Correlation Neglect and Information Redundancy in Stock Evaluations

Alessandra Guerrero

I. Executive Summary

Stock traders and investors often receive information from different sources which is correlated, instead of mutually independent. This phenomenon can lead to double counting of certain information, where those who do not account for this redundancy can lead to biased and distorted beliefs. Using a survey approach, we tested whether correlation neglect (failure to account for the double counting of information) can be alleviated by informing participants of its existence. Our evidence demonstrates that correlation neglect is more prevalent in the face of potential losses, but can be improved through drawing the focus of participants into the redundancy of signals in the face of potential wins.

II. Introduction

Stock traders and those who work in the financial sector regularly receive immense amount of information from countless separate sources. These sources, however, often tend to be interconnected; that is, one source may report what has been already reported by another source. Hence, traders often receive what is called "correlated" information. If traders fail to account for this correlation, they may receive a biased view of the truth and retrieve lesser returns than is optimal, which would reduce their performance. For example, if a trader reads that a certain stock is bound to fall by 20% in the next couple of weeks from two separate sources, without realizing that the second source is re-reporting this potential fall from the first source, they would be double counting that information.

However, if that trader knows both sets of predictions came from the same source and was only re-reported, they would be able to account for this correlation, and hence be able to appropriately respond and reduce the potential bias. Our purpose in conducting this research is then to answer the following question: Do traders account for double counting if they are aware the information comes from the same source (the information is correlated)? That is, can correlation neglect in investing be corrected by informing traders about the extent of information redundancy? We will follow this section by explaining our approach to answering this question, the issues with using said approach, and the results of our empirical exercise.

III. Approach

A. Data Generating Process: Survey Approach

In order to gather the data to answer our research question, we created a survey of potential stock evaluations. The survey consisted of three distinct parts. First was a set of demographic questions about age, gender, and educational level (highest degree achieved). Then, we asked a set of control questions on stock evaluations to set a baseline for a individual participant. In this control section, each participant was asked about four different valuation predictions of equally trustworthy and completely *independent* computers along with the valuation of their own trustworthy computer prediction. These two control questions are crucial, as they allow us to see whether the participant has an inclination to trust their own computer or simply take the average of all predictions.

Lastly, we asked a set of five treatment questions to the same participant. It was important for us to follow the survey process designed by Laudenbach, Ungeheuer, and Weber 2017b and have the same participant answer both the set of control questions and the set of treatment questions. In the treatment section, we had five different questions where it was made clear to the participant that they computers are not independent and may re-report the same valuation predictions. The first computer showed an independent prediction, while the three remaining computers showed a prediction derived from the average of its own prediction and that of the first computer's. With the responses to these treatment questions, we can conclude whether participants would rather take the simple average of all of the shown predictions or calculate the results of each individual prediction and choose the one with the highest potential return (given that all predictions are equally trustworthy). Hence, we can arrive at both the existence of correlation neglect and see whether informing the participant of information redundancy decreases correlation neglect.

The stock valuation predictions were all posed in terms of percentage increases and decreases. We asked stock valuation predictions for both increasing and decreasing predictions (with potential gains and losses), as well as differentiating between potentially small increases against potentially large increases (and repeating with decreases), to test if correlation neglect shifts according to perceived losses or perceived differences in gains. The written set up of the survey and the following questionnaire with specific can be found in Appendix section C.

B. Issues with a Survey Approach

As with survey approaches within the scope of this type of research, there were some limitations that need to be accounted for and explained. Feedback from selected participants revealed that the average of valuation predictions was relatively easy to calculate in the control questions, however it became more difficult with treatment questions. In the treatment questions the predictions were not in multiples of ten, which made quick calculations more difficult than in the control questions. We do see some response attrition with the latter treatment questions, leading us to conclude that this perceived difficulty reduced the willingness of some participants to both calculate potentials accurately and respond appropriately (i.e. some may have simply guessed or chosen randomly), which could have skewed the results.

This might be less of an issue for participants with a Bachelor's or Master's degree in a STEM related field, but could pose an issue for participants with a lower education or those not mathematically-inclined.

