MEASURING AMBIGUITY ATTITUDES FOR ALL
(NATURAL) EVENTS

AURÉLIEN BAILLON\textsuperscript{a}, ZHENXING HUANG\textsuperscript{b}, ASLI SELIM\textsuperscript{a}, & PETER P. WAKKER\textsuperscript{a}

a: Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands, Baillon@ese.eur.nl, Wakker@ese.eur.nl; asliselim@gmail.com
b: School of Economics, Shanghai University of Finance and Economics, and Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai, P.R. China, ybhzxadam@gmail.com

November, 2017

Measurements of ambiguity attitudes have so far focused on artificial events, where (subjective) beliefs can be derived from symmetry of events and then can be controlled for. For natural events as relevant in applications, such a symmetry and corresponding control are usually absent, precluding traditional measurement methods. This paper introduces two indexes of ambiguity attitudes, one for aversion and the other for insensitivity/perception, for which we can control for likelihood beliefs even if those are unknown. Hence, we can now measure ambiguity attitudes for natural events. Our indexes are valid under many ambiguity theories, do not require expected utility for risk, and are easy to elicit in practice. We use our indexes to investigate time pressure under ambiguity. People do not become more ambiguity averse under time pressure, but become more insensitive (perceive more ambiguity). These findings support the validity of our indexes.

JEL-CLASSIFICATION: D81, C91

KEYWORDS: ambiguity aversion; Ellsberg paradox; sources of uncertainty; time pressure
1. INTRODUCTION

Ambiguity (unknown probabilities) is central in many practical decisions (Keynes 1921; Knight 1921). Ellsberg’s paradox (1961) showed that fundamentally new models are needed to handle ambiguity. Gilboa (1987), Gilboa and Schmeidler (1989), and Schmeidler (1989) introduced such new models, with many to follow.\(^1\) Ambiguity theories are now widely applied (Easley and O’Hara 2009; Guidolin and Rinaldi 2013; Shaw 2016). However, measurements of ambiguity have been lagging behind, usually employing artificial laboratory events as in Ellsberg’s paradox rather than the natural events that occur in practice.

To properly measure ambiguity aversion we need to control for likelihood beliefs in the events of interest, so as to calibrate the benchmark of ambiguity neutrality. But this control is difficult to implement for natural events. To illustrate this point, consider someone preferring to receive $100 under the ambiguous event \(A\) of an increase in copper price of at least 0.01% next week rather than receiving $100 under the risky event \(K\) of heads in a coin toss \((p = 0.5)\) next week. This preference need not reflect ambiguity seeking. The person may be ambiguity neutral but assign a higher likelihood belief to \(A\) than to \(K\). Without proper control of likelihoods, we therefore cannot know people’s ambiguity attitudes. However, how to control for likelihoods using revealed preferences in a tractable manner has as yet been unknown for naturally occurring events.

Controlling for likelihoods is easy for artificial events generated in the lab. Such events concern Ellsberg urns with color compositions kept secret from the subjects, or subjects only being informed about experimenter-specified probability intervals. Then likelihood beliefs can be derived from symmetry of colors or from symmetry about the midpoints of intervals. This explains why measurements of ambiguity have as yet focused on artificial events.

Several authors warned against the focus on artificial ambiguities, arguing for the importance of natural events (Camerer and Weber 1992 p. 361; Ellsberg 2011 p. 223; Theoretical surveys include Etner, Jeleva, and Tallon (2012), Gilboa and Marinacci (2016), Machina and Siniscalchi (2014), and Marinacci (2015).

\(^1\) Theoretical surveys include Etner, Jeleva, and Tallon (2012), Gilboa and Marinacci (2016), Machina and Siniscalchi (2014), and Marinacci (2015).
difficulty to identify (revealed preference based) likelihood beliefs of such events has as yet been a problematic obstacle though. This paper introduces a simple method to measure ambiguity attitudes for natural events. The solution to the aforementioned problem is surprisingly easy: we control for likelihoods not by directly measuring them but by making them drop from the equations irrespective of what they are. Our method is tractable and easy to implement, as we demonstrate in an experiment. Hence, it can for instance be easily used as an add-on in large-scale surveys and field studies. Using natural events will increase external validity (Camerer and Weber 1992 p. 361).

Empirical studies, discussed later, have shown that ambiguity is a rich phenomenon. Hence, two indexes are needed to capture ambiguity descriptively. The first measures the well-known aversion to ambiguity and is often taken to be normative. The second captures the degree of ambiguity, i.e., the perceived level of ambiguity. Dimmock et al. (2015) therefore called their version of this index perceived level of ambiguity. The higher this level is, the less the decision maker discriminates between different degrees of likelihood, and the more these degrees are treated alike, as one blur. The second index thus also captures insensitivity toward likelihood changes, which is why we use the term a(mbiguity generated)-insensitivity. Empirical studies have found that (uncorrected) ambiguity aversion is likelihood dependent, even with prevailing ambiguity seeking, rather than aversion, for unlikely gain-events (Trautmann and van de Kuilen 2015). That is, ambiguity aversion even predicts in the wrong direction for such events. This illustrates the desirability to use the second index to correct for this likelihood dependence.

To summarize, relative to their predecessors, our indexes: (a) correct for subjective likelihoods also if unknown and, hence, which is our main novelty: (b) can be used for all events, both artificial and natural; (c) correct for likelihood dependence of ambiguity aversion. Further, as discussed later, our indexes (d) are directly observable; (e) are valid under many ambiguity theories, unifying preceding indexes; in particular, they (f) retain validity if expected utility for risk is violated. Our paper also shows that the ambiguity aversion index can better be related to matching probabilities than to nonadditive weighting functions as done before (Dow and Werlang 1992); see §5.
Our indexes will be defined for three-fold partitions. The follow-up paper Baillon, Li, and Wakker (2017), a theoretical counterpart to this paper, provides a theoretical foundation of our indexes, and generalizations to general partitions. It shows that our indexes are valid for a large number of ambiguity theories, being all that evaluate prospects \( \gamma_E 0 \) (yielding one nonzero outcome, \( \gamma \), under event \( E \), and 0 otherwise) by a product \( W(E)U(\gamma) \).

