DRAFT – PLEASE DO NOT QUOTE OR CITE WITHOUT PERMISSION FROM THE AUTHORS.

Predicting Outcomes in a Sequence of Binary Events:
Belief Updating and Gambler's Fallacy Reasoning

Kariyushi Rao and Reid Hastie
Department of Behavioral Science, The University of Chicago Booth School of Business

April 20, 2020

Kariyushi Rao https://orcid.org/0000-0002-5027-8233

Correspondence should be addressed to Kariyushi Rao, Department of Behavioral Science, Chicago Booth School of Business, 5807 S Woodlawn Avenue, Chicago IL 60637. Email: kariyushi.rao@chicagobooth.edu.

Acknowledgements: This research was supported by funds from the Chicago Booth Graduate School of Business. The authors would like to thank the Chicago Booth community of scholars and many colleagues, most notably Maya Bar-Hillel, Diag Davenport, Samuel Hirshman, Emir Kamenica, Joshua Klayman, Andrew Meyer, Joshua B. Miller, Richard Thaler, Oleg Urminsky, Chen Xia, Kazuo Yamaguchi, and Nick Epley. We also wish to recognize our research assistants for their help categorizing our qualitative data: Helena Karas, Nicholas O'Donnell, and Leah Umanskiy. The first author is greatly indebted to Chris Wickens and Philipp Chapkovski for their timely and thoughtful guidance on programming experiments in oTree.

Individual Differences

After completing the experimental procedure, we collected additional information from each participant in the present studies, including demographics, responses to general knowledge questions, and participants' qualitative descriptions of their own prediction strategies. In Section 1, we present an analysis of the relationship between individual differences and participants' prediction strategies in each of the present studies. In Section 2, we present a summary of participants' qualitative descriptions of their prediction strategies.

Section 1: Individual Difference Measures and Participant Strategies

In this section, we consider the relationship between participants' prediction strategies and individual difference measures. We first describe the individual difference measures collected, as well as statistics that capture participants' prediction strategies. We then present descriptive statistics and pairwise correlations for each study separately. After completing the experimental procedure, participants were asked to respond to a brief demographics questionnaire, which asked for participants' age, gender, and highest degree obtained. Three additional questions tested participants familiarity with probability theory (Probability Questions), and two questions tested participants' financial literacy (Financial Literacy Questions). The full text of the Probability and Financial Literacy follows below.

Probability Questions

1. Suppose you have a fair coin. Each time you flip the coin, there is a 50% chance the coin lands son heads, and a 50% chance the coin lands on tails. What is the probability that the coin lands on heads three times in a row? (Question designed by authors of present article.)

2. The chance of getting a viral infection is 0.0005. Out of 10,000 people, about how many of them are expected to get a viral infection? (Adapted from Peters and colleagues, 2006. 56% of participants were able to answer this question correctly in Peters and colleagues' study.)

3. In the Acme Publishing Sweepstakes, the chance of winning a car is 1 in 1,000. What percent of tickets in the Acme Publishing Sweepstakes win a car? 1%, 10%, or 0.1%? (Adapted from Peters and colleagues, 2006. 46% of participants were able to answer this question correctly in Peters and colleagues' study.)

Financial Literacy Questions

1. Suppose you owe $1,000 on a loan and the interest rate you are charged is 20% per year compounded annually. If you didn't pay anything off, at this interest rate, how many years would it take for the amount you owe to double?
   - o Less than 2 years
   - o At least 2 years but less than 5 years
   - o At least 5 years but less than 10 years
   - o At least 10 years

   (Adapted from FINRA 2015 survey. 75% of participants answered this question correctly on the FINRA survey.)

2. Fill in the blank. Buying a single company's stock is usually _____ buying a stock mutual fund.
   - o more safe than
   - o the same as
   - o less safe than

   (Adapted from FINRA 2015 survey. 46% of participants answered this question correctly on the FINRA survey.)

Participants were also asked for a subjective rating of their own knowledge of the stock market, and of the frequency with which they gamble.

1. How well do you feel you understand the stock market, compared to the average person?
   - o Better than average (coded as 1)
   - o About average (coded as 0)
   - o Worse than average (coded as –1)

2. How often do you feel you gamble, compared to the average person?
   - o More than average (coded as 1)
   - o About average (coded as 0)
   - o Less than average (coded as –1)

In addition to the individual difference measures collected at the end of the experimental procedure, we also calculated statistics that capture individual participants' prediction strategies. For participants in the Continuous Versions of Studies 1–3 (1A, 2A, 3A), we calculated the

average probability each participant assigned to repetition of the terminal streak at the end of each Target sequence (Average "Repeat"). We also performed individual participant-level linear regressions to estimate the coefficient (Slope) of Streak Length – the unit increase in probability of repetition for each unit increase in Streak Length.

For participants in the Binary Versions of Studies 1–3 (1B, 2B, 3B), we summed each participant's "Repeat" predictions across the Target sequences (Number "Repeat"). We also performed individual participant-level binary logistic regressions to estimate the coefficient (Slope) of Streak Length - the percent increase or decrease in the odds a participant predicts a streak will repeat for each unit increase in Streak Length. For example, a participant following a "hot hand" prediction strategy is more likely to predict repetition as Streak Length increases. The log odds (β) coefficient of Streak Length obtained from a binary logistic regression over such a participant's predictions might be something like 1.07. In this case, the odds (e^{β}) that this participant predicts a streak of length 3 will repeat would be 2.93 ($e^{1.07}$) times greater than the odds that this participant predicts a streak of length 2 will repeat. The odds this participant predicts a streak of length x will repeat are thus 193% [100 $\times$ (2.93–1)] greater than the odds this participant predicts a streak of length x–1 will repeat. (A detailed explanation of the interpretation of these coefficients is provided in Section 1.7 below.) A brief summary of the individual difference measures, as well as the statistics we used to capture participants' prediction strategies, can be found in Table 1.

Table OS1

Description of Individual Difference Measures and Statistics Summarizing Participants' Prediction Strategies

AG: Age	Participant's age, in years
FE: Female	Participant's gender identity (1 = Female; 0 = Male/Other)
HD: Highest Degree	1 = No Degree, 2 = High School Diploma, 3 = 2-Year College Degree or Skilled Trade Program, 4 = 4-year College Degree or higher
NP: # Accurate Probability	Number of probability questions answered correctly (out of 3)
NF: # Accurate Financial	Number of financial literacy questions answered correctly (out of 2)
KS: Knows Stocks	"Better than average" = 1; "About average" = 0; "Worse than average" = –1
GO: Gambles Often	"More than average" = 1; "About average" = 0; "Less than average" = –1
AR: Average "Repeat"	**Studies 1A, 2A, 3A:** Mean probability assigned to repetition of terminal streak across Target sequences.
NR: Number "Repeat"	**Studies 1B, 2B, 3B:** Sum of "Repeat" (= 1) predictions across Target sequences.
SL: Slope	**Studies 1A, 2A, 3A:** Slope of individual participants' linear regression lines, including predictions over streaks of length 1 (fillers), and of lengths 2, 3, 4, 5, 6, and 7 (targets).
SL: Slope	**Studies 1B, 2B, 3B:** The percent increase or decrease in the odds a participant predicts a streak will for each unit increase in Streak Length.

Section 1.1: Study 1A Results

In Study 1A, none of the individual difference measures were significantly correlated with either of the statistics capturing participants' prediction strategies (AR: Average Repeat, SL: Slope). Table 2 presents the summary statistics for individual difference measures and participants' prediction strategies. Table 3 presents the Pearson Product Moment Correlation Matrix for these variables.

Table OS2

Study 1A: Summary Statistics for Individual Difference Measures and Prediction Strategies

	AnalystUnknown	StockUnknown	BingoUnknown
Mean Age	34.74 (11.06)	37.43 (13.15)	36.06 (9.14)
% Female	0.42	0.36	0.42
Mean Highest Degree	3.36 (0.84)	3.07 (0.99)	2.98 (0.91)
Mean # Accurate Probability	1.68 (0.95)	1.66 (1.07)	1.68 (1.01)
Mean # Accurate Financial	1.42 (0.64)	1.46 (0.62)	1.54 (0.64)
Mean "Know Stocks"	–0.02 (0.69)	–0.09 (0.63)	–0.26 (0.66)
Mean "Gambles Often"	–0.46 (0.67)	–0.61 (0.53)	–0.50 (0.54)
Mean Average Repeat	66% (16%)	70% (14%)	60% (19%)
Mean Slope	6.20 (4.45)	6.97 (5.22)	5.92 (5.70)
N	50	44	50

Note: Numbers in parentheses are standard deviations

Table OS3

Study 1A: Pearson Product Moment Correlation Matrix for Variables of Interest

	AG	FE	HD	NP	NF	KS	GO	AR	SL
AG: Age		0.23	0.03	0.02	0.08	–0.00	–0.25	–0.01	0.09
FE: Female			–0.03	**–0.30****	–0.12	**–0.31****	0.05	–0.14	–0.00
HD: Highest Degree				0.18	0.10	0.11	0.11	–0.02	0.04
NP: # Acc. Probability					**0.30****	0.07	–0.19	0.14	0.12
NF: # Acc. Financial						–0.07	–0.12	0.04	0.11
KS: Know Stocks							0.24	0.04	–0.23
GO: Gambles Often								–0.16	–0.10
AR: Average Repeat									**0.31****
SL: Slope									

Note: The Pearson product moment correlation evaluates the linear relationship between two continuous variables. It assumes two variables change together at a constant rate. This measure is not appropriate for comparing a continuous measure (Slope) with non-continuous measures (e.g. Highest Degree). The alternative Spearman rank-order correlation is also inappropriate, because we are comparing a continuous variable (Slope) with ordinal and categorical variables

(e.g. Female). The above table is only provided to give the reader a sense of the global patterns we observed in the data.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, with Bonferroni correction for multiple comparisons.

