How to improve consumers' understanding of online legal information: insights from a behavioral experiment

Alexander J. Wulf[1] · Ognyan Seizov[1]

Abstract

Past research has shown that online information notices often fail to inform consumers well, even if transparency-enhancing measures are implemented. However, the studies in question have employed research designs that were restricted to pre-contract conclusion scenarios and ad hoc, text-only attempts to optimize disclosures. While these results point to the general limitations of disclosures, they leave open whether optimizing information notices can be of substantial value to consumers in other settings. Our study tests the effectiveness of multimodal disclosure optimization techniques in both the pre- and post-contract conclusion scenarios. The post-contract conclusion scenario is the situation where a consumer has a dispute with a business. While this setting is not the primary target of disclosure legislation, it is a more realistic instance of the actual use of legal information online. Here the consumer has a real incentive to obtain information about his or her rights and obligations. We show that under these conditions, consumers do in fact read, retain and understand more when the attempt has been made to optimize disclosures.

1 Introduction

Information disclosures have been an essential part of the legislator's regulatory toolbox for decades. Protecting consumers by providing information and restoring the knowledge imbalance between businesses and consumers by means of detailed disclosures is an approach that is universally appealing from both the theoretical and political perspectives. As a result, there has been a massive increase in the use of

✉ Alexander J. Wulf
alexander.wulf@srh.de

1 SRH Berlin University of Applied Sciences, Berlin, Germany

these strategies in various jurisdictions (Ben-Shahar & Schneider, 2014). However, the implementation of information obligations has been fraught with challenges. The breadth and depth of most disclosures render them unintelligible to most consumers (Bakos et al., 2014) and the typical vocabulary of *terms and conditions* and *privacy policies* is ambiguous and obfuscating (Pollach, 2005). Consumers lack the literacy and numeracy to engage with disclosure texts (Mak, 2012) and they often resort to other, easier ways of evaluating a business's integrity, such as reputation, user reviews and recommendations or endorsements (Furnell & Phippen, 2012). Businesses therefore often end up disclosing vast amounts of information in complicated legal language, more for the sake of compliance than for the benefit of the consumer (Becher & Benoliel, 2023; Milne & Culnan, 2004). Worse yet, some businesses deliberately misrepresent legal statutes, thus unfairly disadvantaging consumers, so that even if the disclosures are read, they do not serve their intended purpose to inform and protect (Furth-Matzkin, 2017; Wilkinson-Ryan, 2017). In a nutshell, efforts to meet information obligations consistently fall short of what they were intended to achieve, and yet there is no clear consensus on a feasible improvement strategy.

Instead of cooperating in the search for a solution, scholarly research and policy seem to be pursuing two separate paths. On the one hand, a multitude of studies from empirical and doctrinal legal research, behavioral economics, communication science, and information design have demonstrated the shortcomings of information disclosures. On the other hand, policymakers are sticking to the approach of demanding the provision of information and thus 'more of the same'. As discussed elsewhere (Seizov et al., 2019), in Europe, for example, the scope of disclosure mandates is constantly expanding, while the transparency requirements for the resulting information notices have remained on the level of platitudes, e.g. 'clear and comprehensible language', since the 1990s.[1]

In the present study we address shortcomings of the current state of research on disclosures and disclosure optimization. Past research has shown that online information notices often fail to inform consumers well, even if transparency-enhancing measures are implemented (e.g. Ben-Shahar & Chilton, 2016; Elshout et al., 2016; Furnell & Phippen, 2012; Marotta-Wurgler, 2012). However, the studies in question have employed research designs that were restricted to pre-contract conclusion scenarios and ad hoc, text-only attempts at disclosure optimization. While these results point to the general limitations of disclosures, they do not show that optimizing information notices does not lead to an increase in value to consumers in other settings.

Our study tested the effectiveness of multimodal disclosure optimization techniques in the pre- and post-contract conclusion scenario. Multimodal optimization techniques go beyond text-only edits. More engaging and effective messages can be delivered if coherent combinations of several communication modes (i.e., text,

---

[1] See, for example, the Council Directive dated 13 June 1990 on package travel, package holidays and package tours (90/314/EEC) OJ L 158/59 (Package Travel Directive) and Directive 97/7/EC of the European Parliament and of the Council of 20 May 1997 on the protection of consumers in respect of distance contracts OJ L 144/19 (Distance Selling Directive).

visuals, graphics, color, etc.) are used. In Seizov and Wulf (2020) we presented techniques that can be used to improve the transparency of online disclosures. These guidelines are based on empirical insights from a variety of disciplines, as outlined in Seizov et al. (2019).

Testing the effectiveness of disclosure optimization techniques in the pre- and post-contract conclusion scenarios is a novel approach (Wulf & Seizov, 2020). The dominant assumption is that the use case for disclosures is consumers in a pre-contract conclusion scenario. This is understandable given that this premise is also at the heart of disclosure legislation, which aims to protect consumers through information and to prevent them from entering into unfavorable transactions and suffering unexpected negative consequences. As mentioned above, previous empirical evidence shows that in this setting the quality of standard information notices has only limited relevance for consumers. This finding is not surprising: if consumers have no incentive to read disclosures, the transparency of the information they are given is irrelevant. We argue that the pre-contract conclusion scenario alone is therefore an inadequate setting for empirically testing the effectiveness of disclosure optimization techniques. Data collected in this setting cannot tell us whether transparency-enhancing measures do in fact make a difference if consumers have a real incentive to read the legal information provided. We therefore do not know whether improving transparency is helpful to consumers in other situations, most importantly in the post-contract conclusion scenario, where consumers have a dispute with a business and require information about their rights and obligations. Although this setting does not correspond to the primary legislative intent for disclosures, it is a more realistic instance of the actual use of online legal information by consumers, as here they have a real incentive to read legal information. In the research presented in this paper we show that in this setting, reading rates, retention and consumer understanding are improved if disclosures are optimized. Our results therefore provide arguments in support of requiring online legal information to be transparent by law and of enforcing these laws. Drawing on our results, we conclude with some recommendations for revising disclosure policy.

The remainder of this paper is structured as follows: in Section B we review empirical studies on the effectiveness of online disclosures and studies on disclosure optimization techniques. Section C introduces our behavioural experiment (N = 835), which examines how a non-transparent disclosure performs against ones that have been optimized textually, visually, and by multimodal techniques, respectively, in the pre- and post-purchase scenarios. In Section D we present our findings, which indicate that enhancing the content of a disclosure linguistically, visually and multimodally improves consumer understanding in the pre-purchase scenario, and even more so in the post-purchase scenario. Section E concludes the paper with a discussion of our results and their implications for more effective online information disclosure and consumer protection.

2 Empirical research on disclosure effectiveness

The overwhelming majority of previous studies on the effectiveness of information disclosures, both offline and online, rest on an *ex-ante* premise, i.e. that it is best to 'read before you proceed'. This premise is also at the heart of disclosure legislation, which aims to protect consumers through information and prevent them from entering into unfavorable transactions and from suffering unexpected negative consequences. In today's fast-paced online world, terms and conditions, privacy policies, and other significant contractual information generally remain unheeded. This fact has led legal scholars such as Ben-Shahar and Schneider (2014), Nordhausen Scholes (2009), Helberger (2013) and Marotta-Wurgler (2012) to cast disclosures in a rather negative light. At the same time, legislators around the world frequently rely on disclosures as a regulatory tool. For example, the European Commission consistently relies on disclosures as a central pillar of its consumer protection and information agenda, as evidenced by a series of Directives and Regulations adopted from the early 1990s to the present day.[2]

There is a broad consensus among scholars that information disclosures are generally presented in overly complex, inaccessible forms and that there is no single solution to this problem. Lotter (2019) speaks of the benefits of "unlocking complexity", rather than merely reducing it. This may be achieved by emphasizing interconnections between data and by motivating individuals to engage with it, rather than fear it. He uses an apt forestry metaphor to describe the age-old urge to reduce complexity: "If you can no longer see the forest for the trees, you get out a chainsaw and cut until only a single tree remains. The forest is no more, but everything is nice and tidy" [translated from German, p. 34]. A viable alternative to simplification would therefore be to "unlock the complexity" of online information notices by emphasizing the interconnections between their different elements, changing how they are framed and improving consumers' motivation to engage with them.

