10-17-2023

Numeracy, Consistency, and the Allais Paradox

Sean J. McGinity
Illinois State University, seanjmcginity@gmail.com

NUMERACY, CONSISTENCY, AND THE ALLAIS PARADOX

SEAN J. MCGINITY

44 Pages

This paper explores the relations between mathematical abilities and preferences, preferences for consistency, and the Allais paradox. To answer this question, 144 participants completed a survey which measured objective numeracy, subjective numeracy, preference for consistency, and consistency of probabilistic choices. The results did not show the predicted relations where consistency would moderate the relations where subjective numeracy mediates the relations between objective numeracy and the consistency of probabilistic choices. The previously indicated relations where subjective numeracy simply mediated the relations between objective numeracy and probabilistic reasoning were also not replicated.

While the hypotheses were not supported overall one interesting finding was that consistency of probabilistic choices was affected by consistency. Additionally, the relations between subjective and objective numeracy while weaker were still present as has been previously found. While the sample size was slightly smaller than the original power analysis suggested, and the participants spent less time on average than would have been expected the results seem to suggest that there may not be a systematic relationship between numeracy and consistency.

KEYWORDS: Allais paradox, numeracy, consistency, decision-making, mediation

NUMERACY, CONSISTENCY, AND THE ALLAIS PARADOX

SEAN J. MCGINITY

A Thesis Submitted in Partial
Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

Department of Psychology

ILLINOIS STATE UNIVERSITY

2023

NUMERACY, CONSISTENCY, AND THE ALLAIS PARADOX

SEAN J. MCGINITY

COMMITTEE MEMBERS:

Matthew Hesson-McInnis, Chair

Julie Campbell

ACKNOWLEDGMENTS

I would like to thank my committee members Dr. Matthew Hesson-McInnis & Dr. Julie Campbell for all the time they have spent helping me through the many revisions of this document, as well as Dr. Alexandra Ilie for ensuring its general quality. I would also like to thank my family for all their support in its many forms. Finally, I would like to thank everyone who has every had the patience to listen to me try and describe this document through conversation.

S.J.M.

CONTENTS

TABLES

FIGURES

CHAPTER I: INTRODUCTION

Numeracy is defined as an "aptitude with probabilities, fractions, and ratios" (Fagerlin et al., 2007). Within the context of this paper (and the scales used) numeracy will generally refer to the interpretation and manipulation of probabilities. At present there is a wealth of literature showing the importance of numeracy in a variety of impactful areas of decision-making. An extensive literature review by Garcia-Retamero et al. (2019) found that higher numeracy was associated with better outcomes in health and finance in both lab and real-world settings. They reported that people with lower numeracy tend to overestimate their risk of getting certain diseases, are less well-equipped to make medical decisions, have more trouble gauging the risks of side effects, and have poorer interactions with their doctors. Those with lower numeracy are prescribed 20% more medications and have a 40% higher risk of comorbid conditions, including myocardial infarction, chronic obstructive pulmonary disease, peptic ulcer disease, liver disease, diabetes, and HIV/AIDS (Garcia-Retamero et al. 2019, p. 3). More troubling is that this is the case after controlling for education, age, ethnicity, and body mass-index.

The authors similarly cited several papers linking financial outcomes to numeracy. One such paper by Estrada-Mejia et al. (2016) found that a one-point increase on the Lipkus Numeracy Scale (an 11-point scale) corresponded to a five percent increase in personal wealth. The authors additionally found that numeracy correlated with a person's ability to correctly calculate change, whether they had an emergency fund, and how likely they were to incur checking fees.

To utilize this link to enhance such outcomes, the surprisingly complicated nature of numeracy and its relation to other variables that may affect decision-making must be more precisely understood. For instance, the use of mathematical abilities like simple addition in everyday life can be affected by a person's own perception of those abilities (Jansen et al., 2016).

In many studies involving the measurement of numeracy and its effects on specific outcomes, there is a presumed understanding that a person's risk-taking is minimized and that their preference for greater rewards is maximized when engaged in tasks like choosing between a more reliable lower payout and a less reliable higher payout. One way to avoid this assumption is to use the Allais paradox. The Allais paradox was originally conceived by the French physicist and economist Maurice Allais as a counterexample to Expected Utility Theory (EUT), which is sometimes called the Von Neumann–Morgenstern utility theorem. EUT is an economic theory meant to explain behavior through four axioms: completeness, transitivity, independence, and continuity. The Allais paradox shows a flaw in the independence axiom. Harman & Gonzalez (2015) state that:

> The independence axiom in EUT states that a rational decision maker should not base his or her preference on outcomes that are identical in amount and probability between gambles. Common outcomes between two gambles should cancel out. (para. 2)

The Allais paradox is the demonstration of a common violation of the independence axiom made by decision-makers in pairs of gambles like the one below:

Gamble pair 1:	Gamble pair 2:
A: \$1000 (p = 1)	A′: \$1000 (p = .11).; \$0 (p = .89)
B: \$1000 (p = .89); \$5000 (p = .1); \$0 (p = .01)	B′: \$5000 (p = .1).; \$0 (p = .9)

In the above set of pairs, option A in gamble pair 1 shows a 100% probability for a person to gain \$1000, while option B shows an 89% probability of gaining \$1000, a 10% probability of gaining $5000, and a 1% chance of gaining $0. The task of the participant is to choose A or B. Once a participant has chosen A or B from the first gamble pair, they are then presented with gamble pair 2. In gamble pair 2, a participant is presented with option A′ and B′. Option A′ gives an 11% probability of gaining $1000 and an 89% chance of gaining $0, while option B′ shows a 10% probability of gaining $5000 and a 90% probability of gaining $0.

Violations of EUT occur when a participant would first select A then B′ or B then A′. With said violations being demonstrated below:

First Choice		**Second Choice**
A	————→	A′
B	————→	B′
A	- - - - - - →	B′
B	- - - - - - →	A′

EUT Consistent ————→ EUT Inconsistent - - - - - - →

Plenty of studies including Da Silva et al. (2013), Harman & Gonzalez (2015), and Kahneman & Tversky (1979) have demonstrated that a significant number of participants violate the assumption of independence when they choose A and B′ rather than a consistent pair like A & A′ or B & B′. This violation happens because the difference of expected utility is consistent between both pairs. Demonstrated by rewriting the pairs like so:

Gamble pair 1:

A: 1000 (p = .89); 1000 (p = .11)

B: 1000 (p = .89); 0 (p = .01); 5000 (p = .1)

Gamble pair 2:

A′: 0 (p = .89); 1000 (p = .11)

B′: 0 (p = .89); 0 (p = .01); 5000 (p = .1)

Moving from the original pair of A & B to A′ & B′ we see that the only difference between the two is that 1000 (p = .89) has been reduced to 0 (p = .89). Even though the difference is consistent between each pair of gambles, a significant proportion of participants still choose A in the original set and B′ in the altered set; thus, illustrating that the change in the outcomes, even though identical for both possibilities, does not have an independent effect on the choice made. The typical explanation for this change rests upon framing effects of the individual gambles where one takes the sure gamble because it looks better and takes the riskier gamble in the second because they feel there is little difference. Prospect theory, introduced in Kahneman & Tversky (1979), states that the risk of loss has an outsized effect on a participant's choices as compared to the potential gain. However, these explanations fail to account for potential personality differences like a person's preference for consistency.

Preference for consistency was formally constructed as a scale under the same name (the PFC for short) by Cialdini et al. (1995). Preference for consistency is a survey which measures how important it is to a person that they behave consistently over time (i.e., do the same thing given the same options), that they are predictable, that their attitudes and beliefs are consistent with their actions, and that they prefer the company of others who display these same kinds of consistency. Relatively little research has been done on the relation between numeracy and consistency. Vestewig (1978), found a marginally significant difference between those who self-reported high consistency on gambling tasks vs. those who self-reported low consistency and their actual consistency between two gambling tasks, however this was a measurement of strategy of gambling tasks not a personality variable like the PFC. Meanwhile Ashby (2017) found that numeracy increased the amount of sampling (practice) done when participants could practice choosing between different scenarios low-risk/high-risk scenarios.

