Behavioural study on disclosure of ADR information to consumers by traders and ADR entities

Final Report

Technical annex

LE Europe

November 2022

Table of Contents

1 Methodology for the experiment analysis

1.1 Overarching analysis methodology

The overarching methodology for the experimental analysis recognised similarities in the analysis across different outcome measures and website types (trader and ADR entity). For each of these, an outcome measure (either a proportion or an average) was compared among two or more treatment groups.

Given that treatments contained up to four variants, pairwise comparisons can grow numerous quickly. This increases the likelihood of type I errors; mistakenly identifying something as statistically significant whereas in reality no effect exists. The main overarching analysis reduced this likelihood by reducing the number of tests conducting for this project in a two-step approach.

In the **first step**, omnibus tests were used to test whether *any* of the treatment variants had any effect on the outcome measure. Outcome measures expressed as proportion (e.g., the proportion going to ADR) were tested using a χ^2 test of independence. Outcome measures expressed as averages (e.g., average score for objective understanding) were tested using ANOVA. Crucially, omnibus tests shed light on whether *any* treatment effect exists, but not *which* treatment variant is driving the result.

The **second step** was conducted if and only if the omnibus test in the first step was statistically significant at, at least, the 5% level. In the second step, the outcomes for the different treatment variants were tested pairwise, e.g. testing TO1 against TO2, TO3 and TO4 respectively[1]. Outcomes expressed as proportions used z-tests. Outcome expressed as averages used t-tests. These pairwise tests provide information on *which* variants drive the results observed in the first step.

1.2 Classification of open responses

The experiment asked open response questions asking respondents what their next step would be after browsing each trader's website. These responses were classified using a string-matching tool.

For this purpose, two lists of target words were produced (per experiment language) signalling an intention to use ADR as a next step. One list contained 'full terms' (e.g., "Alternative Dispute Resolution"), whereas the other contained acronyms and initialisms (e.g., "ADR"). The lists of target words were agreed with the Commission and devised to be concise and contain terms that clearly related to ADR. Specifically, the terms (in English) included were 'alternative dispute resolution' and 'online dispute resolution' and their respective acronyms, as well as the name of the fictitious ADR entity mentioned on the website ('Dispute Resolution Centre') and, for the lab experiments, 'dispute resolution body' (since writing this would be clearly indicative of pursuing ADR, and it appeared on both additional pages accessible to respondents in lab experiments).

The national language lists used are presented in Table 1.

[1] The website structure treatments in the online experiment (TO) had an explicitly baseline treatment against which all other treatment variants could be tested. This was not the case for all treatments. Where no baseline exists, all pairwise tests were conducted.

Table 1 Target words by country

Country	Full terms	Acronyms and initialisms[1]
Austria (online experiment)	alternative streitbeilegung, online streitbeilegung, das streitbeilegungszentrum	AS, OS
Italy (online experiment)	risoluzione alternativa delle controversie, risoluzione delle controversie online, il centro di risoluzione delle controversie	ADR, ODR
Poland (online experiment)	alternatywne metody rozstrzygania sporów, internetowego rozwiązywania sporów, centrum rozstrzygania sporów	ADR, ODR
Sweden (online experiment)	alternativ tvistlösning, tvistlösning på nätet, tvistlösningscentret	- (no official ADR/ODR acronyms are used in Sweden)[2]
Germany (lab experiment)	alternative streitbeilegung, online streitbeilegung, das streitbeilegungszentrum, streitbeilegungsstelle	AS, OS
Spain (lab experiment)	resolución alternativa de litigios, resolución de litigios en linea, centro de resolución de disputas, organismo de resolución de litigios	RAL, RLL

Note: [1] for acronyms the fuzzy matching process was case sensitive. [2] Acronyms were taken from the ADR Directive. However, for Sweden the ADR Directive does not give any acronyms for ADR or ODR.

The lists were matched using a "fuzzy" matching procedure that generates a score between 0 to 100 signifying how well the respondents' text matched to the target words. The two lists were analysed separately since small spelling mistakes in short acronyms have a disproportionate impact on the fuzzy.

The fuzzy matching algorithm identified the two closest terms found in the respondents' answers to the specified target words. This was done via the Levenshtein Distance[2]. This distance calculates the similarities between words as, informally, the minimum number of single-character edits (i.e., insertions, deletions or substitutions) required to change one word into the other.[3]

Responses were classified as indicating that ADR would be the consumer's next step if the match exceeded a certain 'success threshold'. This was set at 60% similarity for the full terms list, and 70% similarity for the list of acronyms. The selection of these thresholds was, by necessity, somewhat subjective. Hence, to decide on suitable thresholds manual inspections of the results were carried out. From these inspections, it was established that setting the thresholds at these levels classified responses as 'successes' where the respondent clearly tried to write that they would do ADR but, for example, may have misspelt one of the words, and classes responses that do not correspond to ADR (e.g. emails the seller) as failures. Raising the thresholds above these levels results in losing

[2] Implemented in Python 3.9 using the 'The Fuzz' package.

[3] https://www.cuelogic.com/blog/the-levenshtein-algorithm

(class as 'failure') many responses that in fact intend to indicate ADR, whereas if we lower them we start classifying as a 'success' responses that do not indicate ADR.

Exact or absolute matching of target words to the respondents' response, was used as a sensitivity check of the fuzzy matching regime. Absolute matching yielded similar treatment effects as the fuzzy match, and was therefore not reported in the main report.

1.3 Analysis deviating from the overarching methodology

Two parts of the analysis deviated from the overarching methodology as highlighted above, to capture specific characteristics.

1.3.1 Vulnerability characteristics

The analysis of treatment effects by vulnerability characteristics skipped the first step in the overarching methodology. Instead, it immediately tested treatment effects within each group defined by vulnerability characteristics, using z-tests and t-tests. The omnibus test was skipped since reporting on this test would create additional complexity while reporting the results. Reporting the results in a way that can be easily understood by a lay person was deemed to outweigh the usefulness of the omnibus test in this particular situation.