As Lotter (2019) has pointed out, simplification is a natural initial response whenever we encounter complexity. Accordingly, many empirical studies have tested the effects of simplifying disclosures on reading rates and consumer understanding. Marotta-Wurgler (2012) used clickstream data to analyze the online purchasing choices of almost 48,000 software buyers and concluded that standard (or 'boilerplate') contract terms received precious little attention from consumers and played no role in their final decisions to purchase a piece of software. Neither increasing the visibility of the terms, making consent before a purchase mandatory nor making the terms buyer-friendly had any influence on consumers' choices, illustrating the

---

[2] Cf. esp. Council Directive of 13 June 1990 on package travel, package holidays and package tours (90/314/EEC) OJ L 158/59 (Package Travel Directive); Directive 97/7/EC of the European Parliament and of the Council of 20 May 1997 on the protection of consumers in respect of distance contracts OJ L 144/19 (Distance Selling Directive); Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market OJ L 149/22 (Unfair Commercial Practices Directive); Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on consumer rights OJ L 304/64 (Consumer Rights Directive). For further details see also Seizov et al. (2019) and Wulf and Seizov (2020).

failure of mandated disclosures to inform and protect. Elshout et al. (2016) carried out three experiments in 12 EU member states (N = 1000 participants per country) to test whether simplifying online terms resulted in an increase in the number of people who read them. They found that more people read the disclosures when they were moderately simplified and shortened. Moreover, a brief comprehension quiz also revealed that the simplified terms were somewhat better understood. The effect of inserting quality assurance cues from national and European consumer organizations was also tested. They were found to have generally beneficial effects on consumer trust in the different shopping scenarios.

Arguably the strongest empirical case against simplification was made by Ben-Shahar and Chilton (2016), who assembled a six-point list of disclosure improvements based on the dominant best practices in consumer information design and tested how various combinations of these recommendations affected understanding. The scenario they used presupposed high respondent sensitivity and attention insofar as the subjects were asked to provide extensive and highly intimate information on their sexual practices for the development of an online dating app. Relying on a representative sample of US adults (N = 1484), the authors found that no single improvement strategy alone nor any combination of strategies resulted in a statistically significant increase in reading time or understanding. Even when responding to highly intrusive questions regarding sensitive personal information, consumers failed to meaningfully engage with the information notices, as evidenced by the prohibitively short reading times and the small number of correct answers across all experimental conditions. While these results by Ben-Shahar and Chilton (2016), along with related research, seem to cast a rather negative light on simplification as a strategy for improving disclosures, it is important to note that virtually all attempts to introduce simplifications applied in previous studies have been textual in nature, largely ignoring a host of non-textual information design principles that have been shown to work in other contexts (Seizov et al., 2019). The present paper goes beyond textual measures to produce a disclosure that has been enhanced along multiple dimensions of document design and whose performance we have tested with regard to both reading time and understanding.

Another aspect of information disclosure that has been studied is consumer motivation and consumer attention. Darolia and Harper (2017) studied the behavior of US college students (N = 9802) taking out student loans. The researchers created the following experimental scenario: they supplied half their subjects with individually tailored 'debt letters' which contained exhaustive information on their supposed financial circumstances, including past debt, their current student loan options, the typical payback schedule and an overall assessment of how a further student loan would affect their financial status in the short to medium term. These subjects made essentially the same borrowing choices as students who did not receive any supplemental information. After in-depth interviews with 27 of the experimental subjects, the researchers concluded that improving the way in which such information is presented cannot incentivize disinterested recipients to read it. Conversely, in a study that looked at how alcohol consumers paid attention to brand labels and health warnings, Kersbergen and Field (2017) found that even subjects who were highly motivated to reduce their alcohol intake did not read and understand the health warnings

printed on containers of alcoholic beverages. The authors nevertheless concluded that the content of the warnings should be improved to better engage motivated consumers. Thus, consumer motivation and information design go hand in hand when it comes to consumer understanding.

Eye-tracking has also been used effectively to demonstrate the relationship between consumers' visual attention and information processing relating to the product. For example, Reale and Flint (2016) conducted a study of the menu choices of subjects (N = 84) primed with different health-related information designs. The findings indicated that using non-textual attention markers (e.g., color coding, logos) promoted healthier choices, confirming that strategies that go beyond the textual realm can raise readers' motivation to read and learn. Visschers et al. (2010) used eye-tracking to study the visual attention paid by shoppers (N = 32) to nutrition labels and concluded that only a combination of health motivation and label design can effectively direct consumer attention and improve understanding. The authors therefore concluded that optimizing both the content and the presentation of the information supplied can have positive motivational and framing effects on consumers, resulting in greater engagement and improved understanding.

There is a considerable literature on the effectiveness of text-only versus visual disclosures used to warn consumers about the effects of tobacco. Beginning with Iceland in 1969, countries around the world adopted mandatory textual health warnings on cigarette packs over the following years and decades, and many of them later supplemented these messages with 'shock images', again following the lead of Iceland, which started doing so in 1985 (Hiilamo et al., 2014). These legislative changes have provided ample scope for empirical research into the effectiveness of alternative modes of consumer education. Pictorial warnings have generally proved to discourage smoking more effectively than text-only messages (see Fong et al., 2009 for an overview of the findings). Specifically within the law and economics literature, Jolls (2013) assessed the effects of the cigarette pack warnings mandated by the US Family Smoking Prevention and Tobacco Control Act of 2009 on smokers' factual misperceptions about the risks associated with their tobacco consumption. Here, too, combined text-and-image warnings were found to have a stronger impact on the survey respondents' attitudes than text-only warning messages.

Consumer motivation can also be effectively manipulated through the context in which information is received, i.e. the conditions in which it is read. The pre-contractual use case for disclosures that policymakers usually have in mind assumes that consumers exhibit high levels of intrinsic motivation which compels them to read a non-transparent information notice attentively from start to finish, even though at that stage, there is no clearly defined learning goal or tangible outcome. Such a context is not particularly conducive to the intake of information. Psychological research teaches us that structured, purposeful reading with a clear aim and feedback in the form of knowledge tests produces much better learning results, which the educational psychology literature refers to as the 'testing effect'. Rowland (2014) conducted a meta-analysis of several decades of experimental research on the testing effect, focusing especially on comparing the learning outcomes of studying with testing vs. studying without testing. The author concluded that the purposeful processing of information, i.e. highly

motivated reading, is a key component of improved understanding and retention, an observation that various classroom studies have confirmed (Pyc & Rawson, 2010; Rawson & Dunlosky, 2011; Roediger et al., 2010). The current information disclosure system thus largely sets consumers up to fail. It requires them to read and 'learn' vast amounts of information without a tangible goal (e.g. a knowledge check, a reward or any other form of feedback) and it does not stimulate their information retrieval mechanisms in any way. By contrast, a post-contract conclusion scenario for the use of online disclosures during a dispute with a seller can capitalize on the testing effect and improve consumers' motivation and ability to learn about their rights and obligations.

The testing effect is further augmented by the difficulty of recalling or retrieving the information (Bjork & Bjork, 1992; Karpicke & Roediger, 2007, 2008; Pyc & Rawson, 2009), the amount of thinking required to arrive at the correct answer (Carpenter, 2009), asking subjects to formulate the correct answer on their own vs. letting them chose from a predefined set of options (Anderson & Bower, 1972; Butler & Roediger, 2007; Carpenter & DeLosh, 2006) and administering a closed-book knowledge test requiring subjects to actively recall information (as opposed to an open-book test) (Agarwal & Roediger, 2011; Agarwal et al., 2008; Rummer & Schwede, 2019). While in a pre-contract conclusion scenario consumers also have all the necessary contractual information available to them, they often skim through information notices far too quickly to be able to learn anything from them (Marotta-Wurgler, 2012), fail to extract the necessary information (Furnell & Phippen, 2012) and underestimate or misunderstand the implications of blatantly unfavorable contract terms (Ben-Shahar & Chilton, 2016). We therefore included a post-purchase scenario in our experiment to test an alternative use case, and we also explored the effects of textual, visual and multimodal transparency enhancements on both pre- and post-purchase disclosure reading rates and understanding.

