WORKING PAPER SERIES

No. 1/2026

The Benefits of Intentions:
Deliberate Harm Avoidance and Consumer Responses

Sheheryar Banuri
School of Economics, University of East Anglia and
Hughes Hall, University of Cambridge (UK)

Christa Brunnschweiler
Department of Economics,
Norwegian University of Science and Technology
and CESifo

Deanna Karapetyan
Financial Conduct Authority (UK)

Department of Economics

◙ Norwegian University of Science and Technology
N-7491 Trondheim, Norway
http://www.ntnu.edu/econ/working-papers

The Benefits of Intentions: Deliberate Harm Avoidance and Consumer Responses

Sheheryar Banuri; Christa Brunnschweiler; Deanna Karapetyan[*]

January 2026

Abstract

This paper investigates why firms engage in costly environmental and ethical practices, focusing on whether consumer responses depend on firms' intentions or outcomes. Existing literature links ESG practices to positive performance and stakeholder rewards, but most evidence is observational and cannot disentangle intentionality from outcomes. Using a controlled experiment, we examine consumer reactions when firms choose between a "clean" technology (avoiding harm at a cost) and a "dirty" technology (higher returns with negative externalities). Two treatments isolate intentionality: Random Choice versus Willful Choice. After observing the firm's choice and the resulting externality, consumers can respond by transferring (taking away) resources to the firm in a give-or-take Dictator Game. We find a pronounced asymmetry in how intentions matter. Consumers punish firms whenever a negative externality is incurred, regardless of intentionality, indicating that punitive responses are largely outcome-driven. By contrast, when harm is avoided, intentions play a central role: firms that deliberately choose to prevent a negative externality are treated with significantly greater leniency than firms for which absence of harm arises randomly, reflected in positive transfers on average. These findings highlight that intentionality affects punitive responses and helps explain why firms may voluntarily adopt costly ethical practices when choices are observable.

JEL Codes: D03; D64; L21; C91
Keywords: Intentionality; Harm Avoidance; Consumer Responses; ESG; Environmental Externalities

1. Introduction

Why do firms engage in ethical practices that are (often) costly for them? A dominant explanation emphasizes demand-side responses: consumers/investors reward ethical practices by firms adhering to ethical principles (broadly understood as firms demonstrating Environmental responsibility, Social equity, and Governance transparency, or the ESG framework). A growing number of papers demonstrate that consumers respond to products carrying ethical labels (Hainmueller, Hiscox, and Sequeira, 2015), that environmentally oriented municipal bonds (adhering to environmentally focused principles) command greater price premiums than closely matched ordinary bonds (Baker, Bergstresser, Serafeim, and Wurgler, 2022), and that investors are willing to withstand relatively worse performance of socially responsible and religiously oriented mutual funds (Banuri, Murgia, and ul Haq, 2024). A common thread amongst this literature is that consumer awareness and interpretation of ESG practices is critical (Tan, Hashim, and Zheng, 2025). The consensus is that strong corporate social responsibility (CSR) practices (Ali et al., 2023), and environmental sustainability especially (Golicic and Smith, 2013; Ardia et al., 2023), are linked to positive firm performance. Nevertheless, much of this evidence relies on observational data and focuses on firm-level outcomes, making it difficult to disentangle the role of firms' intentions from the outcomes they generate. As a result, while these findings are consistent with the view that stakeholders value ethical practices, they do not cleanly identify whether stakeholders respond to ethical outcomes per se, or to firms' deliberate choices to engage in such practices when doing so is costly.

Sustainable consumption and production patterns, encompassing environmental, social, and economic sustainability, are central to climate change mitigation and form the basis of the United Nations' twelfth Sustainable Development Goal. They are also fundamental to governments' commitments under the Paris Agreement on Climate Change. Reflecting this importance, sustainability considerations have become increasingly salient for firms. For example, the Dow Jones Sustainability Indices, launched in 1999, now evaluate the sustainability performance of thousands of publicly listed companies worldwide, while Fairtrade International reports that certified producers received over €201 million in Fairtrade Premium payments in 2021 alone (Fairtrade International, 2023).

Environmental sustainability provides a particularly salient context in which firms often face a trade-off between private profitability and harm reduction. Survey evidence suggests that

environmental sustainability is of particular concern to consumers: a 2019 survey found that 77% of UK grocery shoppers had avoided, or would consider avoiding, products based on a brand's environmental policy, and a global survey of 24,000 respondents reported that 59% were willing to boycott firms perceived as harming the environment (dentsu and Microsoft, 2021). At the same time, a large majority of consumers report difficulty in identifying genuinely sustainable producers, highlighting the importance of observable and credible firm choices. These patterns raise a key behavioural question: whether and how consumers' responses depend not only on environmental outcomes, but also on firms' deliberate choices to avoid environmental harm when doing so is privately costly.

In this paper, we shed light on this question using a lab experiment that focuses on consumer behaviour and cleanly disentangles intentions from outcomes. We thereby shed light on the question of when and how consumers respond to ESG practices. The experiment simulates a firm that chooses between two production technologies, a "clean" technology that generates returns with no externality imposed on consumers, and a "dirty" technology that generates higher returns to the firm but imposes a negative externality on consumers. We conduct two treatments, a "no intentionality" baseline where the production technology is randomly imposed (labelled as Random Choice), and an "intentionality" treatment where firms can choose the technology freely (labelled as Wilful Choice). Our focus is squarely on the consumers, who learn about the choice and externality cost (if present) and then participate in a simple give-or-take dictator game where they can transfer funds to, or take funds away from, the firm.

Our main finding is that consumers respond fundamentally differently when firms deliberately choose to avoid imposing a negative externality. In the Wilful Choice condition, firms that incur a private cost to prevent harm are met with substantial moral leniency by consumers (manifested in positive transfers on average) relative to firms that deliberately choose to impose the externality. Importantly, this leniency is not driven by outcomes alone. When comparing across treatments, firms that deliberately choose to prevent harm are treated more leniently than firms for which the absence of harm arises randomly, indicating that agency plays a central role in shaping consumer responses.

We also observe that consumers punish firms whenever a negative externality is incurred, regardless of whether it is imposed intentionally or randomly, highlighting that experienced outcomes remain a primary determinant of punitive behaviour. Taken together, these results

suggest that while self-interest and outcome-based punishment are pervasive, deliberate harm avoidance can soften punitive responses and even elicit positive transfers, providing a behavioural foundation for why firms may voluntarily engage in costly ethical practices when their choices are observable. This asymmetry indicates that intentionality primarily affects consumer responses in the absence of harm, whereas punishment for harm is largely outcome driven.

Our paper contributes to the literature on voluntary corporate social responsibility activities by shedding new light on how consumers respond to information about firms' ethical behaviour, and on the role of agency in shaping those responses. Early work by Creyer and Ross (1997) provides stated-preference evidence that consumers value ethical firm behaviour and are willing to pay a premium for ethically produced goods, while expecting products from unethical firms to be offered at lower prices. Subsequent studies show that consumers distinguish between different dimensions of corporate social responsibility, such as social versus environmental responsibility, and may react differently to information along these dimensions (Catlin et al., 2017; Hosta and Zabkar, 2021), though these studies likewise rely on stated behaviour. Govind et al. (2019) document a gap between consumers' stated ethical attitudes and their actual purchasing behaviour, highlighting the difficulty of inferring underlying mechanisms from survey responses alone.