This implies, in particular, that we assume only one utility function \( U \), for risk and all sources of uncertainty. Here we deviate from utility-based models of ambiguity, such as the popular smooth model (Klibanoff, Marinacci, and Mukerji 2005). Using tools of Izhakian and Brenner (2011), Baillon, Li, and Wakker (2017) shows that our indexes are nevertheless also useful under the smooth model. Our indexes thus unify and generalize many existing indexes (point (e) above). Baillon, Li, and Wakker (2017) also shows that our two indexes capture orthogonal, i.e., completely distinct, components of the data. This mathematical separation homeomorphically supports the psychological interpretation of the indexes as distinct components (motivational versus cognitive) and the empirical desirability to consider both. This paper will support these points empirically.

---

2 As yet, indexes proposed in the literature concerned only one ambiguity theory. In the same way as no risk aversion index can be valid for all risk theories, our ambiguity indexes cannot be valid for all ambiguity theories. Yet, they are for many. Such theories include biseparable utility (Ghirardato and Marinac 2002), which in turn includes Choquet expected utility (Schmeidler 1989), prospect theory for gains (Tversky and Kahneman 1992), and the \( \alpha \)-maxmin model (Ghirardato, Maccheroni, and Marinacci 2004; Jaffray 1994; our Eq. 5.1 shows how to derive \( \alpha \) from our indexes). Some nonbiseparable theories that are included: separate-outcome weighting theories \( x_E y \rightarrow W(E)U(x) + W(E^c)U(y) \); Einhorn and Hogarth 1985), Chateauneuf and Faro’s (2009) confidence representation if the worst outcome is 0, and Lehrer and Teper’s (2015) event-separable representation.


4 That is, our indexes provide local ambiguity premiums in probability units that are analogous to Pratt’s (1964 Eq. 5) local risk premium, which was in money units. This result also holds for a subclass of Maccheroni, Marinacci, and Rustichini’s (2006) variational model.

5 That is, not only observable results but also underlying processes in the model agree with real processes.
Because a-insensitivity has been less known than ambiguity aversion, and different interpretations are possible for our insensitivity index, we test how our two indexes react to cognitive manipulations. For this purpose, we use time pressure (TP). TP has received special attention in the psychological literature because it provides a good context for manipulating cognitive limitations, in addition to its practical relevance (Ariely and Zakay 2001; Essl and Jaussi 2017). De Paola and Gioia (2015) and Spiliopoulos and Ortmann (2017) argued for the usefulness of TP and, relatedly, response time, as a tool in experimental economics and for its relevance in economic applications. Despite the many studies of TP under risk (known probabilities; references are in §4.4) there have not yet been studies of TP under ambiguity. Providing the first such study is an additional contribution of our paper. Our findings corroborate the interpretation of the indexes, supporting the validity of our method. In particular, they illustrate the usefulness of our second index.

The outline of this paper is as follows. Section 2 gives formal definitions of our ambiguity indexes and informal arguments for their plausibility. We present the indexes without assuming any decision theory so that empirically oriented readers can readily use them with no need to study such a theory. This also shows that the indexes have intuitive appeal without requiring a commitment to one of the many ambiguity theories popular today. Sections 3-4 demonstrate the validity of our indexes empirically, and Sections 5-6 discuss and conclude. Experimental details are in the appendix, with further details in an Online Appendix.

2. MEASURING AMBIGUITY ATTITUDES WITHOUT MEASURING SUBJECTIVE LIKELIHOODS: DEFINITIONS OF OUR INDEXES

We focus on gain outcomes throughout this paper. Formally speaking, ambiguity does not concern just a single event $E$, but a partition, such as $\{E, E^c\}$, or, more generally, a source of uncertainty. We assume a minimal degree of richness of the sources of uncertainty considered: there should be three mutually exclusive and exhaustive nonnull events $E_1, E_2, \text{ and } E_3$. In many situations where we start from a partition with two events we can extend it by properly partitioning one of those two
events. For example, in the two-color Ellsberg urn we can involve other features of
the ball to be drawn, such as shades of colors or numbers on the balls. In our
experiment, the events refer to the AEX stock index. For instance, in Part 1 of the
experiment, \( E_1 = (-\infty, -0.2), E_2 = [-0.2, 0.2], \) and \( E_3 = (0.2, \infty) \), where intervals
describe percentage increases of the AEX index during the experiment. Thus, they
cornor natural events with uncertainty that really occurred and that was of practical
relevance to financial traders. \( E_{ij} \) denotes the union \( E_i \cup E_j \), where \( i \neq j \) is implicit.
We call every \( E_i \) a \textit{single event} and every \( E_{ij} \) a \textit{composite event}.

Dimmock, Kouwenberg, and Wakker (2016, Theorem 3.1) showed that matching
probabilities are convenient for measuring ambiguity attitudes. Early applications
include Kahn and Sarin (1988) and Viscusi and Magat (1992). They are a central
theoretical tool in Izhakian’s (2017) ambiguity theory. Karni (2009) discussed their
empirical elicitation. Matching probabilities entirely capture ambiguity attitudes, free
from any complications regarding risk attitudes, because those drop from the
equations and need not be measured. In particular, our measurements are not affected
by the large heterogeneity of risk attitudes (Bruhin, Fehr-Duda, and Epper 2010). We
will, therefore, use matching probabilities. An admitted drawback of matching
probabilities is that their assessment is cognitively harder than, for instance, of
certainty equivalents (Bleichrodt, Pinto, and Wakker 2001 p. 1505; Callen et al. 2013
p. 136; Halter and Beringer 1960 p. 124)—a pro is that they can be measured for all
kinds of outcomes, also if nonquantitative. For any fixed prize (€20 in our
experiment), we define the matching probability \( m \) of event \( E \) through the following
indifference:

Receiving €20 under event \( E \) is equivalent to receiving €20 with probability \( m \). (2.1)

In both cases it is understood that the complementary payoff is nil. Under ambiguity
neutrality, the matching probability of an event, say \( m(E_i) \), and its complement,
\( m(E_{23}) \), will add to 1, but under ambiguity aversion the sum will fall below 1. Its
difference with 1 can then be taken as the degree of aversion. We will take the
average of this difference over the three events. We write \( m_i = m(E_i), m_{ij} =
m(E_{ij}), \bar{m}_s = (m_1 + m_2 + m_3)/3 \) for the average single-event matching probability,
\( \bar{m}_c = (m_{12} + m_{13} + m_{23})/3 \) for the average composite-event matching probability,
and define:
DEFINITION 2.1. The ambiguity aversion index is

\[
b = 1 - \overline{m_c} - \overline{m_s}.
\]  

