

The Virtual Bingo Blower: An open-source tool to generate ambiguity and risk in experiments.*

Ola Andersson[†] Geoffrey Castillo[‡] Erik Wengström[§]

25th September 2024

We propose the Virtual Bingo Blower (VBB) as a way to generate credible risk and ambiguity in computerized experiments. Using a physics engine—a computer simulation of a physical system—the VBB simulates a conventional bingo blower. Different aspects of the VBB, such as the number of balls, their color, and their speed, can be easily modified. In an online experiment, we measure ambiguity attitudes and vary the source of ambiguity, using either the VBB or natural events. We find that the VBB and natural events result in a similar degree of ambiguity aversion. Further, we find that, by manipulating the number of balls, the VBB can be used to manipulate the perceived level of ambiguity.

Keywords: ambiguity, risk, bingo blower, online experiment, experimental tools, randomization

JEL Codes: C91, D81

*We gratefully acknowledge financial support from the Jan Wallander and Tom Hedelius Foundation and Riksbankens Jubileumsfond. We are grateful for comments from John Hey, Gianna Lotito, Anna Maffioletti, Andrea Morone, and Peter Wakker, as well as seminar participants at ESA World 2023 in Lyon, NBEE in Lund, MiddExLab virtual experimental economics seminar, Department of Psychology at Uppsala University, and ESA Europe 2024 in Helsinki.

[†]Department of Economics, Uppsala University. ola.andersson@nek.uu.se

[‡]Department of Economics, Nottingham Trent University. geoffrey.castillo@ntu.ac.uk

[§]Department of Economics, Lund University, and Hanken School of Economics. erik.wengstrom@nek.lu.se

1 Introduction

In this paper, we present a novel tool to generate ambiguity in experiments: the Virtual Bingo Blower (VBB). In a bingo blower, balls of different colors move continuously, and a ball selected at random determines the winning color. The constant motion of the balls makes it difficult to assess their proportions, thus generating ambiguity.

Bingo blowers have recently become popular for generating ambiguity in experiments (Georgalos, 2019, 2021; Hey et al., 2010; Hey and Pace, 2014; Kothiyal et al., 2014). They are, however, bulky and noisy, as well as difficult to use for online experiments. The VBB solves these issues by running in web browsers. It is thus adapted to computerized experiments in general, and to online experiments in particular.

The VBB relies on a real-time physics engine to simulate the movement of the balls. A key advantage of using a physics simulation is that the VBB is not deterministic. As a consequence, no one, including the experimenters, can predict how the balls move and thus which ball will be drawn. Subjects can verify this by manipulating the balls within the VBB and seeing how the physics simulation reacts to their manipulations.

To demonstrate the usefulness of the VBB, we validate it in an online experiment in which we measure ambiguity attitudes using the method of Baillon et al. (2018). We compare ambiguity attitudes measured when ambiguity is generated with the VBB and when ambiguity is generated with natural events. We show that the VBB and natural events lead to a similar degree of ambiguity aversion. We also show that the perceived level of ambiguity can be manipulated by increasing the number of balls in the VBB.

Our contribution is thus to offer a validated tool to generate ambiguity, particularly suited to online experiments. Compared to other methods, a main advantage of the VBB is that it is easy to change the number of balls, their proportions, their colors, or even their speed. Contrary to a physical bingo blower, which would need to be stopped, emptied, and refilled with balls, with the VBB these changes are simple and can even be done trial-to-trial.

Moreover, while we focus on ambiguity, the VBB can also generate risk by using a small, easily countable number of balls, or by revealing the proportions

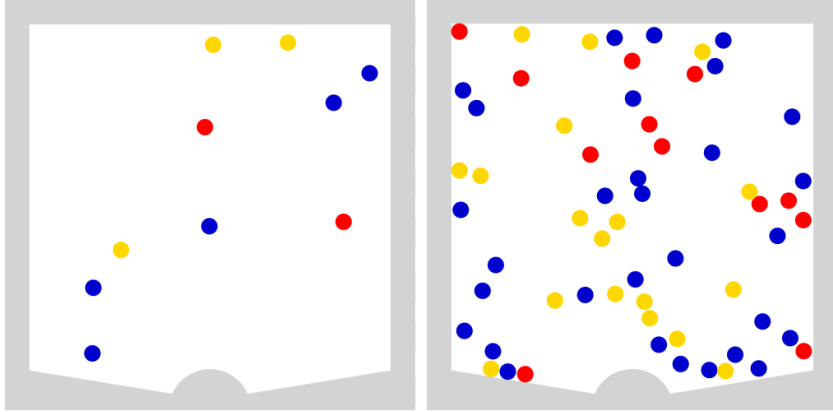


Figure 1: VBB with 10 balls and with 60 balls.

of the different colors to subjects. This way, experimenters can use the VBB to, for example, play the lottery selected for payment at the end of a risk elicitation experiment, or to choose which round of a public goods game is relevant for payment. By doing so, experimenters do not have to rely on random number generators in their code, which are not transparent to subjects.

Figure 1 shows screenshots of the VBB, for a small and a large number of balls. To see the VBB in action, a demo is available at <https://geoffreycastillo.com/bingo-blower-js-demo>. The VBB runs in all modern web browsers, including on mobile. As such, it can be used in any web-based experimental software, such as Qualtrics, oTree (Chen et al., 2016), or LIONESS (Giamattei et al., 2020). Apart from being able to add balls of different colors, the parameters of the physics simulation, such as the air resistance or the bounciness of the balls, can also be adjusted. The VBB is open-sourced and available as `bingo-blower.js` on GitHub at <https://github.com/geoffreycastillo/bingo-blower-js>, where we also provide documentation.

In Section 2, we present the online experiment we conducted to validate the VBB. In the experiment, we elicit ambiguity attitudes using the method of Baillon et al. (2018). This method measures ambiguity attitudes via two indices: one captures ambiguity aversion, while the other captures the perceived level of ambiguity. If there is no ambiguity at all, which would happen if subjects know the proportions of the different colors, then both indices should be equal to 0.

Between treatments, we generate ambiguity using either the VBB or natural

events. We further manipulate the number of balls in the VBB while keeping the proportions of the different colors constant. For the natural events, we use the evolution of the Dow Jones Industrial Average.

We present our results in Section 3. In line with our preregistered hypotheses, we find that ambiguity aversion is similar across treatments. On average, we find evidence for slight ambiguity seekingness, both with the VBB and with natural events, the latter being in line with the results of Baillon et al. (2018). This result suggests that subjects perceived the VBB as a natural source of ambiguity.

Further, we find that the VBB gives rise to a lower perceived level of ambiguity compared to natural events, and that increasing the number of balls in the VBB increases the perceived level of ambiguity. The increase is small, but note that our manipulation—increasing the number of balls—is simple. More complex manipulations, such as increasing the speed of the balls, making their movement more erratic, or choosing colors that are less distinguishable, may be able to generate even greater differences. Previous research has also found that the level of ambiguity varies between different ambiguity sources, but that it is hard to manipulate the level within a given source (Garcia et al., 2020).

We contribute to the literature by offering a tool that shares the positive features of bingo blowers and adds the ease of computerized implementation. As Hey et al. (2010) explain, bingo blowers solve many of the issues associated with other methods used to generate ambiguity (see Camerer and Weber, 1992, and Trautmann and van de Kuilen, 2015, for reviews of the ambiguity literature). For example, they do not rely on information asymmetry between experimenters and subjects, as is the case with unknown urns (made famous by Ellsberg, 1961). Subjects can always look at the balls moving inside, and can witness how a ball is selected for payment. Bingo blowers also do not rely on second-order probability distributions to generate ambiguity (e.g., MacCrimmon and Larsson, 1979), which generates risk instead (Abdellaoui et al., 2011). Further, unlike natural events, bingo blowers are not affected by specific knowledge that subjects may have, which would need to be controlled for.