IV. Results

A. Measuring Correlation Neglect

We will begin our main analysis with a review of the descriptive statistics. We analyzed a sample of 52 respondents with an average age of 24 years, of which the majority were male (69%), and held a Bachelor's degree (73%). On average, the respondent time was approximately 7 minutes. Table I shows the descriptive statistics of all questions, sectioned by *Control* and

Treatment. In addition to the descriptive statistics, we have three columns demonstrating the rational belief, the correlation neglect belief, and the base signal of each question's valuation prediction.

Table I shows that in the treatment questions #1, #2, and #5, the mean valuation prediction is closer to the rational belief. This signifies that, on average, participants were able to calculate the averages correctly and therefore acted as sophisticated agents (i.e. they were able to account for information redundancy and adjust accordingly). On the other hand, for questions #3 and #4, the participants' valuation predictions were on average closer to the correlation neglect belief, thus they calculated the average without taking into account information redundancy with Computer A. Similar to Enke and Zimmermann 2017, we find that providing less potential valuation predictions (e.g. alternative potentials as in question #5) to the participant reduces correlation neglect.

Following the methodology of Laudenbach, Ungeheuer, and Weber 2017b, we created a naivete parameter. When this parameter is equal to 1, there is full correlation neglect. When it is equal to 0, then there is no correlation neglect and the agent (participant) is sophisticated or Bayesian. Using this naivete parameter, we create kernel density estimate plots for each treatment question with normalized beliefs. Figure 6 is the plot of kernel density estimates of median naivete parameter (MNP) for all questions. This plot revels that, in general, individuals are heterogeneous in their beliefs and tend to behave with correlation neglect. The density distribution follows a hump-shape, with the highest kernel density between 0.5 and 1, hence participants behaved with correlation neglect more than they behaved sophisticatedly.

We repeat the above kernel density exercise with each of the treatment questions to test how much participants behave with correlation neglect in each one. We can see that with questions #1 and #2, individuals are still heterogenous but the highest kernel density lies at 0. Thus, for these questions, individuals behave sophisticatedly. On the other hand, for questions #3 and #4, we see the highest kernel density lies at 1. Thus, for these questions, individuals behave with almost complete correlation neglect. Notable, these questions ask for valuation predictions with potential decreases, leading us to conclude that correlation neglect is more prevalent where a high degree of loss aversion is present. Question #5 demonstrates highly heterogenous beliefs where two predictions are provided, but the highest kernel density lies at 0 still.

B. Regression Results

In addition, we ran basic linear regressions for each question of the survey with the demographic variables as the regressands, in order to ascertain differences in beliefs according to the demographic controls. For each question, we ran the following regression:

$$Q_i = \beta_0 + \beta_1 \times \text{Education Level}_i + \beta_1 \times \text{Gender}_i + \beta_3 \times \text{Duration}_i + \mu_i$$

The intuition behind the regression was that participants with a higher education level would be more sophisticated, and more able to perceive information redundancy. In theory, there should be no discernable difference between the beliefs of males and females. In addition, people taking more time (higher duration) answering the survey should be able to make better computation and demonstrate less correlation neglect.

Table II shows the above regression for each question in the *Control* and *Treatment* sections. We can see that there is no demonstrated relationship between any demographic variables and the beliefs in valuation prediction in each question (all results are highly statistically insignificant aside from gender in *Treatment* question #2, which appears to be a result of circumstance rather than a systematic relationship). The same is true of the regressions in Table III, which demonstrate the results of the same regression for normalized beliefs. None of the coefficient estimates in Table III are statistically significant. Hence, the evidence shows there is no relationship between any of our demographic variables and behaving either with correlation neglect or as a sophisticated agent.

C. Alleviating Correlation Neglect

Some studies have described ways in which correlation neglect can be alleviated or improved. On one hand, Laudenbach, Ungeheuer, and Weber 2017a find that sampling of historical or simulated returns can potentially help investors reduce the amount of correlation neglect and achieve a more optimal allocation of assets. If investors are shown the record of historical returns of a portfolio, then they are more likely to see the correlation and information redundancy, as opposed to simply having the information that stocks might be correlated. This policy would be simple and cost-effective to implement for brokers and would make it easier for investors to gain information without having to read through extensive risk documents.