Finally, another branch of academic literature discounts the effectiveness of disclosures altogether, even when consumers do read them. In a study of several dozen apartment leases in Massachusetts, Furth-Matzkin (2017) identified numerous unenforceable terms which, intentionally or not, disadvantaged the tenants by misrepresenting their rights and imposing undue burdens. The author contends that in the event of a dispute, most tenants perceive lease terms as binding and do not consider challenging them, thus foregoing their legal rights. A study by Wilkinson-Ryan (2017) confirmed these fears. In two experiments, the author found that "subjects believed policies embedded in a contract were more likely to be legally enforceable, judged those policies as more fair, and imagined that they would be less likely to challenge those policies in court". Furth-Matzkin and Sommers (2020) arrived at a similar observation: even in cases where consumers entered into a contract based on false promises or deceptive advertising, they were unlikely to take legal action or even to express their discontent publicly because they blindly believed in the enforceability of small print. Thus if information disclosures are to be improved, better oversight and legal enforcement and more effective consumer education are also required.

With this in mind, the rest of this paper focuses on the first obstacle to better consumer protection through information, i.e. the problem of improving the reading rates and comprehensibility of online consumer disclosures. Our contribution to the

literature consists in (a) a comparative test of the effectiveness of information notices in a pre- and a post-purchase scenario and (b) a test of empirically motivated, comprehensive textual, visual and multimodal optimization. Our research both draws on multidisciplinary expertise in creating clear and effective modes of communication, which has been lacking in previous disclosure improvement efforts, and incorporates into the experimental design the higher levels of motivation that consumers experience during a dispute with an online seller. We thus for the first time provide a novel test of evidence-based optimization techniques in two real-life scenarios.

3 Research design

We designed a behavioral experiment that immersed participants in either of two scenarios, to which they were assigned on a random basis. In the pre-purchase scenario, half of them were shopping for a set of custom-made drinking glasses for a friend's birthday. The ordering process was designed to be as realistic as possible, including selecting the glasses, stating the shipping address, providing a discount code (if available) and selecting a shipping method. Once the purchasing process was completed, the participants were automatically taken to the online shop's Terms and Conditions. In the pre-purchase scenario we refrained from nudging the participants to read the information notice, so that they would behave as they normally would in an online shopping situation. In the post-purchase scenario, we told the other half of the participants that they had already purchased a set of custom-made drinking glasses to give to a friend as a birthday present. Regrettably the glasses had arrived too late for the occasion, as shipping took five business days, rather than one or two, as advertised. To raise their motivation to read the contract information, we highlighted the possibility that the subjects could request some kind of compensation. To give them an incentive to process the ensuing information notice in a purposeful way, we explicitly stated that we would ask them a few knowledge questions regarding their purchases. We then took the participants to the Terms and Conditions of the online shop, which were identical to those used in the pre-purchase scenario. Participants in the post-purchase conditions were thus given several reasons to read the information notice: they were facing a problem, they might be entitled to compensation and they would have to answer questions about the disclosure. Through these multiple incentives we aimed to recreate the high motivation levels that consumers would have in a real-life post-purchase dispute.

In both the pre- and the post purchase scenarios we randomly presented the subjects with one of four variants of the Terms and Conditions ('disclosure types'): A—a densely written text-only disclosure ('non-transparent disclosure'), B—a linguistically optimized and neatly structured text-only disclosure ('textually optimized disclosure'), C—a visually formatted one-pager that presented the main contractual stipulations in graphic form ('visually optimized disclosure'), or D—a combination of C and then B ('visually and textually optimized disclosure'). The 'non-transparent disclosure; pre-purchase' was adopted from Elshout et al. (2016), with some minor modifications to fit our experimental conditions. We then optimized that disclosure

to create the three other versions, following the multidisciplinary guidelines presented in Seizov et al. (2019) and Seizov and Wulf (2020).

We recorded the time that the subjects spent on reading the disclosures, with no time limit being imposed. Once the participants had clicked 'Next', there was no going back to the disclosure. Having viewed one of the four disclosures in either of the two purchasing scenarios, the participants were asked three multiple-choice knowledge questions based on the information contained in the Terms and Conditions. One of them concerned the post-purchase problem (the right to compensation in case of a shipping delay), while the other two were of a more general nature (product defects and returns). Together, the questions addressed three core concerns consumers have in e-commerce settings: when will my product arrive, what if it is defective, and how long can I return it? In order not to force participants to guess and thus risk inadvertently correct responses, we offered a 'Don't know' answer option.

The experiment thus aimed to test how, in both a pre- and a post-purchase scenario, linguistic, visual and multimodal disclosure optimization affects the participants' (1) attention span when reading information disclosures and (2) understanding and retention of information disclosures. We summarize our eight experimental conditions in Table 1.

3.1 Hypotheses on reading time

We first tested whether our textual and visual transparency enhancements helped participants to grasp the content of the disclosures faster than the dry, text-only form of information disclosures that is currently the standard on the internet.

Hypothesis I-a: The more transparent a disclosure, the less time consumers need to engage with it, as evidenced by lower reading times. We hypothesize that the ranking of the disclosure types in terms of transparency is $D > C > B > A$. Accordingly, the ranking in terms of reading time should be $C < B < A$.

Note disclosure Type D is not included in the hypothesized ranking of reading time because it is a compound of two different formats and thus longer than any of the other disclosures. There is thus no clear theoretical basis for deciding where to include it in the ranking.

In previous research, which utilizes pre-purchase, low-motivation scenarios and limited, text-only disclosure modifications, disclosure reading rates and times were uniformly low. In the post-purchase, high-motivation scenario that we also tested, the participants were incentivized to read the terms and conditions thoroughly. We therefore also formulated:

Hypothesis I-b: Participants in the post-purchase scenario will pay more attention to the information disclosure than participants in the pre-purchase scenario, as evidenced by longer reading times.

Table 1 The eight conditions used in the behavioral experiment

Disclosure characteristics	1—"Pre-purchase"-Scenario *Participants make an online purchase and read the shop's Terms & Conditions as part of the checkout procedure*	2—"Post-purchase"-Scenario *Participants face a transaction problem and read the shop's Terms & Conditions to determine their consumer rights*
A—"Non-transparent" *tightly formatted, linguistically complex text*	A1: "Non-transparent disclosure; pre-purchase"	A2: "Non-transparent disclosure; post-purchase"
B—"Textually optimized" *linguistically optimized text with a clear layout*	B1: "Textually optimized disclosure; pre-purchase"	B2: "Textually optimized disclosure; post-purchase"
C—"Visually optimized" *visual one-pager with thematic icons and a minimum of text*	C1: "Visually optimized disclosure; pre-purchase"	C2: "Visually optimized disclosure; post-purchase"
D—"Visually and textually optimized" *combination of B and C*	D1: "Visually and textually optimized disclosure; pre-purchase"	D2: "Visually and textually optimized disclosure; post-purchase"

3.2 Hypotheses on consumer understanding

In the opinions of many legal experts, businesspeople and consumers (Seizov & Wulf, 2020), the dry, text-only form of information disclosures is at least partially responsible for the fact that they are rarely or only cursorily read and not very effective. Our textual and visual transparency enhancements went well beyond previous attempts to improve disclosures and promote information acquisition. Incorporating visual and graphical elements into the disclosure can harness the communicative power of multimodality and produce superior documents in terms of both consumer attention and learning outcomes. We thus formulated two further hypotheses.