Section 1.2: Study 1B Results

In Study 1B, there was a small positive correlation between subjective ratings of stock market knowledge and the number of times participants predicted streaks would repeat. None of the other individual difference measures were significantly correlated with either of the statistics capturing participants' prediction strategies (NR: Number Repeat, SL: Slope).

Table OS4

Study 1B: Summary Statistics for Individual Difference Measures and Prediction Strategies

	AnalystUnknown	StockUnknown	BingoUnknown
Mean Age	34.05 (10.60)	36.37 (13.93)	35.80 (10.65)
% Female	0.42	0.54	0.47
Mean Highest Degree	3.43 (0.81)	3.23 (0.95)	3.38 (0.89)
Mean # Accurate Probability	1.92 (0.95)	2.08 (0.92)	1.97 (1.21)
Mean # Accurate Financial	1.46 (0.67)	1.41 (0.67)	1.43 (0.64)
Mean "Know Stocks"	–0.02 (0.58)	–0.08 (0.62)	–0.15 (0.68)
Mean "Gambles Often"	–0.45 (0.66)	–0.46 (0.68)	–0.50 (0.70)
Mean Number Repeat	3.72 (1.84)	4.00 (1.68)	3.15 (2.02)
Mean Slope (% Δ Odds)	68% (76%)	84% (81%)	51% (79%)
N	95	97	108

Note: Numbers in parentheses are standard deviations. Slope: $[100 \times (e^{\beta}-1)]$ The percent change in the odds that a participant predicts a streak of length x will repeat versus a streak of length x–1. Separation was observed while running participant-level logistic regressions on 127/300 participants' predictions over target sequences. Firth's procedure was applied to all of the participant-level logistic regressions to resolve the separation issue, producing less biased coefficients (see Heinze & Schemper, 2002).

Table OS5

Study 1B: Pearson Product Moment Correlation Matrix for Variables of Interest

	AG	FE	HD	NP	NF	KS	GO	NR	SL
AG: Age		**0.23****	0.11	–0.10	0.09	0.01	–0.18	–0.04	0.01
FE: Female			0.12	–0.15	–0.04	**–0.34*****	**–0.28*****	–0.13	0.02
HD: Highest Degree				0.09	0.07	0.14	–0.05	0.10	0.07
NP: # Acc. Probability					**0.33*****	0.17	0.01	0.14	0.03
NF: # Acc. Financial						0.14	–0.12	0.14	0.12
KS: Know Stocks							**0.22****	**0.21****	0.00
GO: Gambles Often								0.02	–0.11
NR: Number Repeat									0.04
SL: Slope									

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, with Bonferroni correction for multiple comparisons.

Section 1.3: Study 2A Results

In Study 2A, none of the individual difference measures were significantly correlated with either of the statistics capturing participants' prediction strategies.

Table OS6

Study 2A: Summary Statistics for Individual Difference Measures and Prediction Strategies

	Analyst50	Stock50	Bingo50
Mean Age	34.06 (9.29)	36.52 (9.46)	36.16 (11.69)
% Female	0.44	0.46	0.43
Mean Highest Degree	3.04 (0.84)	3.19 (0.94)	3.23 (0.85)
Mean # Accurate Probability	1.98 (0.96)	2.02 (0.98)	1.82 (1.16)
Mean # Accurate Financial	1.42 (0.61)	1.56 (0.62)	1.30 (0.74)
Mean "Know Stocks"	-0.19 (0.60)	-0.13 (0.57)	0.09 (0.69)
Mean "Gambles Often"	-0.52 (0.73)	-0.54 (0.68)	-0.41 (0.68)
Mean Average Repeat	55% (20%)	58% (20%)	44% (18%)
Mean Slope	2.90 (5.65)	4.54 (6.52)	0.24 (7.20)
N	52	48	56

Note: Numbers in parentheses are standard deviations

Table OS7

Study 2A: Pearson Product Moment Correlation Matrix for Variables of Interest

	AG	FE	HD	NP	NF	KS	GO	AR	SL
AG: Age		0.17	0.02	–0.09	0.07	–0.06	–0.17	–0.03	–0.02
FE: Female			0.02	**–0.34*****	**–0.30****	–0.21	–0.04	–0.03	0.00
HD: Highest Degree				0.11	–0.07	0.19	0.02	–0.03	–0.04
NP: # Acc. Probability					**0.49*****	0.17	0.01	–0.04	–0.09
NF: # Acc. Financial						0.09	–0.11	–0.12	–0.12
KS: Know Stocks							**0.43*****	0.00	–0.10
GO: Gambles Often								0.04	0.07
AR: Average Repeat									**0.58*****
SL: Slope									

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, with Bonferroni correction for multiple comparisons.

Section 1.4: Study 2B Results

In Study 2B, none of the individual difference measures were significantly correlated with either of the statistics capturing participants' prediction strategies.

Table OS8

Study 2B: Summary Statistics for Individual Difference Measures and Prediction Strategies

	Analyst50	Stock50	Bingo50
Mean Age	37.93 (13.33)	33.77 (11.83)	35.17 (12.14)
% Female	0.51	0.44	0.47
Mean Highest Degree	3.45 (0.83)	3.27 (0.95)	3.18 (0.97)
Mean # Accurate Probability	2.07 (0.92)	1.85 (0.83)	1.94 (0.91)
Mean # Accurate Financial	1.54 (0.61)	1.47 (0.62)	1.32 (0.71)
Mean "Know Stocks"	–0.15 (0.58)	–0.23 (0.60)	–0.17 (0.63)
Mean "Gambles Often"	–0.60 (0.62)	–0.44 (0.71)	–0.55 (0.66)
Mean Number Repeat	2.35 (1.93)	2.48 (2.10)	1.54 (1.67)
Mean Slope (% Δ Odds)	53% (84%)	47% (75%)	16% (61%)
N	97	103	101

Note: Numbers in parentheses are standard deviations

Table OS9

Study 2B: Pearson Product Moment Correlation Matrix for Variables of Interest

	AG	FE	HD	NP	NF	KS	GO	NR	SL
AG: Age		0.08	0.16	0.02	0.03	–0.02	–0.12	–0.10	0.03
FE: Female			–0.02	**–0.24*** **	–0.15	–0.30	–0.16	0.04	0.04
HD: Highest Degree				0.06	0.09	**0.23** **	0.04	0.06	–0.03
NP: # Acc. Probability					**0.40*** **	0.14	–0.03	–0.09	0.03
NF: # Acc. Financial						0.11	–0.12	–0.03	0.03
KS: Know Stocks							**0.19***	0.04	0.06
GO: Gambles Often								–0.01	0.02
NR: Number Repeat									**0.32*** **
SL: Slope									

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, with Bonferroni correction for multiple comparisons.

Section 1.5: Study 3A Results

In Study 3A, none of the individual difference measures were significantly correlated with either of the statistics capturing participants' prediction strategies.

Table OS10

Study 3A: Summary Statistics for Individual Difference Measures and Prediction Strategies

	Analyst25-50-75	Stock25-50-75	Bingo25-50-75
Mean Age	31.90 (7.68)	35.94 (12.35)	34.42 (9.37)
% Female	0.50	0.48	0.50
Mean Highest Degree	3.30 (0.84)	3.26 (0.90)	3.32 (0.82)
Mean # Accurate Probability	1.98 (1.04)	1.86 (0.88)	1.90 (0.86)
Mean # Accurate Financial	1.48 (0.65)	1.52 (0.58)	1.38 (0.67)
Mean "Know Stocks"	–0.12 (0.69)	–0.04 (0.60)	–0.24 (0.59)
Mean "Gambles Often"	–0.54 (0.68)	–0.46 (0.61)	–0.38 (0.75)
Mean Average Repeat	72% (12%)	65% (19%)	65% (15%)
Mean Slope	6.99 (3.94)	5.58 (5.12)	6.13 (5.11)
N	50	50	50

Note: Numbers in parentheses are standard deviations.

Table OS11

Study 3A: Pearson Product Moment Correlation Matrix for Variables of Interest

	AG	FE	HD	NP	NF	KS	GO	AR	SL
AG: Age		0.03	–0.04	0.01	0.18	0.12	–0.05	–0.06	–0.02
FE: Female			0.18	–0.10	–0.19	–0.19	–0.16	–0.02	–0.05
HD: Highest Degree				0.04	–0.02	0.22	0.03	–0.07	0.03
NP: # Acc. Probability					**0.32****	0.18	–0.05	0.05	–0.11
NF: # Acc. Financial						0.10	–0.14	0.12	–0.08
KS: Know Stocks							**0.32****	–0.01	–0.04
GO: Gambles Often								–0.13	–0.11
AR: Average Repeat									0.24
SL: Slope									

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, with Bonferroni correction for multiple comparisons.

Section 1.6: Study 3B Results

In Study 3B, none of the individual difference measures were significantly correlated with either of the statistics capturing participants' prediction strategies.

