RATIONALITY & COMPETITION
CRC TRR 190

Motivating Crowdworkers with Nonmonetary Incentives and Payment Framing—Evidence from a Large-Scale Experiment

Maja Adena (WZB Berlin, TU Berlin)
Patrycja Janowska-Widomska (University of Warsaw)
Julian Harke (WZB Berlin)

Discussion Paper No. 576

June 05, 2026

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, Ludwig Maximilian University of Munich, 80539 Munich, Germany
info@rationality-and-competition.de

Motivating Crowdworkers with Nonmonetary Incentives and Payment Framing—Evidence from a Large-Scale Experiment

Maja Adena (WZB & TU Berlin), Patrycja Janowska-Widomska (UW), and Julian Harke (WZB)*

Abstract:

An increasing number of individuals worldwide are participating in crowdwork and telework. They often perform tasks such as AI training and content moderation. While these tasks are typically conducted in large quantities and often involve routine elements, their nature makes them inherently demanding. They require high levels of engagement or creativity and produce outputs of subjective quality that are difficult to measure. In a preregistered field experiment involving over 5,500 crowdworkers, we examined the impact of automated recognition and work-appreciation phrases and payment framing on motivation, performance, and job satisfaction. The results indicate that recognition—automated phrases: *Great work! You did a good job! Nice job! Well done!*—positively influences subjective job satisfaction, and that loss-framed payment is somewhat more effective than gain framing. Overall, the treatments have little effect on objective and subjective performance and motivation.

Keywords: crowdworkers, complex tasks, routine, automated motivation, nonmonetary incentives, recognition and appreciation, loss and gain framework, experiment

JEL: J33, J24, M54, M52, D83, C93

* Maja Adena (0000-0002-6378-3466): Wissenschaftszentrum Berlin für Sozialforschung, Reichpietschufer 50, D-10785 Berlin, Maja.Adena@wzb.eu; Patrycja Janowska-Widomska (0000-0002-0853-2170): University of Warsaw Faculty of Economic Sciences, Długa 44/50, 00-241 Warsaw, Poland, p.janowska2@uw.edu.pl, Julian Harke: Wissenschaftszentrum Berlin für Sozialforschung, Reichpietschufer 50, D-10785 Berlin, julian.harke@gmx.de. We are grateful to Katharina Dorn, Lena Simmat, Steffen Mayer, Kea von Daak, Anna Zemskov, and Marius Werz for excellent research assistance, and many others for help in conducting the field experiment. We preregistered the design, hypotheses, construction of the variables, and the analysis of the experiment before any data collection at the AEA RTC registry (https://doi.org/10.1257/rct.6744). We obtained the IRB approval from the WZB Research Ethics Committee (IRB Approval Number 2019/5/82) before the data collection started. Support by Deutsche Forschungsgemeinschaft through project number 417014946, through CRC TRR 190 (project number 280092119) and through FOR 5675 (project number 518302089), and by Leibniz (P146/2022 and K595/2024) is gratefully acknowledged (Maja Adena).

1. Introduction

Crowdworking—that is, working through online platforms—is a relatively recent phenomenon that is attracting a substantial and growing number of individuals.[1] Crowdsourcing, by contrast, is a growing practice of transferring tasks traditionally done by a company's or institution's employees to a broad, unspecified network of people through an open call (Howe 2006). Crowdworking and crowdsourcing are widely used in AI settings, where millions of workers, particularly in the Global South, are paid to manually annotate and process vast quantities of data (Gonzalez-Cabello et al. 2024). Although this work often involves repetitive tasks, it increasingly demands engagement, cognitive effort, or creativity, as technological advancements reduce reliance on human labor for purely routine tasks. In many such settings, workers follow standardized procedures while also making judgments in cases where outputs are difficult to verify objectively. This shift towards remote work[2] and crowdworking has enhanced information asymmetry, meaning that the worker understands the extent of their commitment, but the employer is less able to assess it. As a result, worker motivation becomes crucial and directly impacts the quality and accuracy of the output. However, the factors that may enhance motivation in nontraditional work settings and realistic modern tasks that are neither purely routine nor purely creative remain insufficiently understood (Hossain and Kauranen 2015).