Several studies focus on the reaction to environmental pollution. In particular, the present paper relates to the experiment by Knobe (2003) who showed that randomly chosen respondents were much more willing to blame a hypothetical company chairman for bad environmental behaviour than to praise them for good behaviour. This has become known as the "Knobe effect" and confirmed in numerous settings, though Utikal and Fischbacher (2014) use a laboratory experiment to find situations where the effect vanishes. Our paper complements this literature by providing controlled experimental evidence on how consumers respond to firms' choices when ethical behaviour is privately costly and externally observable. We show that consumers react not only to whether harm occurs, but crucially to whether a firm deliberately chooses to avoid imposing a negative externality.

This paper also contributes to the large experimental literature on the role of intentions in social preferences. A wide range of studies demonstrate that individuals condition their behaviour on perceived intentions, both in punishing bad intentions (Blount, 1995; Brandts and Solà, 2001; Charness and Rabin, 2002; Nelson, 2002; Offerman, 2002; Charness, 2004; Charness and

Rabin, 2005; Charness and Levine, 2007a; Falk, Fehr, and Fischbacher, 2003; Falk, Fehr, and Fischbacher, 2008) and in rewarding good intentions (Charness and Levine, 2007; Falk, Fehr, and Fischbacher, 2008). Whitson et al. (2015) further show that direct recipients of actions respond differently to intentions than third-party observers, while Bicchieri and Maras (2022) demonstrate that intentionality plays an important role in third-party punishment decisions. Our results are consistent with the importance of intentions highlighted in this literature but refine it in an important way. In our setting, consumers punish firms whenever a negative externality is incurred, regardless of whether the harm is imposed intentionally or randomly, indicating that punishment is largely outcome-driven. By contrast, when harm is avoided, intentions matter: consumers grant moral credit to firms that deliberately choose to prevent externalities, in the form of positive transfers. This asymmetry suggests that intentionality primarily shapes moral leniency rather than punitive behaviour. While prior studies debate the relative importance of intentions versus distributional fairness (Nelson, 2002; Charness and Levine, 2007; Stanca, 2010; Bone and Raihani, 2015), our evidence indicates that good intentions play a decisive role specifically in the absence-of-harm domain, whereas responses to harm itself remain largely governed by outcomes.

2. Experimental Design

The overall study consists of two distinct parts involving an effort task, and two treatments. The experiment involved two types of players, "firms" (neutrally labelled as "blue players") and "consumers" ("yellow players"). Firms engage in an effort task that can generate a negative externality on consumers. Consumers then engage in a similar effort task, but their earnings can potentially be reduced by negative externalities imposed by firms. After the effort tasks are completed, consumers then engage in a give-or-take Dictator game with the firm as the recipient.

The experiment was conducted entirely online, using oTree (Chen et al., 2016) and the Prolific subject pool. As the experiment was online-only, firms and consumers participated asynchronously. Sessions were conducted with subjects in the role of firms (N = 61), followed by sessions with subjects in the role of consumers (N=300), two weeks apart. Subjects were fully informed about this, with subjects in the role of firms acknowledging that they may receive additional bonus payments at the end of the study period (which they then received in

their Prolific accounts) following the decisions by consumers. The sample was balanced on gender, and subjects were recruited from the US, UK, and Ireland subject pools.

2.1 Firm tasks

A total of 61 subjects were recruited into the role of firms. These subjects were provided instructions on an effort task (described below), and then were asked to undertake the task for three rounds. The sequential stages of the experiment for the firm are shown in Figure 1.
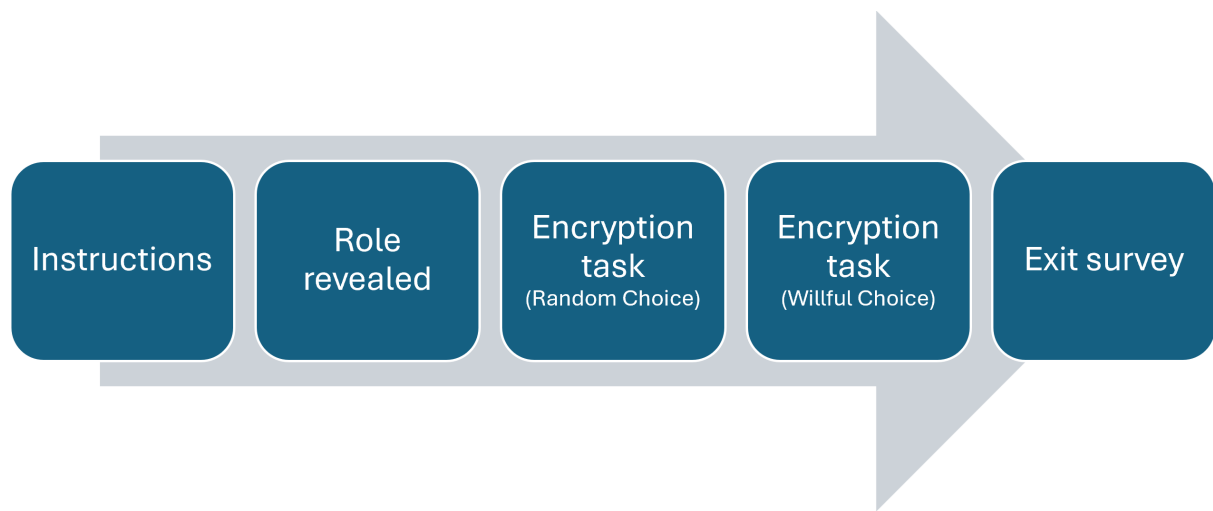

Figure 1: Experimental Stages for firms

The first round had subjects face one of two types of contracts, either the "clean" contract, which meant they were asked to undertake the task with a piece rate of 1.5 tokens (£0.0375) for each word encrypted, or the "dirty" contract where they were asked to undertake the task with a piece rate of 2 tokens (£0.05) for each word encrypted and a reduction of 1 (£0.025) token off a matched consumer's payoff. Note that the contract was randomly selected, and all contracts were neutrally labelled.

In the second round, subjects undertook the task with the contract that was not assigned in the first round. Finally, the third round offered subjects a choice: they could freely choose between the two contracts they had been exposed to earlier. This method has two interesting properties. First, as mentioned in Banuri and Keefer (2016), this introduces subjects to the two contracts, and since they experience both, the contract choice in the final round makes the trade-offs to

the subjects clear. Second, by randomizing the order of the contract presented in the first round, we avoid anchoring the subjects on either contract type. This is especially important since one of the contracts generates a negative externality on other subjects.

The data generated by the firms is then used for subjects in the role of consumers. One treatment faced by consumers (Wilful Choice) uses the data from round 3 (where subjects can choose not to impose an externality, at a cost to themselves), while the control (Random Choice) uses the data from the first two rounds, where the externality is imposed without choice of the firm. We discuss this in more detail in the next subsection.