Due to the appeal of bingo blowers, other attempts have been made to translate them into computerized environments. For example, Levati and Morone (2013) used a 3D animation of a bingo blower cage projected on a wall to generate ambiguity in their experiments. Closer to us, Morone and Caferra (2024) also developed

a new tool, the Ambiguity Box, which aims at adapting the bingo blower to computerized experiments. The Ambiguity Box was used by Caferra et al. (2023) to study dynamic inconsistency under ambiguity. A crucial difference between the Ambiguity Box and the VBB is that the VBB uses a real-time physics engine to simulate the movement of the balls inside the bingo blower. By contrast, the Ambiguity Box is a sequence of pre-rendered static images displayed as a GIF, where the proportion of squares of different colors corresponds to the underlying probabilities. The VBB is thus less abstract, closer to the experience of a real bingo blower. Another implication of using a real physics simulation is that subjects can themselves trigger and witness the drawing of a ball from the bingo blower to determine their earnings. Neither subjects nor experimenters can predict which ball is going to be drawn. Experiments using the Ambiguity Box instead have to rely on a random number generator to play out the chosen lottery at the end of the experiment, which reduces transparency to subjects and could introduce mistrust, something the use of the bingo blower initially aimed at avoiding.

2 Experimental design: Validating the VBB

The aim of the experiment is to validate the VBB in a setting with ambiguity. To do so, we measure ambiguity attitudes in three treatments that differ in how we generate ambiguity. We then compare the measured ambiguity attitudes between these treatments.

2.1 Measuring ambiguity attitudes

We follow the method introduced by Baillon et al. (2018) to measure ambiguity attitudes. We refer to that paper for details.

The method starts by partitioning the set of possible events into at least three mutually exclusive and exhaustive, non-null, single events: $\{E_1, E_2, E_3\}$. The unions of these events are called composite events: $\{E_1 \cup E_2, E_1 \cup E_3, E_2 \cup E_3\}$. Then, for each event $E \in \{E_1, E_2, E_3, E_1 \cup E_2, E_1 \cup E_3, E_2 \cup E_3\}$, subjects are asked to report the matching probability m that makes them indifferent between receiving x if E is realised and receiving x with probability m . Throughout our experiment, we use $x = \pounds 3$.

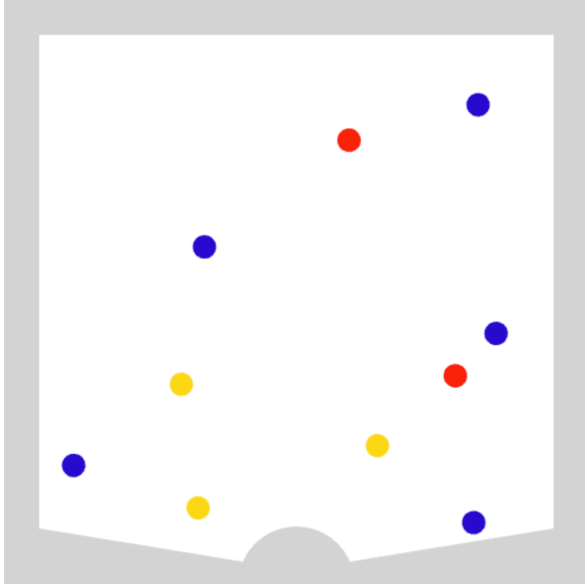
Option A You win £3 if one of the following balls is drawn:	Your choice	Option B You win £3 with the following chance:
	A B	0 %
	A B	20 %
	A B	35 %
	A B	40 %
	A B	45 %
	A B	50 %
	A B	55 %
	A B	60 %
	A B	65 %
	A B	70 %
	A B	75 %
	A B	80 %
	A B	85 %
	A B	90 %
	A B	93 %
	A B	95 %
	A B	97 %
	A B	98 %
	A B	99 %
	A B	100 %

Figure 2: Example of a task (VBB10 treatment).

Baillon et al. (2018) elicit matching probabilities using choice lists, which we also do. For each event, subjects face 10 choices presented as rows in a table. In each choice, subjects choose between an ambiguous option—winning £3 if the event is realized—and a risky option that offers a probability p of winning £3. Across all 10 choices, the ambiguous option remains the same while p varies between 0% and 100%. Since there are six events and 10 choices per event, this procedure creates 60 questions organized in six tables, one for each event. Figure 2 shows an example of such a table. As in Baillon et al. (2018), the tables are centered around 33% for single events and around 66% for composite events.

Based on these matching probabilities, Baillon et al. (2018) define two indices: the b -index and the a -index. The b -index measures aversion to ambiguity. As such, it captures motivational components of ambiguity attitudes. The a -index measures likelihood insensitivity, which generalizes ambiguity perception. It thus should be higher when one perceives a higher level of ambiguity. This index captures cognitive

Table 1: Treatments.

Treatment	Source	Events
VBB10	VBB 10 balls	Ball drawn $\in \{\text{Blue, Red, Yellow}\}$
VBB60	VBB 60 balls	Ball drawn $\in \{\text{Blue, Red, Yellow}\}$
Natural	Dow	% change $\in \{(-100, -51), (-51, 0.11), (0.11, \infty)\}$

components of ambiguity attitudes and is thus more malleable.

$b > 0$ indicates ambiguity aversion, $b < 0$, ambiguity seekingness, and $a > 0$, insensitivity. If one is ambiguity neutral, or if all events have known probabilities—i.e., there is no ambiguity—then $b = a = 0$.

2.2 Treatments

In our treatments, we manipulate the source of ambiguity. Table 1 summarizes our treatments. As can be seen, two of them, VBB10 and VBB60, use the VBB to generate ambiguity. The last treatment, Natural, uses movements of the Dow Jones Industrial Average Index.

2.2.1 The VBB treatments: VBB10 and VBB60

In the VBB treatments, an event is the color of the ball drawn from the VBB. We use three different colors: red, blue, and yellow. The six events are thus: $\{\text{red, blue, yellow, red} \cup \text{blue, red} \cup \text{yellow, blue} \cup \text{yellow}\}$.

The proportion of each color is the same in the two VBB treatments: 20% red, 50% blue, and 30% yellow. The two treatments differ in the total number of balls in the VBB: 10 balls for VBB10, and 60 for VBB60.

2.2.2 The Natural treatment

In the Natural treatment, an event is the daily percentage change of the Dow Jones Industrial Average Index on a specific day in the future. We use historical data on the daily movements of the Dow to calibrate the single events to be approximately equal in terms of probability to the single events in the VBB treatments. The partition that results from this calibration is the intervals of percentage changes: $(-100, -0.51]$, $(-0.51, 0.11]$, and $(0.11, \infty)$. The six events

are thus: $\{(-100, -0.51], (-0.51, 0.11], (0.11, \infty), (-100, -0.51] \cup (-0.51, 0.11], (-100, -0.51] \cup (0.11, \infty), (-0.51, 0.11] \cup (0.11, \infty)\}$.

2.3 Hypotheses

Our hypotheses relate to how the treatments affect the b - and the a -indices.

If the b -index strictly captures motivational components of ambiguity attitudes, it should not be affected by changes in the source of ambiguity. However, it is plausible that ambiguity aversion, and consequently the b -index, could vary with the treatment. We thus adopt a conservative stance and assume the b -index will not be affected by our treatments.

For the a -index, we expect that, moving from VBB10 to VBB60 and then to Natural, the level of perceived ambiguity to increase: while it may be possible to count the number of balls in VBB10, it is impossible in VBB60, and there is no probability distribution in Natural.

This reasoning can be summarized in the following two preregistered hypotheses, where we let $a(i)$ and $b(i)$ denote the a - and the b -index in treatment i .

Hypothesis 1 (b -index). Ambiguity aversion, as measured by the b -index, does not vary between treatments: $b(\text{VBB10}) = b(\text{VBB60}) = b(\text{Natural})$.

Hypothesis 2 (a -index). The perceived level of ambiguity, as measured by the a -index, increases from the VBB10 treatment to the VBB60 treatment, and reaches its highest level in the Natural treatment: $a(\text{VBB10}) < a(\text{VBB60}) < a(\text{Natural})$.