Hypothesis II-a: The more transparent a disclosure, the better participants will understand it. In line with the transparency ranking purported above, we thus expect the following ranking according to the number of correct responses to the knowledge questions: $D > C > B > A$.

In previous research on the testing effect in learning situations, the purposeful processing of information in preparation for a test yielded better learning outcomes than unmotivated perusal of the same material—a principle that offers further support for our post-purchase transaction dispute scenario. We thus added:

Hypothesis II-b: In each of the four post-purchase scenarios, the participants will perform better on the knowledge test than the participants in the corresponding pre-purchase scenarios, as evidenced by the fact that they answered more knowledge questions correctly.

To test our hypotheses, we drew on the results of an online survey of 835 UK residents that we conducted as part of a larger research project. The participants were recruited by the leading service provider Prolific. We describe the sample in Table 2. While our sample was restricted to a single country, the average demographics exhibited by the respondents are quite similar to those of e-commerce participants across the EU (Eurostat, 2021), so arguably our findings can be generalized beyond the UK to some extent.

4 Findings

We first investigated each of our hypotheses using descriptive statistics, followed by a more thorough analysis using regression models. Since previous research has often first focused on reading time (Ben-Shahar & Chilton, 2016; Marotta-Wurgler, 2012), we also began by looking at how reading times varied across our eight experimental conditions (Hypothesis I—a & b).

4.1 Disclosure optimization and reading time

To investigate Hypothesis I—a & b, we first compared how long the respondents took to read each of the four different disclosure types in the pre- and post-purchase scenarios. The descriptive results are shown in Table 3. The mean and median reading times range from around half a minute to just over two minutes.

Table 2 Summary statistics on the sample

Sample distributions for qualitative variables

Gender	Male 33.2%	Female 66.5%	Other 0.4%		
Country of residence	England 84.2%	Scotland 7.6%	Wales 5.2%	N. Ireland 1%	Other 2.1%
Education (highest completed)	High school 39.8%	Bachelor 38.7%	Master 14.5%	Doctorate 2.8%	Vocational 2.6% / Other 1.7%
Occupation	Self-employed 9.9%	Employee 54.7%	Student 10.4%	Working Student 3.2%	Unemployed 11% / Other 10.7%
Frequency of online shopping	Very Rarely 0.4%	Rarely 1.9%	Sometimes 20.4%	Often 45.6%	Very Often 31.7%
Daily online time	<1 h 1%	1 to 3 h 27.4%	4 to 7 h 48%	>8 h 23.6%	
Government role in consumer protection	Let the market decide 5.1%	Policy with business input 70.5%	Policy without business input 24.3%		
Dispute with an online business in the past 2 years	Yes 41.2%	No 58.8%			

N=835 for all variables. Average age of participants: $M = 36.15$, $SD = 13.22$

Table 3 Reading times for each experimental scenario (in seconds)

Disclosure type/scenario	N	SD	Min	Mean	Median	Max
A1—Non-transparent; pre-purchase	104	66.85	3.61	65.47	38.91	344.92
A2—Non-transparent; post-purchase	103	102.87	7.93	132.21	118.90	690.44
B1—Textually optimized; pre-purchase	102	22.47	4.39	41.64	35.72	132.04
B2—Textually optimized; post-purchase	116	185.58	3.91	109.16	78.63	1895.54
C1—Visually optimized; pre-purchase	103	21.78	5.21	32.00	27.85	179.86
C2—Visually optimized; post-purchase	100	46.85	3.47	45.15	37.09	368.01
D1—Visually and textually optimized; pre-purchase	102	51.82	7.90	66.48	52.80	338.80
D2—Visually and textually optimized; post-purchase	105	172.70	14.41	127.06	91.05	1668.83

For the pre- and post-purchase scenarios, the mean and median reading times for the textually (Type B) and visually optimized (Type C) disclosures were lower than for the reference text, i.e. the non-transparent disclosure version (Type A). When we compared the times for the textually optimized (Type B) and visually optimized (Type C) disclosures, we found that the latter performed slightly better in the pre-purchase scenario and much better in the post-purchase scenario. In sum, the descriptive results thus support Hypothesis I-a, i.e. the ranking by reading time is Type C < Type B < Type A. The disclosure that had been optimized both visually and textually (Type D) required similar reading times as the non-transparent reference text (A), presumably since it combined two different formats and was thus longer than any of the other disclosures.

To establish descriptive statistics for Hypothesis I-b, we looked at how long the respondents took to read each of the four disclosure types in the pre- versus the post-purchase scenario. In support of Hypothesis I-b we found that the post-contractual reading time (attention) exceeded the pre-contractual reading time by a considerable margin for all disclosure types.

We then tested Hypothesis I—a & b more formally using OLS regression models. This enabled us to control for any confounding influences of the respondents' personal characteristics, as shown on an aggregate basis in Table 2 above. In each of the models summarized in Table 4 (below), the dependent variable is *Reading Time*. Since the residuals of Model (1) proved to be skewed and the Breusch-Pagan/Cook-Weisberg test (p-value < 0.01) indicated some heteroskedasticity, Models (2) and (3) use a logged dependent variable, which remedied both problems. We then tested for multicollinearity between all explanatory variables using Variance Inflation Factors, but found no sign of multicollinearity (mean VIF value < 2).[3] All models feature robust standard errors.

---

[3] The categories of two ordinal control variables (*Frequency of online shopping* and *Daily online time*) showed high VIF values, resulting in a mean VIF value of 10.71. In line with common practice (see, for example, Wulf 2014: 195–196), we therefore included these two control variables in our models as continuous rather than as categorical variables. This did not affect any of our results. However, it substantially reduced the high VIF values for these variables and led to a lower mean VIF value for our models.

Table 4 Reading time models

Dependent variable: Reading time (in seconds)	(1)	(2)	(3)
Post-purchase	52.12***	0.62***	0.46***
	(7.30)	(0.05)	(0.04)
Disclosure type B: Textually optimized	− 24.34*	− 0.27**	− 0.32***
	(11.12)	(0.08)	(0.07)
Disclosure type C: Visually optimized	− 61.84***	− 0.75***	− 0.82***
	(6.84)	(0.08)	(0.07)
Disclosure type D: Visually and textually optimized	0.73	0.11	− 0.10
	(11.05)	(0.08)	(0.07)
Observations	827	827	708
R-squared	0.13	0.29	0.32
Mean VIF	1.74	1.74	1.83
Log of dependent variable	No	Yes	Yes
Control variables	Yes	Yes	Yes

Constant term included but not reported. ***, ** and *indicate significance at the 0.1%, 1% and 5% levels, respectively. *Reading Time* is the reading time for a disclosure in seconds. The reference category for *Post-Purchase* is *Pre-Purchase*. The reference category for *Disclosure Type* is *A: Non-transparent*. Model (3) excludes outlier participants with reading times of less than 20 and more than 700 seconds. See Table 2 above for a description of the respondents' personal characteristics that are used as control variables

The regression results are highly consistent with the descriptive statistics presented above—which is to be expected, given that with sufficiently large samples and random assignment of the respondents to the treatments, their personal characteristics should not make any systematic difference. Hypothesis I-b is again supported in that the reading times in the post-purchase scenario are almost one minute longer than in the pre-purchase scenario (see Model (1), where the coefficient for *Post-Purchase* is highly significant at the 0.1% level). The significance level of this central result holds across all specifications. In Model (2), the dependent variable was log transformed for greater robustness. In Model (3), we also excluded outliers with very short (less than 20 s) and very long (more than 700 s) reading times, which improves the fit of the model but leaves the results largely unaffected.

Regarding Hypothesis I-a, the regression results also allowed us to gauge the extent to which the different disclosure types affected reading times. The results are consistent with the descriptive analysis of the sample means above. As evidenced by the statistically significant coefficients obtained across all specifications, *Reading Time* was shorter for the *textually optimized* disclosure (Type B), and shorter yet for the *visually optimized* disclosure (Type C). However, as already suggested by the descriptive results, the coefficient of the *visually and textually optimized* (Type D) disclosure is not statistically significant.

