variance lecture slide 5.

NPV Knowledge

A repeated-measures ANOVA was conducted to investigate the effects of experiment phase condition on participants' NPV knowledge rating. Figure B22 shows these data. The main effect of phase was significant, $F(2.43, 128.59) = 7.80$, $p < .001$, $\hat{\eta}^2_p = .128$. The post-explanation rating was significantly higher than the pre-explanation rating, $\Delta M = -0.59$, 95% $\text{CI}_{\text{Tukey}(5)}$ $[-0.92, -0.26]$, $t(53) = -5.07$, $p_{\text{Tukey}(5)} < .001$. However, there were no significant differences in rating between any of the later phases.

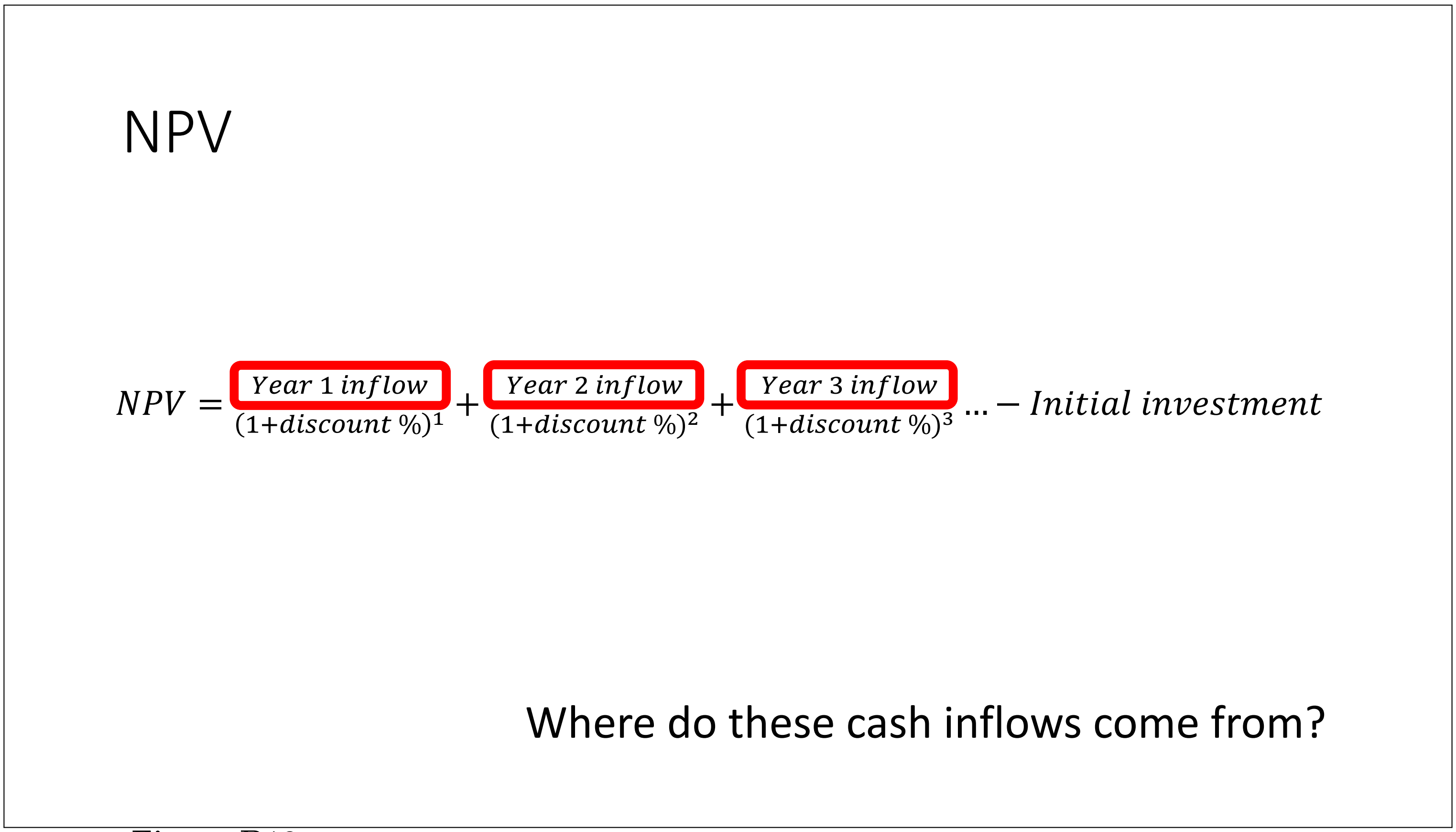


Figure B10

Variance lecture slide 6.

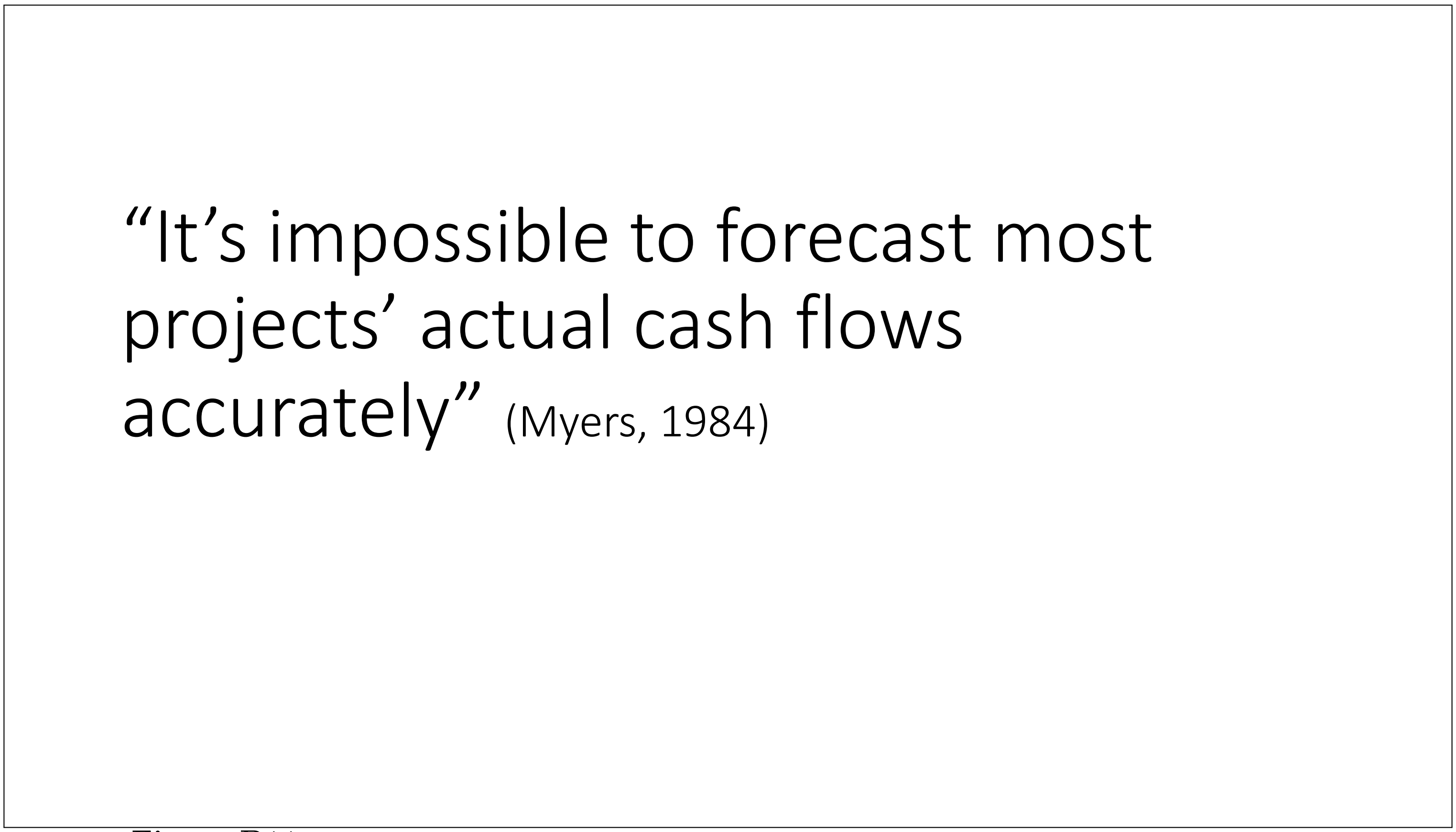


Figure B11

Variance lecture slide 7.