A measure of an individual's preference for consistency as well as their actual consistency on a probabilistic instrument like the Allais paradox (with its variety of probabilistic choices) allows us to understand what effect this preference for consistency may have on probabilistic and risk-based instruments such as those choices they would encounter in the medical or financial sectors. Some individuals may have little preference to behave consistently on such measures (and perhaps in the real world) and this would obfuscate any relation between their understanding of risks and benefits of such choices and their actual selections. For example, in a medical survey that asks patients to make a series of choices about whether they would prefer one-time surgery or long-term medication such individuals may (while having a very good understanding of the risks and benefits of each) alternate answers, in a way which suggests that they do not understand the material, because they do not seem to have a defined preference. This risk seems highly likely for surveys that are meant to gauge the general choices of a medical population where a defined preference may be assumed. To demonstrate a clearer link between probabilistic choices and an individual's actual preferences, the specific relations between a preference for consistency and consistency of choices may need to be more profoundly developed.

CHAPTER II: LITERATURE REVIEW

One of the best explorations into the general nature of numeracy as a construct was done in Peters & Bjalkebring (2015). Their paper included several studies involving two previously validated scales of numeracy on gambling tasks, which are frequently used in measuring mathematical ability because of the need to understand probability. The objective numeracy scale (ONS) is used as a measure of mathematical ability, including items related to statistics and probability (Weller et al., 2013). The subjective numeracy scale (SNS) is a measure of a person's self-perceived competence of and preference for use of mathematical information (Fagerlin et al., 2007). Whereas the SNS was originally developed as a more pleasant, less stressful alternative to the ONS, it is generally considered to be a related, but separate construct (Liberali et al., 2012).

In understanding numeracy as a construct, it is important that a measure of objective ability, such as ONS, is not conflated with one's own perceived ability, such as SNS. The SNS was originally developed as a shorter more pleasant correlate of the ONS meant to keep participants from experiencing math anxiety (Fagerlin et al., 2007). The SNS did so by featuring fewer items and none that required any calculation or computation. Participants merely had to reflect on their mathematical preferences. The two scales of numeracy have been found to be distinct in later studies. Liberali et al. (2012) conducted a factor analysis using ONS, SNS, and the Cognitive Reflection Test (CRT). They found that although the ONS & SNS are similar, the ONS shares more similarities with the CRT, another cognitive measure of numeracy, than with the more self-reported nature of the one's own cognitive ability found in the SNS.

Peters & Bjalkebring (2015) used structural equation modeling to show that ONS is more highly predictive of SNS, and that SNS acted as a mediator between ONS and several tasks involving the gambling tasks, indicating the general relation of the two variables. Their finding is in line with Jansen et al. (2016) as the relation between mathematical ability and its use appears to be mediated by how a person perceives their own abilities.

In Jansen et al. (2016), using two multiple mediated models, the authors found that the relation between better addition skills (related to mathematical ability and objective numeracy) and using math in everyday life were positively mediated by perceived mathematical competence. Math anxiety negatively mediated the relationship between addition skills and use of math in everyday life, but only for women. Their findings provide support for the idea that perceived mathematical competence leads to higher use of math for those who are higher in objective numeracy.

Misuraca et al. (2016) found that more numerate individuals used more effortful decision-making in the form of maximizing (using a more exhaustive search strategy when deciding between alternatives). It may be that more numerate individuals are likely to try to fully understand their options before deciding, which may allow them to be more consistent. Similarly, participants may be less likely to make consistent choices as opposed to changing their preference when their cognitive capacity is reduced (Olschewski et al., 2018). Those with higher numeracy may use their cognitive resources more efficiently when making mathematical decisions, thus allowing for greater consistency. Ashby (2017) found that higher objective numeracy decreased the amount of alternating between options for consecutive trials. Their explanation was that more numerate participants may have been able to look more closely at differences between choices since outcome of each trial was only displayed for 500ms. High objective numeracy also increased the number of risky options explored, as well as consistency from a measure where subjects were given descriptions of risk-based scenarios and successive trials in one study where a set number of trials was given, but not in another where the trials were only limited by the participants. However, the predicted Expected Value (EV) maximization of participants with high objective numeracy was not found.

The lack of expected value maximization of high numeracy participants seems to be due to higher objective numeracy only predicting more effort in examining gambles when the difference between gambles is substantial enough. Traczyk et al. (2018) found that high SNS participants maximized expected value (EV) on every problem, but that participants who had high scores on the Berlin Numeracy Task (BNT), a measure of objective numeracy, only maximized EV where there was a substantial difference in pay-off. Traczyk et al. (2020), found that those who had higher numeracy based on the BNT had more consistency in probability processing and therefore more consistency across consecutive bets in gambling tasks where participants were asked to wager within some range of money (e.g., from $0 to $500) when the probability of winning was provided. Participants who scored higher on the BNT had less variance in their wagers with similar probabilities than those with lower scores on the BNT. The authors suggested that the higher consistency in probability processing (being less swayed by affective context like loss-frames and feedback) led more numerate individuals to behave more consistently across gambles. However, there does not appear to be evidence that this higher consistency was due to more numerate individuals being more likely to explicitly calculate expected value over all their wagers.

Most studies have found that objective numeracy and similar constructs are associated with better gambling choices, but some articles have also shown the complicated nature of these relations. A survey of a variety of different studies conducted by the Peters et al., (2006), to understand decision-making of those with high and low numeracy found such mixed results. Four studies were used to explore a variety of framing effects. The first found that those with lower numeracy had more susceptibility to negative/positive framing effects (i.e., "Susie missed 22% of the points on her test." vs. "Susie got 78% of the points on her test."), where the negative/positive presentation was manipulated between-subjects. The effect was more pronounced for those with lower numeracy but remained in the high numeracy group. The second study found that in risk presentation (i.e., "10% of 100 patients will die." vs. "10 out of 100 patients will die.") lower numeracy participants rated the 10% as less risky than 10, with the higher numeracy participants rating not showing a difference. The third study found that less numerate participants were more likely to choose sub-optimally between a 9 in 100 (labelled as 9%) chance vs. a 1 in 10 chance (labelled as 10%), which may be linked to their less well-formed feelings about the two choices. Peters et al.'s fourth study was a loss-bets task with a between-subjects design where one group evaluated a no loss gamble, with the possibility of winning $9 with a probability of 7/36 and another group was asked to evaluate a gamble where they could potentially win $9 with a probability of 7/36 or lose $.05 with a probability of 29/36. Those with high numeracy rated the loss gamble more favorably than the no loss gamble, and scores were otherwise roughly equivalent. The authors propose that the reason could be that those in the high numeracy are able to use the -$.05 loss as a relevant cue in relation to the $9 to see the overall value of the gamble but are unable to do so otherwise.

Peters et al., (2019), greatly extended research into this paradox using four different studies with a variety of samples (actuaries, accountants, and a general sample taken from mechanical Turk). The loss-bet paradox occurs when separate sets of participants are asked to evaluate a bet with no loss condition ($9 in 7/36) or a bet with a loss condition ($9 in 7/36; -$.05 in 29/36). On average, participants in the loss condition evaluate it more favorably, but interestingly this increases with their numeracy. The authors confirmed that this occurred under a variety of frames (loss emphasized, win emphasized, order switched, etc.) and found that it held in all of these. When participants were asked to compare the difference between the $9 win and the $.05 loss, the interaction increased, and the difference increased the more numerate the individual was. These findings should hopefully not interfere with consistency choice on the Allais paradox as all options will involve comparing gains and losses and there was no interaction between it and presentation of gambles. However, it does demonstrate the rather complicated relation that objective numeracy can have with gambling tasks.