We then performed one-tailed Wald tests for the equality of the coefficients as a systematic test of Hypothesis I-a. The tests for the six pairs of coefficients were

Table 5 Number of correct answers for the different experimental scenarios

Disclosure type/scenario	N	SD	Min	Mean	Median	Max
A1—Non-transparent; pre-purchase	104	0.92	0	1.13	1	3
A2—Non-transparent; post-purchase	103	0.88	0	1.75	2	3
B1—Textually optimized; pre-purchase	102	0.98	0	1.37	1	3
B2—Textually optimized; post-purchase	116	0.88	0	1.88	2	3
C1—Visually optimized; pre-purchase	103	0.95	0	1.59	2	3
C2—Visually optimized; post-purchase	100	0.80	0	2.23	2	3
D1—Visually and Textually optimized; pre-purchase	102	1.05	0	1.80	2	3
D2—Visually and Textually optimized; post-purchase	105	0.80	0	2.22	2	3

based on Model (2).[4] They confirmed that the reading times were significantly different for the following pairs of coefficients: Type A vs. Type B (p-value 0.004), Type A vs. Type C (p-value < 0.001), Type B vs. Type C (p-value < 0.001), Type B vs. Type D (p-value < 0.001) and Type C vs. Type D (p-value < 0.001). However, there was no significant difference between the reading times for Type A and Type D (p-value 0.577). In sum, these tests support Hypothesis I-a insofar as the ranking by reading time is Type C < Type B < Type A.

4.2 Disclosure optimization and consumer understanding

Another crucial measure of the effectiveness of a disclosure is the amount of information that subjects obtain from it (Ben-Shahar & Schneider, 2014; Furnell & Phippen, 2012; Milne & Culnan, 2004). To investigate Hypothesis II–a we compared how the eight experimental conditions performed in a quick, three-question knowledge test of the information contained in the disclosed terms and conditions. Table 5 shows descriptive statistics on the number of correct answers to these questions.

For the textually optimized (Type B), visually optimized (Type C) and both visually and textually optimized (Type D) disclosures, the mean numbers of correct answers per scenario were higher than for the reference text, i.e. the non-transparent disclosure (Type A). The descriptive results thus support Hypothesis II-a that the ranking based on the number of correct responses to the knowledge questions is Type D > Type C > Type B > Type A, with the limitation that Types C and D performed roughly equally in the post-purchase scenario. Thus, the descriptive data indicate that respondents who read the optimized disclosures gave a higher number of correct answers in both scenarios.

To gauge the extent and possible benefits of the testing effect (Anderson & Bower, 1972; Carpenter, 2009; Rowland, 2014) in the post-contract conclusion scenario, we explored how each of the four post-purchase scenarios performed against their pre-purchase counterparts in regard to information retention (Hypothesis II-b). Since the mean number of correct answers in any post-purchase scenario was always

---

[4] We adjusted the p-values using the Bonferroni correction method to account for multiple comparisons. This was the most convenient and conservative way to control for the family-wise error rate in a way that is robust to heteroscedasticity and non-normality. Moreover, unlike other methods, this one does not require the tests to be independent.

Table 6 Consumer understanding models

Dependent variable: consumer understanding	(1)	(2)	(3)	(4)
Post-purchase	0.60***	0.62***	0.49***	0.53***
	(0.07)	(0.09)	(0.08)	(0.06)
Disclosure type B: Textually optimized	0.21*	0.45***	0.27**	0.18*
	(0.10)	(0.13)	(0.10)	(0.09)
Disclosure type C: Visually optimized	0.54***	0.73***	0.57***	0.48***
	(0.10)	(0.13)	(0.11)	(0.09)
Disclosure type D: Visually and textually optimized	0.68***	0.90***	0.66***	0.59***
	(0.11)	(0.14)	(0.11)	(0.09)
Observations	827	827	708	827
(Pseudo) R-squared	0.10	0.13	0.10	0.17
Estimation	Poisson	Poisson	Poisson	OLS
Control Variables	Yes	Yes	Yes	Yes

Constant term included but not reported. ***, ** and * indicate significance at the 0.1%, 1% and 5% levels, respectively. *Consumer Understanding* measures the number of correct answers to a three-question knowledge test on the content of the disclosures. In Model (3), the first test question was excluded as a robustness test (see below). The reference category for *Post-Purchase* is *Pre-Purchase*. The reference category for *Disclosure Type* is *A: Non-transparent*. See Table 2 above for a description of the respondents' personal characteristics used as control variables

higher than that for the respective pre-purchase scenario, the descriptive evidence clearly supports Hypothesis II-b.

We then tested Hypotheses II—a & b more formally using truncated Poisson regression models. This again allowed us to control for any confounding influences of the respondents' personal characteristics (cf. Table 2). In each of the models summarized in Table 6 below, the dependent variable is *Consumer Understanding*, i.e. the number of correct answers to the three-question multiple-choice knowledge test. We chose Poisson regression to accommodate the count nature of the dependent variable. A histogram of the dependent variable showed no evidence of zero inflation, and there was no systematic reason in the data generation process to believe that this could be the case. The variance of our dependent variable is almost equal to its mean, so that a central assumption of the Poisson model is met. Because our dependent variable cannot take a value greater than three (correct answers), we used an appropriately truncated model. Finally, we fitted all models with robust standard errors.

The results for the *Consumer Understanding Models* are highly consistent with the descriptive statistics presented above. Model (1) is our preferred baseline model, which we will use below to discuss our hypotheses. We also conducted the following robustness tests. Model (4) is identical to Model (1) but was estimated using the OLS method. Model (2) employs an alternative dependent variable that excludes the responses to the first of the three multiple-choice questions, which only applied to the post-purchase scenario (the right to compensation in case of shipping delays) and may therefore have disadvantaged those participants who were allocated to the pre-purchase scenario. Mirroring the corresponding model for *Reading Time*, Model (3) again excludes outliers with very short (< 20 s) and

very long (> 700 s) reading times. Overall, the main results described below are stable across all model specifications.

Regarding Hypothesis II-a, the regression results from our preferred baseline model (1) allowed us to test whether optimizing disclosures leads to better consumer understanding. The coefficient for the textually optimized (Type B) disclosure is statistically significant at the 5% level; the coefficients for the visually optimized (Type C) and the visually and textually optimized (Type D) disclosures are significant at the 0.1% level. Thus *Consumer Understanding* was better for all three optimized disclosure variants than for the reference text, the non-transparent disclosure (Type A). This result holds across all model specifications. Compared to disclosure Type A, *Consumer Understanding* was predicted to increase by an average of 23% for Type B, 72% for Type C and 98% for the Type D disclosure, controlling for all other variables.[5]

We then performed one-tailed Wald tests of the equality of the coefficients to systematically test Hypothesis II-a. The tests for the six pairs of coefficients were based on Model (1).[6] They confirmed that *Consumer Understanding* did in fact differ significantly between the following pairs of coefficients: Type A vs. Type B (p-value 0.084), Type A vs. Type C (p-value < 0.001), Type A vs. Type D (p-value < 0.001), Type B vs. Type C (p-value < 0.003) and Type B vs. Type D (p-value < 0.001). However, *Consumer Understanding* did not differ significantly after reading Type C vs. Type D (p-value < 0.598). In sum, these results support Hypothesis II-a insofar as the ranking on the basis of *Consumer Understanding* is Type D/ Type C > Type B > Type A. However, part of Hypothesis II-a, Type D > Type C, is not supported by the evidence.

Regarding Hypothesis II-b, the regression results for Model (1) allowed us to test whether participants in the post-purchase scenario performed better on the knowledge test than participants in the pre-purchase scenario. This was confirmed by the coefficient for *Post-Purchase,* which was highly significant at the 0.1% level. Thus switching from the *Pre-Purchase* to the *Post-Purchase* scenario was found to increase *Consumer Understanding* by 83% on average, controlling for all other variables. The significance level of this result holds across all specifications. This evidence thus supports Hypothesis II-b that participants in any post-purchase scenario will perform better on the knowledge test than participants in the pre-purchase scenario. This finding is a direct indication that there is a 'testing effect' involved in the understanding and retention of disclosures, and it supports the view expressed in Wulf and Seizov (2020) that the post-contract conclusion use case of information disclosures deserves further attention from researchers and policymakers alike.