The analysis of treatment effects by vulnerability characteristic further assessed whether treatments effects differed by those with and without a particular potential vulnerability. This was assessed using logit regression analysis. For the online experiment, logit regressions were conducted including:

- dummy variables for the relevant group of treatments (e.g. the traders' website structure treatments); **baseline:** treatment 1 (e.g. TO1);
- dummy variables indicating the two groups defined by a vulnerability characteristic (e.g. elderly and non-elderly); **baseline:** non-vulnerable group; and,
- a full set of interaction dummies between the previous two variables.

The assessment whether treatment effects differed by between potentially vulnerable and non-vulnerable consumers, was based on the coefficient and statistical significance of the interaction terms.

1.3.2 Funding source of the ADR entity

The question on the funding source of the ADR entity randomly allocated respondents to "government", "trader" or "trade association". Respondent then could assess their attitude towards the ADR entity using a five-point scale ranging from "completely positive" to "completely negative".

Since respondents could provide five different answers, differences in attitude generated by different funding sources were tested by comparing the **distribution** of answers. More precisely, a χ^2 test of independence was used to pairwise compare each of the three pairs of funding sources ("government/trader", "government/trade association", "trader/trade association"). Each χ^2 test provide evidence whether the distribution of attitudes differed among the two relevant sources.

1.4 Bonferroni correction

As mentioned, the overarching methodology used a two-step approach to reduce the likelihood of type I errors. These can occur frequently if many tests are conducted, which is known as the multiple comparisons problem. Another way of correcting for this is to consider "family-wise error rate", e.g. via Bonferroni corrections.

A Bonferroni correction calculates the "target" significance level of any individual test by taking the desired "family-wise" error rate (α) and the number of tests within each family (m). A family of tests could, for example, be all pairwise z-tests for a particular outcome variable. The Bonferroni correction then calculates the individual significance level as α/m.

The families of tests in the analysis conducted for this report consisted of around 5 tests each. With a target "family-wise error rate" of 5%, this suggests that individual tests should be evaluated at the 1% significance level.

2 Calculation of minimum detectable effect sizes

This section highlights the ability of the methodology to detect treatment effects in the experiments. More precisely, for each type of test highlighted above, this section provides the minimum detectable effect size given the sample sizes obtained during fieldwork.

To judge the effect sizes, the following types of statistics were calculated:

- χ^2 **test:** effect size w
- **z-test:** minimum detectable difference in proportions, in percentage points
- **ANOVA:** effect size f
- **t-test:** Cohen's d

For these calculation, the following inputs were assumed:

- 99% significance level, in line with the Bonferroni correction
- 80% power level
- *For z-tests only:* baseline proportions of 25%[4]

To help interpret these calculations, the following can be considered:

- w **from a** χ^2 **test:** typically considered small if $w = 0.1$, medium if $w = 0.3$, large if $w = 0.5$[5]
- f **from an ANOVA:** typically considered small if $f = 0.1$, medium if $f = 0.25$, large if $f = 0.4$[6]
- **Cohen's d:** typically considered small if $d = 0.2$, medium if $d = 0.5$, large if $d = 0.8$[7]

2.1 Online experiment

The calculations here are based on the fieldwork obtaining 4,050 completed responses. With this in mind, a number of scenarios need to be accounted for in the calculation of the minimum detectable effect size, namely:

- **Questions asked twice of respondents on the traders' website:** the open-ended question on the traders' website was asked twice of each respondent. These responses were pooled in the analysis. Relevant tests are the χ^2 test and z-tests only.

  This yields the following sample sizes for the calculation:
  - **Total sample:** 8,100
  - **Per website structure treatment (TO: 4 variants):** 2,025
  - **Per text treatment (text: 2 variants):** 4,050

[4] This percentage seems reasonable as a general assumption given the data obtained during both pilot and main fieldwork.

[5] https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095743480

[6] Idem.

[7] https://rpsychologist.com/cohend/

- **Questions asked once of respondents on the traders' website:** the closed-ended question and the objective understanding question were asked only once. Relevant tests are the χ^2 test, z-tests, ANOVA and t-tests.

  This yields the following sample sizes for the calculation:

  - **Total sample:** 4,050
  - **Per website structure treatment (TO: 4 variants):** 1,012
  - **Per text treatment (text: 2 variants):** 2,025
- **ADR entity website:** each part of the ADR entity website was only completed once by each respondent. Relevant tests are the χ^2 test, z-tests, ANOVA and t-tests.

  This yields the following sample sizes for the calculation:

  - **Total sample:** 4,050
  - **Per ADR entity treatment (EO: 4 variants):** 1,012
- **Funding source of ADR entity:** this question was asked only once to each respondent. Relevant test is only the χ^2 test.

  This yields the following sample sizes for the calculation:

  - **Total sample:** 4,050
  - **Sample per pairwise comparison (2/3 of the total sample):** 2,700

The table below provides the minimum detectable effect sizes given the scenarios above.

Table 2 Minimum detectable effect size: online experiment

	w from χ^2 test	Percentage points from z-test[a]	f from ANOVA	Cohen's d from t-test[a]
Questions asked twice: traders' website	**S:** 0.044 (df = 3) **T:** 0.038 (df = 1)	**S:** 4.43pp **T:** 3.11pp	N/A	N/A
Questions asked once: traders' website	**S:** 0.062 (df = 3) **T:** 0.054 (df = 1)	**S:** 6.33pp **T:** 4.43pp	**S:** 0.062 **T:** 0.054	**S:** 0.152 **T:** 0.107
ADR entity website	0.062 (df = 3)	6.33pp	0.062	0.152
Funding source of ADR entity	0.079 (df = 4)	N/A	N/A	N/A

Note: S refers to the "website structure treatments" (TO) only. This represents 4 treatments. T refers to the "text treatments" (text) only. This represents 2 treatments.