Harman & Gonzalez (2015) found that descriptive framing reduced probability of inconsistent choices from Allais paradox and that these inconsistent choices are also eliminated with enough experience of the gambling pairs. Additionally, the authors found that participants who were more consistent explored more gambles to gain more experience. It is likely that those with higher numeracy are more likely to put more effort into understanding those gambles and would be likely to understand those gambles sooner than those with lower numeracy. The same relation of higher ability objective numeracy and more effortful subjective numeracy strategies would emerge if the authors had measured these constructs as well.

Of potential concern is that the scale being used to measure a person's preference for consistency, the Preference for Consistency Scale (PFC) constructed in Cialdini (1995) was later explored as a bipolar scale, with three categories. Bator & Cialdini (2006) did several studies on preference for consistency; based on theory and previous results, they suggested that the PFC scale could be used to divide individuals into three groups (anticonsistent, aconsistent, and consistent preferring). However, they found only marginal support for use as a nominal variable rather than scale. So, hopefully data from the scale can be assumed to be approximately linear, so as not to further complicate the model.

Wedell & Böckenholt (1990), as well as Vestewig (1978), found a marginally significant difference between those who self-reported high consistency on gambling tasks vs. those who self-reported low consistency and their actual consistency between two gambling tasks. The current study, with a mediated model, will be better able to detect reported consistency (in this case an abbreviated PFC) and actual consistency in the Allais gambles.

Yechiam (2020) found that rates of switching between choices were more highly correlated between different tasks and sessions than actual choice selections. Those findings present a potential problem for the current study design as well as a potential boon for its theoretical value. The current study attempts to see if choice switching is lower for individuals with lower PFC.

Other concerns are related generally to framing effects, as some studies have found that they may be generally troublesome for the implications of many studies of mathematical decision-making. Mandel (2014) found that differences in numeracy could be at least partially attributed to how exact descriptions are in gambles like the Allais paradox and that individual differences may be partly accounted for by differing assumptions about quantifiers. Additionally, Levy et al. (2014), found that differences in ONS performance could be at least partially accounted for by what domain the questions pertained to (i.e., health/medical question, financial question). Those findings could pose a problem with generalizing from the ONS.

The relatively small number of studies that have measured the link between objective numeracy and consistency have only done so in terms of internal consistency or consistency between measures on related dependent variables (Ashby, 2017). One possible explanation for this link would be that people with higher objective numeracy tend to be more consistent in general. Another is that objective numeracy allows people with a preference for consistency to behave in more mathematically consistent ways.

The Current Study

More numerate individuals should be able to understand the gambles more deeply than less numerate individuals in a manner previously described by Peters & Bjalkebring (2015) where SNS mediates the relation between ONS and a mathematical task. Hypothesis 1 was that the mediation by SNS between ONS and the Allais task would be significant in replicating Peters & Bjalkebring (2015; Figure 1). Additionally, those with a lesser preference for consistency may not be motivated to answer consistently in the gambling pairs even if they have higher ONS and SNS, so the mediation should be moderated by their preference for consistency, with a higher preference leading to more consistent gambles. Hypothesis 2 was therefore that the preference for consistency would positively moderate the relation between the ONS and number of consistent gambles in the Allais paradox task, which was moderated by SNS, positively (Figure 2).

CHAPTER III: METHODS

Power Analysis

Harman & Gonzalez (2015) was used to determine the appropriate effect sizes. They reported 9 chi-squared tests for the Allais Paradox. The chi-squared tests were weighted and averaged and then converted to R^2 using the method from Lenhard, W. & Lenhard, A. (2016). The R^2 of 0.12 was then used with G-power to create the table below:

GPower Analysis

		Power		
		0.8	0.85	0.9
Variables	4	96	107	122
	5	103	115	131
	8	122	135	153

Based on the results of this table, the maximum number of participants needed to determine significance was 153.

Participants

Two-hundred participants were requested through MTurk, with only those who indicated being in the United States and over the age of 18 able to see the request. A total of 199 valid records were able to be acquired. Due to failed attention checks 55 of the participants in that sample were excluded, leaving the total number of participants available for analysis at 144. While it was planned to collect demographic data including gender, age, and socio-economic status the final version of the survey, thanks to human oversight, did not ask participants for any of that information. So, exact demographics of the sample were not collected.

A study by Douglas et al. (2023) was conducted specifically to compare data quality from online sources like MTurk. They recruited 500 adult participants in the United States from

MTurk and found that workers had an average age of 38.75 (*SD* = 11.53). Participants were 36.20% female, 63.40% male, and 0.40% preferred not to say. Respondents were 75.40% Caucasian, 15.00% African American, 5.80% Latinx or Hispanic, 4.00% Asian, 1.20% American Indian or Alaskan Native, 0.40% Middle Eastern or North African, with 0.60% holding another identity and 1.00% preferring not to say. Most participants (60.00%) had a four-year degree, 21.80% had a masters, 7.00% had some college, 5.40% had a two-year degree, 5.00% had a high school degree, 0.60% had a doctorate, and 0.20% had not finished high school. Family income is broken down in Table 1.

Measures

Four total variables were used in the study. The first was objective numeracy, which is a measure of how well individuals use probabilistic and mathematical concepts. It was measured using the Rasch-based numeracy scale (RNS) developed by Weller et al. (2013; Appendix A). This abbreviated numeracy scale is approximately normal and only 8-items long (scored from 0 to 8 with one point for each correct answer) combining elements of the Lipkus' ONS and the Frederick's CRT scales. It has good internal consistency (α = .69) and has a better ability to differentiate numeracy between different levels than any other objective numeracy scale. These psychometric properties are why it was preferred. It should be noted there are some concerns that CRT and ONS measure separate constructs. Liberali et al. (2012) found that some of the items between the two scales loaded onto separate factors. However, Weller et. al (2013) found that the items did not load onto separate factors and were non-orthogonal.

The second variable was subjective numeracy, which is a measure of a person's mathematical and probabilistic self-efficacy and preference for using mathematics in decision-making. The scale that was used to measure this variable was the Subjective Numeracy Scale developed by Fagerlin et al. (2007; Appendix B). It has 8 items, each with a 6-point Likert type scale and has very good internal consistency with a Cronbach's alpha of .83.

The third variable was preference for consistency, a measure of how much a person prefers that their own actions, as well as the actions of those they associate with, are consistent over time and with their beliefs. Preference for consistency was measured by the abbreviated preference for consistency scale (PFC-B) developed by Cialdini et al. (1995; Appendix C). The original preference for consistency scale (PFC; $\alpha = .89$) showed only slight improvement over the abbreviated version (PFC-B; $\alpha =.84$). So, the abbreviated 9-item scale will be used instead of the 18-item scale in the interest of time.

Finally, consistency of risk-taking was measured by consistency of the Allais Paradox. The scale used was the one developed by Kahneman & Tversky (1979) and updated and reformatted by Da Silva et al. (2013; Appendix D). This scale consists of six pairs of questions with two options each (12 questions total) where answering both A & C is consistent according to Expected Utility Theory (EUT) and answering both B & D is also consistent. The total possible number of consistent choices made by a participant ranged from 0 to 6.

Procedures

Participants were recruited using Amazon's Mechanical Turk survey website where they received $0.50 if they successfully completed the survey. They were first informed of the risks involved with the individual tasks. They were then informed that the study was anonymous and that the only risk is potential distress for anyone with a history of math anxiety. Additionally, they were asked if they understood and consented to participating in the study. If they proceeded, each of the individual instruments was randomly given. During each of the four survey instruments one or two attention checks were given (i.e., “choose 3 for this question”) to ensure that participants are reading each question. After the completion of the surveys, participants were debriefed on the potential benefits of the survey and given contact information for any questions they may have had about the study.