(2.2)

Note that no statistical claims or randomness assumptions are made at this stage in this definition. We use a deterministic calculation here, recoding direct observations. Under ambiguity neutrality\(^6\), \(m_i = P(E_i)\) and \(m_{ij} = P(E_i) + P(E_j)\) for some additive subjective probability measure \(P\). Then \(\overline{m_c} = 1/3\) and \(\overline{m_c} = 2/3\), implying \(b = 0\). We have thus calibrated ambiguity neutrality, providing control for subjective likelihoods even though we do not know them. This happens because the subjective likelihoods drop from the equations irrespective of what they are. This observation is key to our method. Maximal ambiguity aversion occurs for \(b = 1\). Then the matching probabilities of all events are 0. Ambiguity aversion is minimal for \(b = -1\), when matching probabilities for all events are 1.

The ambiguity aversion index can also be defined if we only consider a two-event partition. We can focus on only one event \(E_i\) and its complement \(E_i^c\), and substitute \(m(E_i)\) for \(\overline{m_s}\) and \(m(E_i^c)\) for \(\overline{m_c}\) in Eq. 2.2, maintaining the control for likelihood. This would reduce the measurement effort—at the cost of reliability. For the insensitivity index defined next we, to the contrary, essentially need three events.

Theoretically, the second index captures the extent to which matching probabilities and event weights regress towards fifty-fifty, with low likelihoods overvalued and high likelihoods undervalued. This leads to reduced differences \(\overline{m_c} - \overline{m_s}\). In the most extreme case of complete ambiguity and, correspondingly, complete insensitivity (Cohen and Jaffray 1980), no distinction at all is made between different levels of likelihood (e.g. all events are taken as fifty-fifty), resulting in \(\overline{m_c} - \overline{m_s} = 0\). These observations suggest that the second index can be interpreted as a cognitive component (Budescu et al. 2014 p. 3; Dimmock et al. 2015; Dimmock, 2016).

\(^6\) When objective probabilities are assumed present in the domain considered, as is our case, then ambiguity neutrality is equivalent to probabilistic sophistication (Dean and Ortoleva 2017 p. 393 footnote 1). If no objective probabilities are present, such as when only considering the unknown Ellsberg two-color urn, then probabilistic sophistication is strictly more general—but then matching probabilities, and our indexes, cannot even be defined.
Kouwenberg, and Wakker 2016; Einhorn and Hogarth 1985; Gayer 2010), an interpretation well supported by our results. For this index, the following rescaling of \( \overline{m_c} - \overline{m_s} \) is convenient.

**Definition 2.2.** The ambiguity-generated insensitivity (a-insensitivity) index\(^7\) is

\[
\alpha = 3 \times \left( \frac{1}{3} - (\overline{m_c} - \overline{m_s}) \right).
\]  

(2.3)

Under ambiguity neutrality, with perfect discrimination between single and composite events, or under absence of ambiguity, \( \overline{m_c} = 2/3 \) and \( \overline{m_s} = 1/3 \), and their difference is 1/3. Index \( \alpha \) measures how much the actual difference falls short of 1/3. We multiplied by 3 to obtain a convenient normalization with a maximal value 1 (maximal insensitivity, with \( \overline{m_c} = \overline{m_s} \)).

Ambiguity neutrality gives \( \alpha = 0 \). We have again calibrated ambiguity neutrality here, controlling for subjective likelihoods by letting them drop from the equations. Empirically, we usually find prevailing insensitivity, \( \alpha > 0 \), but there are subjects with \( \alpha < 0 \). Hence it is desirable for descriptive purposes to allow \( \alpha < 0 \), which we do. The \( \alpha \)-maxmin model, however, does not allow \( \alpha < 0 \) (Baillon, Li, and Wakker 2017), which is no problem for normative applications that take \( \alpha < 0 \) to be irrational.

Our two indexes are orthogonal (Baillon, Li, and Wakker 2017). If one is 0, suggesting ambiguity neutrality, the other may still deviate from 0, showing ambiguity attitude. In particular, contrary to what has sometimes been suggested in the literature, if there is no aversion to ambiguity, then ambiguity may still play an important role through insensitivity.\(^8\)

Dimmock et al. (2015) referred to their version of the second index as perceived level of ambiguity. Dimmock et al.’s term, and the multiple priors model underlying it, may serve best for applications that, unlike this paper, have normative aims. Their assumption of expected utility for risk, and their restriction \( \alpha \geq 0 \), are problematic for descriptive applications though. For risk, insensitivity (i.e., inverse-S probability weighting) has been commonly found (Fehr-Duda and Epper 2012; Wakker 2010

---

\(^7\) Under multiple prior theories, this index can be called “perceived level of ambiguity.”

\(^8\) For instance, in Schmeidler (1989), with \( W \) denoting the weighting function, \( W(E) + W(E^c) = 1 \) may hold for all events \( E \), while there may still be strong insensitivity.
§9.5). Our second index naturally extends this insensitivity found under risk to ambiguity, where empirical studies have found that it is usually amplified (Trautmann and van de Kuilen 2015; Wakker 2010 p. 292). Hence, we use the term ambiguity-generated insensitivity (a-insensitivity) to refer to it. Insensitivity was central in the early Einhorn and Hogarth (1985). Gonzalez and Wu (1999) gave an illuminating discussion of its psychological interpretations, for risk. The textbook Wakker (2010, §10.4.2 and §11.8) presents the concept for ambiguity.

3. EXPERIMENT: METHOD

This section presents the experiment. Appendix A gives further details. We investigate the effect of time pressure (TP) on ambiguity. The ambiguity concerns the performance of the AEX (Amsterdam stock exchange) index. Using our method, we can study TP for natural events.

Hypotheses

It is natural to expect that TP will reduce cognitive understanding and, hence, increase the insensitivity index. This is the hypothesis we test. We had no prior prediction about the impact of TP on ambiguity aversion. Ambiguity aversion reflects how much more there is dislike for uncertainty than for risk. We saw no reason for this difference to become bigger or smaller.

Subjects

N = 104 subjects participated (56 male, median age 20). They were all students from Erasmus University Rotterdam, recruited from a pool of volunteers. They were randomly allocated to the control and the TP treatment.