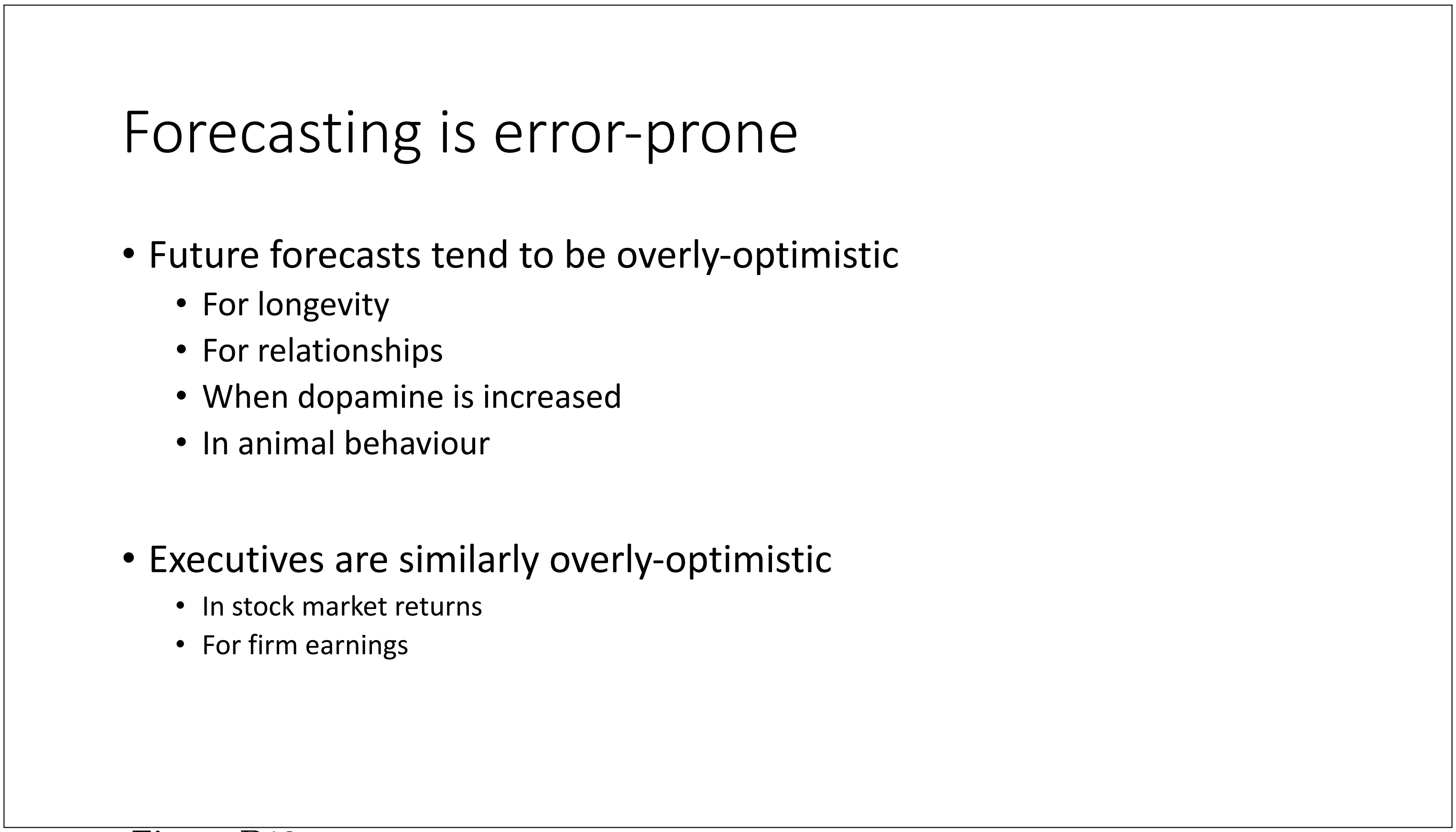


Figure B12

Variance lecture slide 8.

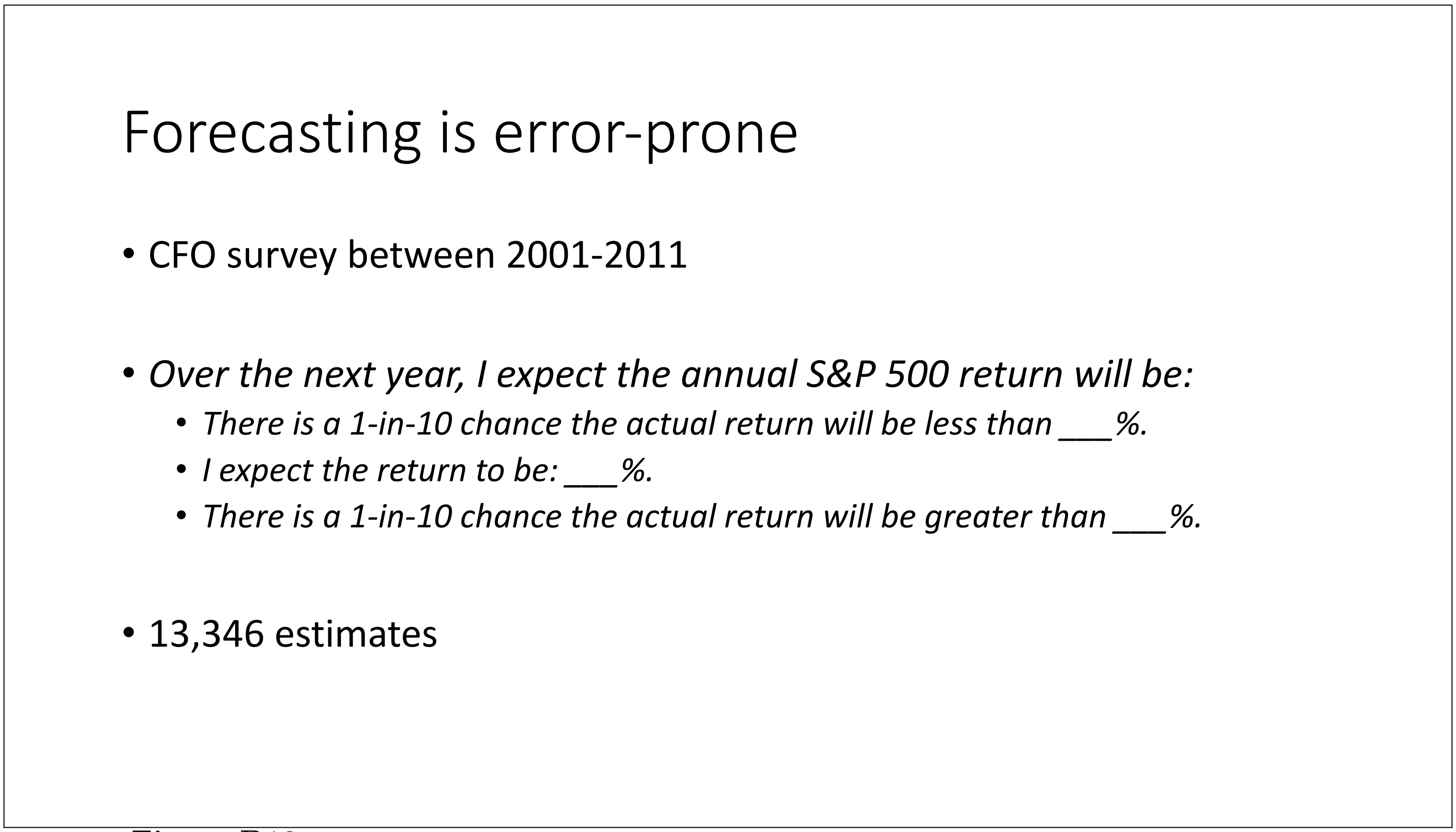


Figure B13

Variance lecture slide 9.