Analyses

The data was analyzed through two regression models which were based on the two hypotheses (Figure 1 & Figure 2). The analysis was completed in R and used the PROCESS function (Hayes, 2013). The first model was a series of relations between objective numeracy (RNS), subjective numeracy (SNS) and consistency of the Allais paradox choices. In this model RNS predicted both SNS and Allais consistency directly while also predicting Allais consistency through an SNS mediated pathway. The significance of the main effect of ONS on the Allais consistency was determined through the regression slope while the effect of the mediation will be determined through bootstrapping a 95% CI, significant if and only if it does not include 0 in its range.

The second model was identical to the first model with the addition of a moderation of a fourth variable (PFC-B) on the pathway between SNS and Allais consistency. The significance of the main effect of ONS on Allais consistency again was determined by the regression slope. The effect of the mediated moderation was determined by bootstrapping a 95% CI (based on 5000 samples) with significance indicated by a range where 0 is not inclusive.

CHAPTER IV: RESULTS

The Allais scores ranged from 0 to 6, had a mean of 3.73 (*SD* = 1.54), a median of 4, and were normally distributed with a skew of -.24. Scores from the Rasch-based Numeracy Scale ranged from 0 to 8, with a mean of 2.57 (*SD* = 1.88), a median of 2, and were somewhat right-tail skewed with a value of .82. The Subjective Numeracy Scale scores ranged from 25 to 48, had a mean of 36.22 (*SD* = 4.39), a median of 36, and were normally distributed with a skew of .19. Scores from the Preference for Consistency ranged from 39 to 81, had a mean of 61.97 (*SD* = 7.58), a median of 63 and were slightly left tail skewed with a value of -.47.

A multiple regression with Allais score predicted by RNS, SNS, and PFC-B with no interactions was used to test assumptions because these were not testable using the model and output generated by the process macro. Independence was tested using a Durbin-Watson statistic, which had a value of 1.93, very close to the standard value of 2, indicating that the assumption of independence was met. A Breusch-Pagan test ($\chi^2(1, 144) = .37$, $p = 0.54$) along with a scatterplot of the predicted values by the standardized residual values (Figure 5) were used to determine that the assumption of homoskedasticity had been met. The scatterplot was also used to verify that the assumption of linearity had been met. The assumption of multicollinearity was met with VIF values of 1.03 for the RNS, 1.11 for the SNS, and 1.09 for the PFC-B, all well below the cutoff of 10. Finally, the assumption of normality was met by creating a pp plot from the standardized residuals which showed an approximately normal distribution (Figure 6).

The results from the first model did not replicate the findings of Peters & Bjalkebring (2015). The effect of RNS on SNS was significant, $b = .16$, $t(144) = 1.50$, $p = .049$, bootstrap 95% CI = [.001, .328]. The effect of SNS on Allais score was not significant, $b = -.05$, $t(144) = -$

0.36, $p = .720$, bootstrap 95% CI = [-.307, .212]. Additionally, the direct effect of RNS on Allais score was not significant, $b = -.13$, $t(144) = -.89$, $p = .376$, bootstrap 95% CI [-.385, .134]. Nor was the indirect effect mediated through SNS, $b = -.01$, boot 95% CI = [-.063, .043]. Given these results the first hypothesis was mostly unsupported. This information is condensed in Figure 3.

The results from the second model were similar and did carry over as expected. The effect of RNS on SNS was again significant, $b = .16$, $t(144) = 1.50$, $p = .049$, bootstrap 95% CI = [.001, .328]. The effect of SNS on Allais score was not significant, $b = -.16$, $t(144) = -1.08$, $p = .28$. The simple effect of PFC-B on Allais score was significant, $b = .29$, $t(144) = 2.21$, $p = .029$. However, the interaction of PFC-B on the mediation of RNS by SNS on Allais score was not significant, $\Delta R^2 = .13$, $F(1, 139) = 1.83$, $p = .178$. Nor was the indirect effect mediated through SNS significant, $b = -.02$, bootstrap 95% CI = [-.079, .013]. Additionally, the direct effect of RNS on Allais score was not significant, $b = -.15$, $t(144) = -1.17$, $p = .243$, bootstrap 95% CI = [-.408, .104]. Given these results the second hypothesis was largely unsupported. This information is condensed in Figure 4.

CHAPTER V: DISCUSSION

Neither model indicates the expected significance of the mediated pathway of objective numeracy's effect on the Allais' paradox's consistency through subjective numeracy. The interaction term between the moderated pathway for preference for consistency was also therefore non-significant. One potential reason for this could have been the relatively weak relation between RNS and SNS within the sample. While still significant, the correlation between the two variables was weak, $r = .16$. Previous studies have shown much higher correlations between the two. In Peters & Bjalkebring (2015) the correlation between the two variables was highly significant with a much more moderate correlation of .46. This pathway is integral to the proposed model as a whole and such a weak correlation presents a problem with detecting an indirect pathway.

Given this, it is not surprising that the model was not replicated even with a similar gambling variable being used as the dependent variable in a similar model. However, it is also possible given the lack of correlation and the surprising direction (Table 2) that neither of the variables (at least in their present forms) were significantly related to the Allais paradox.

The second hypothesis was also therefore not supported overall, given that any interaction between preference for consistency hinged on the indirect pathway being established in the simpler model. Surprisingly, the simple effect of PFC-B on Allais score was in fact significant, indicating some sort of relation between the two variables. This in and of itself is an interesting finding that shows that the model was able to detect that if a person reported having less desire to be consistent overall, their decisions inside of the Allais task followed suit.

It is difficult to determine why the models failed to support the hypotheses, as there were several limitations within the study. Fifty-five of the participants failed at least one attention check, with 29 of those only failing one, and 17 of those only failing two. However, only allowing for those who passed all checks, the sample was brought down to 144, a mere 9 participants short of the indicated minimum to provide sufficient power, and while it is possible that this may have contributed, it is unlikely that this alone created the deviation between expectation and reality. It could have been that the measurements together did not provide sufficient power. The Rasch-based numeracy scale, as well as the abbreviated preference for consistency, were selected due to the smaller number of items. Neither had previously been used within a model like this one.

Additionally, it could have been that the sample which was taken from an online marketplace (which pays participants for their time and attention) could have proven to be less controlled than the settings used in previous studies. Of particular concern is that in Olschewski et al. (2018), participants, when asked to perform more complicated simultaneous tasks, made choices that were more inconsistent than when asked to perform simpler tasks simultaneously. Given the nature of online surveys and the ability of participants to complete them anywhere, this may mean that the many distractions participants in the current study may have encountered could have reduced their ability to give consistent answers.

In contrast to this a meta-analysis using 90 samples by Walter et al. (2019), suggests that online sources of data are of similar quality to those from more conventional sources. Similarly, a meta-analysis by Keith et al. (2023) indicates that while mean and variance do not differ between MTurk and non-MTurk overall, domain affects variance at least for tasks involving perception. The researchers did not examine decision-making specifically, so it is possible that variance differs for those tasks as well. However, a more limited paper by Wolfson & Bartkus (2013) seems to indicate that behavioral economic findings can be replicated on MTurk, at least on three of the four experiments they conducted. Of interest was an experiment that replicated Kahneman and Tversky's Prospect Theory, a well-supported solution to the Allais paradox.

One particular concern was with the amount of time participants spent on the survey. Based on early testing of the survey it was determined that participants would generally need roughly 7 minutes to complete the survey while comprehending each question. Participants spent between 1 minute and 23 seconds and 30 minutes and 8 seconds with a mean time of 6 minutes 41.59 seconds and a standard deviation of 5 minutes and 53.74 seconds. Meanwhile the median time was 3 minutes and 57.5 seconds, making the distribution highly right tail skewed, 1.59. So, over half of the participants spent considerably less time than expected for adequate attention to have been given. However, this did not appear to affect the participants' ability to pass attention checks as the relation between time and passed attention checks for all 199 participants was practically non-existent ($r = -.01$, $p = .87$).