The experiment consisted of two parts, Parts 1 and 2 (Table 1), consisting of eight questions each. They were preceded by a training part (Part 0) of eight questions, to familiarize subjects with the stimuli. All subjects faced the same questions, except that subjects in the TP treatment had to make their choices in Part 1 under time pressure.
There were 42 subjects in the control treatment and 62 in the TP treatment. The TP sample had more subjects because we expected more variance there.

**TABLE 1: Organization of the experiment**

<table>
<thead>
<tr>
<th>Stimuli: Within- and between-subject treatments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control treatment</td>
</tr>
<tr>
<td>Within subject</td>
</tr>
<tr>
<td>No time pressure</td>
</tr>
<tr>
<td>Part 1</td>
</tr>
<tr>
<td>Time pressure treatment</td>
</tr>
<tr>
<td>Time pressure</td>
</tr>
<tr>
<td>Part 2</td>
</tr>
<tr>
<td>No time pressure</td>
</tr>
</tbody>
</table>

**Stimuli: Choice lists**

In each question, subjects were asked to choose between two options.

**OPTION 1:** You win €20 if the AEX index increases/decreases by … between the beginning and the end of the experiment (which lasted 25 minutes on average), and nothing otherwise.

**OPTION 2:** You win €20 with p% probability and nothing otherwise.

We used choice lists to infer the probability p in Option 2 that leads to indifference between the two options; see the Appendix for details. This p is the matching probability of the AEX event. For the TP treatment, a 25-second time limit was set for each choice in Part 1.

**Stimuli: Uncertain events**

In each part we consider a triple of mutually exclusive and exhaustive single events and their compositions; see Table 2.

**TABLE 2: Single AEX-change events for different parts (unit is percentage)**

<table>
<thead>
<tr>
<th>Event E₁</th>
<th>Event E₂</th>
<th>Event E₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part 1</td>
<td>(−∞,−0.2)</td>
<td>[−0.2,0.2]</td>
</tr>
<tr>
<td>Part 2</td>
<td>(−∞,−0.1)</td>
<td>[−0.1,0.3]</td>
</tr>
</tbody>
</table>

*In the training Part 0, the events were (−∞,−0.4), [−0.4,0.1], and (0.1,∞).*
We chose the partition in Part 2 somewhat differently to make subjects choose afresh rather than erroneously speculate on relations between the questions. For each part, we measured matching probabilities of all six single and composite events, of which two were repeated to test consistency. The order of the eight questions was randomized for each subject within each part.

Stimuli: Further questions
At the end of the experiment, subjects were asked to report their age, gender, and nationality.

Incentives
We used the random incentive system. All subjects received a show-up fee of €5 and one of their choices was randomly selected to be played for real; see the Appendix for details. This implies, as usual, that the truth of the events must be observable for payment, so that uncertainties in a far future, for instance, cannot be incentivized this way.

Analysis
We computed ambiguity aversion and a-insensitivity indexes as explained in §2. Five subjects in the TP treatment did not submit one of their matching probabilities on time and were therefore excluded from the analysis, leaving us with 99 subjects.

We ran OLS regressions to study the impact of TP on a-insensitivity and ambiguity aversion. Because we obtain two values of each index per subject (one for each part), we clustered standard errors at the individual level. Furthermore, because the residuals of the regression on $a$ were correlated with the residuals of the regression on $b$, we used Seemingly Unrelated Regressions. In the baseline model (Model 1 in the result tables), we take Part 1 in the control treatment as the reference group and consider three dummy variables: part 2*control, part 1*TP and part 2*TP, where each variable takes value 1 if the observation is from the specific part in the specific treatment. We then add control variables (age, gender, and nationality in Model 2) to assess the robustness of the results.

We analyze responses time to verify that subjects answered faster in the TP treatment. To do so, we run OLS regressions for the response time with clustered
standard errors, as for the indexes. For some events we elicited the matching probabilities twice to test for consistency, since TP can be expected to decrease consistency. For each treatment and each part, we compare the first and second elicitation of these matching probabilities using t-tests with the Bonferroni correction for multiple comparisons. In the rest of the analysis, we only use the first matching probability elicited for each event.

By set-monotonicity, the matching probability of a composite event should exceed the matching probability of either one of its two constituents. Thus, we can test set-monotonicity six times in each part. Weak monotonicity is defined by $\bar{m}_c \geq \bar{m}_s$. It ensures $a \leq 1$. Violations of weak monotonicity entail very erratic answers. We nevertheless kept all answers in the analysis. Excluding the indexes when weak monotonicity is violated does not affect our conclusions (see the full results in the Online Appendix) unless we report otherwise. We will run non-parametric analysis (Wilcoxon tests and Mann-Whitney U tests) to test whether time pressure had an impact on the number of set-monotonicity and weak monotonicity violations.

4. EXPERIMENT: RESULTS

Table 3 gives descriptive results for the ambiguity indexes.

<table>
<thead>
<tr>
<th></th>
<th>a (part 1)</th>
<th>a (part 2)</th>
<th>b (part 1)</th>
<th>b (part 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>.15</td>
<td>.17</td>
<td>-.07</td>
<td>-.11</td>
</tr>
<tr>
<td>standard deviation</td>
<td>.44</td>
<td>.41</td>
<td>.21</td>
<td>.24</td>
</tr>
<tr>
<td>standard error</td>
<td>.07</td>
<td>.06</td>
<td>.03</td>
<td>.04</td>
</tr>
<tr>
<td>median</td>
<td>.07</td>
<td>.20</td>
<td>-.08</td>
<td>-.10</td>
</tr>
<tr>
<td>N</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td><strong>TP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>.34</td>
<td>.17</td>
<td>-.09</td>
<td>-.06</td>
</tr>
<tr>
<td>standard deviation</td>
<td>.44</td>
<td>.45</td>
<td>.24</td>
<td>.24</td>
</tr>
<tr>
<td>standard error</td>
<td>.06</td>
<td>.06</td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td>median</td>
<td>.35</td>
<td>.11</td>
<td>-.08</td>
<td>-.11</td>
</tr>
<tr>
<td>N</td>
<td>57</td>
<td>57</td>
<td>57</td>
<td>57</td>
</tr>
</tbody>
</table>
In what follows, we discuss only differences that are significant, with the significance level indicated in the corresponding tables.