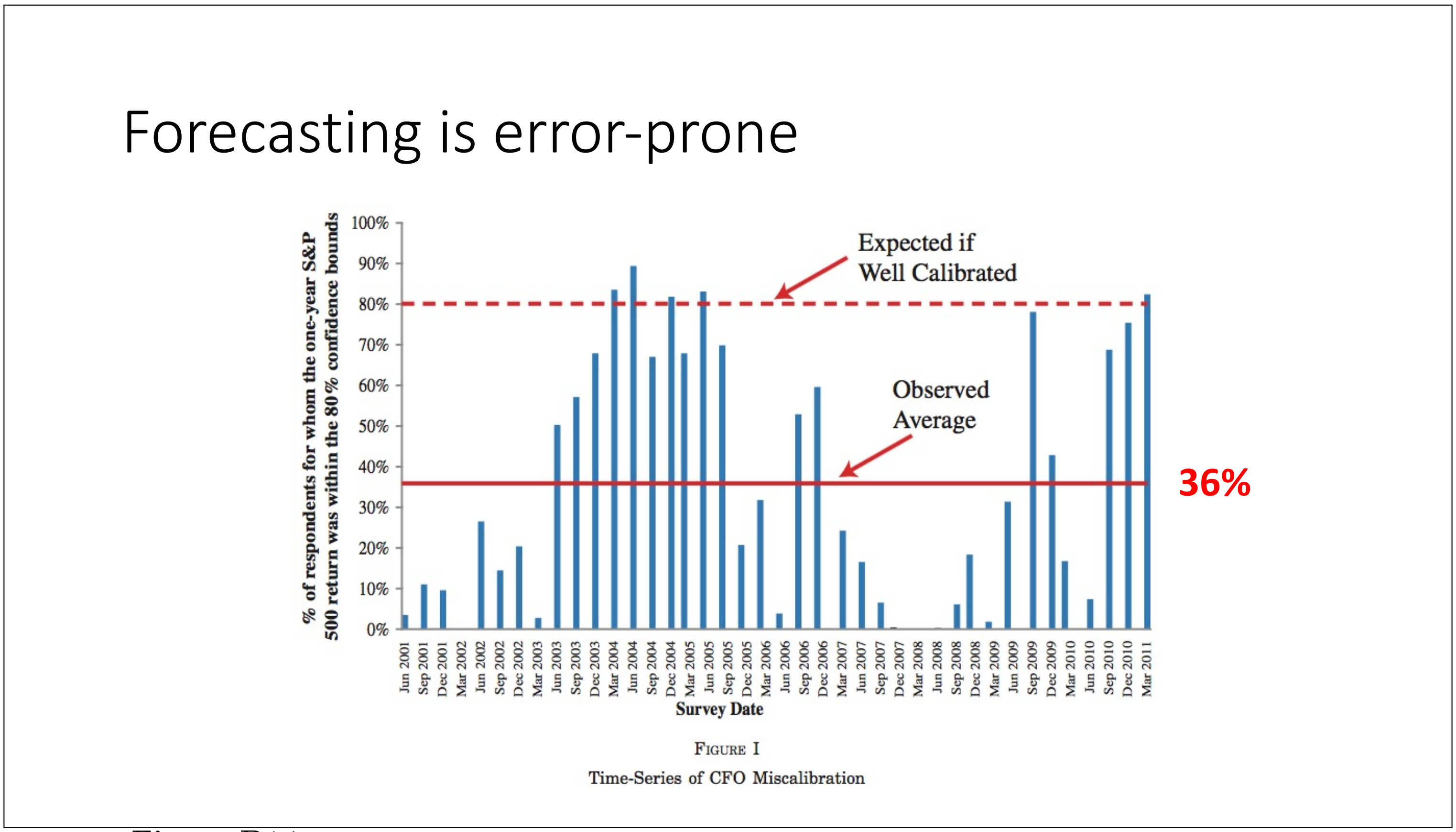


Figure B14

Variance lecture slide 10.

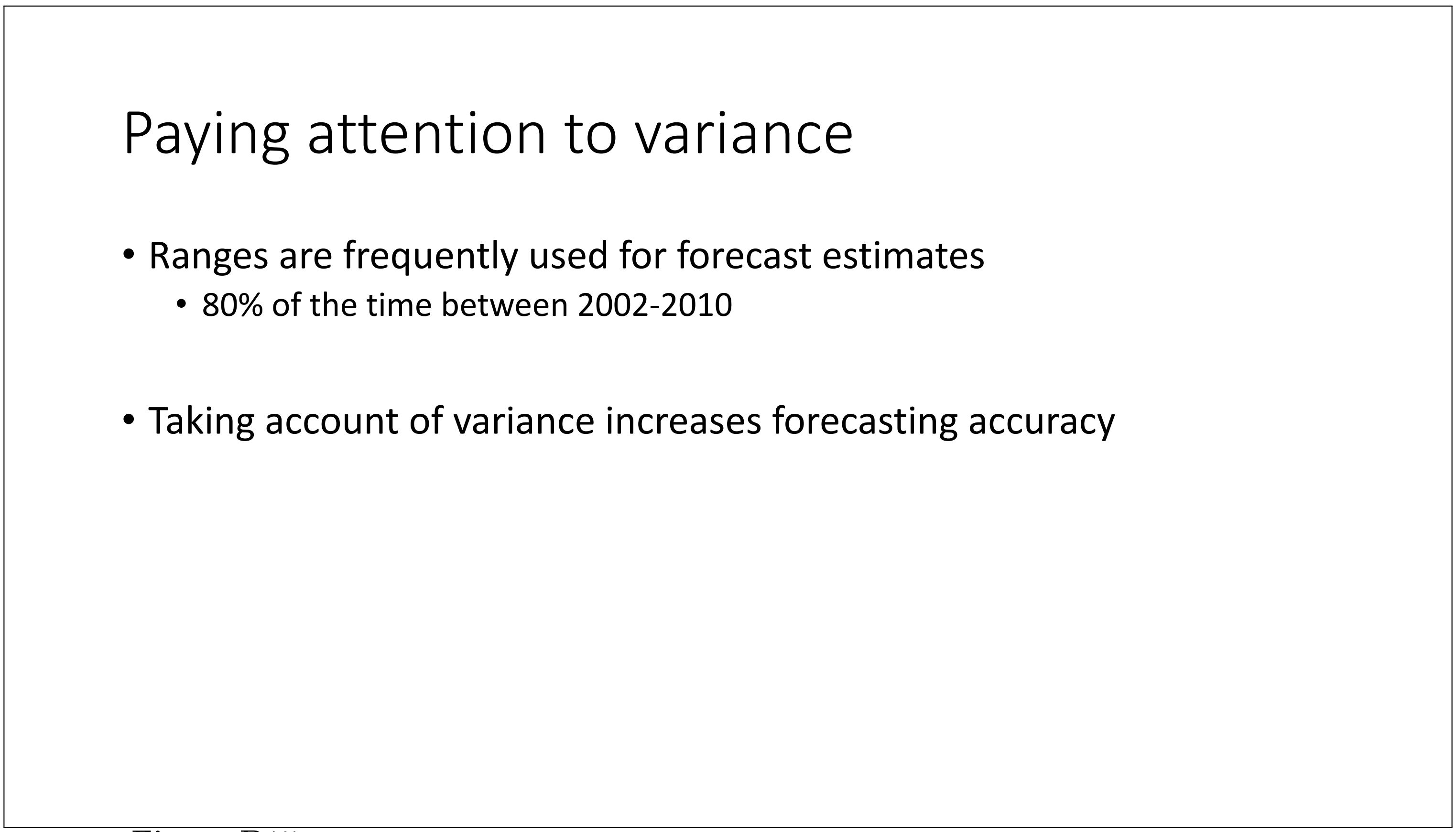


Figure B15

Variance lecture slide 11.