The inclusion of attention checks was not considered to be a potential issue at the time the study was designed; however, Keith et al. (2017), note that it may need to be considered a double-edged sword. In their paper examining MTurk samples in organizational research they found that attention checks may alter demography and exclude otherwise valid participants. They also note that ~ 88% of the MTurk population should be able to complete simple attention checks (e.g., 'select Strongly Disagree' or 'Type "BLUE"') which were the only kind included in the survey. They go on to say that completeness and data quality may be improved by these attention checks and recommend using several complex kinds. While it is possible that the analysis may have been hindered by the exclusion of certain participants based on completeness and attention checks, it is also possible that their inattention would have diminished the quality. Either way the analysis was performed as planned with the sample limited to those who completed the full survey and all attention checks.

A potentially troubling finding was that the total amount of time a participant spent on the survey was significantly negatively related to the consistency of their Allais answers ($r = -.22$, $p = .01$). Conversely the RNS scores were positively correlated with the amount of time spent on the survey ($r = .28$, $p < .001$). One potential explanation is that most participants simply did not spend very much time on the RNS section. The median score was 2 (with a potential range of 0-8), showing that participants underperformed overall. Additionally, participants correctly answered between 11.11% and 50.69% of the RNS questions, potentially showing that little attention was paid to even the simplest of questions. Comparatively, participants fared much better on the Allais questions where the median score was 4 (out of a possible 0-6). Given this, it is possible that most participants chose to focus their efforts on an easier section since payment was not contingent upon correctness, allowing them to move onto other activities.

In any event, the results likely indicate that if the connection between probability-based decision-making, such as that found in the Allais paradox, is moderated by objective numeracy through subjective numeracy, then any interaction with personal consistency would need a large, better controlled sample to be determined. However, given that it is unlikely that the source of the sample was the sole cause of the lack of support for the hypotheses, it is possible that there is little or no connection between numeracy and consistency. The tools used here could certainly be augmented, but there is no guarantee that the predicted relations would be supported elsewhere later.

REFERENCES

Ashby, N. J. S. (2017). Numeracy Predicts Preference Consistency: Deliberative Search Heuristics Increase Choice Consistency for Choices from Description and Experience. *Judgment and Decision Making*, *12*(2), 128–139. http://dx.doi.org/10.1017/S1930297500005672

Bator, R. J., & Cialdini, R. B. (2006). The nature of consistency motivation: Consistency, aconsistency, and anticonsistency in a dissonance paradigm. *Social Influence*, *1*(3), 208–233. https://doi.org/10.1080/15534510600823638

Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality and Social Psychology*, *69*(2), 318–328. https://doi.org/10.1037/0022-3514.69.2.318

Da Silva, S., Baldo, D., & Matsushita, R. (2013). Biological Correlates of the Allais Paradox. *Applied Economics*, *45*(4–6), 555–568. http://dx.doi.org/10.1080/00036846.2011.607133

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE*, *18*(3), e0279720. https://doi.org/10.1371/journal.pone.0279720

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring Numeracy without a Math Test: Development of the Subjective Numeracy Scale. *Medical Decision Making*, *27*(5), 672–680. https://doi.org/10.1177/0272989X07304449

Garcia-Retamero, R., Sobkow, A., Petrova, D., Garrido, D., & Traczyk, J. (2019). Numeracy and Risk Literacy: What Have We Learned so Far? *The Spanish Journal of Psychology*, *22*, E10. https://doi.org/10.1017/sjp.2019.16

Harman, J. L., & Gonzalez, C. (2015). Allais from Experience: Choice Consistency, Rare Events, and Common Consequences in Repeated Decisions. *Journal of Behavioral Decision Making*, *28*(4), 369–381. https://doi.org/10.1002/bdm.1855

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* Guilford Press.

Jansen, B. R. J., Schmitz, E. A., & van der Maas, H. L. J. (2016). Affective and Motivational Factors Mediate the Relation between Math Skills and Use of Math in Everyday Life. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00513

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Keith, M. G., Stevenor, B. A., & McAbee, S. T. (2023). Scale Mean and Variance Differences in MTurk and Non-MTurk Samples: A Meta-Analysis. *Journal of Personnel Psychology*, *22*(1), 1–12. https://doi.org/10.1027/1866-5888/a000309

Lenhard, W. & Lenhard, A. (2016). *Computation of effect sizes*. Retrieved from: https://www.psychometrica.de/effect_size.html. Psychometrica. DOI: 10.13140/RG.2.2.17823.92329

Levy, H., Ubel, P. A., Dillard, A. J., Weir, D. R., & Fagerlin, A. (2014). Health Numeracy: The Importance of Domain in Assessing Numeracy. *Medical Decision Making*, *34*(1), 107–115. https://doi.org/10.1177/0272989X13493144

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making*, *25*(4), 361–381. https://doi.org/10.1002/bdm.752

Mandel, D. R. (2014). Do framing effects reveal irrational choice? *Journal of Experimental Psychology: General*, *143*(3), 1185–1198. https://doi.org/10.1037/a0034207

Misuraca, R., Teuscher, U., & Carmeci, F. A. (2016). Who are maximizers? Future oriented and highly numerate individuals: WHO ARE MAXIMISERS? *International Journal of Psychology*, *51*(4), 307–311. https://doi.org/10.1002/ijop.12169

Olschewski, S., Rieskamp, J., & Scheibehenne, B. (2018). Taxing cognitive capacities reduces choice consistency rather than preference: A model-based test. *Journal of Experimental Psychology: General*, *147*(4), 462–484. https://doi.org/10.1037/xge0000403

Peters, E., & Bjalkebring, P. (2015). Multiple numeric competencies: When a number is not just a number. *Journal of Personality and Social Psychology*, *108*(5), 802–822. https://doi.org/10.1037/pspp0000019

Peters, E., Fennema, M. G., & Tiede, K. E. (2019). The loss-bet paradox: Actuaries, accountants, and other numerate people rate numerically inferior gambles as superior. *Journal of Behavioral Decision Making*, *32*(1), 15–29. https://doi.org/10.1002/bdm.2085

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and Decision Making. *Psychological Science*, *17*(5), 407–413. https://doi.org/10.1111/j.1467-9280.2006.01720.x

Reyna, V. F., & Brust-Renck, P. G. (2020). How representations of number and numeracy predict decision paradoxes: A fuzzy-trace theory approach. *Journal of Behavioral Decision Making*. https://doi.org/10.1002/bdm.2179

Traczyk, J., Fulawka, K., Lenda, D., & Zaleskiewicz, T. (2020). Consistency in probability processing as a function of affective context and numeracy. *Journal of Behavioral Decision Making*, bdm.2206. https://doi.org/10.1002/bdm.2206

Traczyk, J., Sobkow, A., Fulawka, K., Kus, J., Petrova, D., & Garcia-Retamero, R. (2018). Numerate decision makers don't use more effortful strategies unless it pays: A process tracing investigation of skilled and adaptive strategy selection in risky decision making. *Judgment and Decision Making*, *13*(4), 372–381. https://doi.org/10.1017/S1930297500009244

Vestewig, R. E. (1978). Cross-response mode consistency in risk taking as a function of self-reported strategy and self-perceived consistency. *Journal of Research in Personality*, *12*(2), 152–163. https://doi.org/10.1016/0092-6566(78)90091-0

Walter, S. L., Seibert, S. E., Goering, D., & O'Boyle, E. H. Jr. (2019). A tale of two sample sources: Do results from online panel data and conventional data converge? *Journal of Business and Psychology*, *34*(4), 425–452. https://doi.org/10.1007/s10869-018-9552-y

Wedell, D. H., & Böckenholt, U. (1990). Moderation of preference reversals in the long run. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(2), 429–438. https://doi.org/10.1037/0096-1523.16.2.429

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, *26*(2), 198–212. https://doi.org/10.1002/bdm.1751

Wolfson, S. N., & Bartkus, J. R. (2013). An assessment of experiments run on Amazon's Mechanical Turk. *Mustang Journal of Business and Ethics*, *5*, 119.