[a] Calculated as the minimum required effect size for a comparison of two treatments only.

Source: LE Europe calculations based on G*Power 3.1.9.2

2.2 Lab experiments

The calculations here are based on the fieldwork obtaining 599 completed responses. With this in mind, a number of scenarios need to be accounted for in the calculation of the minimum detectable effect size, namely:

- **Questions asked twice of respondents on the traders' website:** the open-ended question on the traders' website was asked twice of each respondent. These responses were pooled in the analysis. Relevant tests are the χ^2 test and z-tests only.

  This yields the following sample sizes for the calculation:

  - **Total sample:** 1,198
  - **Per trader's website treatment (TL: 4 variants):** 299

- **Questions asked once of respondents on the traders' website:** the closed-ended question and the objective understanding questions were asked only once. Relevant tests are the χ^2 test, z-tests, ANOVA and t-tests.

  This yields the following sample sizes for the calculation:

  - **Total sample:** 599
  - **Per trader's website treatment (TL: 4 variants):** 150
- **ADR entity website:** each part of the ADR entity website was only completed once by each respondent. Relevant tests are the χ^2 test, z-tests, ANOVA and t-tests.

  This yields the following sample sizes for the calculation:

  - **Total sample:** 599
  - **Per ADR entity treatment (EL: 3 variants):** 200
- *Funding source of ADR entity was not included in the laboratory experiment*

The table below provides the minimum detectable effect sizes given the scenarios above.

Table 3 Minimum detectable effect size: laboratory experiment

	w from χ^2 test	Percentage points from z-test[a]	f from ANOVA	Cohen's d from t-test[a]
Questions asked twice: traders' website	0.114 (df = 3)	11.95pp	N/A	N/A
Questions asked once: traders' website	0.161 (df = 3)	17.20pp	0.161	0397
ADR entity website	0.152 (df = 2)	14.78pp	0.153	0.343

[a] Calculated as the minimum required effect size for a comparison of two treatments only.

Source: LE Europe calculations based on G*Power 3.1.9.2

3 Tabulations of experimental results

3.1 Impact of trader website treatments

3.1.1 Intention to use ADR – online experiment

Table 4 to Table 7 present the results of omnibus and z-tests (outlined in section 1.1) performed on the treatments for the trader websites in the online experiment, with respect to their effect on the intention to use ADR.

Website structure treatments

Table 4 Effects of TO2, TO3 & TO4 vs. TO1 on intention to use ADR (open responses)

Omnibus test				
χ^2 statistic	65.4386			
p-value	< 0.001			
z-tests				
Treatment	**TO1**	**TO2**	**TO3**	**TO4**
Percentage indicating ADR	10.1%	17.7%	17.1%	12.7%
Treatment effect[a]		7.63pp	7.04pp	2.63pp
P-value		< 0.001	< 0.001	0.010
N	2,016	2,012	2,060	2,012

[a] Comparing against TO1.

Source: LE Europe calculations using online experiment data

Table 5 Effects of TO2, TO3 & TO4 vs. TO1 on intention to use ADR (closed responses)

Omnibus test				
χ^2 statistic	74.2096			
p-value	< 0.001			
z-tests				
Treatment	**TO1**	**TO2**	**TO3**	**TO4**
Percentage indicating ADR	31.2%	42.7%	39.2%	31.8%
Treatment effect[a]		11.48pp	7.96pp	0.60pp
P-value		< 0.001	< 0.001	0.778
N	1,008	1,006	1,030	1,006

[a] Comparing against TO1.

Source: LE Europe calculations using online experiment data

Text treatments

Table 6 Effects of the text treatments on the trader websites on intention to use ADR (open response)

Omnibus test		
χ^2 statistic	0.0047	
p-value	0.9468	
z-tests		
Treatment	**Text 1: Plain text**	**Text 2: Highlighting benefits text**
Percentage indicating ADR	14.4%	14.4%
Treatment effect[a]		-0.05pp
P-value		*N/A*
N	4,058	4,042

[a] Comparing against Text 1.

Source: LE Europe calculations using online experiment data

Table 7 Effects of text treatments on the intention to use ADR (closed response)

Omnibus test		
χ^2 statistic	1.9770	
p-value	0.9312	
z-tests		
Treatment	**Text 1: Plain text**	**Text 2: Highlighting benefits text**
Percentage indicating ADR	36.9%	35.5%
Treatment effect[a]		-1.46pp
P-value		*N/A*
N	2,029	2,021

[a] Comparing against Text 1.

Source: LE Europe calculations using online experiment data

3.1.2 Intention to use ADR – lab experiment

Table 8 to Table 11 present the results of omnibus and z-tests (outlined in section 1.1) performed on the treatments for the trader websites in the lab experiment, with respect to their effect on the intention to use ADR.

Table 8 Effects of TL2, TL3 & TL4 vs. TL1 on intention to use ADR (open answer responses)

Omnibus test				
χ^2 statistic	68.0346			
p-value	< 0.001			
z-tests				
Treatment	**TL1**	**TL2**	**TL3**	**TL4**
Percentage indicating ADR	5.7%	25.9%	25.8%	31.9%
Treatment effect[a]		20.2pp	20.2pp	26.2pp
P-value		< 0.001	< 0.001	< 0.001
N	300	294	306	298

[a] Comparing against TL1.

Source: LE Europe calculations using lab experiment data

Table 9 Effects of TL3 & TL4 vs. TL2 on intention to use ADR (open answer responses)

Omnibus test			
χ^2 statistic	*N/A*		
p-value	*N/A*		
z-tests			
Treatment	**TL2**	**TL3**	**TL4**
Percentage indicating ADR	25.9%	25.8%	31.9%
Treatment effect[a]		-0.03pp	6.03pp
P-value		0.9925	0.1055
N	294	306	298

[a] Comparing against TL2.