### 4.1. Ambiguity Aversion Index $b$

To first illustrate the general nature of our data, Figure 1 presents all $b$ indexes of Part 2 as a function of the $b$ indexes of Part 1. Spearman correlations are high ($\rho = 0.77$ for the control treatment and $\rho = 0.85$ for the TP treatment) and most dots are in the lower left quadrant or in the upper right quadrant. It shows that subjects are consistently ambiguity averse or consistently ambiguity seeking across parts.

**Figure 1: ambiguity aversion indexes $b$**

![Graph showing ambiguity aversion indexes $b$ for control and TP treatments.](image)

Percentages of observations above and below the diagonal have been indicated in the figures. Spearman correlations $\rho$ are in the panel titles. The histograms represent the distribution of the index difference between Part 1 and Part 2.
Table 4 displays the results of the panel regressions for the \( b \) indexes. In Part 1, the control subjects are slightly ambiguity seeking (\(-0.07\)), with the dots in panel A slightly to the left. Regarding our main research question: the null hypothesis that TP has no effect cannot be rejected. The index \( b \) in TP does not differ significantly from that in the control in Part 1, with dots in panel B not more or less to the left than in panel A. The only effect we find is a learning effect for the control treatment, where Part 2 is a repetition of Part 1.\(^{10}\) Here ambiguity aversion is lower in Part 2 than in Part 1. There is no learning effect for the TP treatment (\( p = 0.14 \)). This may be because TP in Part 1 prevented the subjects to familiarize further with the task.

All aforementioned effects, and their levels of significance, are unaffected when we control for age, gender, and nationality (Dutch / non-Dutch) in Model 2. To test if ambiguity aversion, while not systematically bigger or smaller under TP, would become more or less extreme, we test absolute values of \( b \), but find no evidence for such effects (see Online Appendix).

**TABLE 4: ambiguity aversion indexes \( b \)**

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>(-0.07^*)</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.06)</td>
</tr>
<tr>
<td><strong>part 1 * TP treatment</strong></td>
<td>(-0.02)</td>
<td>(-0.03)</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>part 2 * control treatment</td>
<td>(-0.04^\dagger)</td>
<td>(-0.04^\dagger)</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>part 2 * TP treatment</td>
<td>0.00</td>
<td>(-0.01)</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>male</td>
<td>(-0.08^\dagger)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Dutch</td>
<td>(-0.07)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>age – 20</td>
<td>0.02</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Chi2</td>
<td>6.42\dagger</td>
<td>18.48\dagger</td>
</tr>
<tr>
<td>N</td>
<td>198</td>
<td>198</td>
</tr>
</tbody>
</table>

\(^{\dagger}p < 0.1, ^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001.\) Point estimates are followed by standard errors between brackets. The impact of TP is in bold. The variable age has been recoded as age \(-20\) so that the intercept corresponds to the \( b \) index of a 20-year-old subject (median age).

\(^{10}\) The learning effect is marginally significant and not significant anymore if we exclude the subjects violating weak monotonicity (see Table OB.1 in Online Appendix). So as to avoid learning effects for the part with TP, we had it precede the part without TP.
4.2. A-Insensitivity Index \( a \)

**FIGURE 2:** a-insensitivity indexes \( a \)

\[
\begin{align*}
\text{Part 2} & \quad \text{Part 1} \\
45\% & \quad 48\%
\end{align*}
\]

\[
\begin{align*}
\text{Part 2} & \quad \text{Part 1 (TP)} \\
29\% & \quad 60\%
\end{align*}
\]

Percentages of observations above and below the diagonal have been indicated in the figures. Spearman correlations \( \rho \) are in the panel titles. The histograms represents the distribution of the index difference between Part 1 and Part 2.

Figure 2 depicts all individual \( a \) indexes of Part 2 as a function of the \( a \) indexes of Part 1. Spearman correlations are again high (\( \rho = 0.77 \) for the control treatment and \( \rho = 0.70 \) for TP). Table 5 displays the results of the panel regressions for the \( a \) index. The insensitivity index is between 0.15 and 0.17 for Parts 1 and 2 of the control treatment (no learning effect and points equally split above and below the diagonal in panel A), and also for Part 2 of the TP treatment. However, there is much more a-insensitivity for the TP questions (Part 1 of TP treatment), with \( a = 0.34 \) and
with two-thirds of the dots in panel B to the right of the diagonal. These findings are robust to the addition of control variables (Model 2). Thus, we find a TP effect but no evidence for a learning effect.

**Table 5: a-insensitivity indexes α**

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.15( ^* )</td>
<td>0.20( ^† )</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.12)</td>
</tr>
<tr>
<td><strong>part 1 * TP treatment</strong></td>
<td><em><em>0.19( ^</em> )</em>*</td>
<td><strong>0.18( ^† )</strong></td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>part 2 * control treatment</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>part 2 * TP treatment</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>male</td>
<td>−0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>−0.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td></td>
</tr>
<tr>
<td>age − 20</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td></td>
</tr>
</tbody>
</table>

\( ^† \) p < 0.1, \( ^* \) p < 0.05, \( ^{**} \) p < 0.01, \( ^{***} \) p < 0.001. Point estimates are followed by standard errors between brackets. The impact of TP is in bold. The variable age has been recoded as age − 20 so that the intercept corresponds to the \( α \) index of a 20 year-old subject (median age).