---

[5] For computational reasons, these percentages are based on the non-truncated version of the estimated model, in which there is no upper limit on the number of correct responses.

[6] As in the previous models, we adjusted the p-values to account for multiple comparisons using the Bonferroni correction method.
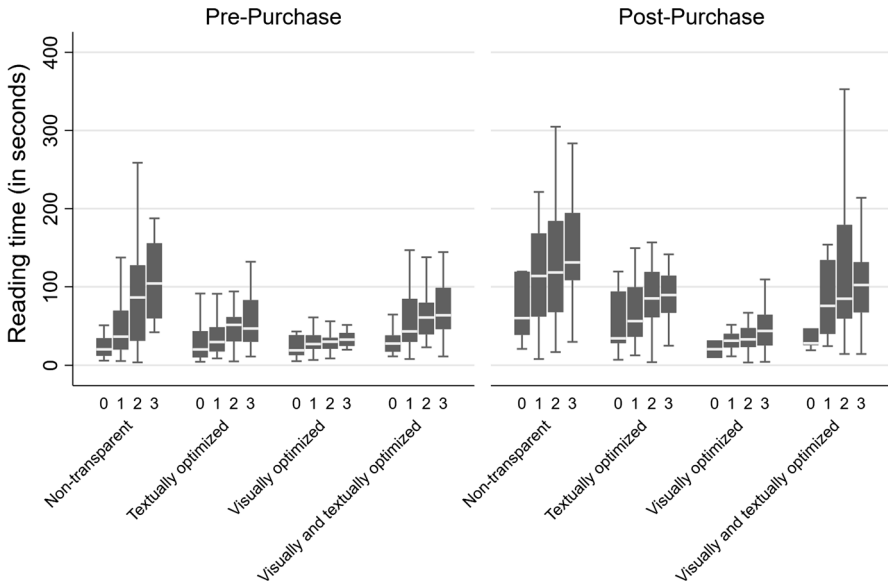
Fig. 1 Boxplots of reading times per experimental condition and number of correct answers. Note: The figure excludes outliers with reading times in excess of 400 seconds. The numbers '0', '1', '2', '3' on the x-axis indicate the number of correct responses to the knowledge questions.

4.3 Relationships between reading time and consumer understanding / disclosure type

Having presented our main results, we want to conclude by investigating the relationships between *Reading Time* and *Consumer Understanding* on the one hand and *Reading Time* and *Disclosure Type* on the other hand in the pre- and post-purchase scenarios in somewhat more detail. Figure 1 shows boxplots of reading times for all eight experimental conditions. The boxplots indicate positive correlations between *Reading Time* and *Consumer Understanding* within each disclosure type. The figure also once more illustrates the result that we obtained for Hypothesis I-b earlier, according to which reading times are considerably longer in the post-contract conclusion scenario for all four disclosure types.

Beyond investigating the relationship between *Reading Time* and *Consumer Understanding* descriptively using boxplots, we also tested whether the two variables were significantly correlated. Table 7 shows positive correlations between reading time and the number of correct answers in the knowledge test for all pre-purchase scenarios. With the exception of the visually optimized disclosure (Type C), all of these correlations were statistically significant, even after accounting for multiple comparisons using the Bonferroni correction method. In contrast, in the post-purchase scenarios the picture is in a sense reversed. Again, apart from the case of the visually optimized disclosure (Type C), these correlations were much smaller, and none of them were statistically significant after application of the Bonferroni correction method. Longer reading times thus

Table 7 Bivariate correlations between reading time and consumer understanding

Disclosure type/scenario	N	Correlation coefficient	P-value
1—"Pre-Purchase Scenario"			
A1—Non-transparent; pre-purchase	104	0.32**	0.009
B1—Textually optimized; pre-purchase	102	0.35**	0.002
C1—Visually optimized; pre-purchase	103	0.17	0.672
D1—Visually and textually optimized; pre-purchase	102	0.29*	0.026
2— "Post-Purchase Scenario"			
A2—Non-transparent; post-purchase	103	0.13	1.000
B2—Textually optimized; post-purchase	116	0.01	1.000
C2—Visually optimized; post-purchase	100	0.23	0.156
D2—Visually and textually optimized; post-purchase	105	− 0.03	1.000

**Significant at the 1% level; * significant at the 5% level. We adjusted the p-values to account for multiple comparisons using the Bonferroni correction method. Resulting p-values above 1 were rounded to 1

greatly improved on the consumer understanding found for the pre-purchase scenarios, where consumers' average reading times were low. In the post-purchase scenarios, however, where the average reading times were higher due to the greater incentive, additional reading time had less of an effect on consumer understanding. Hence, the time spent reading a disclosure appears to have a decreasing marginal benefit for consumer understanding.

However, this relationship does not appear to apply to the Type C disclosure. Since it contains very little text, this disclosure is the only one whose contents the participants could conceivably process in full within the median reading time (28 s in the pre-contractual scenario C1). Accordingly, the increased incentives in the post-contractual scenario caused the smallest additional reading time for this disclosure type (median of 37 s for C2)—after all, why would readers study it more intensively if they understood the content straight away? In other words, if most respondents did in fact absorb the full content of this disclosure, there is no reason to think that additional reading time would yield more correct answers. By contrast, the situation was quite different for the other disclosure types, where the relationship between reading times and document length suggests that most respondents must only have glanced over the texts or stopped reading at some point. If that was the case, it makes sense that those who spent more time reading would tend to be the ones who scored higher on the knowledge test.

To investigate these ideas more thoroughly we checked for interaction effects in our regression models (cf. Table 8). In Model (1), with *Reading Time* as the dependent variable, we found evidence of an interaction effect between *Post-Purchase* and *Disclosure Type* which is statistically significant at the 0.1% level (p-value < 0.001) for the visually optimized disclosure (Type C). The interaction effect was not statistically significant at any conventional level of significance for any of the other disclosure types. Thus, the general effect of higher *Reading Times* in the post-purchase scenario compared to the pre-purchase scenario was significantly smaller for the

Table 8 Interaction effects models

Dependent variable: Reading time (in seconds)	(1)
Post-purchase	62.30***
	(12.14)
Disclosure type B: Textually optimized	− 27.55***
	(7.35)
Disclosure type C: Visually optimized	− 37.74***
	(7.24)
Disclosure type D: Visually and textually optimized	0.37*
	(8.71)
Post-purchase * disclosure type B: Textual optimized	5.58
	(22.29)
Post-purchase * disclosure type C: Visual optimized	− 48.12***
	(13.27)
Post-purchase * disclosure type D: Visually and textually optimized	1.02
	(22.30)
Observations	827
(Pseudo) R-squared	0.14
Estimation	OLS
Log of dependent variable	No
Control variables	Yes

Constant term included but not reported. ***, ** and * indicate significance at the 0.1%, 1% and 5% levels, respectively. *Reading Time* is the reading time for a disclosure in seconds. When included as an explanatory variable, we transformed the variable into minutes for better readability, as otherwise the coefficient becomes almost zero. The reference category for *Post-Purchase* is *Pre-Purchase*. The reference category for *Disclosure Type* is *A: Non-transparent*. See Table 2 above for a description of the respondents' personal characteristics used as control variables

visually optimized disclosure (Type C). This supports the intuitive reasoning outlined in the paragraph above.