Source: LE Europe calculations using lab experiment data

Table 10 Effects of TL2, TL3 & TL4 vs. TL1 on intention to use ADR (closed answer responses)

Omnibus test				
χ^2 statistic	46.1977			
p-value	< 0.001			
z-tests				
Treatment	**TL1**	**TL2**	**TL3**	**TL4**
Percentage indicating ADR	30.7%	48.3%	52.3%	59.5%
Treatment effect[a]		17.6pp	21.6pp	28.8pp
P-value		0.0017	< 0.001	< 0.001
N	150	147	153	149

[a] Comparing against TL1.

Source: LE Europe calculations using lab experiment data

Table 11 Effects of TL3 & TL4 vs. TL2 on intention to use ADR (closed answer responses)

Omnibus test			
χ^2 statistic	*N/A*		
p-value	*N/A*		
z-tests			
Treatment	**TL2**	**TL3**	**TL4**
Percentage indicating ADR	48.3%	52.3%	59.5%
Treatment effect[a]		4.0pp	11.2pp
P-value		0.4911	0.0632
N	147	153	149

[a] Comparing against TL2.

Source: LE Europe calculations using lab experiment data

3.1.3 Understanding of ADR – online experiment

Table 12 and Table 13 present the results of omnibus and z-tests (outlined in section 1.1) performed on the treatments for the trader websites in the online experiment, with respect to their effect on understanding of ADR.

Website structure treatments

Table 12 Effects of TO2, TO3 & TO4 vs. TO1 on understanding following the trader website

Omnibus test				
F statistic	3.409			
p-value	0.017			
t-tests				
Treatment	**TO1**	**TO2**	**TO3**	**TO4**
Average score	1.798	1.928	1.911	1.866
Treatment effect[a]		0.13	0.11	0.07
P-value		0.004	0.015	0.144
N	1,008	1,006	1,030	1,006

[a] Comparing against TO1.

Source: LE Europe calculations using online experiment data

Text treatments

Table 13 Effects of text treatments on the understanding of ADR

Omnibus test		
F statistic	0.9506	
p-value	0.3296	
t-tests		
Treatment	**Text 1: Plain text**	**Text 2: Highlighting benefits text**
Average score	1.861	1.891
Treatment effect[a]		0.03
P-value		*N/A*
N	2,029	2,021

[a] Comparing against Text 1.

Source: LE Europe calculations using online experiment data

3.1.4 Understanding of ADR – lab experiment

Table 14 and Table 15 present the results of omnibus and z-tests (outlined in section 1.1) performed on the treatments for the trader websites in the lab experiment, with respect to their effect on understanding of ADR.

Table 14 Effects of TL2, TL3 & TL4 vs. TL1 on understanding following the trader website

Omnibus test				
F statistic	5.3464			
p-value	0.0012			
t-tests				
Treatment	**TL1**	**TL2**	**TL3**	**TL4**
Average score	1.973	1.939	2.216	2.315
Treatment effect[a]		-0.035	0.242	0.342
P-value		0.7327	0.0161	0.0063
N	150	147	153	149

[a] Comparing against TL1.

Source: LE Europe calculations using lab experiment data

Table 15 Effects TL3 & TL4 vs. TL2 on understanding following the trader website

Omnibus test			
F statistic	*N/A*		
p-value	*N/A*		
t-tests			
Treatment	**TL2**	**TL3**	**TL4**
Average score	1.939	2.216	2.315
Treatment effect[a]		0.277	0.377
P-value		0.0048	0.0022
N	147	153	149

[a] Comparing against TL2.

Source: LE Europe calculations using lab experiment data

3.2 Impact of ADR entity website treatments

3.2.1 Decisions to go to ADR and submit a claim – online experiment

Table 16 to Table 18 present the results of omnibus and z-tests (outlined in section 1.1) performed on the treatments for the ADR entity websites in the online experiment, with respect to their effect on going to ADR and submitting a claim.

Table 16 Effects of EO2, EO3 & EO4 vs. EO1 on the decision to go to ADR

Omnibus test				
χ^2 statistic	2.0944			
p-value	0.553			
z-tests				
Treatment	**EO1**	**EO2**	**EO3**	**EO4**
Percentage going to ADR	70.6%	69.3%	72.2%	70.8%
Treatment effect[a]		-1.25pp	1.60pp	0.24pp
P-value		*N/A*	*N/A*	*N/A*
N	1,000	1,021	1,004	1,025

[a] Comparing against EO1. Where "N/A" is given in space of the p-value, a pairwise z-test was not conducted as the chi-square test of homogeneity found statistical insignificance across all treatments.

Source: LE Europe calculations using online experiment data

Table 17 Effects of EO2, EO3 & EO4 vs. EO1 on the decision to submit a claim

Omnibus test				
χ^2 statistic	1.4573			
p-value	0.692			
z-tests				
Treatment	**EO1**	**EO2**	**EO3**	**EO4**
Percentage submitting a claim[a]	83.2%	83.9%	80.9%	84.1%
Treatment effect[b]		0.64pp	-2.33pp	0.88pp
P-value		*N/A*	*N/A*	*N/A*
N	712	714	731	734

[a] Of those going to ADR. [b] Comparing against EO1. Where "N/A" is given in space of the p-value, a pairwise z-test was not conducted as the chi-square test of homogeneity found statistical insignificance across all treatments.

Source: LE Europe calculations using online experiment data

Table 18 Effects of EO2, EO3 & EO4 vs. EO1 on submitting a claim correctly

Omnibus test				
χ^2 statistic	1.193			
p-value	0.311			
z-tests				
Treatment	**EO1**	**EO2**	**EO3**	**EO4**
Percentage submitting a claim correctly[a]	36.3%	38.7%	33.4%	35.6%
Treatment effect[b]		2.39pp	-2.94pp	-0.76pp
P-value		*N/A*	*N/A*	*N/A*
N	593	598	592	618

[a] Of those submitting an ADR claim. [b] Comparing against EO1. Where "N/A" is given in space of the p-value, a pairwise z-test was not conducted as the chi-square test of homogeneity found statistical insignificance across all treatments.