### 4.3. **Response Time, Consistency, and Monotonicity**

The average response time in the training part is more than 25 seconds, but it gets much lower in Part 1 and then again in Part 2 for both the control and the TP treatment. Understandably, subjects needed to familiarize with the task. In Table 6, the benchmark model (Model 1) shows that the average response time of the control subjects in Part 1 is about 17s per matching probability. It is about 4s longer than for subjects under TP, even though the TP-treatment subjects could spend up to 25s to answer. In Part 2, the control subjects answered faster than in Part 1.
TABLE 6: Response time

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>16.63**</td>
<td>16.66***</td>
</tr>
<tr>
<td></td>
<td>(1.33)</td>
<td>(1.44)</td>
</tr>
<tr>
<td><strong>part 1 * TP treatment</strong></td>
<td><strong>-4.13</strong></td>
<td><strong>-4.44</strong></td>
</tr>
<tr>
<td></td>
<td>(1.40)</td>
<td>(1.44)</td>
</tr>
<tr>
<td>part 2 * control treatment</td>
<td>-2.33*</td>
<td>-2.33*</td>
</tr>
<tr>
<td></td>
<td>(1.14)</td>
<td>(1.14)</td>
</tr>
<tr>
<td>part 2 * TP treatment</td>
<td>-1.77</td>
<td>-2.08</td>
</tr>
<tr>
<td></td>
<td>(1.74)</td>
<td>(1.79)</td>
</tr>
<tr>
<td>male</td>
<td>-1.45</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>(1.21)</td>
<td>(1.19)</td>
</tr>
<tr>
<td>Dutch</td>
<td>0.99</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>(1.19)</td>
<td>(0.32)</td>
</tr>
<tr>
<td>age – 20</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.32)</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.02**</td>
<td>0.03*</td>
</tr>
<tr>
<td>N</td>
<td>1584</td>
<td>1584</td>
</tr>
</tbody>
</table>

† $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Point estimates are followed by standard errors between brackets. The impact of TP is in bold. The variable age has been recoded as age – 20 so that the intercept corresponds to the response time of a 20 year-old subject (median age). The stars reported next to $R^2$ refers to F-tests of overall significance.

We next analyze the consistency of the matching probabilities by comparing repeated elicitations of matching probabilities for some events. Pairwise comparisons for each pair of matching probabilities with the Bonferroni correction indicate one difference, in one of the two tests in Part 1 for the TP treatment: the second matching probability $m_{13}$ is higher than the first one (mean difference = 0.04; $p = 0.01$). The other differences are not significant.

A similar pattern is found in the set-monotonicity tests. Out of 6 monotonicity tests, the average number of violations is 0.58 in Part 1 for the TP treatment, while it is only 0.30 in Part 2 for the same treatment and 0.36 and 0.24 in Parts 1 and 2, respectively, for the control treatment. The difference between Parts 1 and 2 in the TP treatment is significant (within-subject Wilcoxon signed-ranks test; $Z = -2.61$, $p = 0.01$) and the difference between the TP and the control treatment in Part 1 is marginally significant (between-subject Mann-Whitney U test; $Z = -1.71$, $p = 0.09$). The percentage of weak monotonicity violations is 5% and 4% in Parts 1 and 2 for the TP treatment, and 5% and 0% in Parts 1 and 2 for the control treatment. None of the differences is significant.
4.4. Summary and Discussion of the Experiment

We summarize the experimental results. TP indeed increases insensitivity (index \( a \)), as predicted. That TP harms cognitive understanding is further confirmed by increased violations of consistency and set-monotonicity. These findings confirm Ariely and Zakay’s (2001) observation that TP aggravates biases and irrationalities. TP does not increase or decrease ambiguity aversion (index \( b \)).

Somewhat similar to our results, Young et al. (2012) found that TP increases insensitivity under risk. The effects of TP on risk aversion are not clear and can go in either direction (Young et al. 2012; Kircher, Pahlke, and Trautmann 2013), consistent with our absence of effect on ambiguity aversion. Kocher, Pahlke, and Trautmann (2013) also found increased insensitivity toward outcomes under TP for risk. Tinglyg et al. (2013) confirmed a more pronounced four-fold pattern of risk, again in agreement with increased insensitivity. Our second index will therefore be useful for future studies, nudging techniques, and policy recommendations regarding TP. Whereas the (ir)rationality of ambiguity aversion has been widely debated, insensitivity clearly reflects cognitive limitation and irrationality. Thus, the increased demand of full insurance and decreased demand for precautionary and partial insurance found by Bajtelsmit, Coats, and Thistle (2015) perfectly fits with insensitivity, as does the decreased quality of decisions in Conte, Scarsini, and Sürüçü (2016), De Paola and Gioia (2015), and Kirchler et al. (2017). Our study therefore supports the desirability to avoid TP for important decisions.

The absence of ambiguity aversion in our results is not surprising in view of recent studies with similar findings, especially because we used natural events rather than Ellsberg urns (Binmore, Stewart, and Voorhoeve 2012; Charness, Karni, and Levin 2013; Kocher, Lahno, and Trautmann 2017; Trautmann and van de Kuilen 2015). An additional experimental advantage of using natural events—that suspicion about experimenter-manipulated information is avoided—may have contributed to the absence of ambiguity aversion in our study. Such suspicion is further excluded because subjects always bet both on events and on their complements. In fact, suspicion may have worked against the risky events, because the natural events were even more verifiable than the risky events. Finally, the increase in preference (index
b) in Part 2 of the control treatment is in agreement with the familiarity bias (Chew, Ebstein, and Zhong 2012; Fox and Levav 2000; Kilka and Weber 2001).

The events in our experiments were natural in the sense of not involving any artificial concealing of information. We did not consider them in an actually occurring natural decision situation or in a field setting, and the decision situations considered were experimental. However, we used uncertainty that actually occurred and that was relevant to financial traders.

5. GENERAL DISCUSSION

Indexes are simplified summaries of complex realities. Our indexes cannot be expected to perfectly capture ambiguity attitudes, in the same way as the well-known index of relative risk aversion (IRR) cannot be expected to perfectly capture risk attitudes for every decision and every theory. The IRR does perfectly describe risk attitudes under expected utility with CRRA utility. For other utility functions, it will only work well on restricted domains of outcomes (Wakker 2008). Similarly, our indexes do perfectly describe ambiguity attitudes under Chateauneuf, Eichberger, and Grant’s (2007) neo-additive event weighting for several ambiguity theories (Baillon, Li, and Wakker 2017). In general, they will work well if none of the events in the partition is very likely or unlikely. Violations of event additivity and neo-additive weighting occur primarily for extreme events where no theory describes the many irregularities very well.11

There have as yet only been a few studies measuring ambiguity attitudes for natural events. Many did not control for risk attitudes and therefore could not completely identify ambiguity attitudes (Baillon et al. 2017; Fox, Rogers, and Tversky 1996; Fox and Tversky 1998; Kilka and Weber 2001). Abdellaoui et al. (2011) measured indexes similar to ours but had to use complex measurements and data fittings, requiring measurements of subjective probabilities, utilities, and event weights. As regards the treatment of unknown beliefs, Brenner and Izhakian (2015)

---

11 Thus, for risk, Kahneman and Tversky (1979 pp. 282-283) explicitly refrained from specifying any shape of probability weighting for extreme probabilities.
and Gallant, Jahan-Parvar, and Liu (2015) are close to us. They do not assume beliefs
given beforehand, but, like Abdellaoui et al. (2011), derive them from preferences.
We do not need such derivations. Brenner and Izhakian (2015) and Gallant, Jahan-
Parvar, and Liu (2015) deviate from our approach in assuming second-order
probabilities to capture ambiguity. They make parametric assumptions about the first-
and second-order probabilities (assuming normal distributions), including expected
utility for risk with constant relative risk aversion, and then fit the remaining
parameters to the data for a representative agent. Maccheroni, Marinacci, and
Ruffino’s (2013) theoretical analysis follows a similar approach. A difficulty in
parametric fittings often concerns what can be taken as ambiguity neutrality.