5 Discussion

Viewed against the backdrop of the bleak image of information disclosures painted by much of the previous research, the key findings of our study give reason for some cautious optimism. With regard to transparency and disclosure optimization, our participants demonstrated that empirically motivated transparency enhancements improve learning and often reduce reading time in both the pre- and post-purchase online shopping scenarios, thus they do in fact render information disclosures both more accessible and more understandable. Unlike ad hoc, text-only edits (Ben-Shahar & Chilton, 2016; Elshout et al., 2016), document content (Pollach, 2005) and document design (Waller, 2017), such improvements lead to significant gains in consumer learning, and including visual and both visual and textual communication elements seems to maximize those gains. Regarding the point in time at which

consumers should be presented with such information, our findings indicate that this can be much more effective post-purchase, as Wulf and Seizov (2020) and others (Gillette, 2004) have proposed. Under these conditions, the participants paid more attention and displayed better learning outcomes. Together with the knowledge and time gains that can ensue from well-designed disclosure optimization, these findings suggest that there can be a more hopeful path ahead for online disclosures. We discuss each key aspect of our results below.

On the topic of improving disclosures, applying the multidisciplinary disclosure optimization measures presented in Seizov et al. (2019) and Seizov and Wulf (2020) had significant positive effects both on the amount of time that the respondents spent reading and considering the disclosures and on the number of correct answers to the knowledge questions. This finding is consistent with the results of the studies on consumer motivation and notice optimization we have reviewed (Darolia & Harper, 2017; Kersbergen & Field, 2017; Reale & Flint, 2016; Visschers et al., 2010). It also confirms that a multidisciplinary, evidence-based approach to improving not only the language, but also the design aspects of information notices can produce significant positive results, in contrast to the text-only modification efforts tested in previous research. This becomes especially clear when comparing our results to those of Ben-Shahar and Chilton (2016), where none of the disclosure versions garnered more than an average of 30 s of respondent attention, and most received considerably less time. That said, while our results show promise, they cannot fully redeem classical disclosures. At best, our participants achieved a mean of 2.23 correct answers out of 3, but in most experimental conditions they averaged less than 2 correct answers. While these levels of consumer understanding far exceeded those obtained in previous studies and the trend was unequivocally in favour of our optimization strategies (see again Table 5), the level of consumer learning we achieved was still not ideal.

One particular aspect of disclosure optimization that we studied concerned the departure from text-only information formats. In both the pre- and post-purchase experimental scenarios, the disclosures that were visual and those that were both visual and textual tended to outperform the text-only variants. Multimodality (Bateman, 2008; Berger-Walliser et al., 2017; Lemke, 2002), or the orchestration of several communication modes—in our case icons, colors, and text—into a coherent document to transmit a unified message, is a promising strategy for making information disclosures more effective, regardless of the situation in which they are being read. Many legal experts and business practitioners have supported departing from the dry and off-putting text formats for information disclosures that the regulators currently prescribe, including not only images and graphics, but also videos and other more sophisticated formats. Our findings indicate that such a step would likely boost consumer motivation and understanding in several ways. It would, however, also require legislative action and a loosening of the rigid text-only prescriptions of the law. However, the flipside of this strategy is that companies often use sophisticated information designs to manipulate consumers and nudge them towards suboptimal choices. The literature refers to this practice as 'dark patterns' of communication (Bösch et al., 2016; Gray et al., 2018; Mathur et al., 2019). Hence, if businesses are permitted to display legal information in multimodal formats, this

must be accompanied by competent and strict oversight, and also by standardization, e.g. a predetermined set of icons or layouts.

Our findings support another idea which would require legislative amendments, namely that presenting information disclosures *after* concluding a contract constitutes a viable and effective use case (Wulf & Seizov, 2020), which is ostensibly better than presenting them beforehand. Our post-contract, high-motivation scenario improves consumer understanding across all four disclosures, i.e., independently of the optimization effects. One way for legislators to capitalize on this finding would be to define two distinct sets of information, requiring businesses to disclose one in the pre-contractual setting and the other in the post-contractual setting. In the former situation, according to our own research, in addition to considerable other evidence gathered by other authors, consumers have a very limited capacity for taking in information and thus tend to fall back on rational ignorance. To prevent information overload and disclosure fatigue, all that consumers should be immediately presented with pre-contract is a 'skinny' disclosure of the cornerstone contract terms (such as price, shipping, cooling off period, essential product or service characteristics). The visual one-pager we employed (Disclosure Type C) could serve as a model for such brief and accessible pre-contractual disclosures of the key contract terms. The full contract terms, which consumers are most unlikely to read at this stage of the transaction, should be merely brought to their attention, and they must be given the option either to access the full terms immediately and/or to save them for later reference in the event that questions or concerns arise. This requires the provision of the full contractual information on the seller's website in an unalterable, versioned and time-stamped format. Ideally, that information will be presented in an improved fashion more akin to our Type B disclosure, rather than in the commonplace, but user-unfriendly style of the Type A format. Since online consumers can then still be aware and in possession of the complete contract terms, this proposal would not fundamentally alter the current disclosure policy. It would retain the core information obligations but compartmentalize them for greater effectiveness and the benefit of consumers. Distributing disclosure information across the pre- and the post-contractual phase in a way that is consistent with consumer needs was supported by a wide range of stakeholders interviewed by Wulf and Seizov (2020).

A common thread of the above proposals is the need to ensure adequate oversight and enforcement of both the design and content of the disclosures. Previous research has shown that consumer contracts sometimes contain blatantly consumer-hostile terms that would never pass legal review, yet consumers tend to abide by them simply because they are part of a contract they have signed (Furth-Matzkin, 2017). After decades of exposure to opaque and voluminous standard terms, consumers rarely view contracts with a critical eye and are likely to accept even harmful or wrongful terms (Wilkinson-Ryan, 2017). Simply optimizing the presentation of information alone will not therefore suffice to ensure that consumers are adequately protected online. Rather, businesses need to be incentivized to "educate rather than obfuscate" (Willis, 2015). Willis advocates applying the approach of performance-based regulation to contract law in order to better align contract terms with consumer expectations and thus to preclude unpleasant surprises. Performance-based regulation prescribes certain results (e.g.

well-informed consumers) that must be achieved by prescribing a policy, rather than the means (e.g., specific pieces of information) by which these results must be achieved (Sugarman, 2009). If well-designed, such a regulatory regime would likely lead to huge efficiency gains, as businesses can be expected to choose those disclosure strategies that achieve the stated objective at the lowest cost. Furthermore, we should see disclosures adjusting almost instantaneously to changing technological and business conditions, instead of traditional legislation constantly lagging behind market developments by many years, forcing companies to work around it as best as they can. It goes without saying that these performance-based regulations can only be successful if the performance targets are well chosen and compliance is closely monitored. If that is the case, in combination with the call by Seizov and Wulf (2020) for businesses to use their marketing intelligence to address contractual information to consumers more effectively, such an approach could capitalize greatly on the present study's findings.

Our study thus charts a viable new path for handling information disclosures. In this it departs from most previous empirical inquiries, which have painted a predominantly negative picture of this consumer protection mechanism. Even so, our results are not without their limitations. To begin with, an online survey is a highly controlled environment that may not be representative of how consumers act in real-life, despite the fact that we have incorporated a number of elements in both the pre- and the post-purchase scenarios to make them feel as 'real' as possible. A follow-up study could employ an incentive-based design in which participants would be rewarded for answering the knowledge questions correctly. Such a set-up could yield more reliable motivation levels that would be less dependent on how sincerely participants engaged with the experimental scenarios. Similarly, despite the size and heterogeneity of our UK-based sample, it may not be representative of the online shopping behaviours of the general populations of other countries. Nevertheless, our study makes a promising case for shifting the presumed context in which information disclosures will be read to a post-contract conclusion scenario and for applying more varied, multimodal tactics to optimize disclosures.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Agarwal, P., Karpicke, J., Kang, S., Roediger, H., & McDermott, K. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22*(7), 861–876.

Agarwal, P., & Roediger, H. (2011). Expectancy of an open-book test decreases performance on a delayed closed-book test. *Memory, 19*(8), 836–852.

Anderson, J., & Bower, G. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79*(2), 97–123.