Yechiam, E. (2020). Robust consistency of choice switching in decisions from experience. *Judgment and Decision Making*, *15*(1), 74–81. https://doi.org/10.1017/S1930297500006914

TABLES

Table 1

Familial Income for Participants from Douglas et al. (2023) with Combined Income Brackets

Family Income	Percentage
Less than $20,000	6.60%
$20,000 - $39,999	20.00%
$40,000 - $59,999	34.40%
$60,000 - $79,999	16.40%
$80,000 - $99,999	14.40%
At least $100,000	8.20%

Table 2

Correlation Matrix of All Variables Used in Both Models

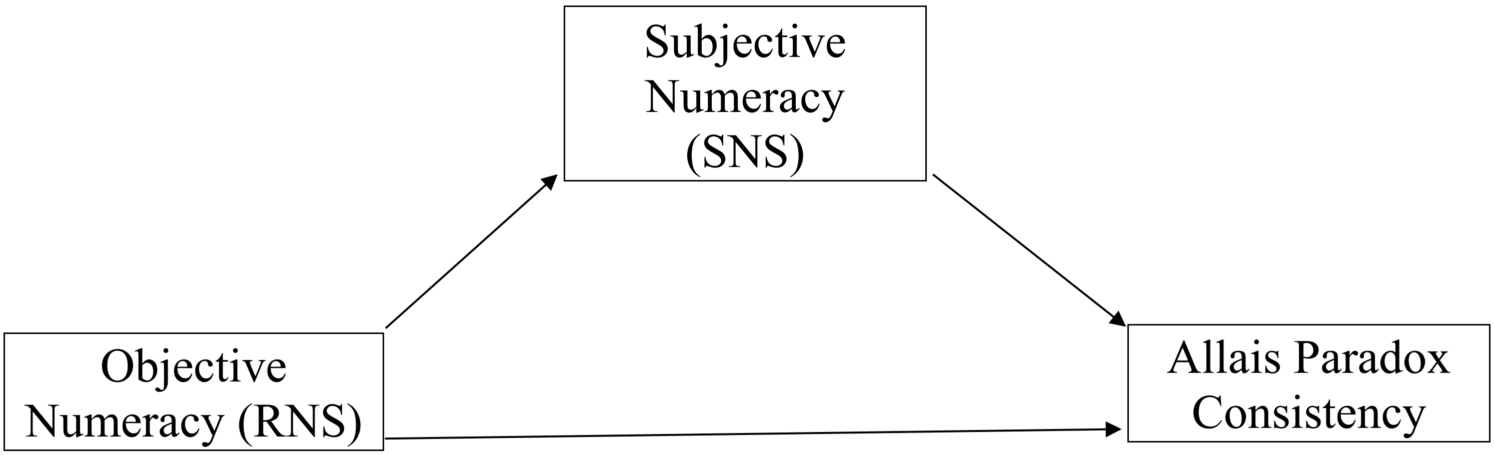
	Allais Score	RNS	SNS	PFC-B
RNS	-.09			
SNS	-.04	.16*		
PFC-B	.15	.10	.28	
Time	-.22**	.28***	.05	.04

Note: *p < .05. **p < .01. ***p < .001

FIGURES

Figure 1

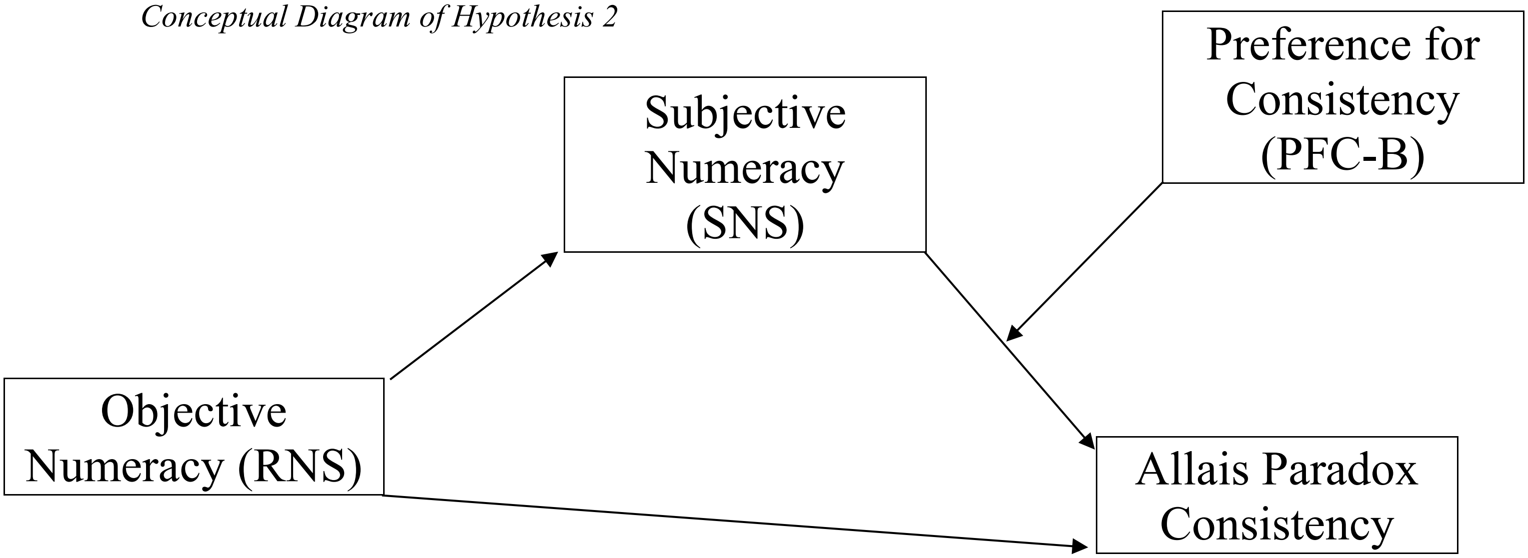
Conceptual Diagram of Hypothesis 1



Note. The mediated model previously found in Peters & Bjalkebring (2015). Specifically, the predicted model shows that the subjective numeracy will mediate the relation between objective numeracy and a mathematical task, in this case the Allais paradox.