Table OS12

Study 3B: Summary Statistics for Individual Difference Measures and Prediction Strategies

	Analyst50	Stock50	Bingo50
Mean Age	36.09 (12.11)	36.85 (11.83)	37.46 (12.53)
% Female	0.44	0.54	0.60
Mean Highest Degree	3.19 (0.92)	3.23 (0.88)	3.23 (0.96)
Mean # Accurate Probability	1.90 (0.94)	2.02 (0.93)	1.94 (1.00)
Mean # Accurate Financial	1.44 (0.66)	1.41 (0.65)	1.36 (0.62)
Mean "Know Stocks"	–0.17 (0.66)	–0.15 (0.64)	–0.23 (0.73)
Mean "Gambles Often"	–0.58 (0.61)	–0.73 (0.51)	–0.68 (0.58)
Mean Number Repeat	2.86 (2.00)	3.18 (2.00)	2.97 (2.12)
Mean Slope (% Δ Odds)	46% (73%)	47% (74%)	33% (74%)
N	98	93	109

Note: Numbers in parentheses are standard deviations.

Table OS13

Study 3B: Pearson Product Moment Correlation Matrix for Variables of Interest

	AG	FE	HD	NP	NF	KS	GO	NR	SL
AG: Age		0.14	0.08	–0.16	–0.06	0.04	–0.05	–0.07	0.05
FE: Female			–0.02	**–0.24****	–0.15	**–0.37*****	–0.20*	–0.10	0.03
HD: Highest Degree				0.14	0.05	0.16	0.00	–0.11	–0.02
NP: # Acc. Probability					**0.32*****	**0.28*****	0.09	0.05	0.03
NF: # Acc. Financial						0.18	0.04	0.05	0.04
KS: Know Stocks							**0.20***	0.04	–0.03
GO: Gambles Often								0.06	–0.07
NR: Number Repeat									**0.25*****
SL: Slope									

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, with Bonferroni correction for multiple comparisons.

Section 1.7: Interpretation of Binary Logistic Regression Coefficients

The independent variable coefficients obtained from a logistic regression can be interpreted as the "change in the log odds that ($Y = 1$) for each unit increase in [independent_variable]," where Y is the response variable, which, in the current application, takes the value 1 when the participant predicts a streak will **repeat**. Exponentiating the log odds coefficients produces the odds ratios, which can be interpreted as the "change in the odds that ($Y = 1$) for each unit increase in [independent_variable]." The odds ratios are difficult to interpret when there is a negative relationship between the odds ratio and a unit increase in the independent variable. Transforming the odds ratios into the percent-change in the odds makes it easier to interpret both positive and negative relationships between the odds ratio and the independent variable. To obtain the percent change in odds for a unit increase in the independent variable, we first subtract 1 from the odds ratio, and then we multiply the result by 100.

Example: For Participant A, the log odds that this participant predicts "repeat" are –0.78. The odds that this participant predicts a streak will repeat are exp(–0.78) = 0.46 times higher for each unit increase in Streak Length. This means that the odds this particular participant predicts "repeat" *decrease* as Streak Length increases. The odds this participant predicts "repeat" *drop* by [100 × (0.46–1)] = –54% for each unit increase in Streak Length.

The following presents a toy example for 2 contrived cases. Participant 1 is *more* likely to predict "repeat" as Streak Length increases. Participant 1 has a "growing" logistic function that *increases* from 0 to 1.

Participant ID	Streak Length	Prediction (0 = Reverse; 1 = Repeat)
1	2	0
1	3	1
1	4	1
1	5	1
1	6	1
1	7	1

When we regress Prediction on Streak Length for Participant 1, we obtain a log odds coefficient of 0.78. Exponentiating this coefficient, we obtain an odds ratio of 2.17. After subtracting 1 from the odds ratio, and then multiplying the result by 100, we obtain a percent-change of 117% in the odds for a 1 unit increase in Streak Length. The substantive interpretation of the percent change in odds for Participant 1 is as follows:

> "The odds that this participant predicts a streak will repeat increase by 117% for each unit increase in Streak length." – or – "The odds that this participant predicts a streak of length *x* will repeat are about 117% greater than the odds this participant predicts a streak of length *x*–1 will repeat."

Participant 2 is *less* likely to predict "repeat" as Streak Length increases. Participant 2 has a "decaying" logistic function that *decreases* from 1 to 0.

Participant ID	Streak Length	Prediction (0 = Reverse; 1 = Repeat)
2	2	1
2	3	1
2	4	1
2	5	0
2	6	0
2	7	0

When we regress Prediction on Streak Length for Participant 2, we obtain a log odds coefficient of –1.13. Exponentiating this coefficient, we obtain an odds ratio of 0.32. After subtracting 1 from the odds ratio, and then multiplying the result by 100, we obtain a percent-change of –68% in the odds for a 1 unit increase in Streak Length. The substantive interpretation of the percent change in odds for Participant 2 is as follows:

> "The odds that this participant predicts a streak will repeat decreases by –68% for each unit increase in Streak Length." - or - "The odds that this participant predicts a streak of length *x* will repeat are about –68% less than the odds this participant predicts a streak of length (*x*–1) will repeat."

Filler Sequences

Participants in the present studies each judged 12 filler sequences, in addition to the 6 target experimental sequences. These filler sequences were randomly selected from a pool of 24 filler stimuli that all ended in a reversal (e.g. Blue-Red, Down-Up). 22 of the filler sequences contained a streak of at least 2 identical signals preceding the reversal at the end of the sequence.[1] In the summary that follows, we present participants' predictions that the next (9th) signal will *repeat* the longest streak of identical signals appearing in each of these 22 filler sequences.[2]

In Studies 1A and 1B, participants' expectations that the longest streak would repeat increased with Streak Length (Figures OS1 and OS2). The patterns of responses were similar across Conditions (Tables OS14 and OS15). In Studies 2A and 2B, participants in the Analyst50 and Stock50 Conditions slightly increased their expectations of repetition as Streak Length increased (Figures OS3 and OS4). Participants in the Bingo50 Condition of Study 2A consistently assigned 50% probability to repetition of the longest streak, and fewer than 50% of participants in Study 2B predicted repetition of the longest streak, across all Streak Lengths (Tables OS16 and OS17). In Studies 3A and 3B, participants' posterior beliefs are consistent with the Bayesian posteriors for each Filler sequence (Figures OS5 and OS6). Participants in Study 3B were somewhat more likely to predict repetition of the longest streak in the Filler sequences than they were to predict repetition of the terminal streak in the Target sequences (Compare Table OS19 to Table 6 in the main text).

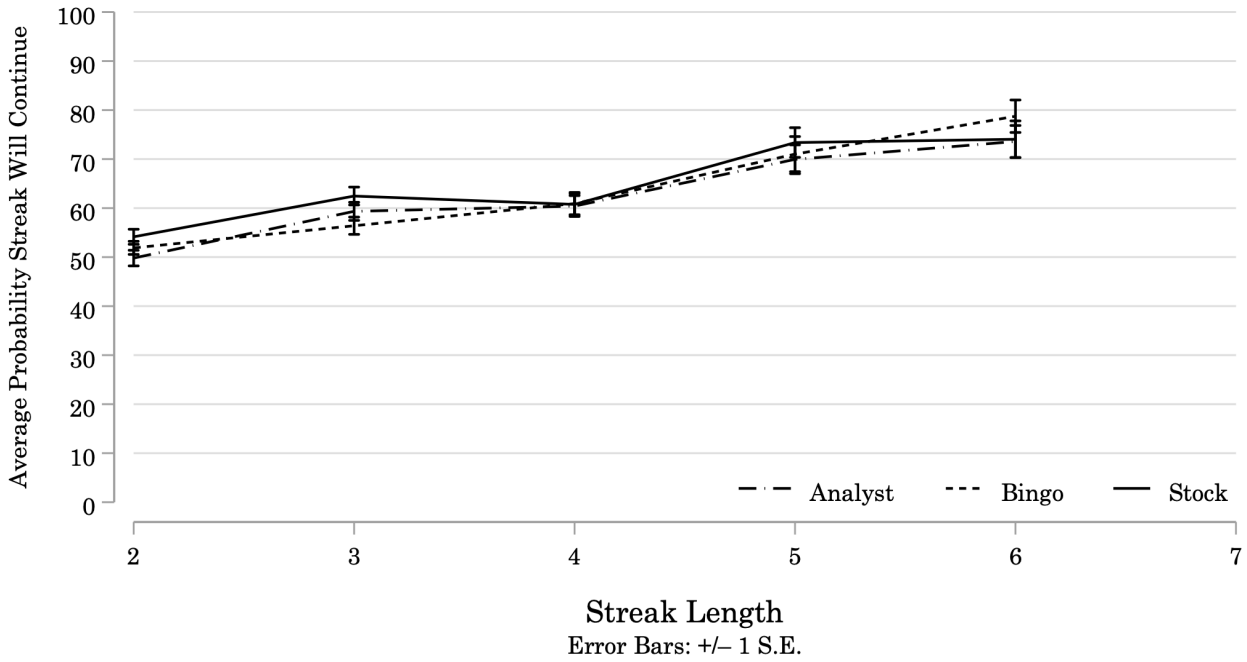
[1] Two filler sequences had an alternation rate of 1.00, and contained no streaks. These two filler sequences were excluded from this analysis.

[2] Several of the filler sequences with a maximum streak length of 2 contained more than one streak of length 2. In these cases, we considered the streak occurring closest to the end of the sequence as the focal streak in the analysis. Focal streaks are highlighted in bold type in the first column of each summary table.

The only differences we might note between predictions for Filler sequences, versus Target sequences, is that participants seem to update somewhat less in response to a streak earlier in the sequence (Fillers) than to streaks at the end of the sequence (Targets).