2.4 Recruitment and experimental procedures

Power calculations showed that we needed 150 subjects in each of the three treatments (significance level $\alpha = 0.05$, power $1 - \beta = 0.8$, effect sizes as in Baillon et al., 2018). Therefore, we aimed to recruit 450 subjects in total.

We recruited participants on Prolific. They were paid a fixed fee of £1 for participating in the experiment. In addition, participants could earn a bonus of £3 if a winning event occurred. Participants were restricted to be living in the UK, speaking English fluently, and having a Prolific approval rating of at least 95%.

The experiment was programmed using oTree (Chen et al., 2016).¹ The full instructions can be found in Appendix C. There we also present screenshots of all tasks. Our design, hypotheses, and exclusion criteria, were registered on the OSF.²

481 subjects finished the experiment. We excluded one subject who did not do the experiment on a laptop or desktop as instructed, as well as 14 subjects whose screen had a refresh rate lower than 50 FPS—both exclusion criteria were pre-registered. That leaves us with 466 subjects: 166 in VBB10, 155 in VBB60, and 149 in Natural. In the regression analyses—which include a gender dummy—we additionally exclude three participants who preferred not to state their gender.

3 Results

In Figure 3 we provide an overview of the data. At first glance, we can see that, on average, the b -index is slightly negative in all treatments, which indicates slight ambiguity seekingness. The a -index is positive, which indicates that subjects perceived ambiguity and were likelihood insensitive.

To test our hypotheses, we follow Baillon et al. (2018) and use seemingly unrelated regressions (SUR). Table 2 reports the regression results. Model (1) does not include controls, and model (2) includes preregistered controls (age and gender). Note that the significance levels reported in this Table refer to two-sided tests. Since Hypothesis 2 is one-sided, we will report the associated one-sided p -values in the text. As a summary of our results, Figure 4 shows the values of a - and b -indices inferred from model (2).

As we can see, we do not observe treatment effects for the b -index. In both models, $b(\text{VBB10})$ is not statistically different from $b(\text{VBB60})$ and from $b(\text{Natural})$. Therefore, in line with Hypothesis 1, we do not reject the null that the b -index is the same across treatments.

On the other hand, we observe an effect on the a -index. $a(\text{VBB60})$ is statistically greater than $a(\text{VBB10})$ (without controls: one-sided $p = 0.090$, with controls: one-sided $p = 0.046$) and than $a(\text{Natural})$ (with and without controls: one-sided $p < 0.001$). Therefore, in line with Hypothesis 2, we reject the null that the a -

¹Our oTree code is available at <https://github.com/geoffreycastillo/otree-virtual-bingo-blower>.

²See <https://osf.io/dm3kq>.

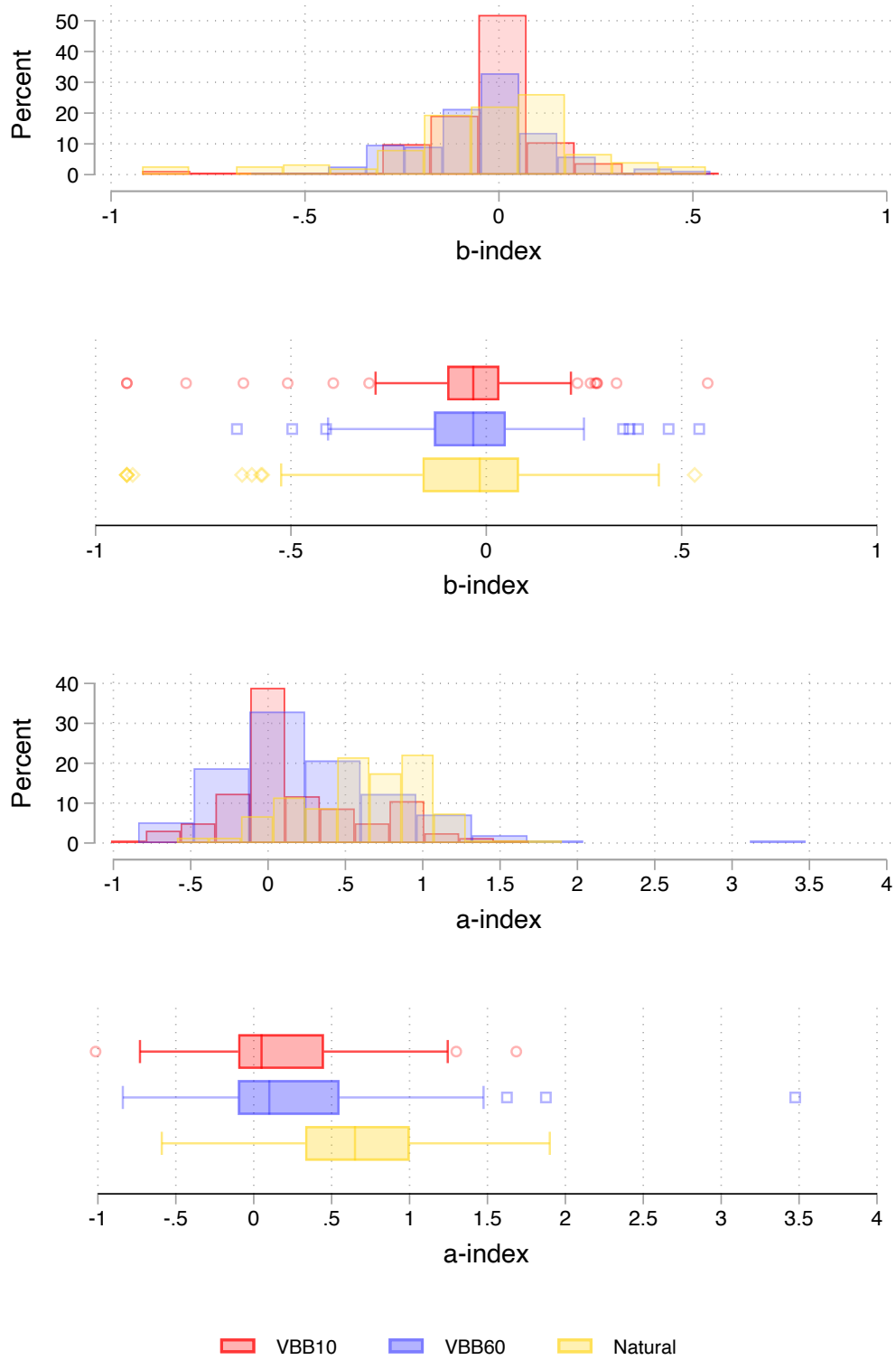


Figure 3: Histograms and box plots for the *b*- and *a*-index.

Table 2: Effect of treatments on a - and b -indices, SUR regression.

	(1)		(2)	
	a -index	b -index	a -index	b -index
VBB60	0.073 (0.055)	0.009 (0.023)	0.091 [•] (0.054)	0.008 (0.023)
Natural	0.449 ^{***} (0.055)	-0.011 (0.023)	0.454 ^{***} (0.055)	-0.009 (0.023)
Age category (ref.: 18-27)				
28-37			0.165 ^{**} (0.063)	0.008 (0.027)
38-47			0.037 (0.069)	-0.027 (0.029)
48-57			0.160 [*] (0.070)	0.011 (0.030)
58+			0.232 ^{**} (0.084)	-0.047 (0.036)
Female			0.037 (0.046)	0.014 (0.019)
Constant	0.176 ^{***} (0.038)	-0.047 ^{**} (0.016)	0.039 (0.064)	-0.050 [•] (0.027)
Observations	463		463	
R^2	0.138	0.002	0.163	0.012
Wald χ^2	74.133	0.733	90.311	5.502
Prob > χ^2	0.000	0.693	0.000	0.599
Breusch-Pagan χ^2	3.298		3.554	
Prob > Breusch-Pagan χ^2	0.069		0.059	

Notes. Standard errors in parentheses. Three participants who answered “other” to the gender question are excluded from the regressions.