Source: LE Europe calculations using online experiment data

3.2.2 Decisions to go to ADR and submit a claim – lab experiment

Table 19 to Table 22 present the results of omnibus and z-tests (outlined in section 1.1) performed on the treatments for the ADR entity websites in the lab experiment, with respect to their effect on going to ADR and submitting a claim.

Table 19 Effects of EL2 & EL3 vs. EL1 on the decision to go to ADR

Omnibus test			
χ^2 statistic	3.1242		
p-value	0.2096		
z-tests			
Treatment	**EL1**	**EL2**	**EL3**
Percentage going to ADR	92.5%	91.9%	96.0%
Treatment effect[a]		-0.62pp	3.44pp
P-value		*N/A*	*N/A*
N	201	198	199

[a] Comparing against EL1. Where "N/A" is given in space of the p-value, a pairwise z-test was not conducted as the chi-square test of homogeneity found statistical insignificance across all treatments.

Source: LE Europe calculations using lab experiment data

Table 20 Effect of EL3 vs. EL1 & EL2 combined on the decision to go to ADR

Omnibus test		
χ^2 statistic	*N/A*	
p-value	*N/A*	
z-tests		
Treatment	**EL1 & EL2 combined**	**EL3**
Percentage going to ADR	92.1%	97.0%
Treatment effect[a]		4.82pp
P-value		0.0528
N	399	199

[a] Comparing against EL1&EL2 combined.

Source: LE Europe calculations using lab experiment data

Table 21 Effects of EL2 & EL3 vs. EL1 on the decision to submit a claim

Omnibus test			
χ^2 statistic	0.0048		
p-value	0.9976		
z-tests			
Treatment	**EL1**	**EL2**	**EL3**
Percentage submitting a claim[a]	97.8%	97.8%	97.9%
Treatment effect[b]		-0.05pp	0.06pp
P-value		*N/A*	*N/A*
N	186	182	191

[a] Of those going to ADR. [b] Comparing against EL1. Where "N/A" is given in space of the p-value, a pairwise z-test was not conducted as the chi-square test of homogeneity found statistical insignificance across all treatments.

Source: LE Europe calculations using lab experiment data

Table 22 Effects of EL2 & EL3 vs. EL1 on submitting a claim correctly

Omnibus test			
χ^2 statistic	0.2278		
p-value	0.8923		
z-tests			
Treatment	**EL1**	**EL2**	**EL3**
Percentage submitting a claim correctly[a]	36.8%	38.8%	39.0%
Treatment effect[b]		1.95	2.22
P-value		*N/A*	*N/A*
N	182	178	187

[a] Of those submitting a claim. [b] Comparing against EL1. Where "N/A" is given in space of the p-value, a pairwise z-test was not conducted as the chi-square test of homogeneity found statistical insignificance across all treatments.

Source: LE Europe calculations using lab experiment data

3.2.3 Understanding of ADR – online experiment

Table 23 presents the results of omnibus and z-tests (outlined in section 1.1) performed on the treatments for the ADR entity websites in the online experiment, with respect to their effect on understanding of ADR.

Table 23 Effects of EO2, EO3 & EO4 vs. EO1 on understanding after the ADR entity website

Omnibus test				
F statistic	1.5320			
p-value	0.2040			
t-tests				
Treatment	**EO1**	**EO2**	**EO3**	**EO4**
Average score	2.34	2.25	2.28	2.33
Treatment effect[a]		-0.09	-0.06	-0.01
P-value		*N/A*	*N/A*	*N/A*
N	1,000	1,021	1,004	1,025

[a] Comparing against EO1. Where "N/A" is given in space of the p-value, a pairwise z-test was not conducted as the chi-square test of homogeneity found statistical insignificance across all treatments.

Source: LE Europe calculations using online experiment data

3.2.4 Understanding of ADR – lab experiment

Table 24 presents the results of omnibus and z-tests (outlined in section 1.1) performed on the treatments for the ADR entity websites in the online experiment, with respect to their effect on understanding of ADR.

Table 24 Effects of EL2 & EL3 vs. EL1 on understanding after the ADR entity website

Omnibus test			
F statistic	1.1396		
p-value	0.3206		
t-tests			
Treatment	**EL1**	**EL2**	**EL3**
Average score	2.92	2.78	2.81
Treatment effect[a]		-0.14	-0.12
P-value		*N/A*	*N/A*
N	201	198	199

[a] Comparing against EL1. Where "N/A" is given in space of the p-value, a pairwise z-test was not conducted as the chi-square test of homogeneity found statistical insignificance across all treatments.

Source: LE Europe calculations using lab experiment data

3.3 Vulnerable consumers in the online experiment

3.3.1 Summary statistics on vulnerability characteristics

Table 25 to Table 29 presents the distribution of respondents across each vulnerability characteristic and how they were categorised as not vulnerable or potentially vulnerable.

Table 25 Distribution over age

Age group	N	Vulnerability category
16-24	497	Not vulnerable
25-34	1,014	
35-44	882	
45-54	888	
55-64	484	Potentially vulnerable
65+	285	

Note: Unweighted data

Source: LE calculations based on online experiment data

Table 26 Distribution over digital literacy score

Score	N	Vulnerability category
9-15	68	Potentially vulnerable
16-21	170	
22-27	468	
28-33	678	
34-39	1,101	Not vulnerable
40-45	943	
46-51	436	
52-57	115	
58-63	32	
64-72	39	

Note: Unweighted data

Source: LE calculations based on online experiment data

Table 27 Distribution over educational attainment

Educational category		Vulnerability category
Low	978	Potentially vulnerable
Medium	1,393	Not vulnerable
High	1,650	
Not known	29	

Note: Unweighted data

Source: LE calculations based on online experiment data

Note in that in Table 28 below, any respondent mentioning "To a great extent" for any category is identified as potentially vulnerable. Respondents could be classified as potentially vulnerable due to multiple categories.