This study seeks to address how simple nonmonetary and monetary interventions—specifically, automated recognition or appreciation phrases and payment framing—affect worker outcomes such as motivation, job satisfaction, and performance in the context of routinized yet complex crowdwork. Although prior research has extensively examined incentive design in clearly routine or clearly creative settings,[3] the predominant reality of work often lies between these extremes. We therefore focus on this hybrid context to address a gap in the literature.

Our focal point lies at the intersection of routinized procedure and subjective judgment. Tasks such as text annotation or AI training often require a blend of routine steps and nonroutine cognitive processing, which makes them particularly interesting from a motivational standpoint. While some aspects of these tasks may be repetitive, workers must frequently exercise judgment, especially in settings where outputs lack objective benchmarks. This combination raises questions about whether motivation-enhancing interventions function similarly when

[1] Brancati et al. (2019) consider various approaches to estimating the size of what they term the platform economy and suggest that up to 1.79% of European workers offer services through online platforms. Using a different definition, Katz and Krueger (2018) estimate that 0.5% of individuals aged 18 and over in the U.S. engage in platform work as their primary occupation.

[2] In addition, there is a long-standing trend towards working from home, significantly amplified by the COVID-19 pandemic. According to Barrero et al. (2023), as of mid-2023, 28% of paid workdays among U.S. citizens aged 20–64 were conducted remotely.

[3] For a review, see Attanasi et al., 2021.

tasks are standardized in procedure but difficult to evaluate objectively. We therefore contribute to the emerging literature that examines how the nature of the task—particularly the combination of routine elements and subjective interpretation—interacts with incentive schemes in shaping worker behavior.[4][5]

For our preregistered experiment, we recruited over 5,500 crowdworkers through Prolific. Their task was to read a newspaper article about a natural disaster and to answer questions related to it, including ones that asked them to identify specific details and select relevant parts of the text. The task was not trivial, as many cases lacked an objective solution and therefore required participants to exert cognitive effort, even though Prolific participants routinely participate in online studies. Each text was evaluated by six participants. Participants received a fixed payment upon task completion and additional bonuses depending on the correct completion of attention checks. We did not deceive participants—their work was needed as a real input to train a machine learning model for use with an even larger set of articles. This is a typical type of task that researchers and AI developers use when standardized procedures must be combined with human judgment. In such cases, subjective interpretations can lead to mistakes that provide erroneous inputs into AI systems trained on this data. Therefore, ensuring that crowdworkers are motivated to provide the highest-quality responses possible is essential. Because the texts differed and no clearly objective correct solution existed, we assessed worker performance using a wisdom-of-crowds approach to evaluate the effectiveness of our interventions.

We chose very simple, automated, and low-cost interventions in both the nonmonetary and monetary domains. In the nonmonetary domain, participants were briefly presented with screens displaying phrases expressing recognition of the work done (*Great work! You did a good job! Nice job! Well done!*) or screens expressing appreciation for their work (*Thank you! Your help makes a difference! We appreciate your support! Your work matters!*), or control blank pages. Such phrases are commonly used in contexts like remote sports classes to motivate participants, even when they are not directly linked to individual performance. In the monetary domain, we studied the effects of framing the payment in terms of gains or losses based on performance in attention checks due to the lack of objective performance measures for the main task. This is analogous to time-tracking systems in traditional workplaces, where compensation

[4] See, for example, Bénabou and Tirole (2003, 2006), Gneezy et al. (2011), Kosfeld et al. (2017), Chandler and Kapelner (2013), and related work showing that incentive effects vary with task characteristics such as routinization, subjective evaluation, and meaning.

[5] We, therefore, emphasize that our task is not creative in the narrow sense of generating novel outputs; rather, it is a hybrid task that combines standardized procedure with bounded interpretation and subjective judgment.

is linked either to recorded hours rather than current effort—which is often unmeasured or even unobservable—or to results, which may emerge only later or as a result of collective work.