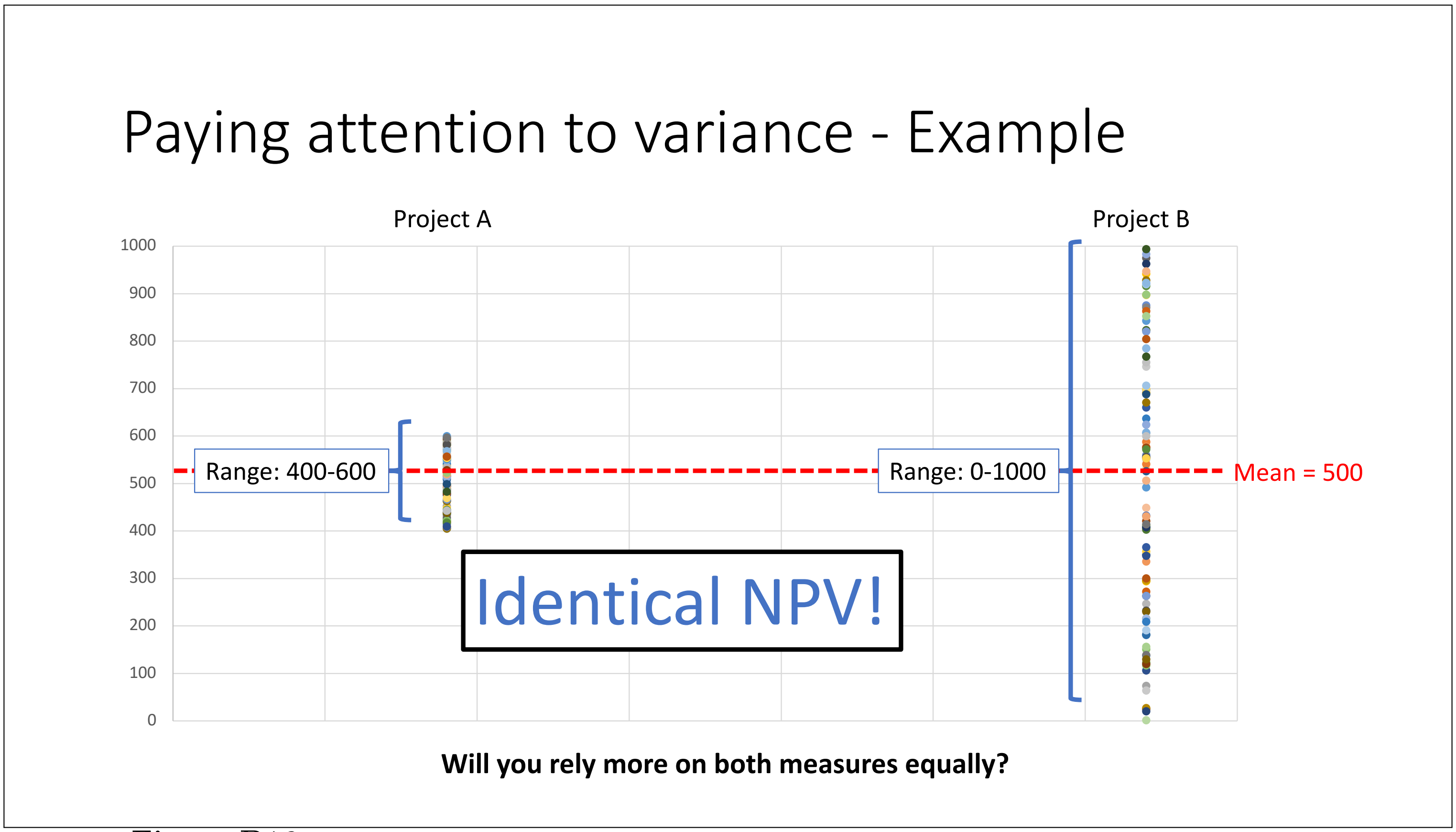


Figure B16

Variance lecture slide 12.

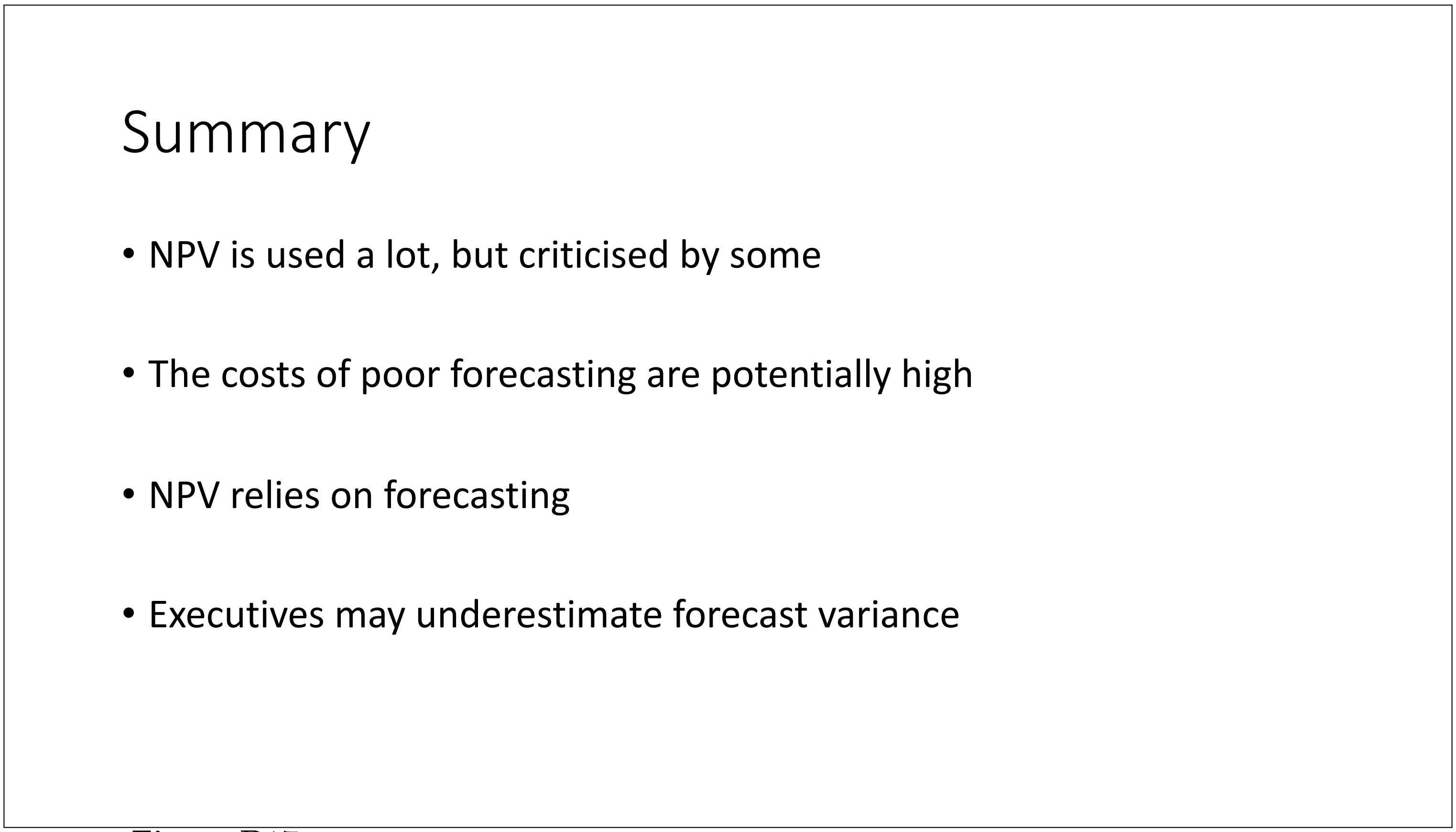


Figure B17

Variance lecture slide 13.