On the other hand, Enke and Zimmermann 2017 find that correlation neglect is largely driven by conceptual problems instead of mathematical ones. Hence, they find that increasing the participant's focus on the correlation and underlying independent signals reduces correlation neglect substantially. This is at odds with the result found by Laudenbach, Ungeheuer, and Weber 2017a, however, this policy is even more simple to implement. It would simply require increasing the investor's focus on the presence of correlation within sources and the degree of correlation between certain sources. The latter solution would be move viable, although those with access to full histories of historical returns may find the first solution more effective.

V. Conclusion

We sought to answer the question: Can correlation neglect in investing be corrected by informing traders about the extent of information redundancy? In addition, we wanted to test if the degree of correlation neglect, or the patterns of correlation neglect prevalence, change if participants are measuring potential losses in stock valuations or change according to the magnitude of the potential gains or losses in valuation predictions. Following the survey approach and econometric methodology of previous literature, we find that correlation neglect is more prevalent when measuring potential losses, and participants behave more sophisticatedly when given less options or when measuring potential gains. We found no evidence that participant behavior changes with demographic features. Our results are in line with those found by the past literature, and lead us to conclude that correlation neglect can be stymied by informing participants of information redundancy (although not in the case of negative valuation predictions). However, the survey approach we used has notable drawbacks that could have skewed our results.

References

Enke, Benjamin and Florian Zimmermann (2017). "Correlation Neglect in Belief Formation". In: *The Review of Economic Studies* 86.1, pp. 313–332. URL: https://doi.org/10.1093/restud/rdx081.

Laudenbach, Christine, Michael Ungeheuer, and Martin Weber (2017a). "How to Alleviate Correlation Neglect?" In: *SSRN Electronic Journal*. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3086722.

– (2017b). "How to Overcome Correlation Neglect?" In: Center for Economic and Policy Research.