Our research aims to shed light on mixed findings regarding how nonmonetary and monetary rewards in nontraditional work environments influence performance, motivation, and job satisfaction in modern blended tasks that are routinized in procedure yet cognitively demanding. Unlike previous studies, (i) we use a realistic context, namely text evaluation as an input for machine learning models, (ii) we incorporate many variants of these texts, ensuring that our results do not depend on a specific version, (iii) we use a task without a clear, objective, correct solution, as is often the case in real-world scenarios, and (iv) we rely on a uniquely large sample compared to previous studies, overcoming power limitations and reducing the risk of false positives. This positioning is also central to the topic of routinization and creativity: the setting allows us to study incentive effects in tasks that combine routine procedures with case-by-case interpretation, rather than treating creativity and routinization as mutually exclusive categories.

In this study, which we conducted in accordance with our preregistration and analysis plan, we find that automated nonmonetary recognition—displaying messages like *Great work!*, *You did a good job!*, *Nice job!*, or *Well done!*—significantly enhanced job satisfaction among crowdworkers. However, it did not translate into measurably higher motivation or performance. Our results also indicate that while framing rewards as losses rather than gains may modestly improve performance on objectively verifiable checks, this pattern is limited and not robust across outcomes. Given the high statistical power of our study, our findings suggest clear limits to what very simple automated methods can achieve in one-off crowdwork settings.

The remainder of the manuscript is organized as follows: Section 2 provides a review of the relevant literature, Section 3 details the study design and hypotheses, Section 4 presents empirical results, and Section 5 offers concluding remarks.

2. Literature

2.1 Incentives in routine and nonroutine tasks

A large body of research studies how monetary and nonmonetary incentives affect motivation, job satisfaction, and performance. While financial incentives can increase effort, their effects are often heterogeneous and may depend on context, task characteristics, and worker preferences (e.g., Gneezy and Rustichini, 2000; Fehr and Goette, 2007; Ariely et al., 2009; Charness et al., 2020). Similarly, nonmonetary incentives—such as recognition, social comparison, reputation systems, and mission alignment—can shape behavior, sometimes increasing effort but also generating unintended effects (e.g., Kosfeld and Neckermann, 2011; Bradler et al., 2016;

Burbano, 2016; Cassar and Meier, 2018; Charness et al., 2023). Related evidence shows that even subtle changes in task framing and communication—such as emphasizing meaning or embedding motivational messages—can affect effort in online labor markets (Chandler and Kapelner, 2013; Kosfeld et al., 2017; Kvaløy et al., 2015). Prior research highlights that incentive effects differ systematically between routine and creative tasks. In routine settings with clearly measurable outputs, incentives can be tightly linked to performance. In contrast, in tasks where effort is difficult to observe or output is not objectively measurable, incentives operate more indirectly through motivation, perceived meaning, and social context (Bénabou and Tirole, 2003, 2006; Ellingsen and Johannesson, 2008). Creative tasks—typically defined as requiring the production of novel or original outputs—respond differently to incentives, with extrinsic rewards sometimes crowding out intrinsic motivation (Deci et al., 1999; Amabile, 1996; see also Attanasi et al., 2021 for a review).

Experimental studies of creativity often involve open-ended production tasks such as writing, design, or problem solving that requires 'thinking outside the box' (e.g., Charness and Grieco, 2019; Gross, 2020; Bradler, Neckermann, & Warnke, 2020). Many real-world tasks—including crowdwork—do not require creativity in the pure open-ended sense but instead combine routinized procedure with bounded creativity, as workers must exercise judgment in the absence of clearly correct answers.[6] Examples include text annotation, content moderation, or classification tasks with ambiguous cases. These tasks combine standardized procedures with subjective evaluation, placing them between purely routine and fully creative work. Compared to purely routine or purely creative tasks, relatively little is known about how incentives operate in such blended environments, where routinization and creativity are intertwined and output is difficult to verify. In a setting closer to ours, DellaVigna and Pope (2018) show that crowdworkers respond to both monetary and psychological motivators, with somewhat stronger—but statistically imprecise—responses to loss-framed incentives. Our study contributes to this literature by focusing on such environments: standardized tasks performed at scale in which output is difficult to evaluate objectively.