The main drawback to this method is that learning may be taking place, which would mean that output in the third and final round can be higher (independent of the impact of choice) than in the first two rounds. This is indeed the case, as we observe higher output as the task progresses (presumably due to learning). In the results section, we discuss how we account for this in the estimations.

2.2 Consumer tasks

The sequential stages of the experiment for the consumers are given in Figure 2. The task for subjects in the role of consumers (N=300) consisted of two parts. The first was an effort task (using the encryption task described below). Subjects were informed that they would be participating in this task with a matched partner (a firm, as described in section 2.1). We use a neutral label to distinguish firms ("Blue players") from consumers ("Yellow players") to avoid priming the subjects. Subjects were informed that they would be paid 2 tokens (£0.05) for each word correctly encrypted. They were also informed that their earnings would potentially be impacted by their matched ("Blue") partner, who had previously participated in this study. After subjects completed the effort task for pay, the amount that they earned was displayed, along with any reductions from the effort of their matched counterpart.
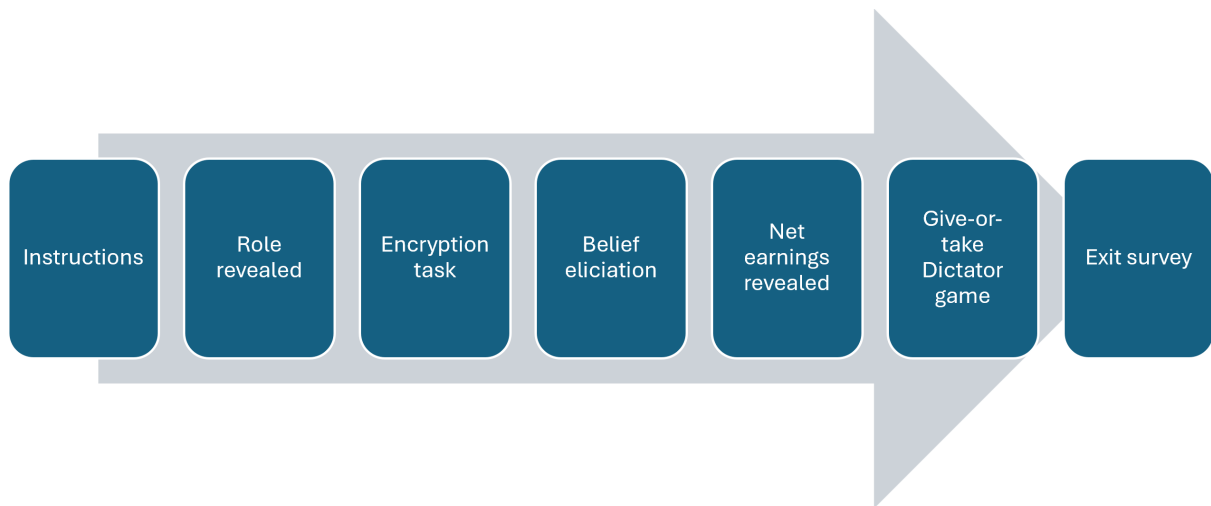
Figure 2: Experimental stages for consumers

Random Choice: As stated earlier, subjects participated under two conditions, a control condition (Random Choice) and a treatment condition (Willful Choice). In the Random Choice condition, subjects were informed that there is a chance their matched partner (firm) engaged in a task that would reduce their earnings. The exact wording used is as follows:

> "Recall that your matched Blue player completed the same Encryption task as you, but that the computer program assigned a payoff scheme that would have either decreased your earnings or the earnings of your matched Blue player."

The program then drew a random number indicating whether an externality was imposed or not. If the externality was imposed, the program used data from the appropriate round (either round 1 or 2 of the firms' task). At the end of the effort task, subjects were informed of the following:

> "The computer program assigned the payoff scheme that WOULD decrease your earnings. This means the Blue player decreased your Encryption task earnings by XX token(s) while exerting effort in the task. Therefore, your earnings are XX tokens."

As the above description makes clear, the program selected the payoff scheme which resulted in earnings being reduced by the actions of the firm ("Blue" player).

If the externality was not imposed, the program used data from the appropriate round (either round 1 or 2 of the firms' randomly assigned task). At the end of the effort task, subjects were informed of the following:

> "The computer program assigned the payoff scheme that WOULD NOT decrease your earnings. This means the Blue player decreased your Encryption task earnings by 0 token(s) while exerting effort in the task. Therefore, your earnings are XX tokens."

To summarize, in the Random Choice condition the instructions make clear to the subjects (consumers) that any externality imposed was due to the program. Once the program selected the payoff scheme, the firms' actions then yielded either a positive reduction in earnings or none at all. Moreover, the extent of the reduction is also revealed to the subjects.

Willful Choice: In the treatment condition (Willful Choice), subjects were informed that their matched counterparts ("Blue" players) were given a choice between two incentive schemes: one with a high piece rate (2 tokens per encrypted word) which imposed a negative externality on the consumers ("Yellow" players), or one with a lower piece rate (1.5 tokens per encrypted word), but which imposed no negative externality. Importantly, the negative externality was lower than the cost to the subject, hence depending on whether consumers focused on individual or joint payoff maximization, the choice of imposing the externality would seem unfair.

If consumers were matched with a firm that chose to impose the externality, they were informed in the following way:

> "The Blue player CHOSE the payoff scheme that WOULD decrease your earnings. This means the Blue player decreased your Encryption task earnings by XX token(s) while exerting effort in the task. Therefore, your earnings are XX tokens."

Similarly, if consumers were matched with a firm that chose not to impose the externality, they were informed in the following way:

> "The Blue player CHOSE the payoff scheme that WOULD NOT decrease your earnings. This means the Blue player decreased your Encryption task earnings by 0 token(s) while exerting effort in the task. Therefore, your earnings are XX tokens."

To summarize, in the Wilful Choice condition the instructions make clear to the subjects (consumers) that any externality imposed was a result of a deliberate choice made by their counterpart. Once the firm made their contract choice, the firms' actions then yielded either a reduction in earnings or not. As before, the extent of the reduction is also revealed to the subjects.

The difference between treatment and control is straightforward. By virtue of the design, the choice of implementing the externality is either made by the program itself (Random Choice) or by the firm (Wilful Choice). The subject in the role of the consumer is informed about the state of the world in which they are in: one where firms either had a choice or not in imposing the externality.

2.3 Key outcome variable

Once subjects in the role of consumers completed the effort task, and the manner and reduction in earnings was revealed to them, and they moved on to the next stage, a give-or-take Dictator Game (List, 2007; Bardsley, 2008). This Dictator Game is played with the dictator role being assigned to the consumer, with the firm as recipient. Consumers are informed that they are endowed with 30 tokens, while their matched counterpart (the same firm as in the first task) is endowed with 15 tokens. In 1 token increments, the Dictator (consumer) can choose to transfer up to 30 tokens to, or take up to 15 tokens away from, their counterpart (firm). As in standard dictator games, dictators have a higher endowment compared to the potential recipients to observe whether the dictators will make endowment levels more equal. The amount that subjects (consumers) give or take from their matched counterparts (firms) is of key interest.