[•] $p < 0.1$, ^{*} $p < 0.05$, ^{**} $p < 0.01$, ^{***} $p < 0.001$.

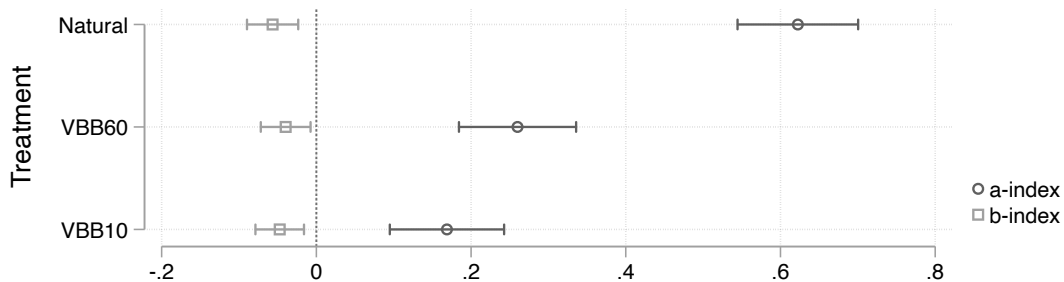


Figure 4: Predicted a -index and b -index estimates in each treatment with 95% confidence intervals (based on model 2, Table 2).

index is the same across treatments: both the Natural treatment and the VBB60 treatment display higher a -index values compared to the VBB10 treatment.

The fact that only the a -index is affected by the treatments further confirms that these two indices measure different components of ambiguity attitudes. Baillon et al. (2018) already found that only the a -index is impacted by time pressure. We find a small negative correlation between the indices, significant at the 10% level (Pearson’s $\rho = -0.090$, $p = 0.051$). However, as can be seen on Figure 5, which displays the a -index and the b -index of each subject as a scatter plot, the indices are widely scattered and lack a clear pattern.

The values of the indices we observe in our experiment are consistent with those reported previously in the literature. For example, compared to Baillon et al. (2018), we also observe a slightly negative b -index, indicating ambiguity seekingness, and a positive a -index, indicating likelihood insensitivity. von Gaudecker et al. (2022) also observe likelihood insensitivity but, using a representative online sample—the LISS panel administered by CentERdata—they observe a slightly positive b -index, indicating ambiguity aversion. Both Baillon et al. (2018) and von Gaudecker et al. (2022) also found considerable heterogeneity in ambiguity attitudes, as we already saw in Figure 5.

In Appendix B we perform a number of additional analyses. First, we report OLS regressions corresponding to the SUR reported in Table 2. We used SUR instead of OLS because the Breusch-Pagan test (also reported at the bottom of Table 2) shows that the residuals from these OLS regressions are correlated, in which case SUR are preferred. Baillon et al. (2018) made the same observation, which also led them to use SUR. The main difference between the SUR and the OLS regressions

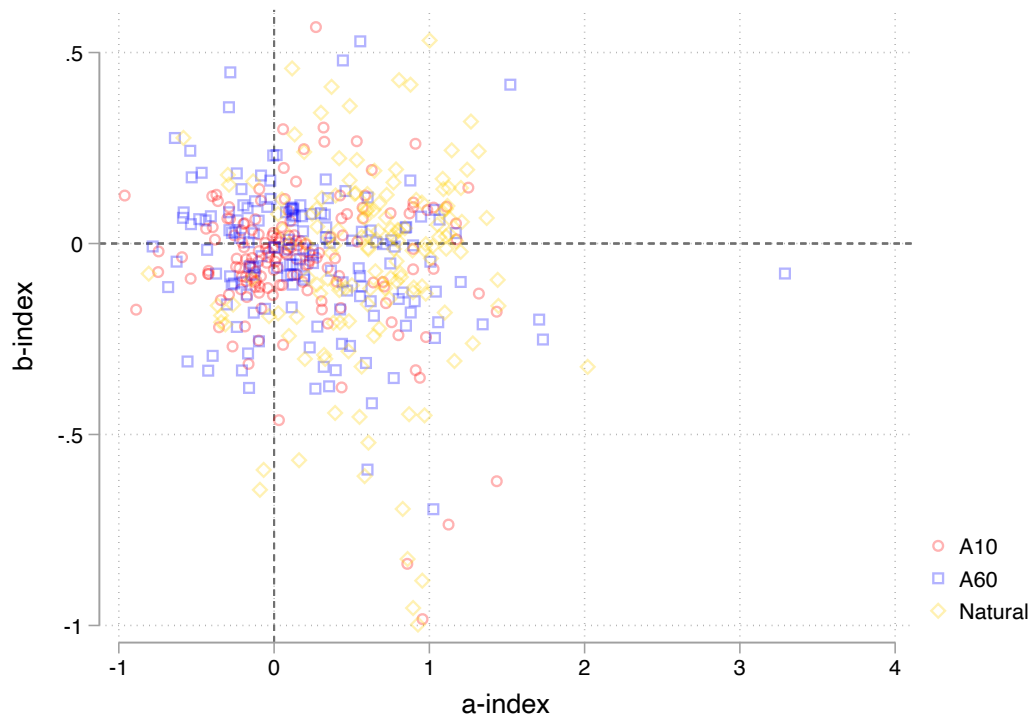


Figure 5: *a*-index and *b*-index estimates by treatment.

is that, under OLS and without controls, we cannot reject the null that the impact of VBB10 and of VBB60 on the a -index is the same (one-sided $p = 0.105$). With controls, as we found with the SUR, we reject the null (one-sided $p = 0.057$).

Second, we also report results from Mann-Whitney U tests, which we also preregistered. The test results corroborate the results from the SUR, except in the case of the a -index where we do not find a difference between VBB10 and VBB60 ($p = 0.454$). One reason for this difference may be that Mann-Whitney U tests are ordinal and thus less sensitive to the dispersion of the indices. We already saw in Figure 3 that the indices are more dispersed in VBB60 compared to VBB10. To test for this explanation, we follow the procedure from Baillon et al. (2018): we take the absolute value of the difference between each index and its treatment mean, and use this as the dependent variable in a regression with the same independent variables as before. This procedure allows us to study how dispersion around the mean varies between treatments. We also report the results from this analysis in the Appendix. We find that indeed the a -index is significantly more dispersed in VBB60 than in VBB10.

Finally, to investigate the perception of the VBB and of natural events, we classify the comments written by subjects at the end of the experiment. Most of the comments are classified as engagement, understanding, or technical issues. We do not observe stark differences across treatments; if anything, subjects find the VBB treatments to be more engaging. We give more details about the method and the results in Appendix B.5.

4 Conclusion

In conclusion, we believe that the VBB is a useful tool when conducting computerized experiments entailing risk or ambiguity. It allows experimenters to introduce transparent randomness into their experiments, which is useful not only for studies focusing on risk and ambiguity but also for any study involving random draws. For example, the VBB can be used to increase transparency in random payment procedures and in the assignment to treatments. The VBB offers an alternative to traditional real-world experimental procedures and other less credible online implementations. To validate the VBB, we conducted an experiment in the context

of ambiguity preference elicitation. We show that a simple manipulation, increasing the number of balls, increases the perceived level of ambiguity. Other more complex manipulations, such as using closely related colors or increasing the speed of the balls, could also be used to further increase the level of ambiguity.