Figure OS1

Study 1A: Average Probability Participants Assigned to Repetition of Longest Streak in Each Filler Sequence, by Streak Length and Condition



Note: The average probability participants assigned to repetition of the longest streak in each Filler sequence (Error Bars: +/– 1 S.E.). Streak Length is the length of the longest streak preceding the reversal at the end of each sequence. Participants increased the probability they assigned to repetition of the longest streak in the Filler sequences as the length of that streak increased. Filler sequences contained streaks that ranged from 2 to 6 identical signals. N = 144.

Table OS14

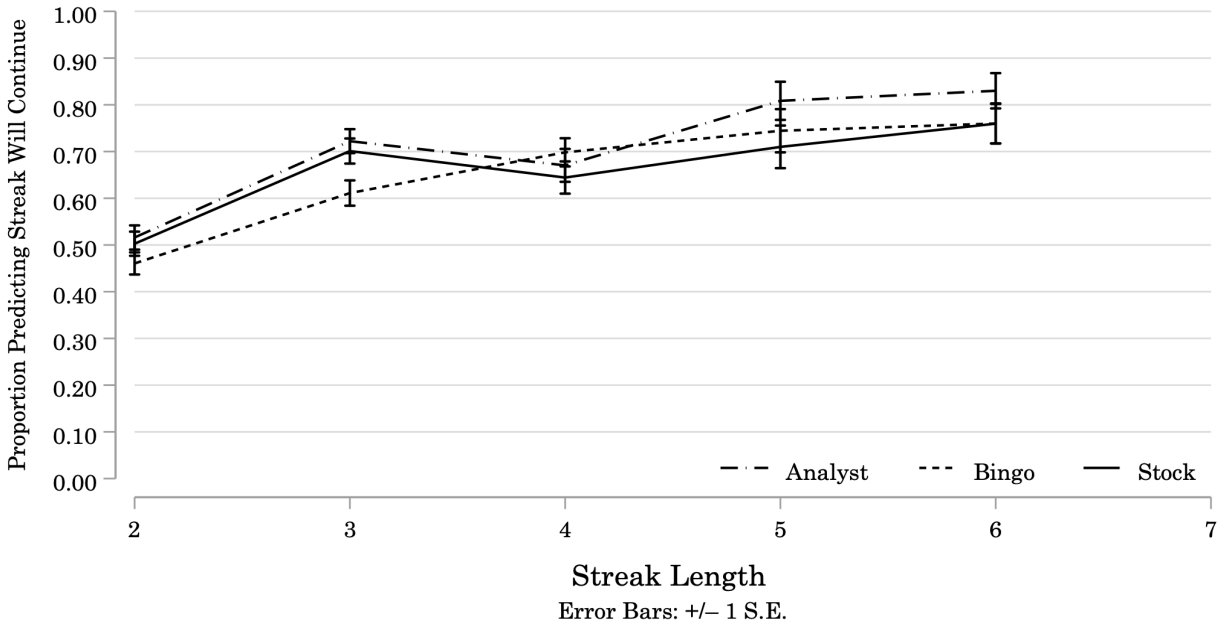
Study 1A: Average Probability Participants Assigned to Repetition of Longest Streak in Each Filler Sequence

	Streak Type	Analyst		Bingo		Stock	
		N	Mean Repeat	N	Mean Repeat	N	Mean Repeat
Streak Length = 2		**178**	**49.80**	**207**	**51.89**	**182**	**54.15**
00**11**0101	1	23	48.61	19	51.84	23	55.04
00110**11**0	1	20	53.20	27	49.59	19	61.42
0101**00**10	0	19	55.21	25	57.32	21	51.29
0**11**01010	1	16	60.88	21	68.48	28	59.54
10100**11**0	1	23	48.30	29	57.66	26	52.96
1100**11**01	1	26	57.73	31	46.13	21	66.67
1101**00**10	0	24	46.50	28	48.18	21	49.81
11011**00**1	0	27	34.52	27	40.56	23	37.17
Streak Length = 3		**158**	**59.34**	**138**	**56.41**	**132**	**62.47**
01**000**101	0	23	66.70	20	62.40	29	65.62
010**111**01	1	27	58.11	21	56.76	21	64.86
0**111**0010	1	23	56.74	21	56.95	22	48.77
1**000**1010	0	32	55.25	25	57.48	22	56.50
11**000**110	0	25	54.76	26	48.15	18	66.67
11101010	1	28	65.39	25	58.36	20	73.25
Streak Length = 4		**110**	**60.42**	**104**	**60.92**	**83**	**60.76**
00001101	0	25	58.72	27	58.67	21	53.33
0101**1110**	1	27	62.44	25	59.52	21	65.71
1**0000**101	0	28	64.64	27	66.19	20	66.70
11110010	1	30	56.07	25	59.08	21	57.57
Streak Length = 5		**46**	**69.96**	**52**	**71.02**	**42**	**73.38**
01**00000**1	0	25	67.08	27	73.85	19	81.79
11111010	1	21	73.38	25	67.96	23	66.43
Streak Length = 6		**54**	**73.59**	**47**	**78.74**	**42**	**74.05**
1**000000**1	0	28	69.29	22	73.55	17	59.00
11111101	1	26	78.23	25	83.32	25	84.28

Note: Several of the sequences with a maximum streak length of 2 contained more than one streak of length 2. In these cases, we considered the streak occurring closest to the end of the sequence as the focal streak in the analysis. Focal streaks are highlighted in bold type in the first column of the table. (Coding: 1 = UP/RED; 0 = DOWN/BLUE.)

Figure OS2

Study 1B: Proportion of Participants Predicting Repetition of Longest Streak in Each Filler Sequence, by Streak Length and Condition



Note: Proportion of participants predicting repetition of the longest streak in each filler sequence (Error Bars: +/– 1 S.E.). Streak Length is the length of the longest streak present in the Filler sequence, prior to the reversal at the end of that sequence. The proportion of participants predicting repetition increased as the length of that streak increased. Filler sequences contained streaks that ranged from 2 to 6 identical signals in length. N = 300.

Table OS15

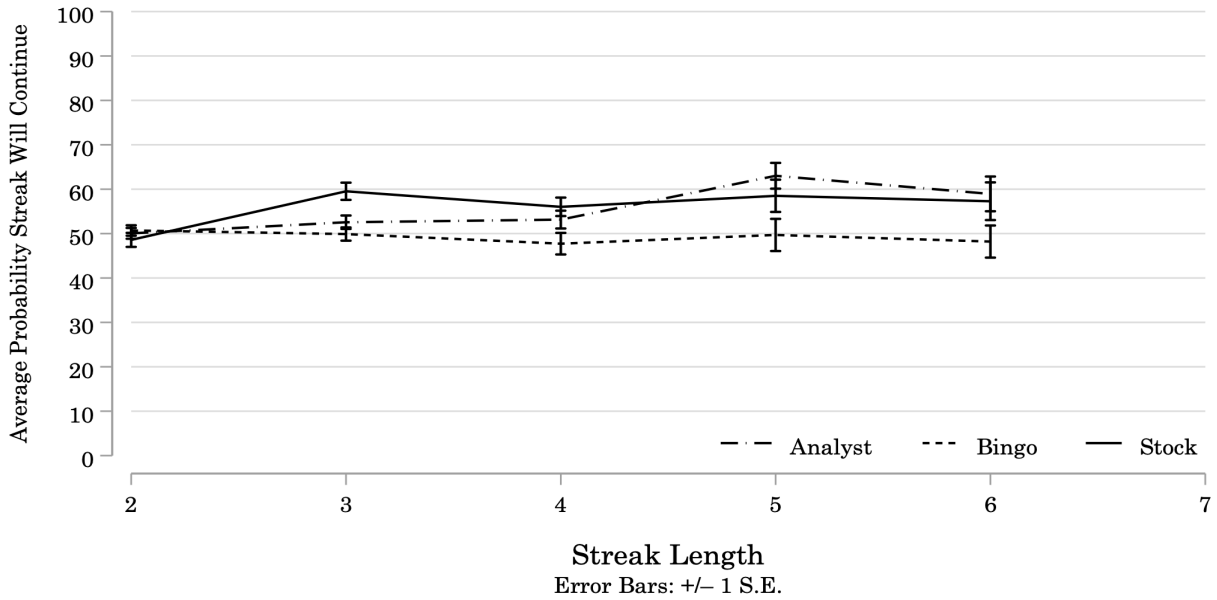
Study 1B: Proportion of Participants Predicting Repetition of Longest Streak in Each Filler Sequence, by Streak Length and Condition

	Streak Type	Analyst		Bingo		Stock	
		N	Prop Repeat	N	Prop Repeat	N	Prop Repeat
Streak Length = 2		**374**	**0.52**	**443**	**0.46**	**374**	**0.50**
00**11**0101	1	50	0.52	48	0.40	56	0.38
00110**11**0	1	50	0.66	57	0.58	39	0.54
0101**00**10	0	47	0.51	64	0.48	53	0.49
0**11**01010	1	56	0.82	57	0.68	44	0.82
10100**11**0	1	40	0.55	54	0.44	46	0.48
1100**11**01	1	38	0.47	54	0.50	46	0.65
1101**00**10	0	50	0.40	57	0.37	49	0.41
11011**00**1	0	43	0.09	52	0.19	41	0.29
Streak Length = 3		**306**	**0.72**	**324**	**0.61**	**291**	**0.70**
01**000**101	0	52	0.73	51	0.76	47	0.77
010**111**01	1	56	0.79	56	0.63	39	0.69
0**111**0010	1	43	0.53	51	0.53	45	0.51
1**000**1010	0	44	0.64	59	0.54	54	0.63
11**000**110	0	54	0.70	55	0.58	53	0.74
11101010	1	57	0.88	52	0.63	53	0.85
Streak Length = 4		**179**	**0.67**	**232**	**0.70**	**194**	**0.64**
00001101	0	49	0.49	53	0.68	56	0.50
010**1111**0	1	46	0.78	58	0.72	44	0.77
1**0000**101	0	39	0.79	65	0.77	46	0.74
11110010	1	45	0.64	56	0.61	48	0.60
Streak Length = 5		**94**	**0.81**	**90**	**0.74**	**100**	**0.71**
01**00000**1	0	44	0.75	41	0.76	53	0.60
11111010	1	50	0.86	49	0.73	47	0.83
Streak Length = 6		**100**	**0.83**	**100**	**0.76**	**104**	**0.76**
1**000000**1	0	52	0.73	50	0.74	54	0.61
11111101	1	48	0.94	50	0.78	50	0.92

Note: Focal streaks are highlighted in bold type in the first column of the table.