2.4 Control variables

Since our online subject pool had varied demographic characteristics, we controlled for age, gender (a dummy variable equal to one if the subject was female), and student status (a dummy variable equal to one if the subject was a full-time student). We also controlled for the externality amount, fairness of the encryption task, and subjects' (consumers) belief discrepancy in the amount of the negative externality. The externality amount, measured in the number of tokens firms decreased consumers earnings by, is used as a control since we might expect a higher externality amount to cause subjects (consumers) to give less/take more from their matched counterpart.

Subjects (consumers) who both experience a negative externality and know that the program generated this externality are more likely to think that experimenter's setup of the encryption task was unfair rather than their matched partner's efforts in encrypting as many words as possible. Thus, we control for subjects' reported fairness of the encryption task by asking them

at the end to rate how fair the encryption task was on a 5-point Likert Scale, from very unfair to very fair (-2, -1, 0, 1, 2).

Belief discrepancy in the amount of the negative externality was measured from subjects' (consumers) belief of what externality amount their matched counterpart imposed subtracted by the actual externality amount imposed. After the net encryption task earnings were revealed but before the externality amount was revealed, subjects (consumers) were asked to state their beliefs about the negative externality by answering the following questions:

> "Note that your earnings in this task were based on your performance and the performance of your matched Blue player only if the Blue player chose the payoff scheme that would decrease your earnings. Remember that if the Blue player chose this payoff scheme, your earnings were reduced by 1 token for each word your matched Blue player encrypted correctly. Now, please answer the following question. Do you think your matched Blue player chose the payoff scheme that would decrease your earnings? Y/N"

> If yes: "Since you answered yes, please answer also answer the following questions. We would like to know how much you think you would have earned if the Blue player's performance had no impact on your earnings. Please express the amount in tokens:____

> Now, we would like to know how much you think your matched Blue player reduced your earnings by. Please express the amount in tokens:___ "

Table 1 provides the summary statistics for the outcome variable of tokens transferred, along with the control variables listed above. There is a significant difference in the externality amount between the Wilful Choice treatment group and the Random Choice control group. This difference is largely due to learning and randomization differences, which are discussed in more detail in Section 3.2 below. Not that when the externality amount is controlled for in the regression analysis presented in Section 3, the results remain robust.

Table 1: Consumer summary statistics

	Willful Choice	Random Choice	Difference	SE
Tokens Transferred	-1.4338	-2.3659	-0.9320	0.7659
age	38.6194	32.4224	-6.1970	11.1495
female	0.5147	0.4756	-0.0391	0.0581
student	0.2279	0.2195	-0.0084	0.0485
externality amount	9.9265	6.2500	-3.6765***	1.0056
externality imposed (dummy)	0.5809	0.4512	-0.1297**	0.0577
belief externality imposed	8.2426	6.5732	-1.6695	1.3632
belief discrepancy	-7.6618	-6.1220	1.5398	1.3709
encryption performance	15.1397	15.5305	0.3908	0.5362
fairness of encryption task	-0.1176	-0.0244	0.0933	0.1457
Observations	300			

Notes: $* \; p < 0.1$, $** \; p < .05$, $*** \; p < 0.01$

2.5 Encryption (effort) task

We implemented the encryption task used by Erkal et al. (2011), which consisted of participants encrypting combinations of letters into numbers. A coding rubric (a table that allocates numbers to letters) is provided for participants, where a number is assigned to each letter of the alphabet in a random order (see Figure 3). Each round of the encryption task lasted 360 seconds (6 minutes). In previously conducted encryption tasks, the best performing participants encrypted just under 6 words per minute (Erkal et al. 2011).

Figure 3: Screenshot of encryption task

2.6 Experimental procedures

Ten online experimental sessions were conducted from October to November 2021. Subjects received detailed online instructions and were given both attention and understanding checks. The full experimental instructions are in the online appendix. If a subject failed more than one attention check, they could not move forward with the experiment (note that most subjects did not fail the attention checks). If a subject answered an understanding check question incorrectly, the computer program redirected them back to the instructions. All understanding check questions needed to be answered correctly for a subject to move on to the next section.

The payoff for each subject was the summation of their participation fee (£2.50), net earnings from the encryption task (based on either 2 tokens (£0.05) or 1.5 tokens (£0.0375) per word encrypted depending on the payment scheme), and money from the Dictator Game based on the decision the victims made (from -15 to 30 tokens). At the end of the experiment and survey, the total payoff is revealed for consumer roles. The payoff revealed for firm roles was based on the same participation fee and money earned in one randomly chosen encryption task round. Firms received money from the Dictator Game only after the consumers completed their sessions.

3. Results

Our primary outcome of interest is how consumers adjust the earnings of their matched firms in the give-or-take Dictator Game. Consumers can either transfer resources to the firm or take resources away from it. We interpret taking away from firms as reflecting punitive behaviour, while positive transfers indicate favourable treatment. Importantly, because consumers are randomly assigned to firms and treatments, systematic differences in transfers across conditions cannot be attributed to income-maximizing considerations but instead reflect differences in consumers' normative responses to firms' behaviour.

Section 3.1 examines how consumers' punitive behaviour and moral leniency toward firms vary across treatments, focusing on the role of firms' choices to impose or avoid a negative externality. Section 3.2 investigates differences in the magnitude of the externality across treatments and presents robustness checks demonstrating that firms' production choices do not mechanically drive the observed patterns in consumer responses.

3.1 Consumers' behaviour

Figure 4 illustrates consumer transfer behaviour in the give-or-take Dictator Game across the two treatments, separately for cases in which a negative externality is avoided (left panel) or imposed (right panel). When no externality is imposed, consumer responses differ markedly across treatments. Under Random Choice, consumers take on average 1.34 tokens from firms, whereas under Wilful Choice, consumers *transfer* on average 1.30 tokens *to firms*. This difference is statistically significant ($p < 0.05$, two-tailed t-test), indicating substantially greater moral leniency toward firms that deliberately choose to avoid harm relative to firms for which the absence of harm arises randomly.

By contrast, when a negative externality is imposed, consumer behaviour is similar across treatments. Consumers take away on average 3.6 tokens under Random Choice and 3.4 tokens under Wilful Choice, a difference that is not statistically significant ($p = 0.80$, two-tailed t-test). This suggests that punitive responses are primarily driven by the experience of harm itself, rather than by whether the harm was imposed intentionally or randomly.