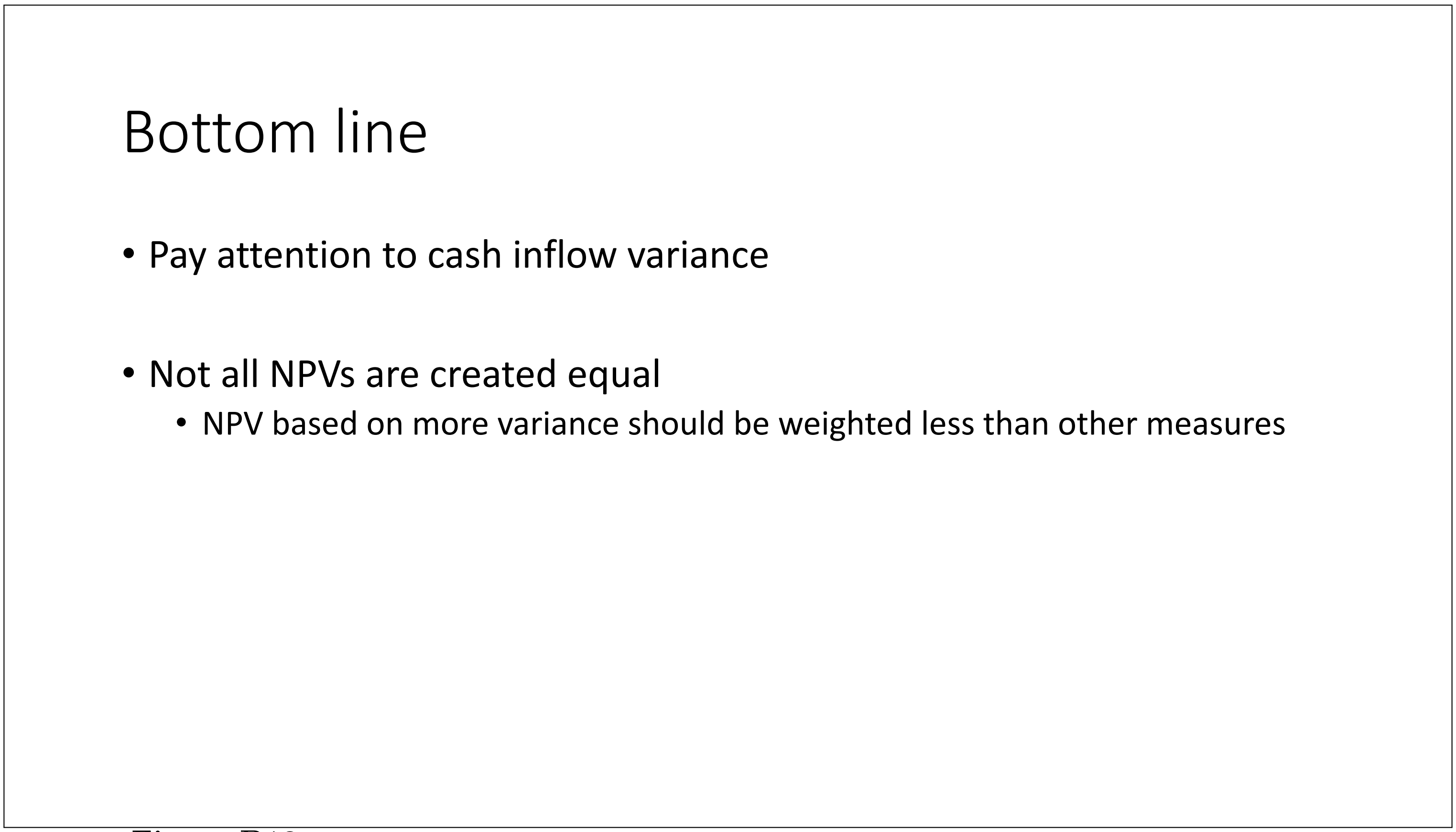


Figure B18

Variance lecture slide 14.

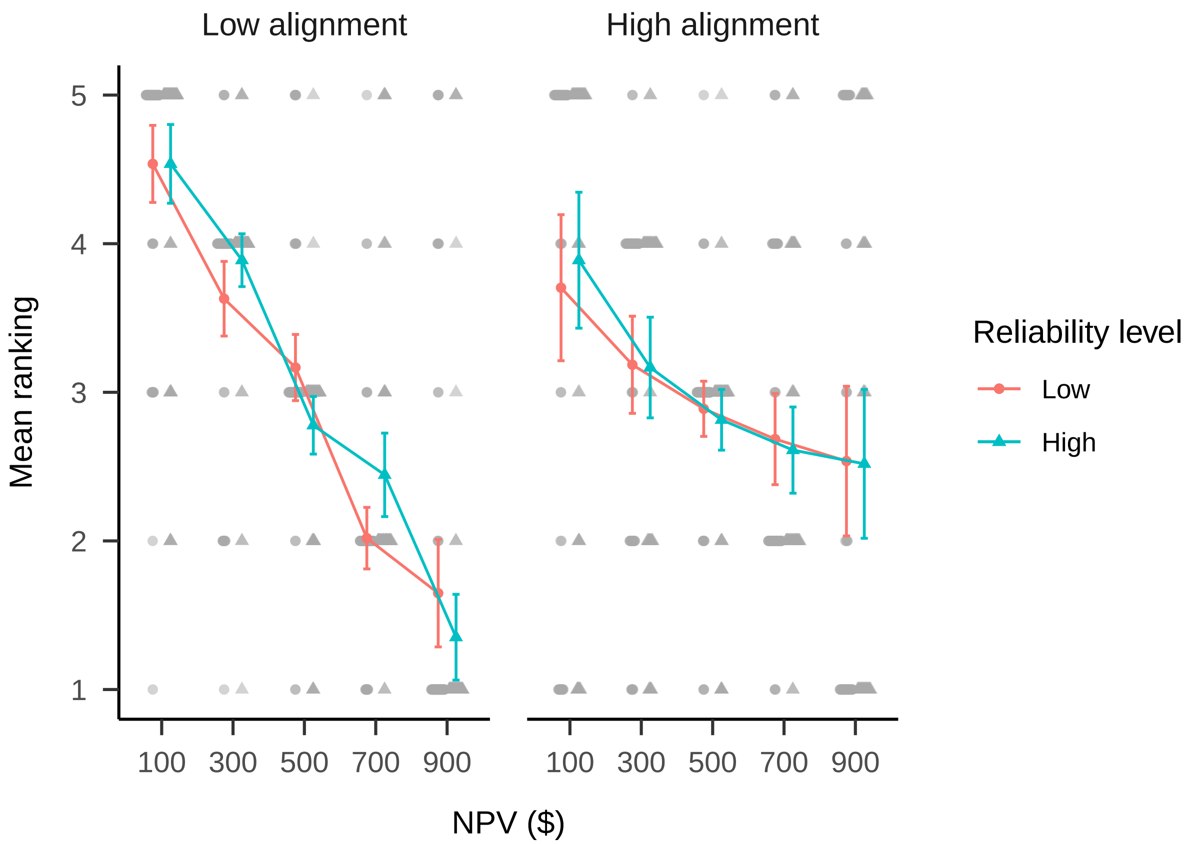


Figure B19

Mean ranking.

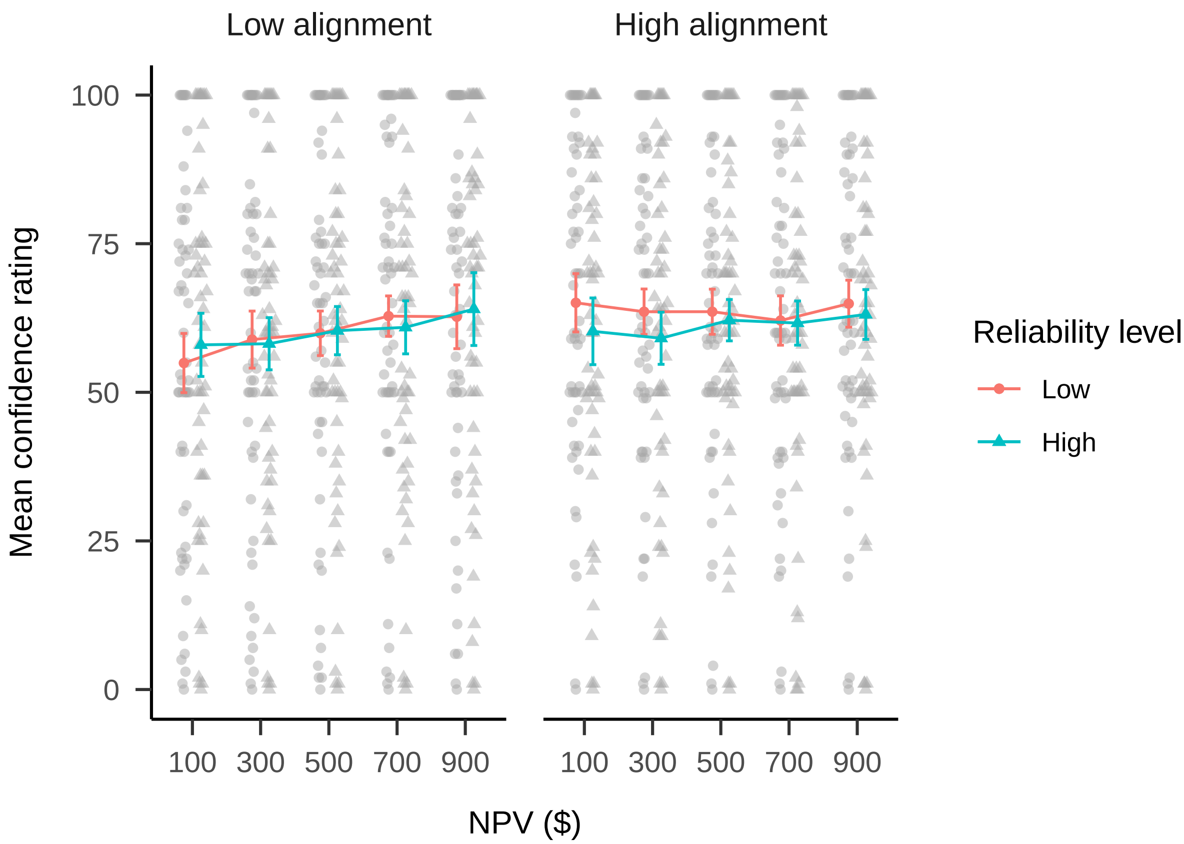


Figure B20

Mean confidence.