Figure OS3

Study 2A: Average Probability Participants Assigned to Repetition of Longest Streak in Each Filler Sequence, by Streak Length and Condition



Note: The average probability participants assigned to repetition of the longest streak in each Filler sequence (Error Bars: +/– 1 S.E.). Streak Length is the length of the longest streak preceding the reversal at the end of each sequence. Participants in the Analyst50 Condition gradually increased the probability they assigned to repetition of the longest streak in the Filler sequences as the length of that streak increased. Participants in the Stock50 Condition initially increase the probability they assign to repetition when Streak Length increases from 2 to 3, but have a flat slope across Streak Lengths 3 to 6. Participants in the Bingo50 Condition have a flat slope across Streak Lengths - they stick pretty close to 50% probability of repetition across Streak Lengths. N = 156.

Table OS16

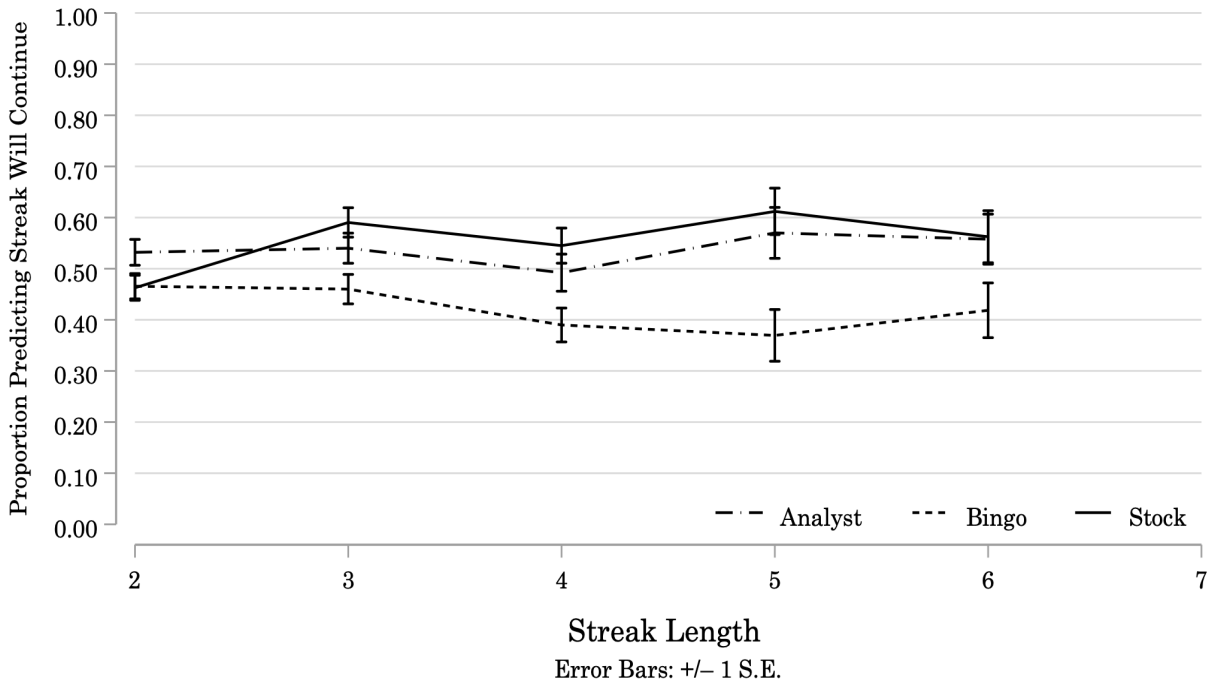
Study 2A: Average Probability Participants Assigned to Repetition of Longest Streak in Each Filler Sequence

	Streak Type	Analyst		Bingo		Stock	
		N	Mean Repeat	N	Mean Repeat	N	Mean Repeat
Streak Length = 2		**218**	**50.06**	**236**	**50.70**	**203**	**48.58**
00**11**0101	1	30	51.60	30	51.60	27	41.89
00110**11**0	1	29	51.52	35	53.63	27	50.30
0101**00**10	0	18	44.22	24	42.46	32	51.34
0**11**01010	1	27	55.59	38	56.26	20	56.90
10100**11**0	1	28	51.36	29	55.14	33	50.12
1100**11**01	1	28	53.00	23	55.96	19	56.11
1101**00**10	0	30	44.87	28	41.89	20	47.50
11011**00**1	0	28	46.61	29	45.69	25	36.88
Streak Length = 3		**143**	**52.56**	**160**	**49.89**	**128**	**59.53**
01**000**101	0	17	60.00	28	43.79	26	59.19
010**11**01	1	32	53.63	20	56.00	20	56.75
0**111**0010	1	20	48.70	30	52.70	23	51.52
1**000**1010	0	24	51.75	29	44.45	19	56.42
11**000**110	0	22	46.73	21	51.33	23	67.78
11101010	1	28	54.86	32	52.78	17	66.47
Streak Length = 4		**103**	**53.16**	**107**	**47.73**	**102**	**56.03**
00001101	0	25	53.68	32	40.53	23	50.83
010**1111**0	1	27	53.67	28	49.71	34	58.09
1**0000**101	0	27	58.07	20	48.85	23	63.26
11110010	1	24	46.50	27	53.37	22	50.73
Streak Length = 5		**53**	**63.02**	**55**	**49.69**	**45**	**58.51**
01**00000**1	0	27	64.11	30	44.83	20	57.20
11111010	1	26	61.88	25	55.52	25	59.56
Streak Length = 6		**51**	**58.94**	**63**	**48.21**	**52**	**57.29**
1**000000**1	0	27	58.48	30	43.73	29	52.79
11111101	1	24	59.46	33	52.27	23	62.96

Note: Focal streaks are highlighted in bold type in the first column of the table.

Figure OS4

Study 2B: Proportion of Participants Predicting Repetition of Longest Streak in Each Filler Sequence, by Streak Length and Condition



Note: Proportion of participants predicting repetition of the longest streak in each filler sequence (Error Bars: +/– 1 S.E.). Streak Length is the length of the longest streak present in the Filler sequence, prior to the reversal at the end of that sequence. Filler sequences contained streaks that ranged from 2 to 6 identical signals in length. The updating pattern for participants submitting Binary Responses in the Analyst and Stock Conditions is quite similar to the pattern we observed among participants submitting Continuous Responses. But, participants submitting Binary Responses in the Bingo Condition show a stronger preference for reversal than those submitting Continuous responses. The proportion of participants predicting repetition increased slightly between Streak Length 2 and 6 in the Analyst and Stock Conditions. In the Bingo Condition, the proportion of participants predicting repetition of the Longest Streak decreased slightly with Streak Length. N = 301.

Table OS17

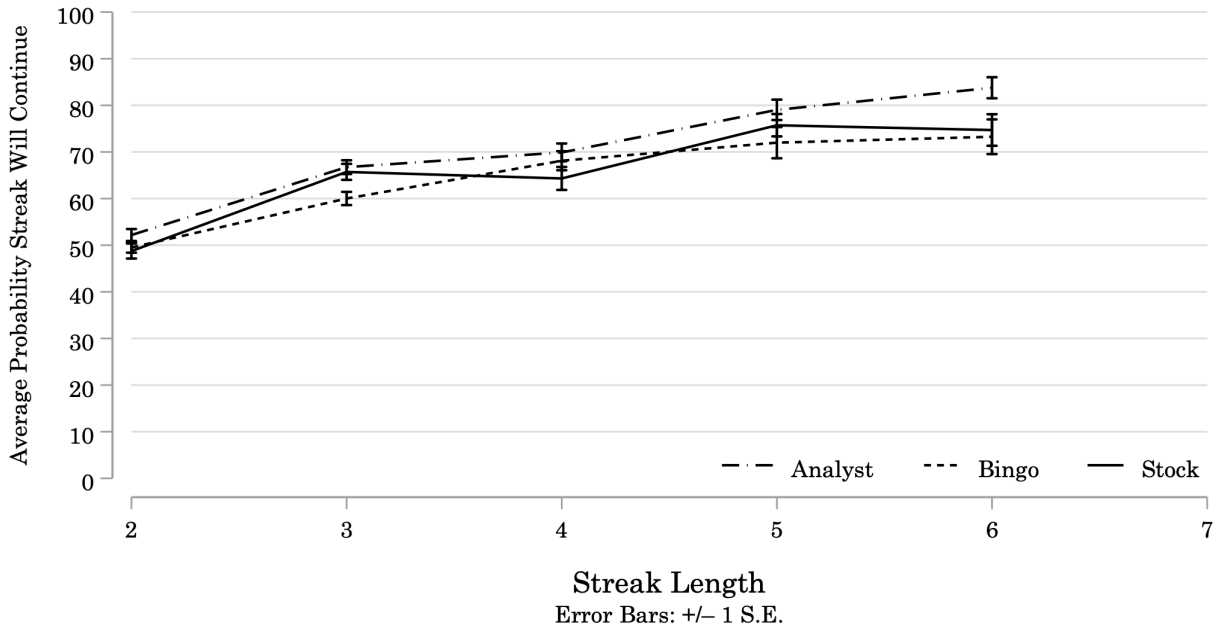
Study 2B: Proportion of Participants Predicting Repetition of Longest Streak in Each Filler Sequence, by Streak Length and Condition