Within the Wilful Choice treatment (i.e., the right sides of the two panels), consumer responses differ sharply depending on the firm's decision. Consumers transfer on average 4.7 fewer tokens when firms deliberately impose an externality compared to when they deliberately avoid it ($p < 0.01$, two-tailed t-test). With other words, within the Wilful choice treatment, transfers are positive on average when the externality is avoided and negative when the externality is deliberately incurred. Under Random Choice (i.e., the left sides of the two panels), the corresponding difference is substantially smaller, though still significantly different: consumers take away 2.26 fewer tokens when an externality is imposed by the computer relative to when it is not ($p < 0.05$, two-tailed t-test). The resulting difference-in-differences of 2.4 tokens is statistically significant ($p < 0.05$), confirming that intentionality amplifies consumer responses in the absence-of-harm domain, but not when harm occurs.
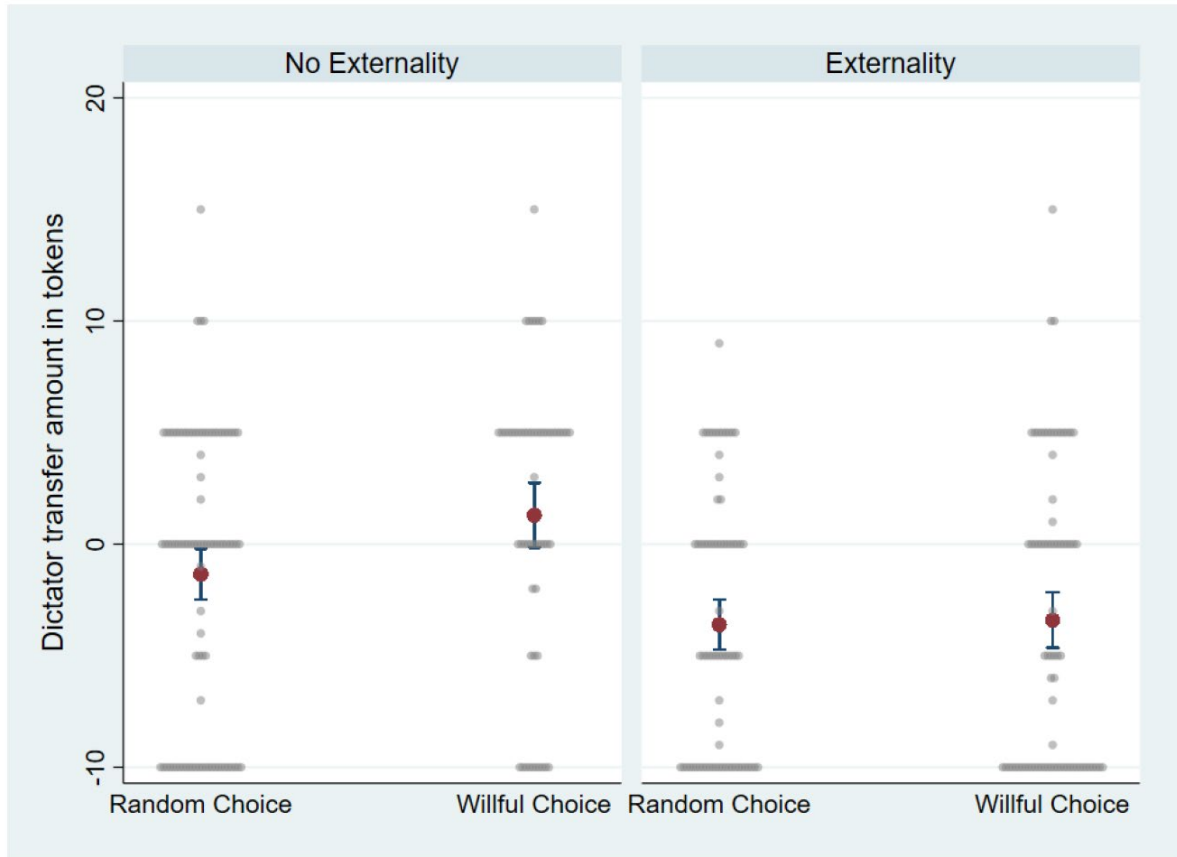
Figure 4: Treatment differences by externality imposition

Notes: Strip plot showing difference in treatment effect under conditions of externality and no externality. The Y-axis displays the number of tokens transferred in the Dictator Game. On the left-hand side of the figure, the mean difference in tokens transferred is shown between treatment (Willful Choice) and control (Random Choice) when consumers do not experience a negative externality. On the right-hand side of the figure, the mean difference in tokens transferred is shown between treatment (Willful choice) and control (Random Choice) when consumers experience a negative externality. This graph displays 90% Confidence Intervals.

Table 2 reports OLS regression results in which the number of tokens transferred by consumers to firms in the give-or-take Dictator Game (ranging from –15 to +30 tokens) is regressed on the interaction between the Willful Choice treatment (equal to one if the firm freely chose the production technology) and an indicator for whether a negative externality was imposed. The coefficient on this interaction term captures whether consumer responses to the imposition of an externality differ when the firm's choice is deliberate rather than random.

Table 2: Regressions results

Dependent Variable:	(1)	(2)	(3)
Tokens Transferred	Main effect	+ Basic controls	+ All controls
willful choice	2.643*	2.826*	2.691*
	(1.324)	(1.415)	(1.466)
externality imposed	-2.264**	-1.854	-1.662
	(0.875)	(1.224)	(1.279)
willful choice × externality imposed	-2.440**	-2.617**	-2.638**
	(0.902)	(1.040)	(1.101)
age		-0.008***	-0.008***
		(0.001)	(0.001)
female		0.489	0.858
		(1.157)	(1.130)
student		-1.440**	-1.394**
		(0.550)	(0.570)
externality amount		-0.0277	-0.00103
		(0.0726)	(0.0756)
fairness of experiment			0.682**
			(0.277)
belief discrepancy			-0.0459
			(0.0275)
Constant	-1.344	-1.026	-1.726
	(0.858)	(1.039)	(1.037)
Observations	300	295	295
R^2	0.077	0.101	0.122

Notes: Willful Choice is a dummy variable equal to one if a consumer was randomly assigned into the treated version of the experiment, where the firm was able to wilfully choose whether a negative externality would be prevented or imposed. Externality imposed is a dummy variable equal to one if a negative externality was imposed on a consumer, whether or not it was randomly assigned or wilfully imposed. The interaction effect between willful choice and externality imposed is therefore needed in order to determine the treatment effects both with and without the negative externality imposed. Column (1) only includes the interaction effect, column (2) also includes the basic control variables, and column (3) includes additional control variables. Standard errors in parentheses. $* \; p < 0.1, \; ** \; p < .05, \; *** \; p < 0.01$

The interaction between Wilful Choice and Externality Imposed is positive and statistically significant, indicating that deliberate choice amplifies the effect of harm on consumer transfers by approximately 2.4 tokens. This estimate mirrors the difference-in-differences pattern observed graphically in Figure 4 and remains robust to the inclusion of both basic demographic and externality amount controls (column 2) and additional fairness and belief controls (column 3). Among the controls, older consumers transfer significantly fewer tokens (p < 0.01), full-time students transfer fewer tokens (p < 0.05), and consumers who perceive the real-effort task as fairer transfer more tokens (p < 0.05).