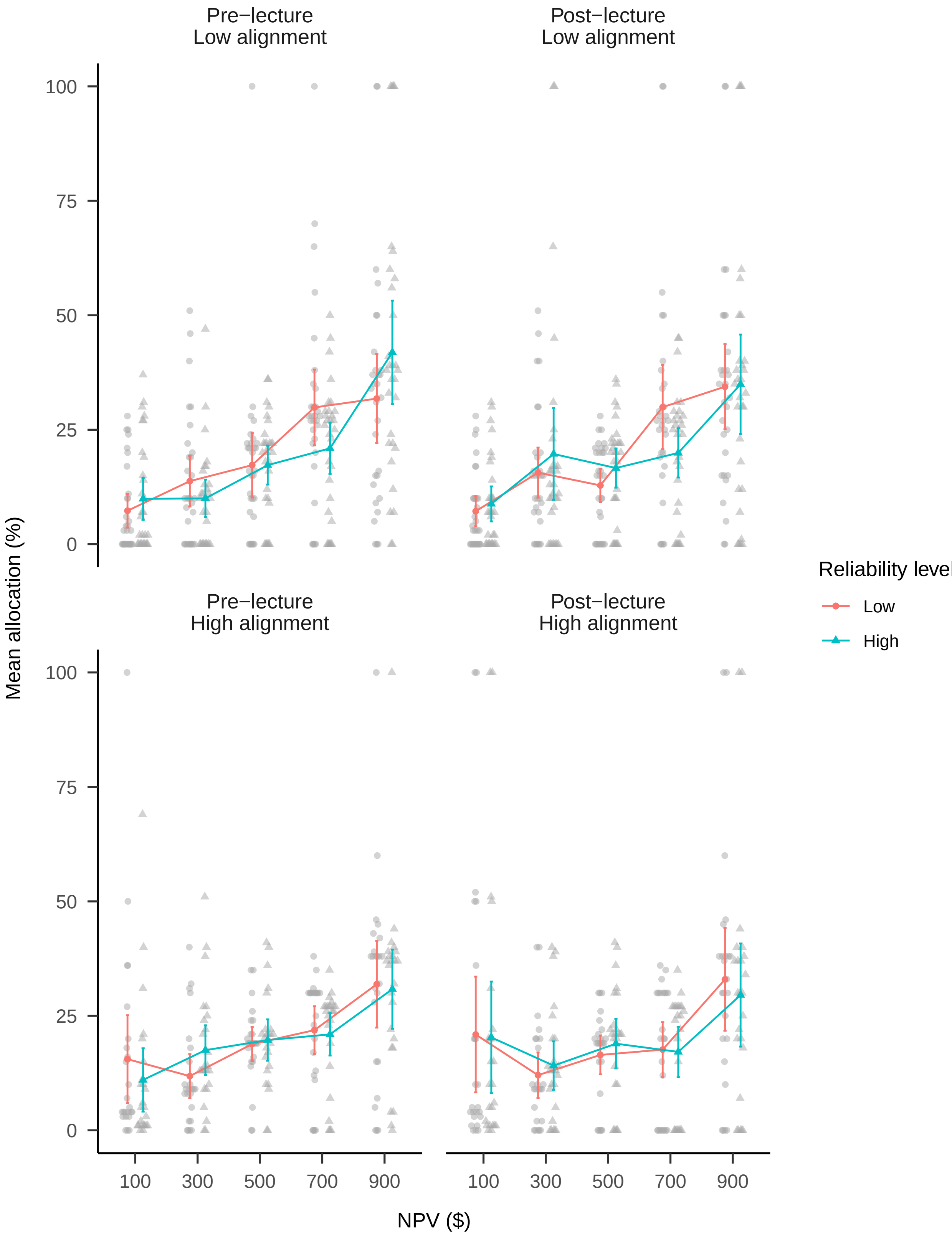


Figure B21

Mean allocation by NPV, reliability level, alignment, and phase.

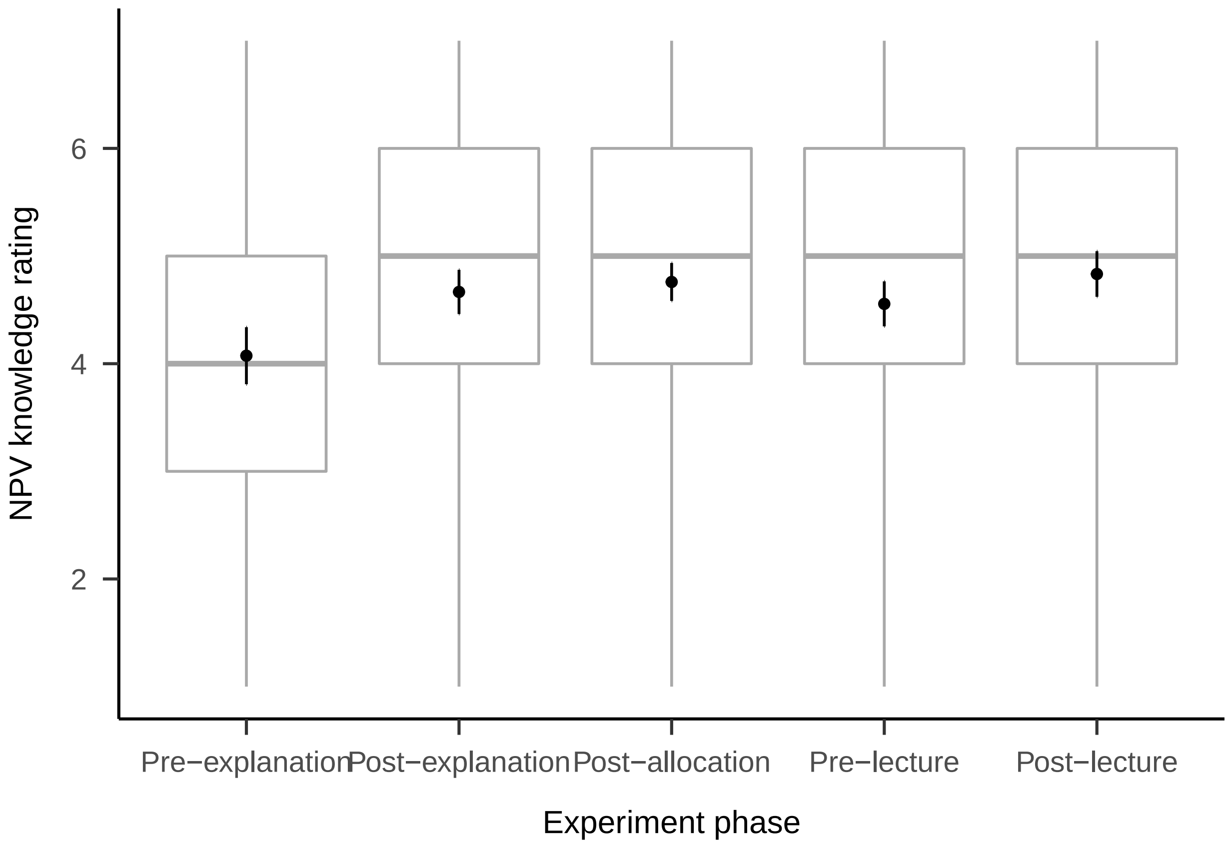


Figure B22

Mean NPV knowledge rating.