	Streak Type	Analyst		Bingo		Stock	
		N	Prop Repeat	N	Prop Repeat	N	Prop Repeat
Streak Length = 2		**391**	**0.53**	**408**	**0.47**	**415**	**0.46**
00**11**0101	1	50	0.46	60	0.37	49	0.43
00110**11**	1	57	0.70	54	0.44	61	0.54
0101**00**10	0	46	0.35	40	0.38	49	0.33
0**11**01010	1	48	0.83	49	0.63	45	0.76
10100**11**0	1	38	0.74	53	0.60	50	0.64
1100**11**01	1	47	0.43	55	0.31	51	0.57
1101**00**10	0	56	0.32	49	0.39	48	0.21
11011**00**1	0	49	0.47	48	0.63	62	0.27
Streak Length = 3		**287**	**0.54**	**300**	**0.46**	**293**	**0.59**
01**000**101	0	45	0.60	54	0.43	42	0.45
010**111**01	1	48	0.48	48	0.23	51	0.57
0**111**0010	1	54	0.56	51	0.59	47	0.68
1**000**1010	0	46	0.37	47	0.26	50	0.38
11**000**110	0	52	0.48	51	0.55	51	0.65
11101010	1	42	0.79	49	0.69	52	0.79
Streak Length = 4		**191**	**0.49**	**218**	**0.39**	**211**	**0.55**
00001101	0	51	0.29	55	0.29	58	0.41
0101**1110**	1	45	0.60	51	0.45	55	0.65
1**0000**101	0	44	0.55	55	0.47	53	0.60
11110010	1	51	0.55	57	0.35	45	0.51
Streak Length = 5		**100**	**0.57**	**92**	**0.37**	**116**	**0.61**
01**00000**1	0	46	0.50	44	0.36	63	0.56
11111010	1	54	0.63	48	0.38	53	0.68
Streak Length = 6		**104**	**0.56**	**86**	**0.42**	**96**	**0.56**
1**0000001**	0	54	0.43	40	0.45	60	0.47
11111101	1	50	0.70	46	0.39	36	0.72

Note: Focal streaks are highlighted in bold type in the first column of the table.

Figure OS5

Study 3A: Average Probability Participants Assigned to Repetition of Longest Streak in Each Filler Sequence, by Streak Length and Condition



Note: The average probability participants assigned to repetition of the longest streak in each Filler sequence (Error Bars: +/– 1 S.E.). Streak Length is the length of the longest streak preceding the reversal at the end of each sequence. Filler sequences contained streaks that ranged from 2 to 6 identical signals. Participants in all 3 Conditions increased the probability they assigned to repetition of the longest streak as the length of that streak increased. Participants in the Analyst25-50-75 Condition assigned slightly higher probability to streaks of length 6, but otherwise predictions were similar across all 3 Conditions. N = 150.

Table OS18

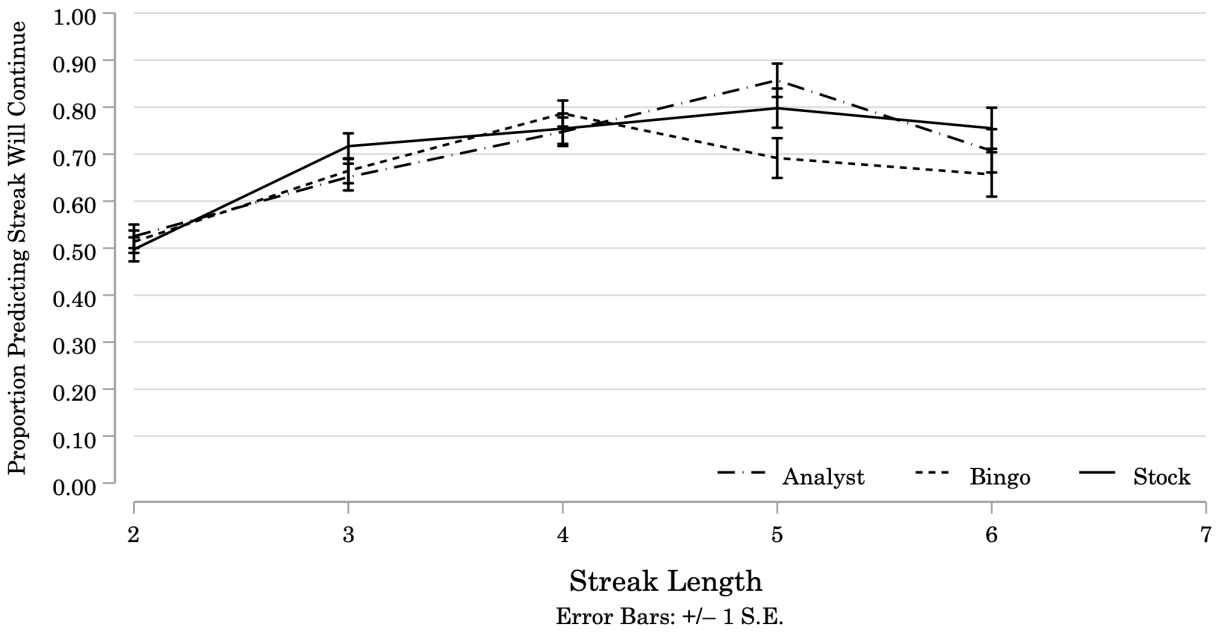
Study 3A: Average Probability Participants Assigned to Repetition of Longest Streak in Each Filler Sequence

	Streak Type	Analyst		Bingo		Stock	
		N	Mean Repeat	N	Mean Repeat	N	Mean Repeat
Streak Length = 2		**203**	**52.19**	**189**	**49.51**	**186**	**48.75**
00**11**0101	1	31	53.23	25	48.48	23	41.61
00110**1**0	1	20	49.55	26	50.85	18	48.83
0101**00**10	0	22	54.73	23	52.17	23	50.78
0**11**01010	1	24	57.71	28	55.61	26	56.00
10100**1**0	1	25	49.24	20	53.15	26	59.42
1100**11**01	1	29	66.66	19	58.11	23	61.70
1101**00**10	0	27	54.19	27	46.41	22	40.45
11011**00**1	0	25	29.48	21	30.76	25	30.12
Streak Length = 3		**148**	**66.74**	**156**	**60.02**	**141**	**65.73**
01**000**101	0	30	72.63	30	66.27	27	74.67
010**111**01	1	25	71.12	26	67.62	19	67.53
0**111**0010	1	17	56.47	21	57.76	22	59.59
1**000**1010	0	26	70.50	29	67.59	21	66.48
11**000**110	0	22	54.18	27	48.04	25	52.52
11101010	1	28	69.14	23	49.87	27	72.19
Streak Length = 4		**98**	**69.90**	**97**	**68.12**	**103**	**64.32**
00001101	0	19	63.21	23	66.48	28	56.54
010**1111**0	1	28	72.50	29	66.62	28	68.93
1**0000**101	0	28	74.57	22	69.86	23	66.13
11110010	1	23	66.57	23	70.00	24	66.29
Streak Length = 5		**52**	**79.04**	**58**	**72.00**	**60**	**75.73**
010**0000**1	0	26	77.23	34	71.06	26	74.65
11111010	1	26	80.85	24	73.33	34	76.56
Streak Length = 6		**58**	**83.78**	**49**	**73.27**	**55**	**74.71**
1**000000**1	0	29	83.66	24	70.71	22	68.77
11111101	1	29	83.90	25	75.72	33	78.67

Note: Focal streaks are highlighted in bold type in the first column of the table.

Figure OS6

Study 3B: Proportion of Participants Predicting Repetition of the Longest Streak in Each Filler Sequence, by Streak Length and Condition



Note: Proportion of participants predicting repetition of the longest streak in each filler sequence (Error Bars: +/– 1 S.E.). Streak Length is the length of the longest streak present in the Filler sequence, prior to the reversal at the end of that sequence. Filler sequences contained streaks that ranged from 2 to 6 identical signals in length. In all 3 Conditions, the proportion of participants predicting repetition increased across Streak Lengths 2 to 4. In the Analyst25-50-75 and Stock25-50-75 Conditions, the proportion of participants predicting repetition continued to increase between Streak Lengths 4 and 5, before tapering off at Streak Length 6. In the Bingo 25-50-75Condition, the proportion of participants predicting repetition declines across Streak Lengths 4 to 6. N = 300.