We examine two classes of incentives applicable in such environments: (i) payment framing (gain vs. loss) of additional payments tied to measurable yet irrelevant output, and (ii) nonmonetary recognition and appreciation, implemented through automated feedback messages. More broadly, evidence on gain- versus loss-framed incentives is mixed, with some studies finding advantages of loss framing and others reporting small or context-dependent differences (de Quidt et al., 2017; Armantier and Boly, 2015; Czibor et al., 2022). By jointly studying

[6] We refer to such tasks as complex cognitive tasks.

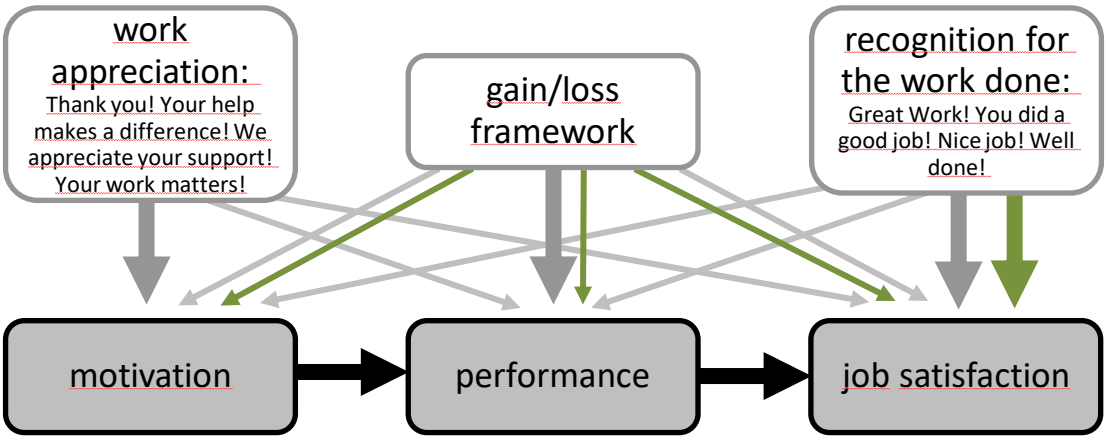
these incentives in a large-scale crowdwork setting, our paper contributes to the literature in three ways. First, it provides evidence on incentive effects in tasks that are neither purely routine nor purely creative but instead sit at the intersection of routinization and creativity, requiring sustained cognitive effort and subjective judgment. Second, it examines how subtle differences in incentive design—such as payment framing and recognition—affect not only performance but also job satisfaction. Third, it extends existing work on incentives in online labor markets by focusing on environments in which performance cannot be objectively benchmarked and must instead be inferred using proxy measures.

2.2. The interdependence of motivation, performance, and job satisfaction

The relationship between motivation, job satisfaction, and job performance has been widely studied in work psychology. A large body of research shows that motivation positively affects both performance and job satisfaction (e.g., Cerasoli et al., 2014; Tamam and Sopiah, 2022; Wu et al., 2020). Monetary incentives may also shape job attitudes, not only effort and performance (e.g., Ockenfels et al., 2015). This relationship is grounded in classic theories such as Herzberg's two-factor theory and Locke's goal-setting theory, which link motivation to improved performance and positive job attitudes (Herzberg et al., 1959; Locke, 1970). Subsequent work further examines the mechanisms and strength of the performance–satisfaction relationship (e.g., Bowling, 2013). Taken together, the literature indicates that motivation significantly influences both performance and job satisfaction.

Combining the insights presented above from the literature with our treatments, we anticipate that the relationships in our study will resemble those illustrated in Figure 1. While some of the literature suggests that job satisfaction affects performance (e.g., Abdullah & Wan, 2013), we expect such causal links to emerge only in long-term, repeated work settings. Therefore, we expect performance to affect job satisfaction; however, our experiment is not designed to test this relationship.

Figure 1: Conceptual framework: Relationships between motivation, performance, and job satisfaction in a one-off worker-employer relationship, as well as the expected influence of recognition, work appreciation, and gain/loss framework.