Table 28 Distribution over vulnerability self-assessment

Vulnerable because of…	To a great extent	To some extent	Hardly at all	Not at all
Health problems	143	636	773	2,498
Financial or employment circumstances	45	1,302	1,013	1,290
Offers, terms or conditions being too complex	361	1,533	1,095	1,061
Age	156	674	1,069	2,151
Other reasons	201	740	1,013	2,096

Note: Unweighted data

Source: LE calculations based on online experiment data

Table 29 Distribution over risk aversion

Risk aversion	N	Vulnerability category
Not at all willing to take risks	378	Not vulnerable
Not very willing to take risks	2,294	
Fairly willing to take risks	1,240	Potentially vulnerable
Very willing to take risks	138	

Note: Unweighted data

Source: LE calculations based on online experiment data

Table 30 provides the tetrachoric correlation between the five vulnerability characteristics. Since all characteristics are identified using binary variables, this type of correlation is appropriate. A positive correlation in the table below shows that a respondent that has one vulnerability is also more likely to have another vulnerability.

Table 30 Tetrachoric correlation between vulnerability characteristics

	Age	Digital literacy	Education	Self-assessment	Risk aversion
Age					
Digital literacy	-0.249				
Education	-0.098	-0.101			
Self-assessment	-0.132	0.100	0.099		
Risk aversion	-0.100	0.157	0.033	-0.012	

Note: Unweighted data

Source: LE calculations based on online experiment data

3.3.2 Experimental results

Table 31 to Table 33 present the treatment effects, with respect to the effect on the intention to use ADR and understanding of ADR, of treatments on the trader websites in the online experiment for each group of consumers. Treatment effects are first presented relative to the baseline (TO1) and then presented as differences within vulnerable groups.

Table 31 Effects of TO2, TO3 & TO4 vs. TO1 on intention to use ADR, by group (open question responses)

a) Treatment effects for different vulnerable characteristics[a]

	TO2	TO3	TO4
Full sample	7.63pp (< 0.001) [2,012]	7.04pp (< 0.001) [2,060]	2.63pp (0.010) [2,012]
Over 55	7.59pp (< 0.001) [396]	5.76pp (< 0.001) [356]	1.46pp (0.4453) [418]
Under 55	7.74pp (< 0.001) [1,616]	7.26pp (0.0103) [1,704]	3.01pp (0.0105) [1,594]
Less digitally literate	7.02pp (< 0.001) [702]	7.23pp (< 0.001) [708]	4.51pp (0.0107) [708]
Digitally literate	8.01pp (< 0.001) [1,310]	6.97pp (< 0.001) [1,352]	1.64pp (0.1909) [1,304]
Less educated	9.52pp (< 0.001) [480]	3.10pp (0.1430) [502]	1.17pp (0.5629) [504]
Educated	6.98pp (< 0.001) [1,532]	8.50pp (< 0.001) [1,558]	3.20pp (0.0071) [1,508]
Vulnerable (self-assessed)	4.16pp (0.0567) [404]	2.20pp (0.2818) [436]	2.06pp (0.3084) [452]
Less vulnerable (self-assessed)	8.48pp (< 0.001) [1,608]	8.38pp (< 0.001) [1,624]	2.83pp (0.0169) [1,560]
Less risk averse	9.74pp (< 0.001) [1,292]	8.92pp (< 0.001) [1,374]	3.40pp (0.0070) [1,346]

More risk averse	3.91pp (0.0307) [720]	3.39pp (0.0629) [686]	1.13pp (0.5209) [666]

[a] Table shows for each group, the treatment effect against TO1 in percentage point (pp), the p-value of the z-test that this treatment effect differs from zero (in parentheses) and the base of each group (N) [in square brackets] Where "N/A" is given in space of the p-value, a pairwise z-test was not conducted as the chi-square test of homogeneity found statistical insignificance across all treatments.

b) Differences in treatment effect within groups of vulnerable characteristics[b]

	TO2	TO3	TO4
Age	-0.186 (0.494)	-0.043 (0.877)	0.019 (0.948)
Digital literacy	0.033 (0.872)	-0.082 (0.685)	-0.189 (0.375)
Education	-0.175 (0.43)	0.462 (0.047)	0.210 (0.383)
Self-assessed vulnerability	0.287 (0.24)	0.445 (0.071)	0.058 (0.816)
Risk aversion	0.393 (0.049)	0.446 (0.027)	0.194 (0.359)

[b] Table shows for each group, the coefficient for the interaction between the group and the respective treatment from a logit regression. The p-value of the coefficient is provided in parentheses. The logit regression included a full set of dummies capturing the treatments (baseline: TO1), a full set of dummies capturing both the potentially vulnerable and non-vulnerable group (baseline: the non-vulnerable group) and the full set of interactions between the two.