Antunez, L., Vidal, L., Sapolinski, A., Gimenez, A., Maiche, A., & Ares, G. (2013). How do design features influence consumer attention when looking for nutritional information on food labels? Results from an eye-tracking study on pan bread labels. *International Journal of Food Sciences and Nutrition, 64*(5), 515–527.

Bakos, Y., Marotta-Wurgler, F., & Trossen, D. R. (2014). Does anyone read the fine print? Consumer attention to standard-form contracts. *Journal of Legal Studies, 43*(1), 1–35.

Bateman, J. (2008). *Multimodality and genre: A foundation for the systematic analysis of multimodal documents*. Springer.

Becher, S., & Benoliel, U. (forthcoming 2023). Dark Contracts. *Boston College Law Review*, *64*.

Ben-Shahar, O., & Chilton, A. (2016). Simplification of privacy disclosures: An experimental test. *Journal of Legal Studies, 45*(2), 41–67.

Ben-Shahar, O., & Schneider, C. (2014). *More than you wanted to know: The failure of mandated disclosure*. Princeton University Press.

Berger-Walliser, G., Barton, T., & Haapio, H. (2017). From visualization to legal design: A collaborative and creative process. *American Business Law Journal, 54*(2), 347–392.

Bjork, R., & Bjork, E. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (pp. 35–67). Hillsdale: Erlbaum.

Bösch, C., Erb, B., Kargl, F., Kopp, H., & Pfattheicher, S. (2016). Tales from the dark side: Privacy dark strategies and privacy dark patterns. *Proceedings on Privacy Enhancing Technologies, 2016*(4), 237–254.

Butler, A., & Roediger, H. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*(4–5), 514–527.

Carpenter, S. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1563–1569.

Carpenter, S., & DeLosh, E. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34*(2), 268–276.

Darolia, R., & Harper, C. (2017). *Information use and attention deferment in college student loan decisions*. University of Missouri.

Elshout, M., Elsen, M., Leenheer, J., Loos, M., & Luzak, J. (2016). *Consumers' attitudes towards terms and conditions (T&Cs): Final report*. Brussels: CHAFEA. Retrieved March 31, 2022, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2847546

Eurostat (2021). E-commerce statistics for individuals. Eurostat, Luxembourg.

Fong, G., Hammond, D., & Hitchman, S. (2009). The impact of pictures on the effectiveness of tobacco warnings. *Bulletin of the World Health Organization, 87*(8), 640–643.

Furnell, S., & Phippen, A. (2012). Online privacy: A matter of policy? *Computer Fraud & Security, 2012*(8), 12–18.

Furth-Matzkin, M. (2017). On the unexpected use of unenforceable contract terms: Evidence from the residential rental market. *Journal of Legal Analysis, 9*(1), 1–49.

Furth-Matzkin, M., & Sommers, R. (2020). Consumer psychology and the problem of fine-print fraud. *Stanford Law Review, 72*(3), 503–561.

Gillette, C. (2004). Rolling contracts as an agency problem. *Wisconsin Law Review, 2004*(2), 679–723.

Gray, C., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. (2018). The dark (patterns) side of UX design. In: *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–14.

Helberger, N. (2013). *Forms matter: Informing consumers effectively.* Retrieved on March 30, 2022, from www.beuc.eu/publications/x2013_089_upa_form_matters_september_2013.pdf

Hiilamo, H., Crosbie, E., & Glantz, S. (2014). The evolution of health warning labels on cigarette packs: the role of precedents, and tobacco industry strategies to block diffusion. *Tobacco Control 23*(1).

Jolls, C. (2013). Product warnings, debiasing, and free speech: The case of tobacco regulation. *Journal of Institutional and Theoretical Economics, 169*(1), 53–78.

Karpicke, J., & Roediger, H. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 704–719.

Karpicke, J., & Roediger, H. (2008). The critical importance of retrieval for learning. *Science, 319*, 966–968.

Kersbergen, I., & Field, M. (2017). Alcohol consumers' attention to warning labels and brand information on alcohol packaging: Findings from cross-sectional and experimental studies. *BMC Public Health, 17*(1), art. 123.

Lemke, J. (2002). Travels in hypermodality. *Visual Communication, 1*(3), 299–325.

Lotter, W. (2019). Die alte Welt wollte Komplexität immer nur reduzieren. Die neue erschließt sie. *Brandeins, 21*(7), 34–39.

Mak, V. (2012). The myth of the 'empowered consumer': Lessons from financial literacy studies. *Zeitschrift für Europäisches Unternehmens- und Verbraucherrecht, 1*(4), 254–263.

Marotta-Wurgler, F. (2012). Does contract disclosure matter? *Journal of Institutional and Theoretical Economics, 168*(1), 94–119.

Mathur, A., Acar, G., Friedman, M., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of ACM Human-Computer Interaction*, pp. 1–32.

Milne, G., & Culnan, M. (2004). Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing, 18*(2), 15–30.

Nordhausen Scholes, A. (2009). Information requirements. In G. Howells (Ed.), *Modernising and harmonizing consumer contract law* (pp. 213–236). Sellier European Law Publishing.

Pollach, I. (2005). A typology of communicative strategies in online privacy policies: Ethics, power and informed consent. *Journal of Business Ethics, 62*(3), 221–235.

Pyc, M., & Rawson, K. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*(4), 437–447.

Pyc, M., & Rawson, K. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330*, 335.

Rauch, S. (2019). UX case study: Tracking EHR automation, scarcity of attention, and transaction hazards. *Online Journal of Public Health Informatics*. https://doi.org/10.5210/ojphi.v11i1.9692

Rawson, K., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology, 140*(3), 283–302.

Reale, S., & Flint, S. (2016). The impact of menu label design on visual attention, food choice and recognition: An eye tracking study. *Journal of Sensory Studies, 31*(4), 328–340.

Roediger, H. A., Kang, S., & Marsh, E. (2010). Benefits of testing memory: Best practices and boundary conditions. In G. Davies & D. Wright (Eds.), *New frontiers in applied memory* (pp. 13–49). Psychology Press.

Rowland, C. (2014). The effect of testing versus restudy on retention: A meta-analytic view of the testing effect. *Psychological Bulletin, 140*(6), 1432–1463.

Rummer, R. S., & Schwede, A. (2019). Open-book versus closed-book tests in university classes: A field experiment. *Frontiers in Psychology, 10*, 463.

Seizov, O., & Wulf, A. (2020). Communicating legal information to customers transparently: A multi-disciplinary multistakeholderist perspective. *Journal of International Consumer Marketing, 33*(2), 159–177.

Seizov, O., Wulf, A., & Luzak, J. (2019). The transparent trap: A multidisciplinary perspective on the design of transparent online disclosures in the EU. *Journal of Consumer Policy, 42*(1), 149–173. https://doi.org/10.1007/s10603-018-9393-0

Sugarman, S. (2009). Performance-based regulation: Enterprise responsibility for reducing death, injury and disease caused by consumer products. *Journal of Health Politics, Policy and Law, 34*(6), 1035–1078.

Visschers, V., Hess, R., & Siegrist, M. (2010). Health motivation and product design determine consumers' visual attention to nutrition information on food products. *Public Health Nutrition, 13*(7), 1099–1106.

Waller, R. (2017). Graphic literacies for a digital age. In A. Black, P. Luna, O. Lund, & S. Walker (Eds.), *Information design: Research and Practice* (pp. 177–203). Routledge.

Wilkinson-Ryan, T. (2017). The perverse consequences of disclosing standard terms. *Cornell Law Review, 103*(1), 117–175.

Willis, L. (2015). Performance-based consumer law. *The University of Chicago Law Review, 82*(3), 1309–1411.

Wulf, A. (2014). Institutional competition of optional codes in European contract law. *European Journal of Law and Economics, 38*(1), 139–162.

Wulf, A., & Seizov, O. (2020). The principle of transparency in practice: How different groups of German stakeholders view EU online information obligations. *European Review of Private Law, 28*(5), 1065–1092.