Baillon and Bleichrodt (2015) used a method similarly tractable as ours. They,
however, used different indexes, and they did not establish a control for likelihood.
Several papers used indexes similar to those presented above but provided no controls
for likelihoods, so that they had to use probability intervals or Ellsberg urns (Baillon,
Cabantous, and Wakker 2012; Dimmock, Kouwenberg, and Wakker 2016; Dimmock
et al. 2015, 2016). Li (2017), a follow-up of this paper, used our method to study
linguistic ambiguities. Such ambiguities are among the most common natural ones.
Her sample of Chinese adolescents had an exceptional spread in wealth, allowing for
a good measurement of wealth dependence of ambiguity attitudes. Li, Turmunk, and
Wakker (2017) used our method to measure the impact of ambiguity attitudes in
strategic situations.

This paper has shown that the ambiguity aversion index can better be related to
matching probabilities than to nonadditive weighting functions \( W \) as done before
(Dow and Werlang 1992; Schmeidler 1989), so as to avoid distortions by risk
attitudes. An additional advantage of matching probabilities is that they are readily
observable. More precisely, we need six indifferences (matching probabilities) to
calculate our two indexes. Measuring a nonadditive function \( W \) (a theoretical
construct: Cozic and Hill 2015) is harder, involving other theoretical constructs \( U \)
and theoretical assumptions. Thus, Abdellaoui et al. (2011) carried out complex
measurements, also involving utilities and beliefs, to obtain their indexes. In this

---

\(^{12}\) They used five event-dependent indexes similar to Kilka and Weber (2001), and based on preference
conditions of Tversky and Wakker (1995), adapting them to matching probabilities.
sense, we make preceding indexes operational (point (d) in intro). Because of point (f) in the introduction (validity also if EU is violated), our method also works for the general Choquet expected utility model in Gilboa (1987) which, unlike Schmeidler (1989), does not assume expected utility for risk; see Baillon, Li, and Wakker (2017). Key is that risk attitudes, including their deviations, cancel out if we use matching probabilities.

Many studies used introspective likelihood measurements (de Lara Resende and Wu 2010; Fox, Rogers, and Tversky 1996; Fox and Tversky 1998; Ivanov 2011) to capture beliefs for natural events. Professional forecasts and survey data are useful for establishing such beliefs (Anderson, Ghysels, and Juergens 2009). But those are not revealed preference based and the beliefs may be nonadditive. Then ambiguity attitudes may be captured partially by those nonadditive stated beliefs and partially by their weighting functions and, thus, ambiguity attitudes cannot be clearly isolated.

Our paper focuses on clearly defined revealed-preference concepts.

In the popular $\alpha$-maxmin model (Ghirardato, Maccheroni, and Marinacci 2004), $\alpha$ is often taken as an index of ambiguity aversion. Baillon, Li, and Wakker (2017) show that, for a popular subclass (priors $(1-\varepsilon)Q + \varepsilon T$ with $Q$ a fixed focal probability measure and $T$ variable and any possible probability measure) $\alpha$ can be recovered from our indexes (where our $\alpha$ is their $\varepsilon$):

$$\alpha = \frac{b}{2a} + \frac{1}{2}. \quad (5.1)$$

Our $b$ can be taken as an index of absolute ambiguity aversion, and $\alpha$ as a relative one, being aversion per perceived unit of ambiguity, renormalized. For readers who prefer the relative index $\alpha$, our method has also shown how to elicit this for natural events with unknown subjective beliefs.

How ambiguity attitudes are related across different sources of uncertainty, and across different persons, is an important topic for future research. The isolation of ambiguity attitudes from likelihood beliefs provided by this paper will be useful for such research.
6. CONCLUSION

Measuring ambiguity attitudes directly from revealed preferences up to now was only possible for artificially created uncertainties because no way was known to correct for unknown likelihood beliefs. We introduce a way to control for such beliefs and, thus, we can define two indexes of ambiguity attitudes that do apply to natural uncertainties as relevant in applications. This increases external validity. Our indexes are valid for many ambiguity theories, unifying and generalizing several existing indexes. In particular, our indexes are valid if expected utility for risk is violated, which is desirable for empirical purposes. Our second index (insensitivity) is needed to capture the likelihood dependence of ambiguity aversion that is usually found empirically.

We apply our indexes in a study on ambiguity under time pressure. Our findings are psychologically plausible, supporting the validity of our indexes: time pressure affects cognitive components (sensitivity/understanding, or level of ambiguity) but not motivational components (ambiguity aversion). Correlations between successive measurements of our indexes are high, confirming the reliability of our method.
APPENDIX A. DETAILS OF THE EXPERIMENT

Procedure
In the experiment, computers of different subjects were separated by wooden panels to minimize interaction between subjects. Brief instructions were read aloud, and tickets with ID numbers were handed out. Subjects typed in their ID numbers to start the experiment. The subjects were randomly allocated to treatment groups through their ID numbers. Talking was not allowed during the experiment. Instructions were given with detailed information about the payment process, user interface, and the type of questions subject would face. The subjects could ask questions to the experimenters at any time. In each session, all subjects started the experiment at the same time.

In the TP treatment, we took two measures to make sure that TP would not have any effects in Parts 0 and 2. First, we imposed a two-minute break after Parts 0 and 1, to avoid spill-over of stress from Part 1 to Part 2. Second, we did not tell the subjects that they will be put under TP prior to Part 1, so as to avoid stress generated by such an announcement in Part 0 (Ordonez and Benson 1997).