APPENDIX A
FIGURES

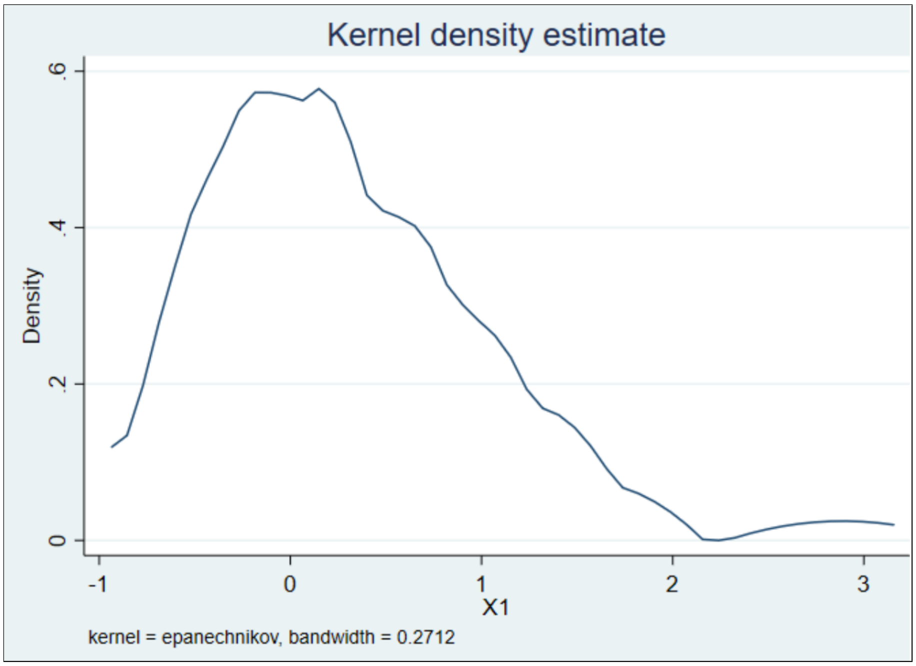

Figure 1: Kernel Density Estimate for Question #1

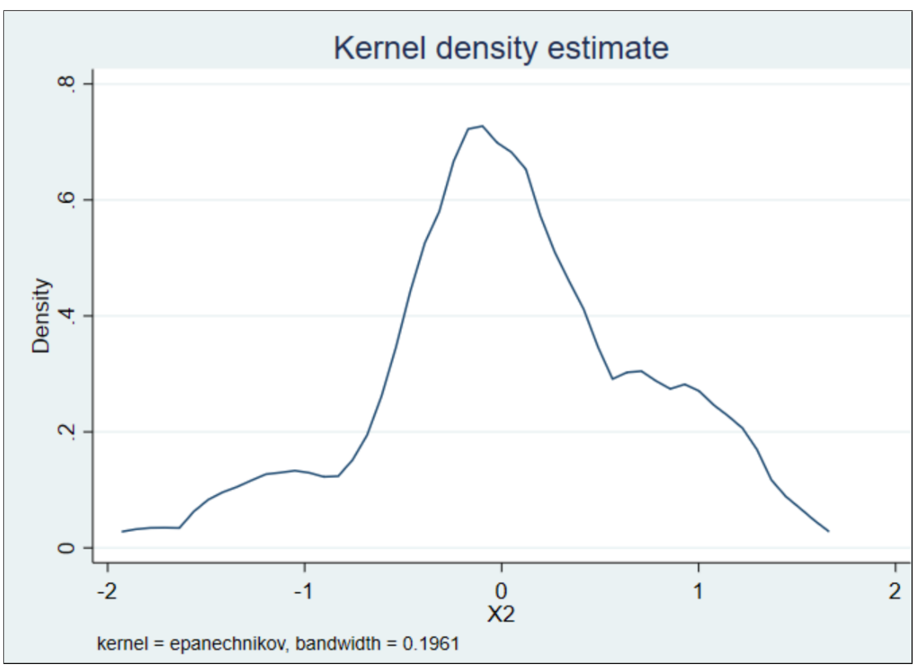

Figure 2: Kernel Density Estimate for Question #2

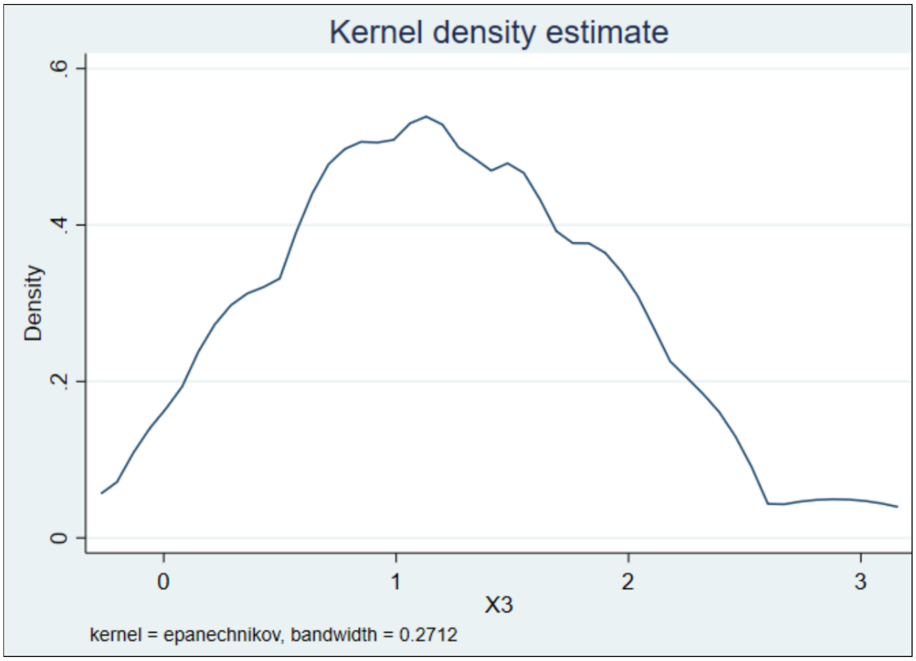

Figure 3: Kernel Density Estimate for Question #3

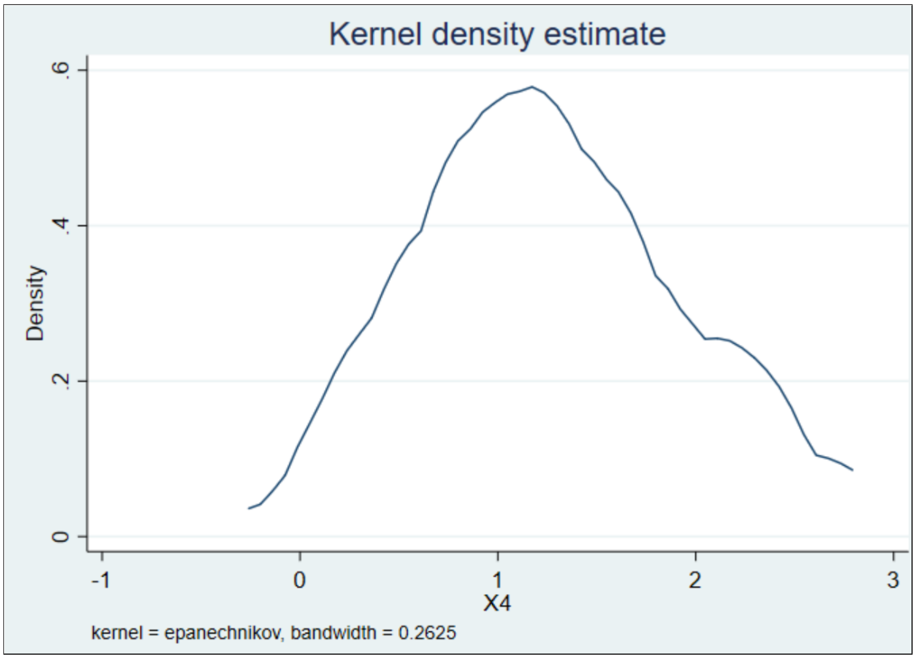

Figure 4: Kernel Density Estimate for Question #4

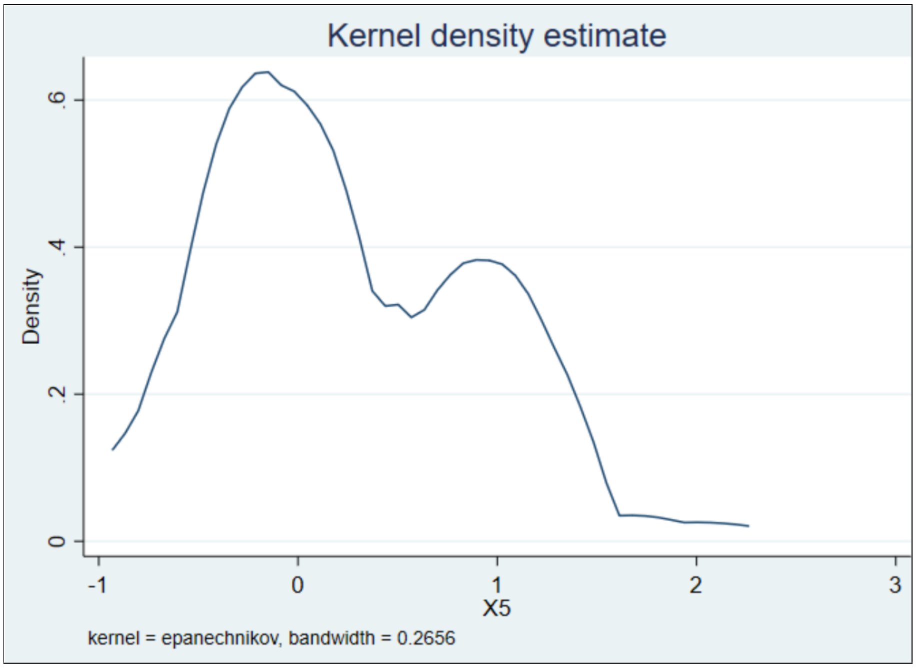

Figure 5: Kernel Density Estimate for Question #5

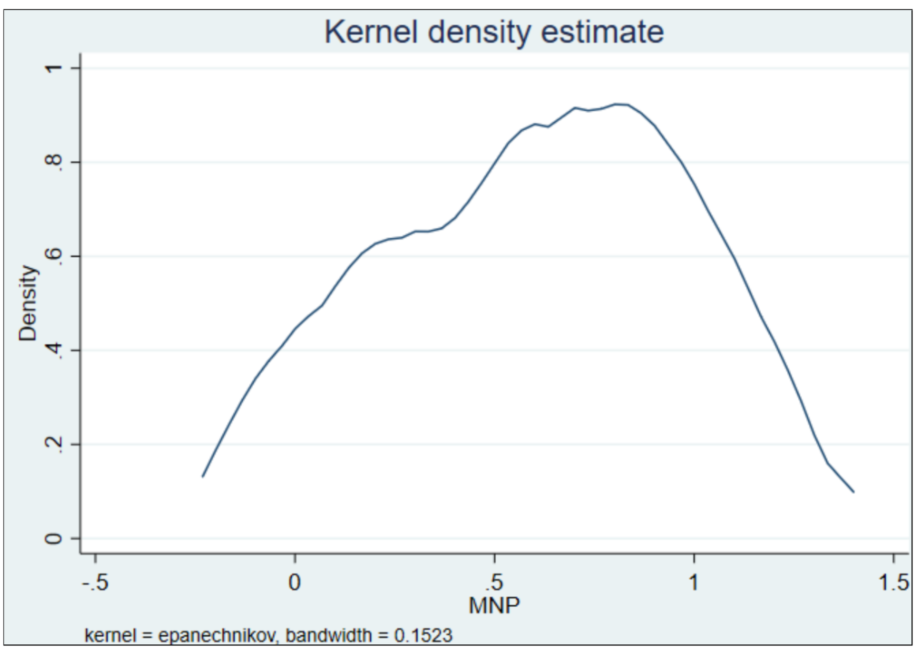

Figure 6: Kernel Density Estimate for All Questions (Combined)

APPENDIX B
TABLES

		Observations	*Mean*	*Std. Deviation*	*Minimum*	*Maximum*	*Rational Belief*	*Correlation Neglect*	*Base Signal*
Control Questions	Question #1	51	21.79	7.48	10	50	25	-	10