References

- Abdellaoui, Mohammed, Aurélien Baillon, Laetitia Placido, and Peter P. Wakker (2011) “The rich domain of uncertainty: Source functions and their experimental implementation,” *American Economic Review*, 101 (2), 695–723, 10.1257/aer.101.2.695.
- Baillon, Aurélien, Zhenxing Huang, Asli Selim, and Peter P. Wakker (2018) “Measuring ambiguity attitudes for all (natural) events,” *Econometrica*, 86 (5), 1839–1858, 10.3982/ECTA14370.
- Caferra, Rocco, John D. Hey, Andrea Morone, and Marco Santorsola (2023) “Dynamic inconsistency under ambiguity: An experiment,” *Journal of Risk and Uncertainty*, 67 (3), 215–238, 10.1007/s11166-023-09418-y.
- Camerer, Colin and Martin Weber (1992) “Recent developments in modeling preferences: Uncertainty and ambiguity,” *Journal of Risk and Uncertainty*, 5 (4), 325–370, 10.1007/BF00122575.
- Chen, Daniel L., Martin Schonger, and Chris Wickens (2016) “oTree—An open-source platform for laboratory, online, and field experiments,” *Journal of Behavioral and Experimental Finance*, 9, 88–97, 10.1016/j.jbef.2015.12.001.
- Ellsberg, Daniel (1961) “Risk, ambiguity, and the Savage axioms,” *The Quarterly Journal of Economics*, 75 (4), 643–669, 10.2307/1884324.
- Garcia, Thomas, Sébastien Massoni, and Marie Claire Villeval (2020) “Ambiguity and excuse-driven behavior in charitable giving,” *European Economic Review*, 124, 103412, 10.1016/j.euroecorev.2020.103412.
- von Gaudecker, Hans-Martin, Axel Wogroly, and Christian Zimpelmann (2022) “The Distribution of Ambiguity Attitudes,” 10.2139/ssrn.4279167.
- Georgalos, Konstantinos (2019) “An experimental test of the predictive power of dynamic ambiguity models,” *Journal of Risk and Uncertainty*, 59 (1), 51–83, 10.1007/s11166-019-09311-7.

- (2021) “Dynamic decision making under ambiguity: An experimental investigation,” *Games and Economic Behavior*, 127, 28–46, 10.1016/j.geb.2021.02.002.
- Giamattei, Marcus, Kyanoush Seyed Yahosseini, Simon Gächter, and Lucas Molleman (2020) “LIONESS Lab: A free web-based platform for conducting interactive experiments online,” *Journal of the Economic Science Association*, 6 (1), 95–111, 10.1007/s40881-020-00087-0.
- Hey, John D., Gianna Lotito, and Anna Maffioletti (2010) “The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity,” *Journal of Risk and Uncertainty*, 41 (2), 81–111, 10.1007/s11166-010-9102-0.
- Hey, John D. and Noemi Pace (2014) “The explanatory and predictive power of non two-stage-probability theories of decision making under ambiguity,” *Journal of Risk and Uncertainty*, 49 (1), 1–29, 10.1007/s11166-014-9198-8.
- Kothiyal, Amit, Vitalie Spinu, and Peter P. Wakker (2014) “An experimental test of prospect theory for predicting choice under ambiguity,” *Journal of Risk and Uncertainty*, 48 (1), 1–17, 10.1007/s11166-014-9185-0.
- Levati, M. Vittoria and Andrea Morone (2013) “Voluntary contributions with risky and uncertain marginal returns: The importance of the parameter values,” *Journal of Public Economic Theory*, 15 (5), 736–744, 10.1111/jpet.12043.
- MacCrimmon, Kenneth R. and Stig Larsson (1979) “Utility theory: Axioms versus ‘paradoxes’,” in Allais, Maurice and Ole Hagen eds. *Expected Utility Hypotheses and the Allais Paradox: Contemporary Discussions of the Decisions under Uncertainty with Allais’ Rejoinder*, 333–409, Dordrecht: Springer Netherlands, 10.1007/978-94-015-7629-1_15.
- Morone, Andrea and Rocco Caferra (2024) “The Ambiguity Box: A new tool to generate ambiguity in the lab,” SSRN Scholarly Paper 4700547, Rochester, NY, <https://papers.ssrn.com/abstract=4700547>.
- Trautmann, Stefan T. and Gijs van de Kuilen (2015) “Ambiguity attitudes,” in *The Wiley Blackwell Handbook of Judgment and Decision Making*, Chap. 3, 89–116: John Wiley & Sons, Ltd, 10.1002/9781118468333.ch3.

Appendices

Appendix A Summary statistics on the sample

Table A1 describes the sample in terms of gender and age. As we can see there are subjects of varying ages present and slightly more females in the sample.

Appendix B Additional analyses

B.1 OLS regressions

We preregistered to run OLS regressions with and without controls. As mentioned in the main text, residuals from the estimation of the two indices reveal that these are correlated, thus warranting a SUR approach. For completeness, we report the results from the regressions here. Results are reported in Table B1 and corroborate the previous results. The only notable difference is that the a -index in VBB10 is not significantly different from VBB60.

B.2 Analysis of dispersion of the indices

We further study the dispersion of the indices by, following the procedure in Baillon et al. (2018), taking the absolute value of each index normalized around its treatment mean, and running a regression to see how the dispersion around the mean varies across treatments. To make it consistent with the previous regression analysis we apply the same specifications. In Table B2 we see a significant difference in dispersion between the two VBB treatments for the a - but not b -index. Clearly, the VBB10 treatment causes less dispersion than the VBB60 treatment.

B.3 Evidence of subjects counting balls in the VBB

Finally, we run SUR-regressions on the absolute distance from zero with added controls for screen performance (FPS) and how long subjects spent on the tasks (Time spent). This is meant to capture any effect of being able to count the balls due to low screen performance, rendering slow-moving balls, or taking screenshots and counting balls, rendering more time spent on each screen on the estimated distance to ambiguity neutral preferences ($a = 0$ and $b = 0$). Results are reported in Table B3. We find no such indications in the data.

Table A1: Proportion of the different demographic categories by treatment

	VBB10	VBB60	Natural
Age:			
18-27	0.22	0.26	0.21
28-37	0.29	0.30	0.28
38-47	0.16	0.21	0.22
48-57	0.21	0.18	0.16
58+	0.12	0.06	0.13
Sex:			
Male	0.41	0.45	0.38
Female	0.59	0.55	0.62
Prefer not to say	0.00	0.00	0.01
Observations	162	155	149

Table B1: Effect of treatments on a - and b -indices, OLS regressions.

	a -index	b -index	a -index	b -index
VBB60	0.073 (0.058)	0.009 (0.020)	0.091 (0.058)	0.008 (0.020)
Natural	0.449 ^{***} (0.050)	-0.011 (0.025)	0.454 ^{***} (0.050)	-0.009 (0.025)
Age category (ref.: 18-27)				
28-37			0.165 [*] (0.065)	0.008 (0.026)
38-47			0.037 (0.062)	-0.027 (0.030)
48-57			0.160 [*] (0.074)	0.011 (0.029)
58+			0.232 ^{**} (0.085)	-0.047 (0.046)
Female			0.037 (0.045)	0.014 (0.019)
Constant	0.176 ^{***} (0.036)	-0.047 ^{***} (0.014)	0.039 (0.063)	-0.050 [•] (0.030)
Observations	463	463	463	463
R^2	0.138	0.002	0.163	0.012

Notes. Robust standard errors in parentheses.“”

• $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B2: Effect of treatments on absolute value of a - and b -indices, SUR regression.

	(1)		(2)	
	$ a - mean $	$ b - mean $	$ a - mean $	$ b - mean $
VBB60	0.076 [*] (0.034)	0.019 (0.017)	0.082 [*] (0.034)	0.024 (0.016)
Natural	-0.019 (0.035)	0.074 ^{***} (0.017)	-0.016 (0.035)	0.072 ^{***} (0.017)
Age category (ref.: 18-27)				
28-37			0.004 (0.040)	-0.024 (0.019)
38-47			-0.027 (0.044)	-0.008 (0.021)
48-57			0.051 (0.044)	-0.017 (0.021)
58+			0.039 (0.053)	0.069 ^{**} (0.025)
Female			0.030 (0.029)	0.023 [•] (0.014)
Constant	0.357 ^{***} (0.024)	0.108 ^{***} (0.012)	0.327 ^{***} (0.041)	0.098 ^{***} (0.019)
Observations	463		463	
R^2	0.018	0.042	0.028	0.080
Wald χ^2	8.356	20.315	13.313	40.312
Prob > χ^2	0.015	0.000	0.065	0.000
Breusch-Pagan χ^2	7.931		7.284	
Prob > Breusch-Pagan χ^2	0.005		0.007	

Notes. Standard errors in parentheses.^{“”}

• $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B3: Effect of treatments on absolute value of a - and b -indices, SUR regression.