Source: LE calculations based on online experiment

Table 32 Effects of TO2, TO3 & TO4 vs. TO1 on intention to use ADR, by group (closed question responses)

a) Treatment effects for different vulnerable characteristics[a]

	TO2	TO3	TO4
Full sample	11.48pp (< 0.001) [1,006]	7.96pp (< 0.001) [1,030]	0.60pp (0.778) [1,006]
Over 55	10.09pp (0.0370) [198]	5.69pp (0.2469) [178]	-5.81pp (0.1895) [209]
Under 55	11.91pp (< 0.001) [808]	8.37pp (< 0.001) [852]	2.22pp (0.3553) [797]
Less digitally literate	11.26pp (0.0030) [351]	8.36pp (0.0267) [354]	2.78pp (0.4537) [354]
Digitally literate	11.57pp (< 0.001) [655]	7.74pp (0.0035) [676]	-0.65pp (0.8025) [652]
Less educated	7.29pp (*N/A*) [240]	1.29pp (*N/A*) [251]	-4.81pp (*N/A*) [252]
Educated	12.96pp (< 0.001) [766]	10.38pp (< 0.001) [779]	2.57pp (0.2878) [754]
Vulnerable (self-assessed)	7.43pp (*N/A*) [202]	4.69pp (*N/A*) [218]	-0.54pp (*N/A*) [226]

Less vulnerable (self-assessed)	12.45pp (< 0.001) [804]	8.86pp (0.0003) [812]	0.99pp (0.6853) [780]
Less risk averse	14.05pp (< 0.001) [646]	9.79pp (0.0002) [687]	0.14pp (0.9578) [673]
More risk averse	6.96pp (*N/A*) [360]	4.41pp (*N/A*) [343]	1.57pp (*N/A*) [333]

[a] Table shows for each group, the treatment effect against TO1 in percentage point (pp), the p-value of the z-test that this treatment effect differs from zero (in parentheses) and the base of each group (N) [in square brackets]. Where "N/A" is given in space of the p-value, a pairwise z-test was not conducted as the chi-square test of homogeneity found statistical insignificance across all treatments.

b) Differences in treatment effect within groups of vulnerable characteristics[b]

	TO2	TO3	TO4
Age	-0.098 (0.685)	-0.189 (0.452)	-0.462 (0.072)
Digital literacy	-0.051 (0.797)	0.001 (0.994)	0.096 (0.635)
Education	-0.236 (0.279)	-0.410 (0.066)	-0.362 (0.105)
Self-assessed vulnerability	-0.205 (0.376)	-0.216 (0.349)	-0.114 (0.627)
Risk aversion	-0.213 (0.278)	-0.239 (0.288)	0.101 (0.618)

[b] Table shows for each group, the coefficient for the interaction between the group and the respective treatment from a logit regression. The p-value of the coefficient is provided in parentheses. The logit regression included a full set of dummies capturing the treatments (baseline: TO1), a full set of dummies capturing both the potentially vulnerable and non-vulnerable group (baseline: the non-vulnerable group) and the full set of interactions between the two.

Source: LE calculations based on online experiment

Table 33 Effects of TO2, TO3 & TO4 vs. TO1 on understanding of ADR, by group

a) Treatment effects for different vulnerable characteristics[a]

	TO2	TO3	TO4
Full sample	0.13 (0.004) [1,006]	0.11 (0.015) [1,030]	0.07 (0.144) [1,006]
Over 55	0.09 (*N/A*) [198]	0.06 (*N/A*) [178]	0.00 (*N/A*) [209]
Under 55	0.14 (0.0057) [808]	0.12 (0.0145) [852]	0.08 (0.1062) [797]
Less digitally literate	0.03 (*N/A*) [351]	0.05 (*N/A*) [354]	0.02 (*N/A)* [354]
Digitally literate	0.17 (0.0023) [655]	0.14 (0.0148) [676]	0.08 (0.1566) [652]
Less educated	0.23 (*N/A*) [240]	0.10 (*N/A*) [251]	0.06 (*N/A*) [252]
Educated	0.09 (*N/A*)	0.12 (*N/A*)	0.07 (*N/A*)

	[766]	[779]	[754]
Vulnerable (self-assessed)	0.05 (*N/A*) [202]	0.02 (*N/A*) [218]	0.13 (*N/A*) [226]
Less vulnerable (self-assessed)	0.15 (0.0036) [804]	0.14 (0.0081) [812]	0.05 (0.3624) [780]
Less risk averse	0.13 (*N/A*) [646]	0.11 (*N/A*) [687]	0.04 (*N/A*) [673]
More risk averse	0.13 (*N/A*) [360]	0.12 (*N/A*) [343]	0.12 (*N/A*) [333]

[a] Table shows for each group, the treatment effect against TO1 in percentage point (pp), the p-value of the z-test that this treatment effect differs from zero (in parentheses) and the base of each group (N) [in square brackets] Where "N/A" is given in space of the p-value, a pairwise z-test was not conducted as the chi-square test of homogeneity found statistical insignificance across all treatments.

b) Differences in treatment effect within groups of vulnerable characteristics[b]

	TO2	TO3	TO4
Age	0.070 (0.537)	0.074 (0.521)	0.103 (0.36)
Digital literacy	0.130 (0.168)	0.105 (0.265)	0.078 (0.407)
Education	-0.129 (0.219)	0.024 (0.814)	-0.010 (0.924)
Self-assessed vulnerability	0.092 (0.402)	0.120 (0.266)	-0.092 (0.39)
Risk aversion	-0.003 (0.973)	-0.023 (0.799)	-0.101 (0.285)

[b] Table shows for each group, the coefficient for the interaction between the group and the respective treatment from a logit regression. The p-value of the coefficient is provided in parentheses. The logit regression included a full set of dummies capturing the treatments (baseline: TO1), a full set of dummies capturing both the potentially vulnerable and non-vulnerable group (baseline: the non-vulnerable group) and the full set of interactions between the two.

Source: LE calculations based on online experiment

Table 34 and Table 35 present the treatment effects (relative to the baseline EO1), with respect to the effect on the decision to go to ADR and understanding of ADR, of treatments on the ADR entity websites in the online experiment for each group of consumers.