Stimuli: Choice lists
Subjects were asked to state which one of the two choice options in §2 they preferred for different values of p, ascending from 0 to 100 (Figures A.1 and A.2). The midpoint between the two values of p where they switched preference was taken as their indifference point and, hence, as the matching probability.

To help subjects answer the questions quickly, which was crucial under TP, the experimental webpage allowed them to state their preferences with a single click. For example, if they clicked on Option 2 when the probability of winning was 50%, then for all $p > 50\%$, the option boxes for Option 2 were automatically filled out and for all $p < 50\%$ the option boxes for Option 1 were automatically filled out. This procedure also precluded violations of stochastic dominance by preventing multiple preference switches. After clicking on their choices, subjects clicked on a “Submit” button to move to the next question. The response times were also tracked.

In Part 1 of the TP treatment, a timer was displayed showing the time left to answer. If subjects failed to submit their choices before the time limit expired, their
choices would be registered but not be paid. This happened only 5 out of the 496 times (62 subjects × 8 choices). In a pilot, the average response time without TP was 36 seconds, and another session of the pilot showed that, under a 30-second time limit, subjects did not experience much TP. Therefore, we chose the 25 seconds limit.

Figure A.1: Screenshot of the experiment software for single event E₃ in Part 0
Stimuli: Avoiding middle bias

The middle bias can distort choice lists: subjects tend to choose the options, in our case the preference switch, that are located in the middle of the provided range (Erev and Ert 2013; Poulton 1989). TP can be expected to reinforce this bias. Had we used a common equally-spaced choice list with, say, 5% incremental steps, then the middle bias would have moved matching probabilities in the direction of 50% (both for the single and composite events). This bias would have enhanced the main phenomenon found in this paper, a-insensitivity, and render our findings less convincing. To avoid this problem, we designed choice lists that are not equally spaced. In our design, the middle bias enhances matching probabilities 1/3 for single events and probabilities 2/3 for composite events. Thus, this bias enhances additivity of the matching probabilities, decreases a-insensitivity, and moves our a-insensitivity index toward 0. It makes findings of nonadditivity and a-insensitivity more convincing.
Table A.1 lists the AEX events that we used. Some questions were repeated for consistency checks. The corresponding events are listed twice.

**Table A.1: List of events on which the AEX prospects were based**

<table>
<thead>
<tr>
<th>Part</th>
<th>Event</th>
<th>Event description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (Training)</td>
<td>$E_1$</td>
<td>the AEX decreases by strictly more than 0.4%</td>
</tr>
<tr>
<td></td>
<td>$E_1$</td>
<td>the AEX decreases by strictly more than 0.4%</td>
</tr>
<tr>
<td></td>
<td>$E_2$</td>
<td>the AEX either decreases by less than 0.4% or increases by less than 0.1%</td>
</tr>
<tr>
<td></td>
<td>$E_3$</td>
<td>the AEX increases by strictly more than 0.1%</td>
</tr>
<tr>
<td></td>
<td>$E_{12}$</td>
<td>the AEX either increases by less than 0.1% or decreases</td>
</tr>
<tr>
<td></td>
<td>$E_{23}$</td>
<td>the AEX either decreases by less than 0.4% or increases</td>
</tr>
<tr>
<td></td>
<td>$E_{13}$</td>
<td>the AEX either decreases by strictly more than 0.4% or increases by strictly more than 0.1%</td>
</tr>
<tr>
<td>1</td>
<td>$E_1$</td>
<td>the AEX decreases by strictly more than 0.2%</td>
</tr>
<tr>
<td></td>
<td>$E_2$</td>
<td>the AEX either decreases by less than 0.2% or increases by less than 0.2%</td>
</tr>
<tr>
<td></td>
<td>$E_2$</td>
<td>the AEX either decreases by less than 0.2% or increases by less than 0.2%</td>
</tr>
<tr>
<td></td>
<td>$E_3$</td>
<td>the AEX increases by strictly more than 0.2%</td>
</tr>
<tr>
<td></td>
<td>$E_{12}$</td>
<td>the AEX either increases by less than 0.2% or decreases</td>
</tr>
<tr>
<td></td>
<td>$E_{12}$</td>
<td>the AEX either increases by less than 0.2% or decreases</td>
</tr>
<tr>
<td></td>
<td>$E_{23}$</td>
<td>the AEX either decreases by less than 0.2% or increases</td>
</tr>
<tr>
<td></td>
<td>$E_{13}$</td>
<td>the AEX either decreases by strictly more than 0.2% or increases by strictly more than 0.2%</td>
</tr>
<tr>
<td>2</td>
<td>$E_1$</td>
<td>the AEX decreases by strictly more than 0.1%</td>
</tr>
<tr>
<td></td>
<td>$E_2$</td>
<td>the AEX either decreases by less than 0.1% or increases by less than 0.3%</td>
</tr>
<tr>
<td></td>
<td>$E_3$</td>
<td>the AEX increases by strictly more than 0.3%</td>
</tr>
<tr>
<td></td>
<td>$E_3$</td>
<td>the AEX increases by strictly more than 0.3%</td>
</tr>
<tr>
<td></td>
<td>$E_{12}$</td>
<td>the AEX either increases by less than 0.3% or decreases</td>
</tr>
<tr>
<td></td>
<td>$E_{23}$</td>
<td>the AEX either decreases by less than 0.1% or increases</td>
</tr>
<tr>
<td></td>
<td>$E_{13}$</td>
<td>the AEX either decreases by strictly more than 0.1% or increases by strictly more than 0.3%</td>
</tr>
</tbody>
</table>

*Incentives*

For each subject, one question (i.e., one row of one choice list) was randomly selected to be played for real at the end of the experiment. If subjects preferred the bet on the stock market index, then the outcome was paid according to the change in the stock market index during the duration of the experiment. Bets on the given probabilities
were settled using dice. In the instructions of the experiment, subjects were presented with two examples to familiarize them with the payment scheme. If the time deadline for a TP question had not been met, the worst outcome (no payoff) resulted. Therefore, it was in the subjects’ interest to submit their choices on time.

ACKNOWLEDGMENT. Roy Kouwenberg, Chen Li, Thorsten Pachur, and Martijn van den Assem made helpful comments. Huang acknowledges support from the China National Science Foundation (Grant no.71703091).

REFERENCES