	Question #2	51	20.64	6.45	0	30	25	-	10
Treatment Questions	Question #1	50	20.68	8.08	10	50	17.5	28.75	40
	Question #2	49	33.4	12.68	0	60	32.5	51.25	70
	Question #3	50	18.65	7.49	0	32.5	32.5	21.25	10
	Question #4	48	33.91	12.24	10	57.5	57.5	38.75	20
	Question #5	49	21.04	8.80	5	50	20	35	50

Table I: Descriptive Statistics and Correlation Neglect by Belief Formation Task.

	Control Questions		*Treatment Questions*				
	Question #1	Question #2	Question #1	Question #2	Question #3	Question #4	Question #5
Constant	22.718^{***}	19.7985^{***}	18.916^{***}	33.995^{***}	18.507^{***}	34.890^{***}	21.451^{***}
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education	-1.860	-1.505	0.246	-2.056	-1.067	-3.918	-0.072
	(0.313)	(0.361)	(0.904)	(0.569)	(0.614)	(0.294)	(0.978)
Gender	0.684	3.881^{**}	1.426	2.429	1.932	2.956	-0.497
	(0.738)	(0.039)	(0.530)	(0.534)	(0.414)	(0.462)	(0.865)
Duration	-0.000	-0.000	-0.000	0.000	0.000	0.000	-0.000
	(0.503)	(0.540)	(0.334)	(0.964)	(0.655)	(0.200)	(0.629)

Table II: Base Regressions for All Questions (Separate). Coefficient estimate p-values indicated below estimates, indicating *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

	Normalized Beliefs				
	$X1$	$X2$	$X3$	$X4$	$X5$
Constant	0.126	0.080	0.251^{***}	1.206^{***}	0.309
	(0.619)	(0.764)	(0.000)	(0.000)	(0.247)
Education	0.022	-0.110	0.087	0.209	0.120
	(0.904)	(0.569)	(0.653)	(0.294)	(0.525)
Gender	0.127	0.130	-0.192	-0.158	-0.265
	(0.530)	(0.534)	(0.374)	(0.462)	(0.212)
Duration	-0.000	0.000	-0.000	-0.000	-0.000
	(0.619)	(0.964)	(0.695)	(0.200)	(0.899)