Figure 2

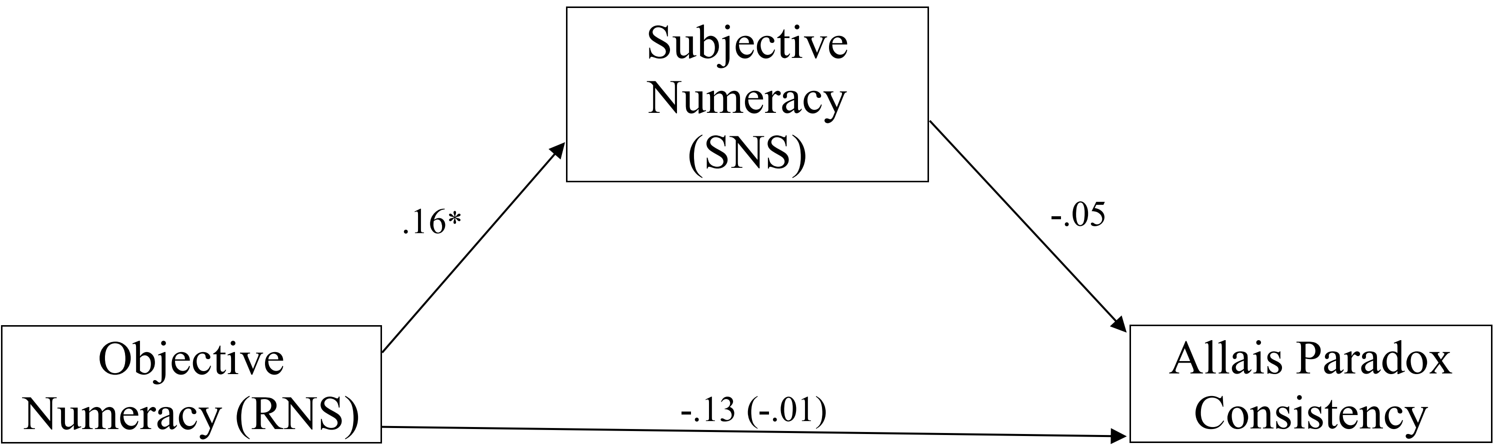
Conceptual Diagram of Hypothesis 2



Note. A moderated mediation model predicting the number of consistent choices an individual will make on the Allais paradox given their preference for consistency, subjective numeracy, and objective numeracy. Specifically, preference for consistency moderates the relations between objective numeracy and Allais paradox consistency which is itself mediated by subjective numeracy.

Figure 3

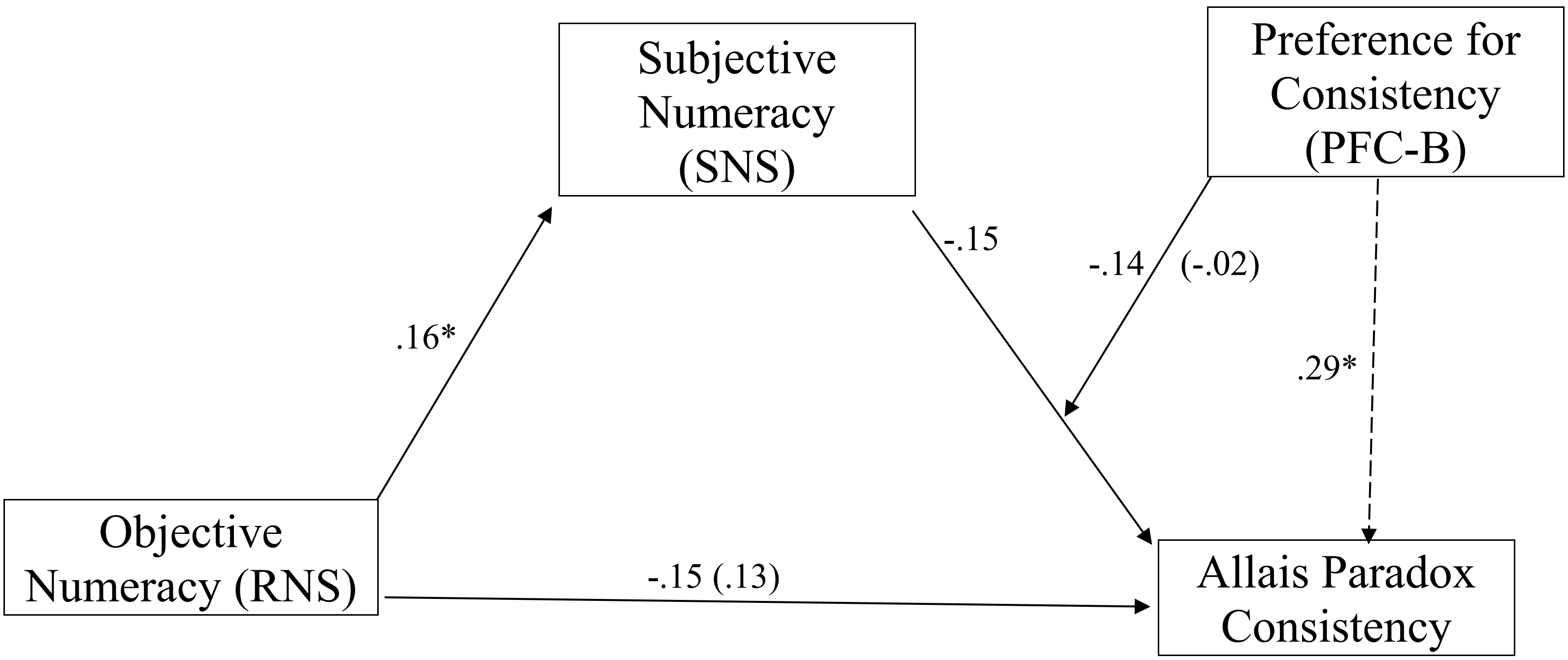
Model Diagram Testing Hypothesis 1



Note. The mediated model previously found in Peters & Bjalkebring (2015). Specifically, the predicted model shows that the subjective numeracy will mediate the relation between objective numeracy and a mathematical task, in this case the Allais paradox.
*$p < .05$

Figure 4

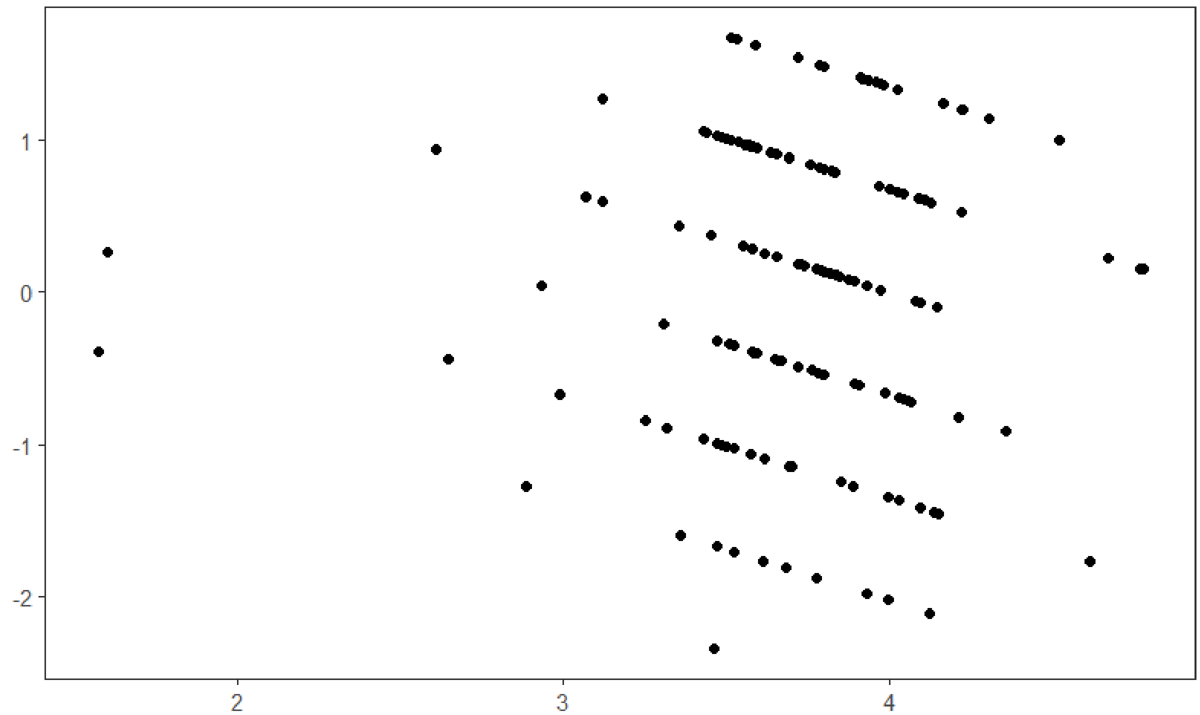
Model Diagram Testing Hypothesis 2



Note. A moderated mediation model predicting the number of consistent choices an individual would make on the Allais paradox given their preference for consistency, subjective numeracy, and objective numeracy. Specifically, whether preference for consistency moderated the relations between objective numeracy and Allais paradox consistency which was thought to have been mediated by subjective numeracy.
*$p < .05$