	(1)		(2)	
	$ a - 0 $	$ b - 0 $	$ a - 0 $	$ b - 0 $
VBB60	0.093 [*] (0.044)	0.015 (0.017)	0.113 ^{**} (0.044)	0.021 (0.017)
Natural	0.325 ^{***} (0.045)	0.067 ^{***} (0.018)	0.330 ^{***} (0.044)	0.066 ^{***} (0.017)
Age category (ref.: 18-27)				
28-37			0.079 (0.050)	-0.034 [•] (0.020)
38-47			0.020 (0.056)	-0.013 (0.022)
48-57			0.142 [*] (0.057)	-0.022 (0.022)
58+			0.162 [*] (0.068)	0.065 [*] (0.027)
Female			0.059 (0.037)	0.025 [•] (0.014)
Time taken			-0.000 (0.000)	-0.000 (0.000)
FPS			0.001 (0.001)	0.000 (0.000)
Constant	0.336 ^{***} (0.031)	0.113 ^{***} (0.012)	0.160 [*] (0.080)	0.106 ^{***} (0.031)
Observations	463		463	
R^2	0.107	0.033	0.135	0.075
Wald χ^2	55.592	15.898	72.323	37.792
Prob > χ^2	0.000	0.000	0.000	0.000
Breusch-Pagan χ^2	10.274		9.082	
Prob > Breusch-Pagan χ^2	0.001		0.003	

Notes. Standard errors in parentheses.“”

• $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B4: *b*-index: means (on diagonal) and p-values from MWU-tests (off diagonal)

	VBB10	VBB60	Natural
VBB10	-0.047		
VBB60	0.774	-0.038	
Natural	0.438	0.657	-0.053

Table B5: *a*-index: means (on diagonal) and p-values from MWU-tests (off diagonal)

	VBB10	VBB60	Natural
VBB10	0.176		
VBB60	0.454	0.249	
Natural	0.000	0.000	0.622

B.4 Mann-Whitney U tests

Following the pre-analysis plan, we apply the Mann-Whitney-U (MWU) test to compare the distributions. None of the pairwise comparisons of treatments are significant, i.e., when considering any pair of random observations from different treatments, the probability that one observation has a higher *b*-index value than the other is not different from 0.5. See Tables B4 and B5.

B.5 Comments categorization

The categorization of comments was initially carried out using a pattern-based classification approach, implemented with a machine learning model (ChatGPT) and Python.³ This approach goes beyond simple keyword spotting and takes into account contextual clues and patterns in the text. The categories were initially suggested by the language model and were subsequently manually reviewed and adjusted by the authors for accuracy. Based on this analysis, comments were assigned to one of the following categories:

- **Understanding:** Comments that mentioned confusion, difficulty in understanding, or explicitly stated incomprehension.
- **Engagement:** Comments that discussed interest, enjoyment, boredom, or difficulty in focusing on the experiment.
- **Technical Issues:** Comments that reported glitches, technical issues, or slow loading times.

³See <https://bit.ly/46gujYR> for the full set of ChatGPT prompts used for the classification.

- **Neutral/No comments:** Comments that explicitly stated no opinion or feedback.

We restrict attention to the 466 subjects used in the main analysis. A total of 87 comments were received in the experiment, out of which 19 were either neutral or had no comment. Table B6 provides a summary of the categorized comments received from participants across different treatment groups. Each category is followed by the number of comments corresponding to it, broken down by treatment type (VBB10, VBB60, Natural). To offer a glimpse into the nature of the comments, representative examples for each category are also included in the table.

Table B6: Categorization of comments

Category	VBB10	VBB60	Natural	Total	Example Comments
Engagement	10	13	6	29	“ I found it interesting! really enjoy these types of surveys.”, 11I found this explanation very confusing”, “I took a shot but was unlucky ”
Neutral/No Comment	9	8	2	19	“ no”, "No further comments", "None. ”
Technical Issues	2	3	3	8	“It was difficult to count the number of each colored ball as they were jumping all the time. There was also slight delay for the program to respond to clicks”, “It was a bit slow to load each page so I was worried it would crash. Otherwise it was fine and pretty fun.”, “Very interesting study. The page seemed to load rather slowly at times (even on pages without the balls present) but everything worked with patience. ”
Understanding	10	8	13	31	“ I had absolutely no idea what was going on despite reading the instructions several times. The description of Option B was completely baffling to me”, “Although I did my best I apologize for struggling to fully understand the instructions. Thank you for giving me the opportunity to participate.”, “confusing ”
Total	31	32	24	87	

Appendix C Instructions

When entering the experiment, subjects first saw a consent form. Then, they saw the instruction page.

C.1 Overall instructions

C.1.1 VBB10 and VBB60 treatments

Welcome! In this experiment you will be asked to make various decisions. Depending on your decisions you may earn a sizable amount of money that will be paid to you after the experiment. This will be composed of the £1 participation fee plus your 'earnings' from the outcome of your decisions.

These earnings will depend partly on the decisions you take during the experiment and partly on chance. So you will need to read these instructions carefully.

Outline of the experiment You will be asked 6 questions of the same type. In each question you will be presented with two options, Option A and Option B.

After you have answered all 6 questions, one of them will be selected at random, and the Option you chose in this question will be played for real.

Therefore, it is in your interest to answer each question as if that were the question to be played out.

The Bingo Blower You will notice a virtual Bingo Blower below. You can see there are balls of three different colors: red, blue and yellow. It is a real physics simulation and not a pre-recorded video, meaning that we have no control on how the balls move.

To convince yourself that this Bingo Blower is simulated in real-time, on this page only you can manipulate the balls by clicking on one, dragging it and seeing how the other balls react. Later on you will not be able to control the balls to prevent you from affecting the outcome.

At the end of the experiment, in order to determine a color, the Bingo Blower will be stopped and the ball closest to the center selected. Since the Bingo Blower is a real physics simulation, we cannot control the color of the ball drawn.

You can use the buttons below to see how drawing a ball will work. After you click on 'Draw a ball', the balls keep tumbling for 3 seconds. Thereafter the balls are frozen and the ball closest to the center is chosen. If you want to try drawing a ball multiple times, click on 'Reset the Blower'.

[Figure C1]

The ball drawn is: *[color of the ball drawn]*

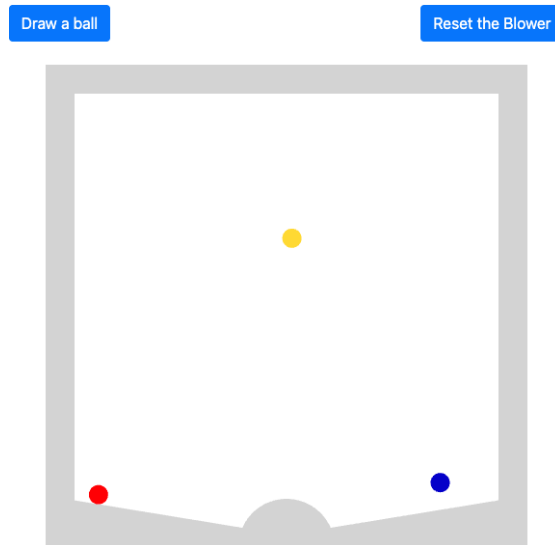


Figure C1

The questions In each question you will choose between two options, Option A and Option B:

- Option A (Bingo Blower option): you win £3 if the ball drawn from the Bingo Blower is of a certain colour, and nothing otherwise.
- Option B (given chance to win option): you have some chance to win £3.

You will be asked to state which one of these two options you prefer for various chances to win in Option B, from 0% to 100%. For example:

- If the chance to win in Option B is 100%, you will most likely prefer Option B because you are then guaranteed to win £3.
- If the chance to win in Option B is 0%, you will most likely prefer Option A because you might win something, as opposed to Option B where there is no chance to win at all.