Table III: Base Regressions for Normalized Beliefs (for each question). Coefficient estimate p-values indicated below estimates, indicating *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

APPENDIX C
SURVEY

Questions: Demographics

(1) What is your age?
(2) What is your gender?
(3) What is the highest degree or level of school you have completed? If currently enrolled in a program, what is your highest degree received?

Questions: Control

(1) "You are an investor in a company that uses a trustworthy computer program to predict the future market value of any given stock. This computer predicts that Google's stock will increase by 10% next month. The current value of the stock is $100. Several other companies utilize equally trustworthy and independent computers A, B, C, D to predict the market value next month."
By what percent do you think the value will increase next month if you know:
  - Computer A predicts it will increase by 10% next month.
  - Computer B predicts it will increase by 20% next month.
  - Computer C predicts it will increase by 30% next month.
  - Computer D predicts it will increase by 40% next month.
(2) "Now suppose that your computer predicts that Google's stock will decrease by 10% next month."
By what percent do you think the value will decrease next month if you know:
  - Computer A predicts it will decrease by 10% next month.
  - Computer B predicts it will decrease by 20% next month.
  - Computer C predicts it will decrease by 30% next month.
  - Computer D predicts it will decrease by 40% next month.

Questions: Treatment

(1) "You are an investor in a company that uses a trustworthy computer program to predict the future market value of any given stock. This computer predicts that Google's stock will increase by 10% next month. The current value of the stock is $100. Several other companies utilize equally trustworthy computers A, B, C, D to predict the market value next month. Predictions by these four computers may or may not be made independently. Any prediction made by your own computer is made independently."
By what percent do you think the value will increase next month if you know:
  - Computer A predicts it will increase by 40% next month.
  - Computer B predicts it will increase by 15% next month, based on the average between Computer B and A.
  - Computer C predicts it will increase by 25% next month, based on the average between Computer C and A.
  - Computer D predicts it will increase by 35% next month, based on the average between Computer D and A.
(2) "You are an investor in a company that uses a trustworthy computer program to predict the future market value of any given stock. This computer predicts that Google's stock

will increase by 10% next month. The current value of the stock is \$100. Several other companies utilize equally trustworthy computers A, B, C, D to predict the market value next month. Predictions by these four computers may or may not be made independently. Any prediction made by your own computer is made independently."

By what percent do you think the value will increase next month if you know:

- Computer A predicts it will increase by 70% next month.
- Computer B predicts it will increase by 35% next month, based on the average between Computer B and A.
- Computer C predicts it will increase by 45% next month, based on the average between Computer C and A.
- Computer D predicts it will increase by 55% next month, based on the average between Computer D and A.

(3) "Now suppose that your computer predicts that Google's stock will decrease by 10% next month."

By what percent do you think the value will decrease next month if you know:

- Computer A predicts it will decrease by 10% next month.
- Computer B predicts it will decrease by 15% next month, based on the average between Computer B and A.
- Computer C predicts it will decrease by 25% next month, based on the average between Computer C and A.
- Computer D predicts it will decrease by 35% next month, based on the average between Computer D and A.

(4) "Now suppose that your computer predicts that Google's stock will decrease by 10% next month."

By what percent do you think the value will decrease next month if you know:

- Computer A predicts it will decrease by 20% next month.
- Computer B predicts it will decrease by 35% next month, based on the average between Computer B and A.
- Computer C predicts it will decrease by 45% next month, based on the average between Computer C and A.
- Computer D predicts it will decrease by 55% next month, based on the average between Computer D and A.

(5) "Now suppose that your computer predicts that Google's stock will increase by 10% next month."

By what percent do you think the value will decrease next month if you know:

- Computer A predicts it will decrease by 50% next month.
- Computer B predicts it will decrease by 25% next month, based on the average between Computer B and A.