Figure 5

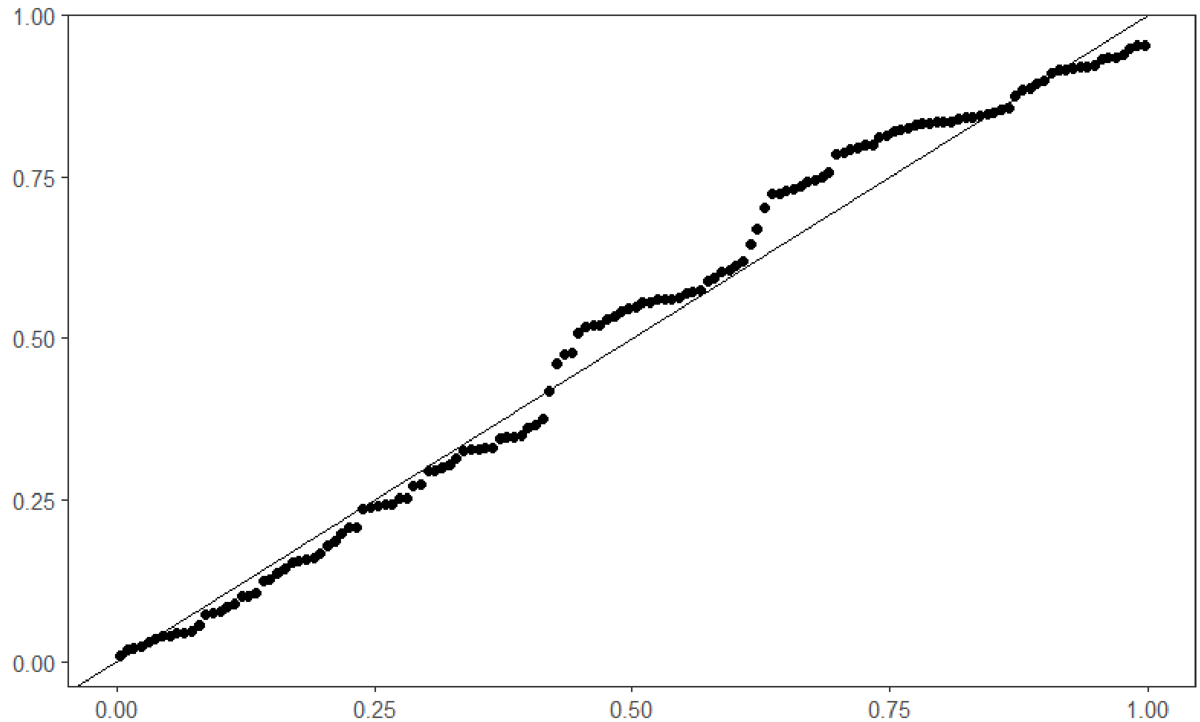
Scatterplot of Unstandardized Predicted Values and Standardized Residuals from Multiple Regression of Allais Score Predicted By RNS, SNS and PFC-B with No Interactions.



Note. Used as an indicator of linearity and homoskedasticity.

Figure 6

PP Plot of Standardized Residuals from Multiple Regression of Allais Score Predicted By RNS, SNS and PFC-B with No Interactions.



Note. Used as an indicator of the assumption of normality.

APPENDIX A: RASCH-BASED NUMERACY SCALE (RNS)

Q12. Suppose you have a close friend who has a lump in her breast and must have a mammography . . . The table below summarizes all of this information. Imagine that your friend tests positive (as if she had a tumor), what is the likelihood that she actually has a tumor?

	tested positive	tested negative	totals
actually has tumor	9	1	10
does not have tumor	9	81	90
total	18	82	100

Q15 (CRT). A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

Q17 (CRT). In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Q3. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?

Q2. In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket from BIG BUCKS?

Q1. Imagine that we roll a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up as an even number?

Q9. If the chance of getting a disease is 20 out of 100, this would be the same as having a _____% chance of getting the disease.

Q8b. If the chance of getting a disease is 10%, how many people would be expected to get the disease? Out of 1000?

APPENDIX B: SUBJECTIVE NUMERACY SCALE (SNS)

Cognitive abilities (1=not at all good, 6=extremely good)

- How good are you at working with fractions?
- How good are you at working with percentages?
- How good are you at calculating a 15% tip?
- How good are you at figuring out how much a shirt will cost if it is 25% off?

Preference for display of numeric information

- When reading the newspaper, how helpful do you find tables and graphs that are parts of a story? (1=not at all, 6=extremely)
- When people tell you the chance of something happening, do you prefer that they use words (‘‘it rarely happens’’) or numbers (‘‘there's a 1% chance’’)? (1=always prefer words, 6=always prefer numbers)
- When you hear a weather forecast, do you prefer predictions using percentages (e.g., ‘‘there will be a 20% chance of rain today’’) or predictions using only words (e.g., ‘‘there is a small chance of rain today’’)? (1=always prefer percentages, 6=always prefer words; reverse coded)
- How often do you find numerical information to be useful? (1=never, 6=very often)

APPENDIX C: PREFERENCE FOR CONSISTENCY ABBREVIATED (PFC-B)

1. It is important to me that those who know me can predict what I will do.
2. I want to be described by others as a stable, predictable person.
3. The appearance of consistency is an important part of the image I present to the world.
4. An important requirement for any friend of mine is personal consistency.
5. I typically prefer to do things the same way.
6. I want my close friends to be predictable.
7. It is important to me that others view me as a stable person.
8. I make an effort to appear consistent to others.
9. It doesn't bother me much if my actions are inconsistent. *b*

Note. Items were scored on a scale with the category designations: Strongly Disagree (1), Disagree (2), Somewhat Disagree (3), Slightly Disagree (4), Neither Agree nor Disagree (5), Slightly Agree (6), Somewhat Agree (7), Agree (8), and Strongly Agree (9).

• *b* Reverse scored.

APPENDIX D: ALLAIS PARADOX

All questions have a set pair which is presented in the same table.

Question 1	
Choose Between	
A	B
$2500 with probability 33% $2400 with probability 66% $0 with probability 1%	$2400 with certainty
Question 2	
Choose Between	
C	D
$2500 with probability 33% $0 with probability 67%	$2400 with probability 34% $0 with probability 66%

Question 3	
Choose between	
A	B
$4000 with probability 80%	$3000 with certainty
Question 5	
Choose between	
C	D
$4000 with probability 20%	$3000 with probability 25%

Question 4	
Choose between	
A	B
A loss of $4000 with probability 80%	A loss of $3000 with certainty
Question 6	
Choose between	
C	D
A loss of $4000 with probability 20%	A loss of $3000 with probability 25%

Question 7	
Choose between	
A	B
A 3-week tour of England, France and Italy with probability 50%	A 1-week tour of England with certainty
Question 8	
Choose between	
C	D
A 3-week tour of England, France and Italy with probability 5%	A 1-week tour of England with probability 10%

Question 9	
Choose between	
A	B
$6000 with probability 45%	$3000 with probability 90%
Question 10	
Choose between	
C	D
$6000 with probability 0.1%	$3000 with probability 0.2%

Question 11	
Choose between	
A	B
A loss of $6000 with probability 45%	A loss of $3000 with probability 90%
Question 12	
Choose between	
C	D
A loss of $6000 with probability 0.1%	A loss of $3000 with probability 0.2%