You can think of the chance to win as the range of winning numbers in a bag containing numbers between 1 and 100. If the chance to win is 50%, it means you would win if a number between 1 and 50 is drawn from the bag, and you would lose if a number between 51 and 100 is drawn.

Example question The questions will look like this:

[Figure C2]

In this example, Option A would lead to a gain of £3 if the ball drawn from the Bingo Blower is red or yellow. Option B is presented with ascending chances to win. For each line we ask you to choose between Option A and Option B.

If you prefer Option A when, for instance, Option B offers a 40% chance to win, then you should also prefer Option A when Option B offers an even lower



Option A You win £3 if one of the following balls is drawn:	Your choice	Option B You win £3 with the following chance:
 	A B	0 %
	A B	1 %
	A B	2 %
	A B	5 %
	A B	10 %
	A B	15 %
	A B	20 %
	A B	25 %
	A B	30 %
	A B	35 %
	A B	40 %
	A B	45 %
	A B	50 %
	A B	55 %
	A B	60 %
	A B	65 %
	A B	70 %
	A B	75 %
	A B	85 %
	A B	100 %

Figure C2

chance to win, say 25% . Similarly, if you prefer Option B with a 60% chance to win, then you should also prefer Option B with an even better chance to win, say 70%. Accordingly, in the figure above Option A is selected for every chance to win between 0 to 45%, and Option B, for every chance to win between 50% and 100%.

To save you the trouble of clicking for each line, you can click on Option A or Option B for one line and the computer will automatically fill your choices for the other lines. For example, in the figure above, button ‘B’ has been clicked only for the line corresponding to 50%.

You can change your choices as many times as you want. Once you are happy with your choices, click the ‘Next’ button at the bottom. You will not be allowed to go back to a question once you have clicked on the ‘Next’ button.

Payment After you have answered all 6 questions, one of them will be randomly selected by the computer. This question will be the question that counts for your payment. Then the computer will also select one of the lines:

- If, on this line, you had selected Option A, you will win £3 if the colour of the ball drawn from the Bingo Blower matches one of the colours displayed in Option A.
- If, instead, you had selected Option B, you will play that option. To do so the computer will randomly draw a number between 0 and 100. For example if Option B is ‘You win £3 with a 74% chance’ then you will win £3 if the number drawn by the computer is strictly smaller than 74, and you will win nothing if it is equal to or larger than 75.

Option A You win £3 if the Dow evolves in one of the following ways:	Your choice	Option B You win £3 with the following chance:
<ul style="list-style-type: none"> The Dow decreases by more than 0.51% The Dow increases by more than 0.11% 	A B	0 %
	A B	1 %
	A B	2 %
	A B	5 %
	A B	10 %
	A B	15 %
	A B	20 %
	A B	25 %
	A B	30 %
	A B	35 %
	A B	40 %
	A B	45 %
	A B	50 %
	A B	55 %
	A B	60 %
	A B	65 %
	A B	70 %
A B	75 %	
A B	85 %	
A B	100 %	

Figure C3

C.1.2 Natural treatment

[Identical to the instructions for VBB10 and VBB60, except:]

The questions In each question you will choose between two options, Option A and Option B:

- Option A (Dow option): you win £3 if the Dow (Dow Jones Industrial Average Index) increases or decreases by some percentages between the opening and closing price on *[date about a week after the experiment]*, and nothing otherwise.
- Option B (given chance to win option): you have some chance to win £3.

[Remaining of section ‘The questions’ identical]

Example question The questions will look like this:

[Figure C3]

In this example, Option A would lead to a gain of £3 if the Dow decreases by more than 0.51% or increases by more than 0.11%. Option B is presented with ascending chances to win. For each line we ask you to choose between Option A and Option B.

[Remaining of section ‘Example question’ identical]

Payment After you have answered all 6 questions, one of them will be randomly selected by the computer. This question will be the question that counts for your



Option A You win £3 if one of the following balls is drawn:	Your choice	Option B You win £3 with the following chance:
 	<input type="checkbox"/> A <input type="checkbox"/> B	0 %
	<input type="checkbox"/> A <input type="checkbox"/> B	1 %
	<input type="checkbox"/> A <input type="checkbox"/> B	2 %
	<input type="checkbox"/> A <input type="checkbox"/> B	5 %
	<input type="checkbox"/> A <input type="checkbox"/> B	10 %
	<input type="checkbox"/> A <input type="checkbox"/> B	15 %
	<input type="checkbox"/> A <input type="checkbox"/> B	20 %
	<input type="checkbox"/> A <input type="checkbox"/> B	25 %
	<input type="checkbox"/> A <input type="checkbox"/> B	30 %
	<input type="checkbox"/> A <input type="checkbox"/> B	35 %
	<input type="checkbox"/> A <input type="checkbox"/> B	40 %
	<input type="checkbox"/> A <input type="checkbox"/> B	45 %
	<input type="checkbox"/> A <input type="checkbox"/> B	50 %
	<input type="checkbox"/> A <input type="checkbox"/> B	55 %
	<input type="checkbox"/> A <input type="checkbox"/> B	60 %
	<input type="checkbox"/> A <input type="checkbox"/> B	65 %
	<input type="checkbox"/> A <input type="checkbox"/> B	70 %
	<input type="checkbox"/> A <input type="checkbox"/> B	75 %
	<input type="checkbox"/> A <input type="checkbox"/> B	85 %
	<input type="checkbox"/> A <input type="checkbox"/> B	100 %

Figure C4

payment. Then the computer will also select one of the lines:

- If, on this line, you had selected Option A, you will win £3 if the evolution of the Dow matches one of the descriptions displayed in Option A. We will use historical data from Yahoo Finance (available by clicking on this link: <https://finance.yahoo.com/quote/%5EDJI/history/>): the opening price is in the column 'Open', and the closing price in the column 'Close*').
- If, instead, you had selected Option B, you will play that option. To do so the computer will randomly draw a number between 0 and 100. For example if Option B is 'You win £3 with a 74% chance' then you will win £3 if the number drawn by the computer is strictly smaller than 74, and you will win nothing if it is equal to or larger than 75.

C.2 Control questions

C.2.1 VBB10 and VBB60 treatments

To test your understanding we have designed the following control questions.

Imagine one of the questions is as follows:

[Figure C4]

Imagine further that this question is selected for payment at the end of the experiment.

Imagine that the line corresponding to a chance to win of 10% is selected. If your choices are as described in the example above, which option would you play at the end of the experiment? *[Single choice]*

- Option A
- Option B

What happens if a red ball is drawn? *[Single choice]*

- I gain £0
- I gain £3
- It is irrelevant: how much I gain depends on the random number selected by the computer to play Option B.

Imagine that the line corresponding to a chance to win of 60% is selected. If your choices are as described in the example above, which option would you play at the end of the experiment? *[Single choice]*

- Option A
- Option B

What happens if a red ball is drawn? *[Single choice]*

- I gain £0
- I gain £3
- It is irrelevant: how much I gain depends on the random number selected by the computer to play Option B.

C.2.2 Natural treatments

To test your understanding we have designed the following control questions.

Imagine one of the questions is as follows:

[Figure C5]

Option A You win £3 if the Dow evolves in one of the following ways:	Your choice	Option B You win £3 with the following chance:
<ul style="list-style-type: none"> • The Dow decreases by more than 0.51% • The Dow increases by more than 0.11% 	A B	0 %
	A B	1 %
	A B	2 %
	A B	5 %
	A B	10 %
	A B	15 %
	A B	20 %
	A B	25 %
	A B	30 %
	A B	35 %
	A B	40 %
	A B	45 %
	A B	50 %
	A B	55 %
	A B	60 %
	A B	65 %
	A B	70 %
	A B	75 %
	A B	85 %
	A B	100 %

Figure C5

Imagine further that this question is selected for payment at the end of the experiment.