Appendix C

Experiment 3

Hypothesised Effects

Figure C1 shows the simulated hypothesised effects for Experiment 3. These effects were constructed as a composite of Experiment 1 data (without the no NPV condition) for the verbal reliability type condition, and data from a pilot study (Dekel, 2021, Appendix B.8) for the numerical reliability type condition. Variance was removed to see the effects clearer.

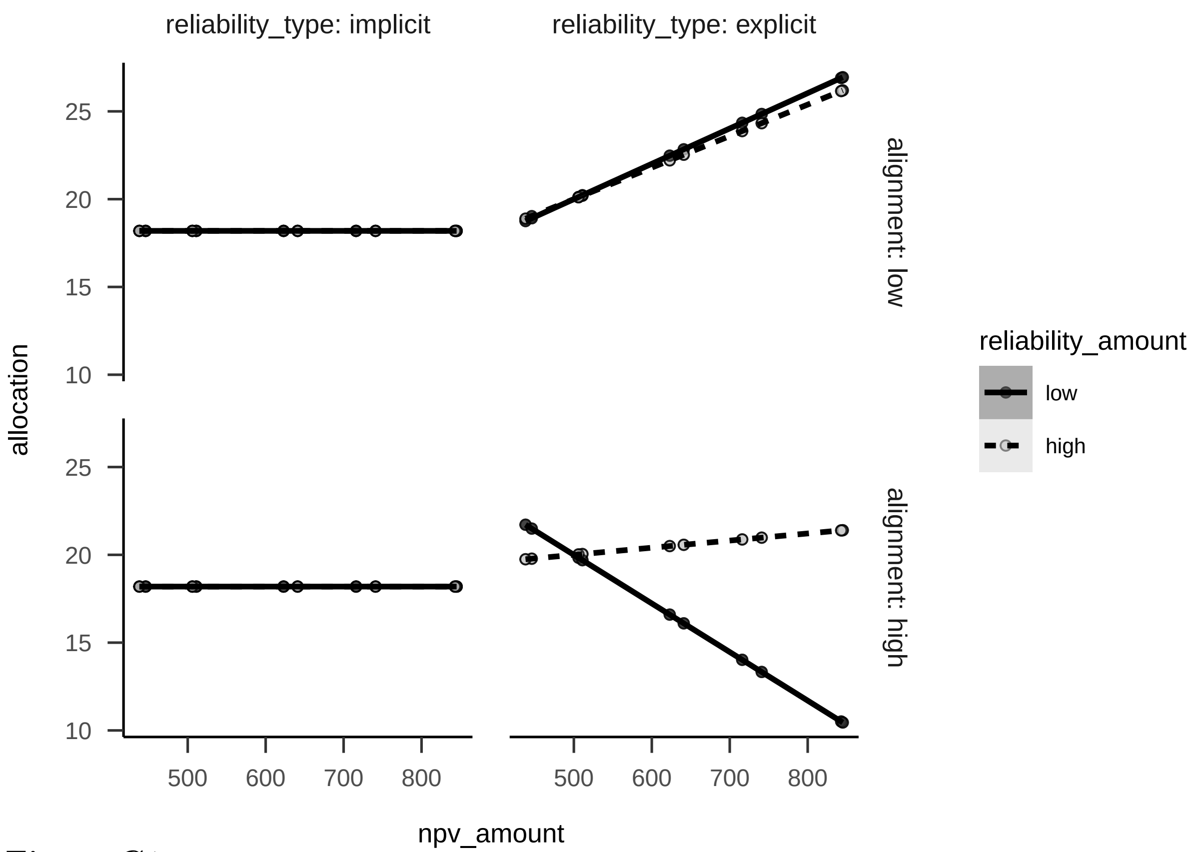


Figure C1

Experiment 3 predicted data.

Power Analysis

A power analysis was conducted through simulation of the effects hypothesised in Experiment 3 (and the simple effects implied by them). The simulated data used the same regression coefficients as Experiment 2 for the explicit condition, no effects for the implicit condition (as shown in Figure C1), and the intercept and residual variance of Experiment 2. The null effects were analysed using the two one-sided tests (TOST) procedure, or *equivalence* testing (Lakens et al., 2018), and setting the smallest effect size of interest to the smallest difference that leads to a significant equivalence between low and high implicit reliability for low alignment in a pilot study (Dekel, 2021, Appendix B.8). Figure C2 shows the resulting power curve. The analysis suggests a total sample size of 448 ($112 \cdot 4$).

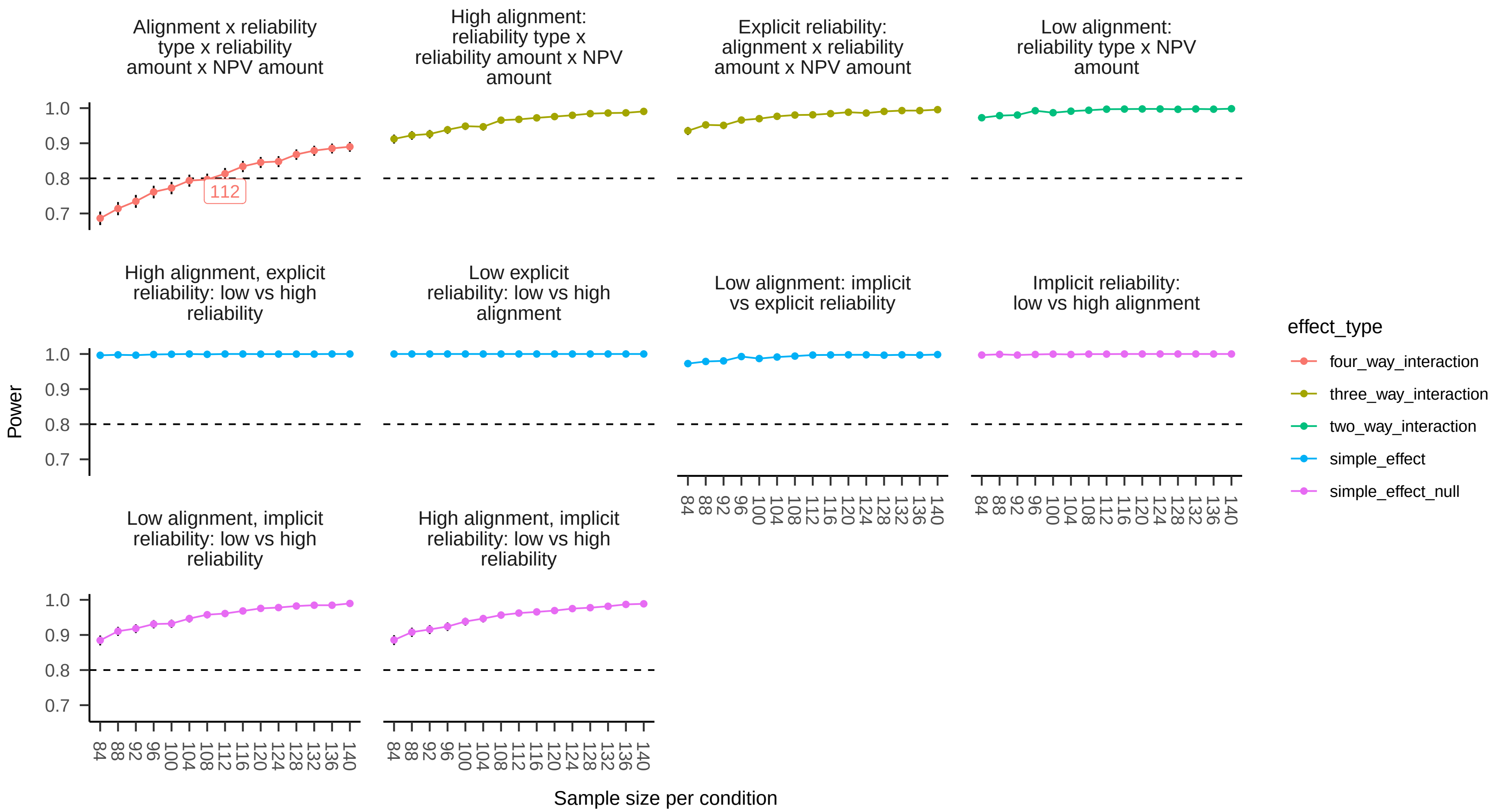
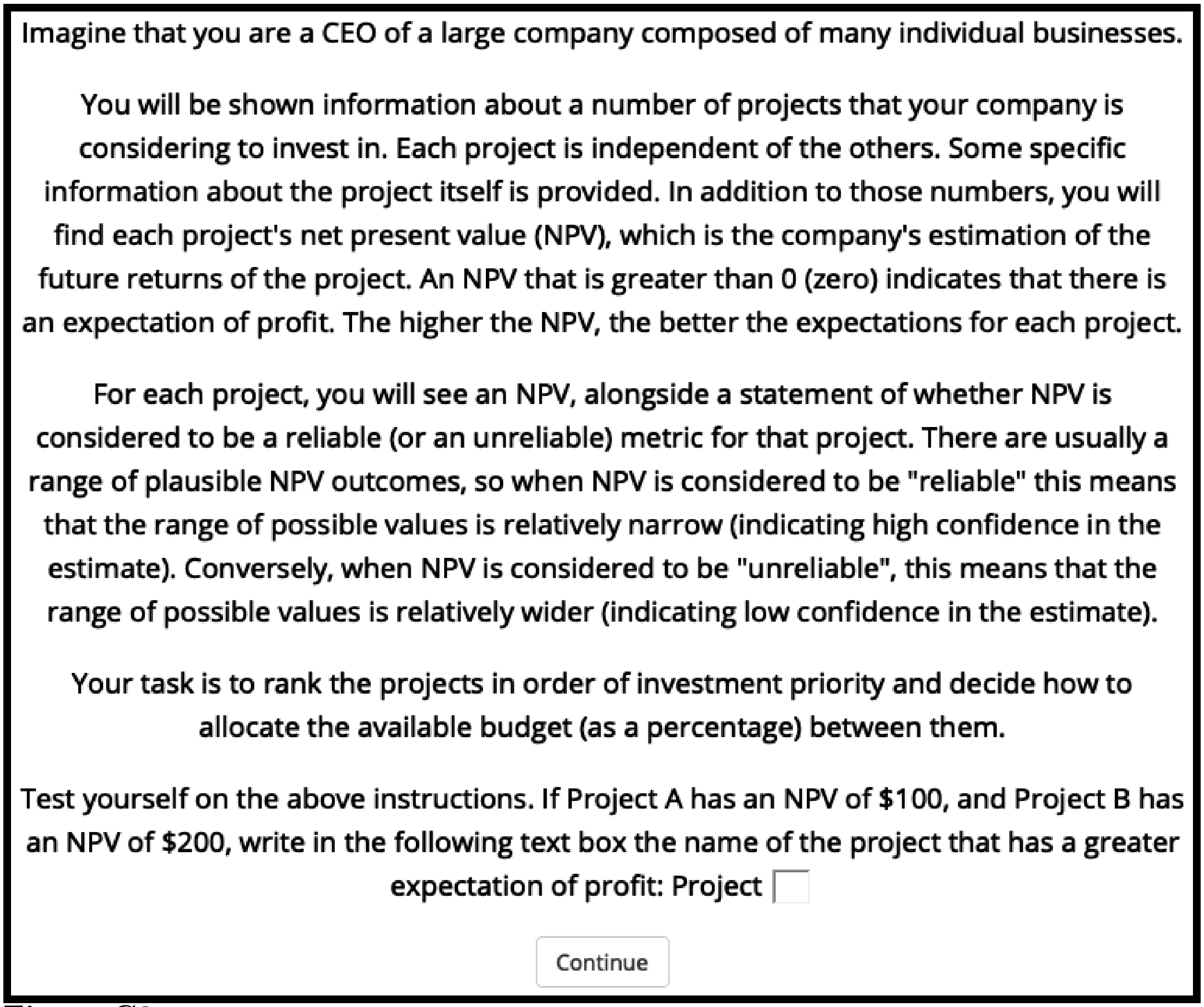