Table OS19

Study 3B: Proportion of Participants Predicting Repetition of Longest Streak in Each Filler Sequence, by Streak Length and Condition

	Streak Type	Analyst		Bingo		Stock	
		N	Prop Repeat	N	Prop Repeat	N	Prop Repeat
Streak Length = 2		**398**	**0.53**	**440**	**0.51**	**384**	**0.50**
00**11**0101	1	52	0.50	60	0.48	43	0.47
00110**11**0	1	46	0.65	54	0.46	52	0.58
0101**00**10	0	55	0.44	52	0.50	54	0.35
0**11**01010	1	56	0.73	50	0.58	44	0.77
10100**11**0	1	46	0.50	59	0.59	41	0.73
1100**11**01	1	47	0.68	61	0.61	50	0.50
1101**00**10	0	51	0.47	57	0.47	45	0.38
11011**00**1	0	45	0.20	47	0.38	55	0.29
Streak Length = 3		**281**	**0.65**	**319**	**0.66**	**272**	**0.72**
01**000**101	0	40	0.65	55	0.78	50	0.80
010**111**01	1	53	0.55	44	0.64	41	0.73
0**111**0010	1	51	0.65	55	0.53	43	0.63
1**000**1010	0	49	0.65	64	0.72	52	0.65
11**000**110	1	41	0.51	48	0.52	41	0.66
11101010	1	47	0.89	53	0.77	45	0.82
Streak Length = 4		**206**	**0.75**	**220**	**0.79**	**179**	**0.75**
00001101	0	54	0.72	54	0.80	47	0.66
010**1111**0	1	49	0.78	55	0.78	42	0.79
1**0000**101	0	59	0.71	46	0.76	44	0.80
11110010	1	44	0.80	65	0.80	46	0.78
Streak Length = 5		**98**	**0.86**	**120**	**0.69**	**94**	**0.80**
01**00000**1	0	52	0.83	61	0.67	48	0.67
11111010	1	46	0.89	59	0.71	46	0.93
Streak Length = 6		**99**	**0.71**	**102**	**0.66**	**98**	**0.76**
1**000000**1	0	49	0.71	49	0.65	43	0.72
11111101	1	50	0.70	53	0.66	55	0.78

Note: Focal streaks are highlighted in bold type in the first column of the table.

Binary Logistic Regression Analysis

The following analysis uses a binary logistic regression model to estimate the population-average effect of Streak Length on the binary outcome variable Prediction, which takes the value "0" when participants predict that the 9th signal *will not match* the 8th signal in the sequence (the streak will *reverse*), and "1" when participants predict that the 9th signal *will match* the 8th signal (the streak will *continue*). The analysis relies heavily on guidelines presented by Szmaragd, Clarke, and Steele (2013), and by Sommet and Morselli (2017).

Section 1 provides an explanation of the model. Section 2 presents the results of the analysis for the Binary version of Study 1, where participants were not told the generator's rate. Section 3 presents the results of the analysis for the Binary version of Study 2, where participants were told the generator's rate was fixed at 50%. Section 4 presents the results of the analysis for the Binary version of Study 3, where participants were told there was an equal chance (33%) that the generator's rate was either 25%, 50%, or 75%.

Section 1: Explanation of the Binary Logistic Repeated Measures Model

If we fit repeated measures data using a standard linear model, the result would be one model equation defined for each participant at each Streak Length. This model assumes that all the residuals across these models are independent of each other, but the residuals in the models defined for a particular subject may in fact be correlated with one another. If we are interested in the population-average effect of Streak Length on Prediction, we can specify a population average model so as to take into account such within-participant autocorrelation. Prediction is a binary variable taking the value "0" (if participants predict a streak will *reverse*) or "1" (if participants predict a streak will *repeat*). We estimate the average log-odds that Prediction = 1 (participant predicts streak will *repeat*) for participant *i* at Streak Length *t* as

$$logitPr(y_{it} = 1|x_{it}) = \beta_0^{PA} + \beta_1^{PA} x_{it}$$

where $logitPr(y_{it} = 1|x_{it})$ is the average log-odds that Prediction = 1 among those participants with predictor variables x_{it}. We can account for within-participant autocorrelation in our estimates of this model by specifying the autocorrelation between the residuals e_{i2}, …, e_{i7}, for each participant *i* and Streak Length 2,...,7. This autocorrelation structure is specified through the choice of a working correlation matrix.[3]

Generalized Estimating Equations (GEE) allow us to estimate the logistic model while allowing for autocorrelation by specifying the structure of a working autocorrelation matrix.

> GEE is a two-stage method in which the autocorrelation structure is treated as a nuisance to be adjusted for. Stage 1 of GEE involves estimating the 'working correlation matrix', the structure of which the user must specify prior to estimation; to specify this matrix correctly, the user must declare the occasion variable. Stage 2 of GEE uses the estimated working correlation matrix to adjust the estimates of the logistic model parameters and standard errors for autocorrelation (Szmaragd et al. 2013, p. 152).

In the subsequent analysis, we specify an unstructured correlation matrix. No assumption is made about the structure of the correlation matrix – the correlation between residuals e_{it} and e_{is} is allowed to vary for each pair of Streak Lengths *t*, *s*, where $t \neq s$. In other words, the correlation between the residual terms in the models for participant *i* at Streak Lengths *t* and *s* is allowed to

[3] Note that in the standard logistic model, there does not appear to be a residual specified. These residuals are "hidden," in the sense that we implicitly assume there is some continuously distributed y^*_{it} that we cannot directly observe, instead we only observe whether the value of y^*_{it} is positive (1) or negative (0). We assume that this hidden outcome variable follows a linear model that depends on the same predictors as in the logistic model, *and* a hidden residual e^*_{it} that is logistically distributed.

be different than that between all other pairs *t'*, *s'*.[4] The log odds of a participant predicting a streak will *repeat* (Prediction = 1) is then estimated as a function of Condition, Streak Length, and the interaction between Condition and Streak Length.[5]

Throughout this analysis, we report the odds ratio (*OR*) as opposed to the log odds, for interpretability. We will refer to the odds ratio as the odds, following convention.

Section 2: Study 1B

The results of the binary repeated measures logistic regression do not differ substantively from the results of the one-way mixed ANOVA. There was a significant main effect of Condition on Prediction ($\chi^2(2) = 11.71$, $p = 0.003$). Pairwise comparisons revealed the odds a participant in the BingoUnknown Condition predicts a streak will *repeat* were significantly lower than the odds for a participant in the AnalystUnknown Condition. However, after applying the Bonferroni correction for multiple comparisons, this difference was no longer significant ($OR_{\text{BINGO/ANALYST}} = 0.65$, $p_{\text{UNADJUSTED}} = 0.034$, $p_{\text{BONFERRONI}} = 0.101$). The odds that a participant in the BingoUnknown Condition predicts a streak will *repeat* were also significantly lower than the odds that a participant in the StockUnknown Condition predicts *repeat*, even after applying the Bonferroni correction ($OR_{\text{BINGO/STOCK}} = 0.65$, $p_{\text{UNADJUSTED}} = 0.001$, $p_{\text{BONFERRONI}} = 0.003$). No significant difference was found between the AnalystUnknown and StockUnknown Conditions ($OR_{\text{STOCK/ANALYST}} = 1.31$, $p_{\text{UNADJUSTED}} = 0.222$, $p_{\text{BONFERRONI}} = 0.666$).

---

[4] Another option would be to use exchangeable or autocorrelation types of correlation matrix structures. In the former case, the correlation between any pair of residual terms is assumed equal to any other pair. In the latter, the correlation between each pair of residual terms e_{it}, e_{it-n} is decreasing in n. We initially chose the autocorrelation structure, because it seems reasonable to assume the correlation between the log-odds a participant predicts a streak of length 7 will *repeat* is likely to be highly correlated with the log-odds a participant predicts a streak of length 6 will *repeat*, but less likely to be highly correlated with the participant's prediction for streaks of length 5 and smaller. There was no difference between the estimates of the model with autocorrelation of error terms and unstructured correlation of error terms. Following Szmaragd et al. (2013) we chose the unstructured model as it is preferred over the alternatives whenever there are few enough observations that it is computationally feasible.

[5] I treat Streak Length as a nominal variable, because within the GEE family of models Stata doesn't have a specific facility for ordinal variables, and the documentation says to just treat ordinals as nominals.

The main effect of Streak Length was significant ($\chi^2(5) = 153.42$, $p < 0.000$). The odds a participant predicts a streak will *repeat* increase as Streak Length increases. Pairwise comparisons revealed only 4 cases where the odds do not increase significantly at longer streak lengths. The odds a participant predicts a streak will *repeat* are not significantly higher for streaks of length 6 than they are for streaks of length 5 ($OR_{\text{STREAK=6/STREAK=5}} = 1.26$, $p_{\text{UNADJUSTED}} = 0.083$), nor are they significantly higher for streaks of length 7 than they are for streaks of length 6 ($OR_{\text{STREAK=7/STREAK=6}} = 1.15$, $p_{\text{UNADJUSTED}} = 0.340$). After applying the Bonferroni correction, two additional pairs emerge as nonsignificant. The odds for streaks of length 4 are not significantly higher than for streaks of length 3 ($OR_{\text{STREAK=4/STREAK=3}} = 1.40$, $p_{\text{BONFERRONI}} = 0.139$), and the odds for streaks of length 7 are not significantly higher than streaks of length 5 ($OR_{\text{STREAK=7/STREAK=5}} = 1.45$, $p_{\text{BONFERRONI}} = 0.205$). At every other combination of Streak Lengths *t* and *t-n*, the odds that a participant predicts a streak will *repeat* are significantly higher (at the $p < 0.001$ level, after applying Bonferroni corrections) for streaks of length *t* than for streaks of length *t-n*. The interaction between Condition and Streak Length was not significant ($\chi^2(10) = 12.44$, $p = 0.257$).