3.2 Robustness checks

As shown in Table 1, the magnitude of the negative externality experienced by consumers is, on average, 59% larger in the Willful Choice treatment than in the Random Choice control. This difference arises primarily because a larger share of firms impose the externality when given the opportunity to choose: 13 percentage points more consumers experience a negative externality in the Wilful Choice treatment than in the Random Choice treatment ($p < 0.05$). In addition, conditional on an externality being imposed, the magnitude of the externality is 23.4% larger when it is imposed deliberately than when it is imposed randomly.

To assess whether these differences reflect systematic behavioural responses rather than confounds, we examine two potential explanations: learning effects across rounds and chance differences in randomization.

We first investigate whether differences in firm ability or performance explain variation in the externality magnitude. All subjects in the "firms" role completed two practice rounds prior to any payoff-relevant tasks; we therefore use performance in the second practice round as a measure of baseline ability. Table 3 reports summary statistics for firm performance across all rounds. We find no significant differences in performance between firms that were randomly assigned to impose an externality and those assigned not to impose one in the forced-choice rounds, indicating that ability is balanced across these conditions.

Table 3: Firm summary statistics

	Mean	SD	Min	Max
age	25.09	7.71	18.00	54.00
female	0.46	0.50	0.00	1.00
externality_choice	0.56	0.50	0.00	1.00
forced externality performance (round 1 or 2)	15.49	4.86	0.00	26.00
forced no externality performance(round 1 or 2)	15.66	4.57	0.00	23.00
choice performance (round 3)	16.32	5.60	0.00	27.00

Performance does, however, improve over time. Firms' performance in the final round—when they are allowed to choose whether to impose an externality—is 6.8% higher than in the first payoff-relevant round, where the externality is randomly assigned ($p < 0.10$). Table 4 further shows a monotonic improvement in performance from the first practice round through the final round. Importantly, average performance in forced-externality rounds is similar to performance in forced no-externality rounds, and Figure 5 illustrates that firm performance is nearly identical across these two conditions.

Table 4: Firm performance in the encryption task

	mean	min	max	count
Practice Round 1	1.590164	0	3	61
Practice Round 2	1.819672	0	4	61
Encryption Task Round 1	15.2623	0	23	61
Encryption Task Round 2	15.59016	0	26	61
Encryption Rask Round 3	16.29508	0	27	61

We also find no significant differences in either baseline ability or final-round performance between firms that choose to impose the externality and those that choose to avoid it. Figure 6 shows that performance under forced externality and chosen externality conditions is comparable. Together, these results indicate that part of the observed difference in externality magnitude reflects general learning over the course of the experiment rather than differential ability across firm types.
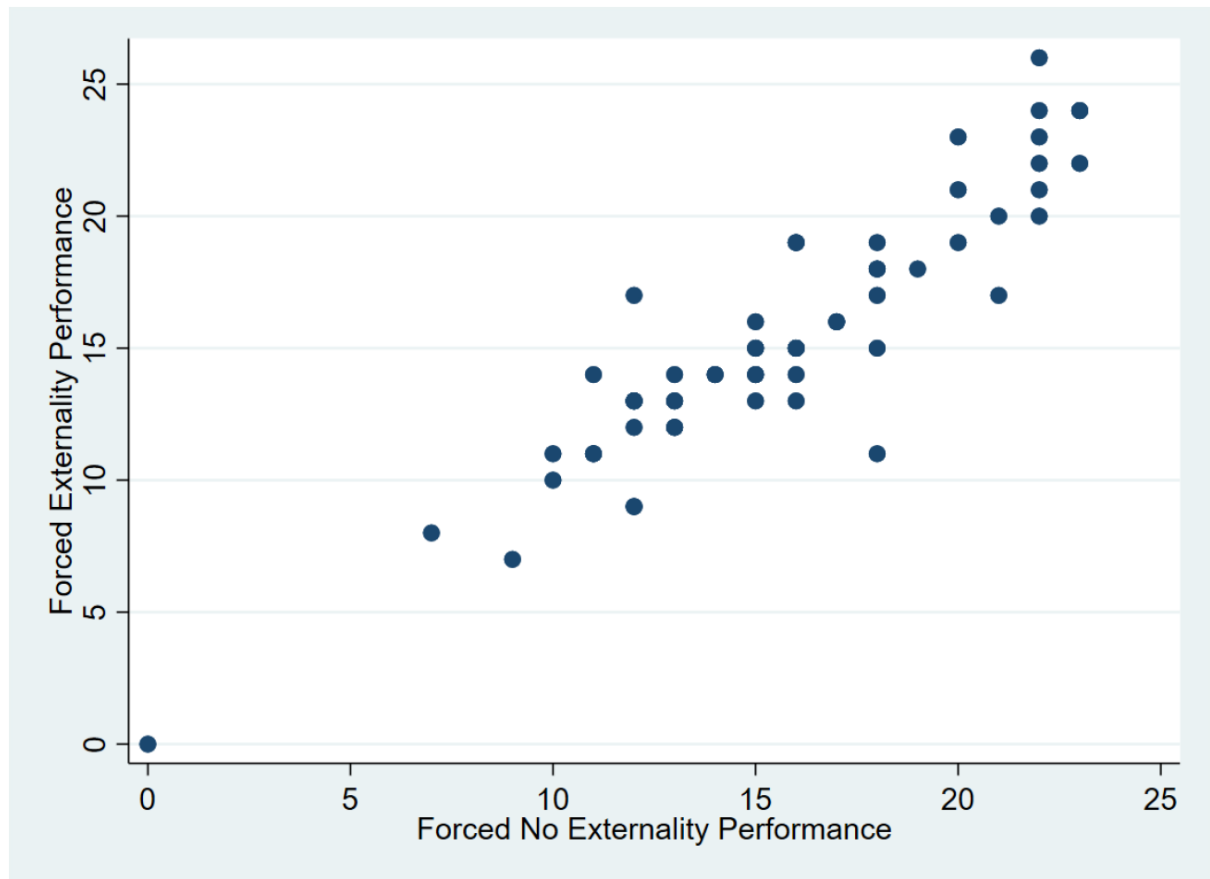

Figure 5: Scatterplot of firm performance under "Clean" (no externality) and "Dirty" (externality) contracts