Imagine that the line corresponding to a chance to win of 10% is selected. If your choices are as described in the example above, which option would you play at the end of the experiment? *[Single choice]*

- Option A
- Option B

What happens if the Dow decreases by more than 0.51%? *[Single choice]*

- I gain £0
- I gain £3
- It is irrelevant: how much I gain depends on the random number selected by the computer to play Option B.

Imagine that the line corresponding to a chance to win of 60% is selected. If your choices are as described in the example above, which option would you play at the end of the experiment? *[Single choice]*

- Option A
- Option B

What happens if the Dow decreases by more than 0.51%? *[Single choice]*

- I gain £0
- I gain £3
- It is irrelevant: how much I gain depends on the random number selected by the computer to play Option B.

C.3 Tasks

Subjects saw 6 questions, randomized independently for each subject.

C.3.1 VBB10 treatment

Table C6 gives an example of a task for a simple event, and Table C7, for a composite event

C.3.2 VBB60 treatment

Tasks in the VBB60 treatments were identical to those in the VBB10 treatment except that that the virtual bingo blower contained 60 balls.

C.3.3 Natural treatment

Table C8 gives an example of a task for a simple event, and Table C9, for a composite event

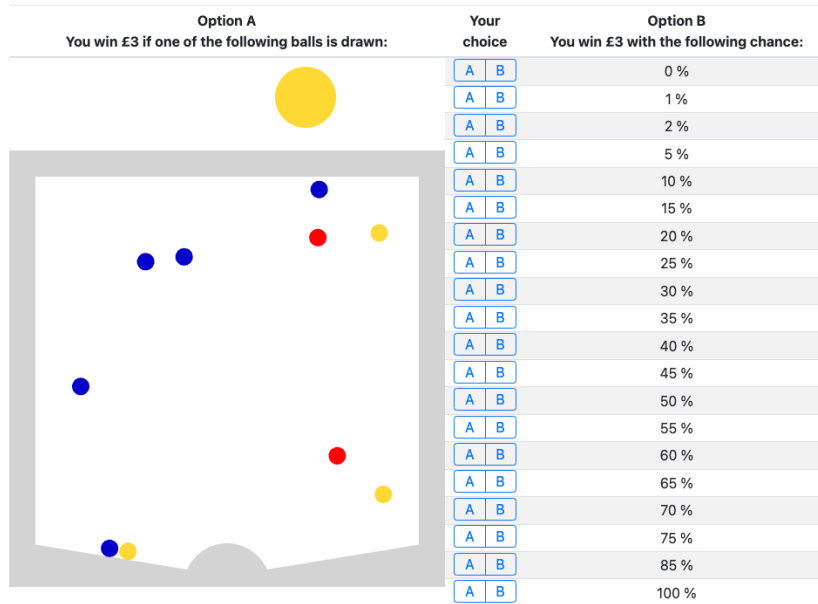


Figure C6: Example of a choice task in VBB10, simple event.

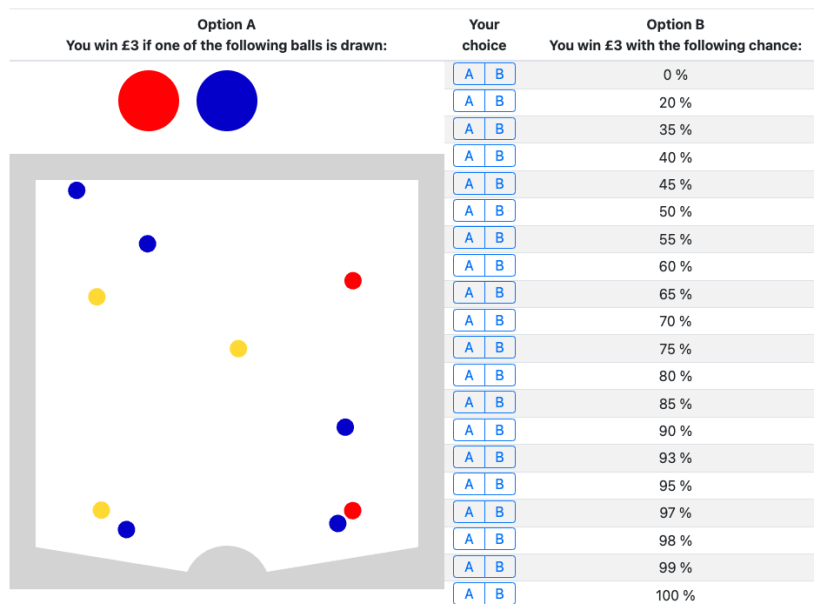


Figure C7: Example of a choice task in VBB10, composite event.

Option A You win £3 if the Dow evolves in one of the following ways:	Your choice	Option B You win £3 with the following chance:
<ul style="list-style-type: none"> The Dow decreases by more than 0.51% The Dow decreases by less than 0.51% or increases by less than 0.11% 	<input type="checkbox"/> A <input type="checkbox"/> B	0 %
	<input type="checkbox"/> A <input type="checkbox"/> B	20 %
	<input type="checkbox"/> A <input type="checkbox"/> B	35 %
	<input type="checkbox"/> A <input type="checkbox"/> B	40 %
	<input type="checkbox"/> A <input type="checkbox"/> B	45 %
	<input type="checkbox"/> A <input type="checkbox"/> B	50 %
	<input type="checkbox"/> A <input type="checkbox"/> B	55 %
	<input type="checkbox"/> A <input type="checkbox"/> B	60 %
	<input type="checkbox"/> A <input type="checkbox"/> B	65 %
	<input type="checkbox"/> A <input type="checkbox"/> B	70 %
	<input type="checkbox"/> A <input type="checkbox"/> B	75 %
	<input type="checkbox"/> A <input type="checkbox"/> B	80 %
	<input type="checkbox"/> A <input type="checkbox"/> B	85 %
	<input type="checkbox"/> A <input type="checkbox"/> B	90 %
	<input type="checkbox"/> A <input type="checkbox"/> B	93 %
	<input type="checkbox"/> A <input type="checkbox"/> B	95 %
	<input type="checkbox"/> A <input type="checkbox"/> B	97 %
	<input type="checkbox"/> A <input type="checkbox"/> B	98 %
	<input type="checkbox"/> A <input type="checkbox"/> B	99 %
	<input type="checkbox"/> A <input type="checkbox"/> B	100 %

Figure C8: Example of a choice task in Natural, simple event.

Option A You win £3 if the Dow evolves in one of the following ways:	Your choice	Option B You win £3 with the following chance:
<ul style="list-style-type: none"> The Dow decreases by less than 0.51% or increases by less than 0.11% 	<input type="checkbox"/> A <input type="checkbox"/> B	0 %
	<input type="checkbox"/> A <input type="checkbox"/> B	1 %
	<input type="checkbox"/> A <input type="checkbox"/> B	2 %
	<input type="checkbox"/> A <input type="checkbox"/> B	5 %
	<input type="checkbox"/> A <input type="checkbox"/> B	10 %
	<input type="checkbox"/> A <input type="checkbox"/> B	15 %
	<input type="checkbox"/> A <input type="checkbox"/> B	20 %
	<input type="checkbox"/> A <input type="checkbox"/> B	25 %
	<input type="checkbox"/> A <input type="checkbox"/> B	30 %
	<input type="checkbox"/> A <input type="checkbox"/> B	35 %
	<input type="checkbox"/> A <input type="checkbox"/> B	40 %
	<input type="checkbox"/> A <input type="checkbox"/> B	45 %
	<input type="checkbox"/> A <input type="checkbox"/> B	50 %
	<input type="checkbox"/> A <input type="checkbox"/> B	55 %
	<input type="checkbox"/> A <input type="checkbox"/> B	60 %
	<input type="checkbox"/> A <input type="checkbox"/> B	65 %
	<input type="checkbox"/> A <input type="checkbox"/> B	70 %
	<input type="checkbox"/> A <input type="checkbox"/> B	75 %
	<input type="checkbox"/> A <input type="checkbox"/> B	85 %
	<input type="checkbox"/> A <input type="checkbox"/> B	100 %

Figure C9: Example of a choice task in Natural, composite event.