Section 3: Study 2B

The results of the binary repeated measures logistic regression do not differ substantively from the results of the one-way mixed ANOVA. There was a significant main effect of Condition on Prediction ($\chi^2(2) = 18.07$, $p = 0.002$). Pairwise comparisons revealed the odds a participant in the Bingo50 Condition predicts a streak will *repeat* are significantly lower than for a participant in the Analyst50 Condition ($OR_{\text{BINGO/ANALYST}} = 0.55$, $p_{\text{UNADJUSTED}} = 0.004$, $p_{\text{BONFERRONI}} = 0.011$). The odds that a participant in the Bingo50 Condition predicts a streak will *repeat* are also significantly lower than the odds that a participant in the Stock50 Condition

predicts *repeat*, even after applying the Bonferroni correction ($OR_{\text{BINGO/STOCK}} = 0.50$, $p_{\text{UNADJUSTED}} = 0.001$, $p_{\text{BONFERRONI}} = 0.003$). No significant difference was found between the Analyst50 and Stock50 Conditions ($OR_{\text{STOCK/ANALYST}} = 1.10$, $p_{\text{UNADJUSTED}} = 0.637$, $p_{\text{BONFERRONI}} = 0.100$).

The main effect of Streak Length was significant ($\chi^2(5) = 62.37$, $p < 0.000$). The odds a participant predicts a streak will *repeat* increase as Streak Length increases, but only for streaks of 4 or longer. Pairwise comparisons revealed 5 cases where the odds do not increase significantly at longer streak lengths. The odds a participant predicts a streak will *repeat* are not significantly higher for streaks of length 3 than they are for streaks of length 2 ($OR_{\text{STREAK=3/STREAK=2}} = 0.89$, $p_{\text{UNADJUSTED}} = 0.455$). The odds are not significantly higher for streaks of length 6 than they are for streaks of length 5 ($OR_{\text{STREAK=6/STREAK=5}} = 1.10$, $p_{\text{UNADJUSTED}} = 0.434$). The odds for streaks of length 7 are not significantly higher than streaks of length 6 ($OR_{\text{STREAK=7/STREAK=6}} = 0.96$, $p_{\text{UNADJUSTED}} = 0.337$), and the difference between the odds for streaks of length 7 are not significantly higher than streaks of length 5 ($OR_{\text{STREAK=7/STREAK=5}} = 1.23$, $p_{\text{UNADJUSTED}} = 0.084$). At every other combination of Streak Lengths *t* and *t-n*, the odds that a participant predicts a streak will *repeat* are significantly higher (at the $p < 0.001$ level, after applying Bonferroni) for streaks of length *t* than for streaks of length *t-n*.

The interaction between Condition and Streak Length was significant ($\chi^2(10) = 25.02$, $p = 0.005$). For streaks longer than 4 signals, the odds a participant predicts a streak will *repeat* are significantly lower in the Bingo50 than in the Analyst50 Condition. There is no significant difference between the Analyst50 and Stock50 Conditions.

Section 4: Study 3B

The results of the binary repeated measures logistic regression do not differ substantively from the results of the one-way mixed ANOVA. The main effect of Condition on Prediction was not significant ($\chi^2(2) = 1.23$, $p = 0.540$). Pairwise comparisons revealed no significant difference between the odds a participant in the Bingo25-50-75 Condition predicts a streak will *repeat* versus a participant in the Analyst25-50-75 Condition ($OR_{\text{BINGO/ANALYST}} = 1.09$, $p_{\text{UNADJUSTED}} = 0.661$), or between the odds a participant in the Stock25-50-75 Condition predicts *repeat* versus a participant in the Analyst25-50-75 Condition ($OR_{\text{STOCK/ANALYST}} = 1.25$, $p_{\text{UNADJUSTED}} = 0.270$). No significant difference was found between the Bingo25-50-75 and Stock25-50-75 Conditions either ($OR_{\text{STOCK/BINGO}} = 1.15$, $p_{\text{UNADJUSTED}} = 0.492$).

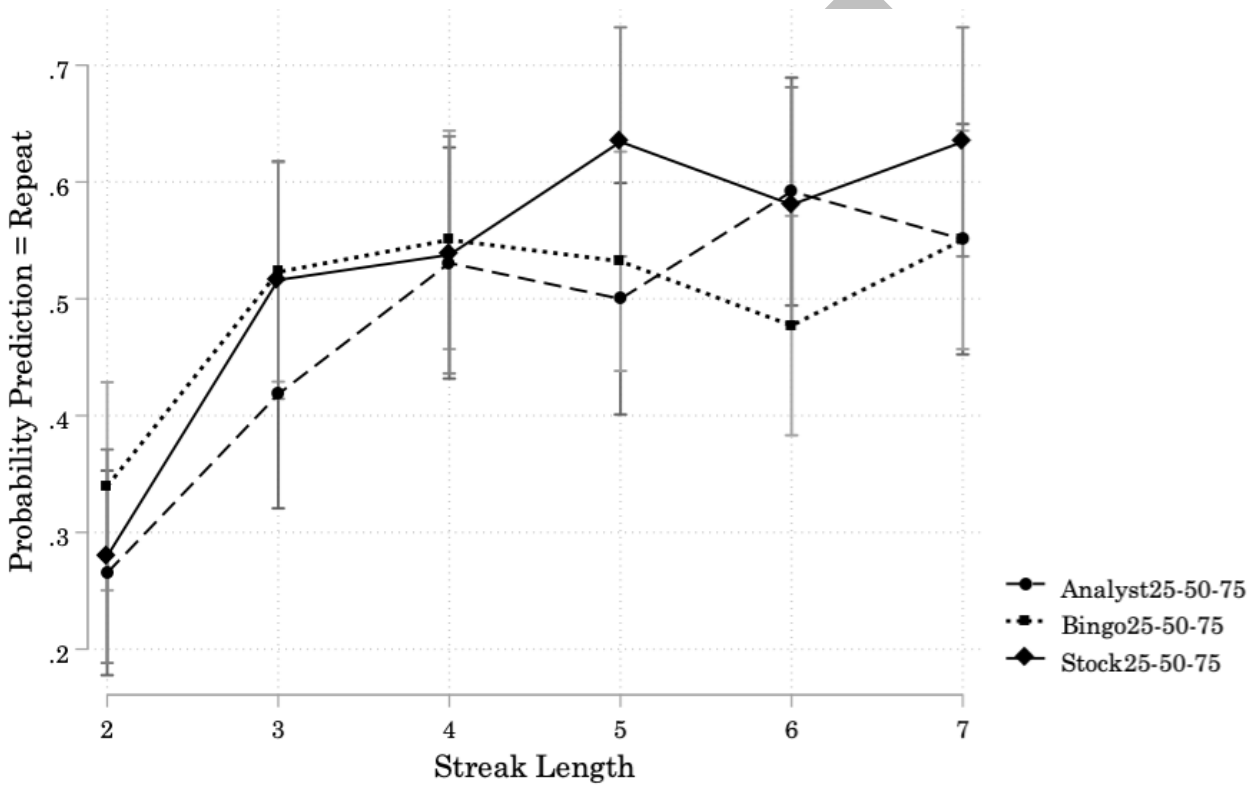
The main effect of Streak Length was significant ($\chi^2(5) = 66.07$, $p < 0.000$). But, it seems this main effect is driven by the difference between predictions about streaks of length 2 versus every other Streak Length. Pairwise comparisons revealed that the odds a participant predicts streaks of length 3, 4, 5, 6, and 7 *repeat* are all significantly higher than the odds for streaks of length 2 (even after applying the Bonferroni correction for multiple comparisons). However, the odds are not significantly higher for streaks longer than 4 compared to shorter streaks (e.g. 4 versus 3, 5 versus 4, 7 versus 6). After an initial increase in odds between Streak Length 2 and 3, the odds a participant predicts a streak will *repeat* seem to stabilize, with no consistent pattern of increasing.

The interaction between Condition and Streak Length was not significant ($\chi^2(10) = 15.92$, $p = 0.102$). But, an inspection of the marginal linear predictions for the probability a participant will predict a streak will *repeat* appear to show a subtly increasing pattern for the

Stock25-50-75 and Analyst25-50-75 Conditions, while predictions in the Bingo25-50-75 Condition seem to stabilize for streaks longer than 3 (Figure OS7).

Figure OS7

Marginal Linear Predictions of Probability Participant Predicts Repeat, by Condition and Streak Length



References

FINRA Foundation. (2015). The National Financial Capability Study (NFCS): A project of the FINIRA Investor Education Foundation (FINRA Foundation). Retrieved from: http://www.usfinancialcapability.org/downloads.php

Heinze, G. and Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine, 21*(16), 2409–2419.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C.K., Mazzocco, K., and Dickert, S. (2006). Numeracy and Decision Making. *Psychological Science, 17(5)*, 407-413.

Sommet, Nicolas, and David Morselli. (2017). Keep Calm and Learn Multilevel Logistic Modeling: A Simplified Three-Step Procedure Using Stata, R, Mplus, and SPSS. *International Review of Social Psychology, 30*(1), 203–218.

Szmaragd, Camille, Paul Clarke, and Fiona Steele. (2013). Subject specific and population average models for binary longitudinal data: a tutorial. *Longitudinal and life course studies, 42*(2), 147-165.