Notes: The Y-axis represents the number of tokens the firm earned in the encryption task when the payoff was associated with a forced externality on the victim (the "Dirty" contract), while the x-axis represents the number of tokens the firm earned in the encryption task when no externality was imposed on the consumer (the "Clean" contract). From this graph, we observe no difference in performance between firms who participated in the encryption task with forced externality vs forced no externality.
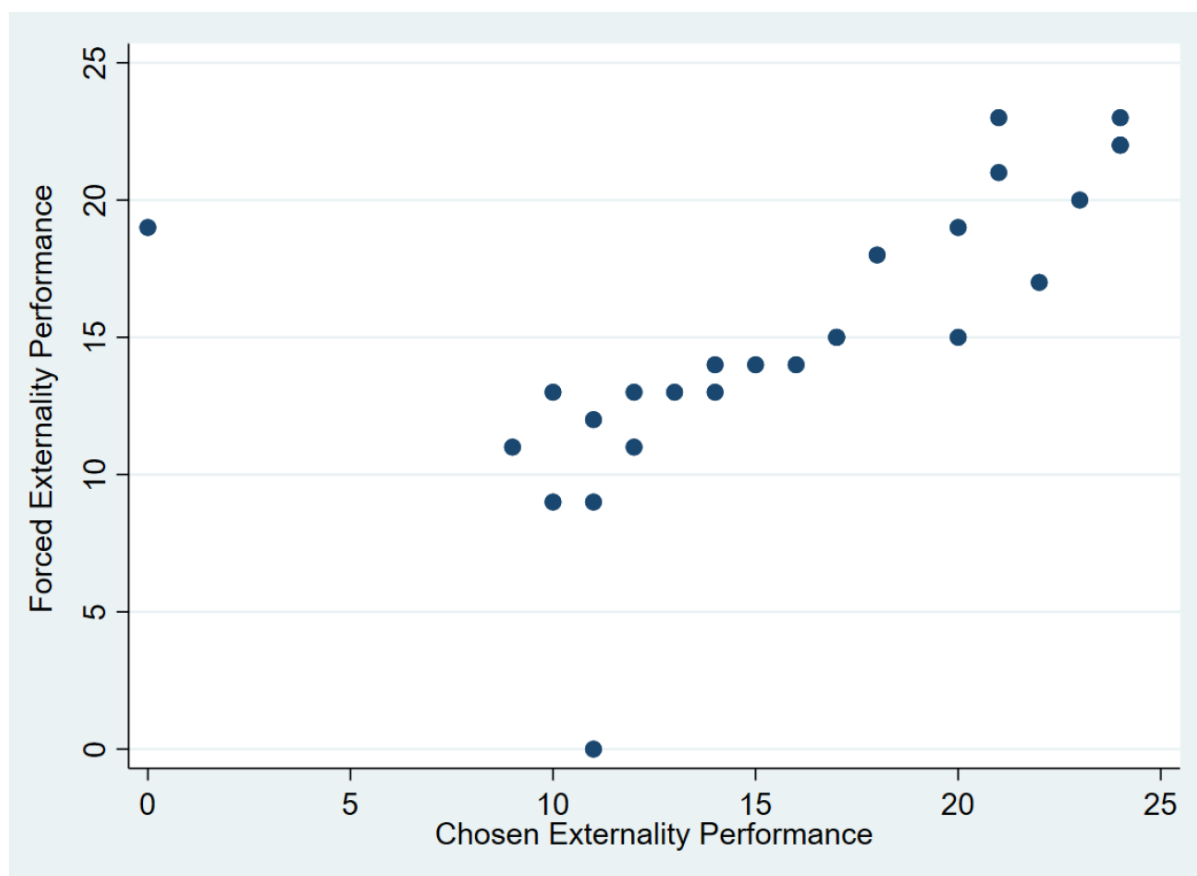
Figure 6: Scatterplot of firm performance under forced "Dirty" (externality) and chosen "Dirty" contracts

Notes: The Y-axis represents the number of tokens the firm earned in the encryption task when the payoff was associated with a forced externality on the consumer, while the x-axis represents the number of tokens the firm earned in the encryption task when the firm wilfully chose a payoff that would impose a negative externality on the consumer. From this graph, we observe that there's no difference in performance between perpetrators who participated in the encryption task with forced externality vs chosen externality.

Second, we examine whether chance imbalances in random assignment contribute to the observed differences. Firms assigned to the Random Choice control group in which the externality was forced have, on average, 15.1% lower baseline ability than the full sample of firms ($p < 0.01$), as measured in the practice rounds. This imbalance carries through to actual task performance, where these firms perform 8.7% worse than average ($p < 0.05$). Because ability is measured prior to treatment assignment, this difference arises purely by chance despite correct randomization procedures.

Taken together, these analyses show that the higher externality magnitude observed in the Wilful Choice treatment reflects a combination of learning effects over time and chance differences in baseline ability across treatment groups. Importantly, these factors do not

undermine our main results on consumer responses, which remain robust once externality magnitude and performance differences are accounted for.

4. Conclusion

Negative externalities can be unintended by-products of economic activity or due to deliberate decisions by firms to avoid harm-reducing costs. Consumers may not always perceive environmentally harmful production as the result of deliberate attempts to impose harm. As a result, consumer responses to environmental degradation are heterogeneous and sometimes inconsistent: similar environmental outcomes may elicit markedly different reactions depending on how firms' actions are perceived. This raises a central question for understanding sustainable production and consumption: whether and how consumers' responses depend not only on environmental outcomes, but also on firms' deliberate choices to avoid harm when doing so is privately costly.

In this paper, we study the role of intentionality in shaping consumer behaviour toward firms that generate, or avoid generating, negative externalities. Using a controlled laboratory experiment, we isolate the effect of firms' deliberate choices from the environmental outcomes by comparing settings in which production technologies are chosen freely with settings in which they are assigned randomly. Consumers observe both the firm's (lack of) choice and the resulting externality, and can respond through a give-or-take Dictator Game, allowing us to measure punitive behaviour and moral leniency in a clean and incentive-compatible way.

Our results reveal a clear asymmetry in how intentions matter. Consumers punish firms whenever a negative externality is incurred, regardless of whether the harm is imposed intentionally or randomly, indicating that punitive responses are largely driven by experienced outcomes. By contrast, when harm is avoided, intentions play a central role. Firms that deliberately choose to prevent a negative externality are treated with significantly greater moral leniency than firms for which the absence of harm arises randomly, manifested in positive transfers on average. No comparable differentiation emerges in the random-choice setting, where consumers consistently take away from firms irrespective of outcomes.

Taken together, these findings provide a behavioural explanation for why firms may voluntarily engage in costly ethical and environmentally sustainable practices when their choices are observable. While consumers do not appear to punish intentional harm more severely than

accidental harm, they do grant moral credit for deliberate harm avoidance in the form of attenuated punishment and even reward, leading us to conclude that there are benefits of good intentions. This asymmetry helps reconcile the prevalence of voluntary sustainability initiatives with the often-limited scope for consumer boycotts and highlights the importance of agency and observability in shaping the economic incentives for ethical production.

References

Ali, W., Bekiros, S., Hussain, N., Khan, S. A., & Nguyen, D. K. (2023). Determinants and consequences of corporate social responsibility disclosure: A survey of extant literature. Journal of Economic Surveys. https://doi.org/10.1111/joes.12556

Ardia, D., Bluteau, K., Boudt, K., & Inghelbrecht, K. (2023). Climate change concerns and the performance of green vs. brown stocks. Management Science. https://doi.org/10.1287/mnsc.2022.4636

Baker, Malcolm, Daniel Bergstresser, George Serafeim, and Jeffrey Wurgler. "The pricing and ownership of US green bonds." Annual review of financial economics 14, no. 1 (2022): 415-437.

Banuri, S., & Keefer, P. (2016). Pro-social motivation, effort and the call to public service. European Economic Review, 83, 139–164. https://doi.org/10.1016/j.euroecorev.2015.10.011

Banuri, Sheryar, Lucia Milena Murgia, and Imtiaz Ul Haq. "The power of religion: Islamic investing in the lab." International Review of Financial Analysis 96 (2024): 103690.