Figure C2

Alignment Experiment 3 power curve. Labels indicate lowest sample size above 80% power.

Instructions

Figures C3 and C4 show the instructions for the verbal and numerical reliability conditions, respectively.



Imagine that you are a CEO of a large company composed of many individual businesses.

You will be shown information about a number of projects that your company is considering to invest in. Each project is independent of the others. Some specific information about the project itself is provided. In addition to those numbers, you will find each project's net present value (NPV), which is the company's estimation of the future returns of the project. An NPV that is greater than 0 (zero) indicates that there is an expectation of profit. The higher the NPV, the better the expectations for each project.

For each project, you will see an NPV, alongside a statement of whether NPV is considered to be a reliable (or an unreliable) metric for that project. There are usually a range of plausible NPV outcomes, so when NPV is considered to be "reliable" this means that the range of possible values is relatively narrow (indicating high confidence in the estimate). Conversely, when NPV is considered to be "unreliable", this means that the range of possible values is relatively wider (indicating low confidence in the estimate).

Your task is to rank the projects in order of investment priority and decide how to allocate the available budget (as a percentage) between them.

Test yourself on the above instructions. If Project A has an NPV of $100, and Project B has an NPV of $200, write in the following text box the name of the project that has a greater expectation of profit: Project

Continue

Figure C3

Experiment 3 verbal reliability instructions.

Interstitial Display

Figure C5 shows an example of an interstitial display.

Imagine that you are a CEO of a large company composed of many individual businesses.

You will be shown information about a number of projects that your company is considering to invest in. Each project is independent of the others. Some specific information about the project itself is provided. In addition to those numbers, you will find each project's net present value (NPV), which is the company's estimation of the future returns of the project. An NPV that is greater than 0 (zero) indicates that there is an expectation of profit. The higher the NPV, the better the expectations for each project.

For each project, you will see a range of possible NPVs alongside a 'midpoint'. The range literally represents the range of plausible outcomes (a uniform distribution), but the midpoint is the best guess, and hence is the same as a single NPV. That is, all values within the range are equally likely, but the midpoint is still the best guess because it is the value that is closest to all the other values.

Your task is to rank the projects in order of investment priority and decide how to allocate the available budget (as a percentage) between them.

Test yourself on the above instructions. If Project A has an NPV of $100, and Project B has an NPV of $200, write in the following text box the name of the project that has a greater expectation of profit: Project

Continue

Figure C4

Experiment 3 numerical reliability instructions.

You will now see project display #1. It is important that you pay attention and read through the task carefully.

To show that you are reading and paying attention, please click on the following checkbox **before** clicking on "Continue":

Continue

Figure C5

An example of an interstitial display in Experiment 3.

Additional Analyses

The three-way interaction (reliability level $\times$ NPV $\times$ reliability type) in the high alignment condition was significant, $\Delta M = 35.43$, 95% CI $[20.74, 50.12]$, $t(444) = 4.74$, $p < .001$. The NPV $\times$ reliability type (averaging over reliability level) in the low alignment condition was significant, $\Delta M = 11.48$, 95% CI $[0.19, 22.77]$, $t(444) = 2.00$, $p = .046$. The association between allocation and NPV for those in the explicit low reliability condition was significantly stronger for those in the low alignment condition, than for those in the high alignment condition, $\Delta M = 35.68$, 95% CI $[22.27, 49.09]$, $t(444) = 5.23$, $p < .001$. The linear NPV trend for those in the low alignment condition was significantly stronger for those in the explicit reliability condition, than for those in the implicit reliability condition (averaging over reliability level), $\Delta M = 11.48$, 95% CI $[0.19, 22.77]$, $t(444) = 2.00$, $p = .046$. The linear NPV trend for those in the implicit reliability condition was not significantly "equivalent" between those in the low and high reliability conditions for both those in the low alignment $\Delta M = 1.64$, 95% CI $[-8.74, 12.03]$, $t(444) = 0.31$, $p = .620$ and high alignment conditions $\Delta M = -1.21$, 95% CI $[-11.59, 9.18]$, $t(444) = 0.22$, $p = .589$. However, this is likely to be because the "lowest effect size of interest" estimate originated from an analysis used before data collection that was different to the one that one used after data collection. Specifically, a univariate linear model was originally used (treating NPV as a continuous predictor), whereas the data were ultimately analysed using a multivariate linear model (treating NPV as a repeated measures factor). In the numerical reliability condition, a pilot experiment (Dekel, 2021, Appendix B.8) suggested that the linear NPV trend would be equivalent between those in the low and high alignment conditions, averaged over reliability level. However, the test of equivalence was not significant, $\Delta M = 15.19$, 95% CI $[3.90, 26.48]$, $t(444) = 2.64$, $p = .996$.