Table 34 Effects of EO2, EO3 & EO4 vs. EO1 on the decision to go to ADR from the ADR entity website, by group

	EO2	EO3	EO4
Full sample	-1.25pp (*N/A*) [1,021]	1.60pp (*N/A*) [1,004]	0.24pp (*N/A*) [1,025]
Over 55	-3.95pp (*N/A*) [189]	1.15pp (*N/A*) [192]	0.66pp (*N/A*) [188]
Under 55	-0.45pp (*N/A*) [832]	1.83pp (*N/A*) [812]	0.37pp (*N/A*) [837]
Less digitally literate	-2.74pp (*N/A*) [346]	3.14pp (*N/A*) [332]	0.92pp (*N/A*) [359]
Digitally literate	-0.46pp (*N/A*) [675]	0.81pp (*N/A*) [672]	-0.13pp (*N/A*) [666]
Less educated	-2.61pp (*N/A*) [245]	6.38pp (*N/A*) [239]	-4.96pp (*N/A*) [237]
Educated	-0.93pp (*N/A*) [776]	-0.28pp (*N/A*) [765]	1.77pp (*N/A*) [788]
Vulnerable (self-assessed)	-4.64pp (0.3342) [223]	4.34pp (0.3804) [196]	8.35pp (0.0754) [224]
Less vulnerable (self-assessed)	-0.67pp (*N/A*) [798]	0.11pp (*N/A*) [808]	-2.35pp (*N/A*) [801]
Less risk averse	-4.08pp (*N/A*) [668]	1.01pp (*N/A*) [674]	1.57pp (*N/A*) [665]
More risk averse	4.33pp (*N/A*) [353]	2.82pp (*N/A*) [330]	-1.63pp (*N/A*) [360]

Table shows for each group, the treatment effect against EO1 in percentage point (pp), the p-value of the z-test that this treatment effect differs from zero (in parentheses) and the base of each group (N) [in square brackets] Where "N/A" is given in space of the p-value, a pairwise z-test was not conducted as the chi-square test of homogeneity found statistical insignificance across all treatments.

Source: LE calculations based on online experiment

Table 35 Effects of EO2, EO3 & EO4 vs. EO1 on understanding following the ADR entity website, by group

	EO2	EO3	EO4
Full sample	-0.09 (*N/A*) [1,021]	-0.06 (*N/A*) [1,004]	-0.01 (*N/A*) [1,025]
Over 55	-0.13 (*N/A*) [189]	-0.06 (*N/A*) [192]	0.07 (*N/A*) [188]
Under 55	-0.08 (*N/A*) [832]	-0.06 (*N/A*) [812]	-0.02 (*N/A*) [837]
Less digitally literate	0.01 (*N/A*) [346]	-0.06 (*N/A*) [332]	0.07 (*N/A*) [359]
Digitally literate	-0.14 (*N/A*) [675]	-0.06 (*N/A*) [672]	-0.05 (*N/A*) [666]
Less educated	-0.12 (*N/A*) [245]	-0.06 (*N/A*) [239]	-0.18 (*N/A*) [237]
Educated	-0.08 (*N/A*) [776]	-0.06 (*N/A*) [765]	0.05 (*N/A*) [788]
Vulnerable (self-assessed)	-0.01 (*N/A*) [223]	-0.04 (*N/A*) [196]	0.15 (*N/A*) [224]
Less vulnerable (self-assessed)	-0.11 (*N/A*) [798]	-0.07 (*N/A*) [808]	-0.06 (*N/A*) [801]
Less risk averse	-0.12 (0.0439) [668]	0.04 (0.5277) [674]	0.08 (0.1987) [665]
More risk averse	-0.03 (0.6894) [353]	-0.25 (0.0036) [330]	-0.16 (0.0616) [360]

Table shows for each group, the treatment effect against EO1 in percentage point (pp), the p-value of the z-test that this treatment effect differs from zero (in parentheses) and the base of each group (N) [in square brackets] Where "N/A" is given in space of the p-value, a pairwise z-test was not conducted as the chi-square test of homogeneity found statistical insignificance across all treatments.

Source: LE calculations based on online experiment

4 Further insights from the experiment and survey

4.1 Trader website closed question responses (online)

Table 36 presents the distribution of responses to the closed question respondents answered following viewing the trader website by treatment variant. Respondents were asked to select the option that best reflected the next steps they would take, with the proportion selecting ADR being used as an outcome measure.

Table 36 Distribution of responses to the closed question by treatment

	Baseline (TO1)	Separate ADR information (TO2)	Salience of ADR information (TO3)	Reducing information overload (TO4)
Do nothing	1.7%	2.5%	2.6%	3.6%
Complaint to the trader again	31.1%	24.0%	25.2%	26.5%
Take the trader to court	5.4%	4.4%	5.5%	5.2%
Contact a consumer organisation	25.3%	20.8%	21.6%	27.9%
Use alternative dispute resolution	31.2%	42.7%	39.2%	31.8%
Do something else	1.1%	2.9%	1.8%	1.7%
Don't know	4.2%	2.6%	4.1%	3.2%

Source: LE calculations based on online experiment

4.2 Funding source of ADR

Table 37 presents the distribution of responses to the question regarding consumers' impressions of ADR under different funding sources (see section 1.3.2). Results of the χ^2 tests are also presented.

Table 37 Impression of ADR entity under different funding sources

	Government	Trader	Trade association
Percentage results			
Positive, of which:	**68%**	**62%**	**63%**
Very positive	22%	20%	18%
Somewhat positive	47%	42%	45%
Neither positive nor negative	**28%**	**30%**	**31%**
Negative, of which:	**4%**	**8%**	**5%**
Somewhat negative	3%	7%	4%
Very negative	1%	1%	1%
N	**1,353**	**1,368**	**1,329**
χ^2 tests[a]			
Government			
Trader	0.0001		
Trade association	0.0262	0.0600	

[a] Shown are the p-value of a χ^2 test of independence between the two funding sources related to the cell, e.g. the bottom-left cell refers to a χ^2 test comparing "government" and "trade association" as a funding source. Since the χ^2 test is symmetric, the p-value for each combination is only provided once.

Source: LE calculations based on online experiment

Index of Tables, Figures and Boxes

Tables

LE Europe

35 rue du Congrès,
1000 Bruxelles, Belgique
info@londoneconomics.co.uk
www.le-europe.eu
@le_europe
+32 2 229 19 02