Bardsley, Nicholas. "Dictator game giving: altruism or artefact?." Experimental economics 11, no. 2 (2008): 122-133.

Bicchieri, C., & Maras, M. (2022). Intentionality matters for third-party punishment but not compensation in trust games. Journal of Economic Behavior & Organization, 197, 205–220. https://doi.org/10.1016/j.jebo.2022.02.026

Blount, S. (1995). When social outcomes aren't fair: The effect of causal attributions on preferences. Organizational Behavior and Human Decision Processes, 63(2), 131–144. https://doi.org/10.1006/obhd.1995.1068

Bone, J. E., & Raihani, N. J. (2015). Human punishment is motivated by both a desire for revenge and a desire for equality. Evolution and Human Behavior, 36(4), 323–330. https://doi.org/10.1016/j.evolhumbehav.2015.02.002

Brandts, J., & Solà, C. (2001). Reference points and negative reciprocity in simple sequential games. Games and Economic Behavior, 36(2), 138–157. https://doi.org/10.1006/game.2000.0818

Braouezec, Y., & Joliet, R. (2019). Time to invest in corporate social responsibility and the value of CSR operations: The case of environmental externalities. Managerial and Decision Economics, 40(5), 539–549. https://doi.org/10.1002/mde.3024

Catlin, J. R., Luchs, M. G., & Phipps, M. (2017). Consumer perceptions of the social vs. environmental dimensions of sustainability. Journal of Consumer Policy, 40, 245–277. https://doi.org/10.1007/s10603-017-9356-x

Charness, G. (2004). Attribution and reciprocity in an experimental labor market. Journal of Labor Economics, 22(3), 665–688. https://doi.org/10.1086/383111

Charness, G., & Levine, D. I. (2007). Intention and stochastic outcomes: An experimental study. The Economic Journal, 117(522), 1051–1072. https://doi.org/10.1111/j.1468-0297.2007.02066.x

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. Quarterly Journal of Economics, 117(3), 817–869. https://doi.org/10.1162/003355302760193904

Charness, G., & Rabin, M. (2005). Expressed preferences and behavior in experimental games. Games and Economic Behavior, 53(2), 151–169. https://doi.org/10.1016/j.geb.2004.09.010

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance, 9, 88–97. https://doi.org/10.1016/j.jbef.2015.12.001

Clementino, E., & Perkins, R. (2021). How do companies respond to environmental, social and governance (ESG) ratings? Evidence from Italy. Journal of Business Ethics, 171, 379–397. https://doi.org/10.1007/s10551-020-04441-4

Creyer, E. H., & Ross, W. T. Jr. (1997). The influence of firm behavior on purchase intention: Do consumers really care about business ethics? Journal of Consumer Marketing, 14(6), 421–432. https://doi.org/10.1108/07363769710185999

Dentsu & Microsoft. (2021). The rise of sustainable media. https://sustainablemedia.dentsu.com/home1/home

Erkal, N., Gangadharan, L., & Nikiforakis, N. (2011). Relative earnings and giving in a real-effort experiment. American Economic Review, 101(7), 3330–3348. https://doi.org/10.1257/aer.101.7.3330

Fairtrade International. (2023). Monitoring the scope and benefits of Fairtrade: Monitoring report (14th ed.). https://www.fairtrade.net/library/monitoring-the-scope-and-benefits-of-fairtrade-14th-edition

Falk, A., Fehr, E., & Fischbacher, U. (2003). On the nature of fair behavior. Economic Inquiry, 41(1), 20–26. https://doi.org/10.1093/ei/41.1.20

Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. Games and Economic Behavior, 62(1), 287–303. https://doi.org/10.1016/j.geb.2007.06.001

Golicic, S. L., & Smith, C. D. (2013). A meta-analysis of environmentally sustainable supply chain management practices and firm performance. Journal of Supply Chain Management, 49(2), 78–95. https://doi.org/10.1111/jscm.12006

Govind, R., Singh, J. J., Garg, N., & D'Silva, S. (2019). Not walking the walk: How dual attitudes influence behavioral outcomes in ethical consumption. Journal of Business Ethics, 155, 1195–1214. https://doi.org/10.1007/s10551-017-3545-z

Hainmueller, Jens, Michael J. Hiscox, and Sandra Sequeira. "Consumer demand for fair trade: Evidence from a multistore field experiment." Review of Economics and Statistics 97, no. 2 (2015): 242-256.

Hosta, M., & Zabkar, V. (2021). Antecedents of environmentally and socially responsible sustainable consumer behavior. Journal of Business Ethics, 171, 273–293. https://doi.org/10.1007/s10551-019-04416-0

Knobe, J. (2003). Intentional action and side effects in ordinary language. Analysis, 63(3), 190–194. https://doi.org/10.2307/3329308

Kuokkanen, H., & Sun, W. (2020). Companies, meet ethical consumers: Strategic CSR management to impact consumer choice. Journal of Business Ethics, 166, 403–423. https://doi.org/10.1007/s10551-019-04145-4

List, J. A. (2007). On the interpretation of giving in dictator games. Journal of Political Economy, 115(3), 482–493. https://doi.org/10.1086/519249

Matten, D., & Moon, J. (2008). "Implicit" and "explicit" CSR: A conceptual framework for a comparative understanding of corporate social responsibility. Academy of Management Review, 33(2), 404–424.

Nelson, W. R. (2002). Equity or intention: It is the thought that counts. Journal of Economic Behavior & Organization, 48(4), 423–430. https://doi.org/10.1016/S0167-2681(01)00245-1

Offerman, T. (2002). Hurting hurts more than helping helps. European Economic Review, 46(8), 1423–1437. https://doi.org/10.1016/S0014-2921(01)00176-3

Stanca, L. (2010). How to be kind? Outcomes versus intentions as determinants of fairness. Economics Letters, 106(1), 19–21. https://doi.org/10.1016/j.econlet.2009.09.013

Tan, Qin Lingda, Sharizal Hashim, and Zhangwei Zheng. "Environmental Social Governance (ESG) in Digitalization research: A bibliometric analysis." SAGE Open 15, no. 1 (2025): 21582440241310953.

Tan, Qin Lingda, Sharizal Hashim, Nor Liza Abdullah, Hafizah Omar Zaki, and Zhangwei Zheng. "Environmental social governance (ESG) and consumer behavior: a systematic review and research agenda." International Journal of Consumer Studies 49, no. 5 (2025): e70105.

Utikal, V., & Fischbacher, U. (2014). Attribution of externalities: An economic approach to the Knobe effect. Economics and Philosophy, 30, 215–240. https://doi.org/10.1017/S0266267114000170

Whitson, J. A., Wang, C. S., See, Y. H. M., Baker, W. E., & Murnighan, J. K. (2015). How, when, and why recipients and observers reward good deeds and punish bad deeds. Organizational Behavior and Human Decision Processes, 128, 84–95. https://doi.org/10.1016/j.obhdp.2015.03.006