Note: Bold arrows indicate expected main channels; grey arrows indicate relationships examined empirically. Green bold arrows indicate confirmed hypotheses, and green thin arrows indicate that some evidence was found. Figure 1 is intended as a conceptual framework summarizing ex ante expectations rather than as a visual summary of confirmed causal effects.

3. Design of the experiment

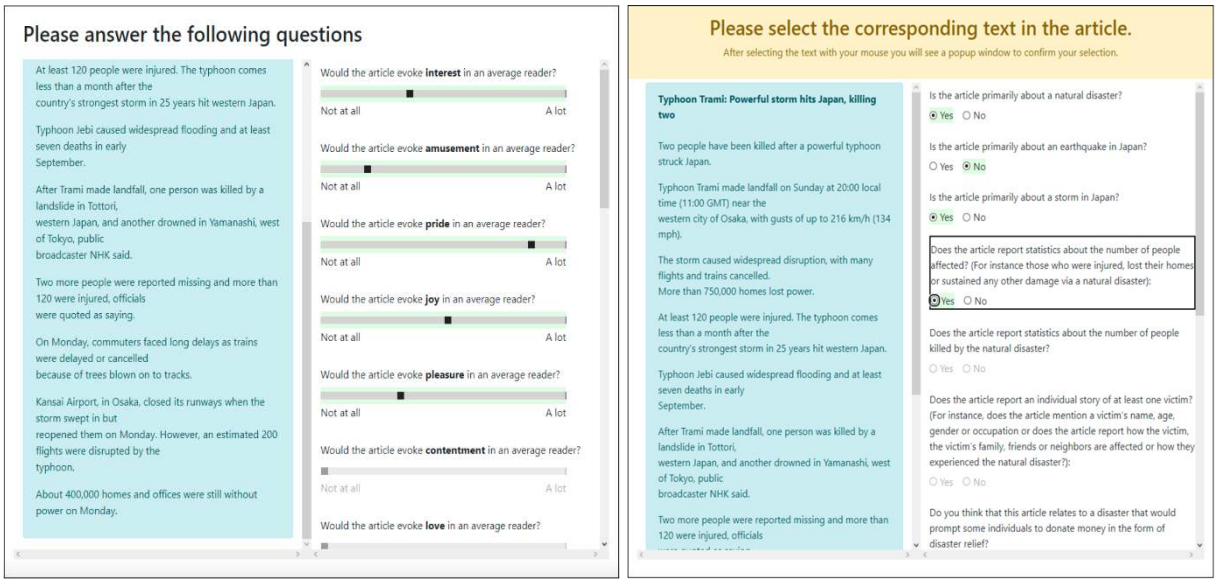
We hired crowdworkers via the Prolific platform, a popular choice among researchers and private companies for AI training, evaluation tasks, surveys, and experiments. Our specific task involved reading and classifying articles on natural disasters. The ultimate objective was to train a machine learning model to analyze how media reporting of natural disasters—particularly the attributes of articles—impacted donations to international relief charities. Given the nature and complexity of the task, we decided to limit the pool to native English speakers currently residing in the UK, holding at least a bachelor's degree, and having approval rates between 99 and 100, see Appendix A.

Participants had to read a newspaper article about a natural disaster and answer an array of questions about it organized into four subtasks on separate screens. After reading the article, on the next screen, participants had to evaluate, for 20 different types of emotions from the Geneva Emotion Wheel (Scherer, 2005), how strongly they would expect the article to evoke each particular emotion in an average reader.[7] For example, we inquired whether the articles would evoke interest, joy, or regret. On two subsequent screens, we asked whether specific

[7] Because the sample size per article was small, we asked participants to assess the emotional impact of the articles from the perspective of an average reader in order to obtain a more objective measure. As this assessment was not incentivized, participants had no reason to bias their responses in any way.

information was included in the article—for example, whether the article reported on an individual or statistics about deaths and injuries. If participants answered these questions affirmatively, they were expected to select the corresponding parts of the article. In the final task, participants were asked to highlight all words related to the natural disaster in the article. Example screenshots are displayed in Figure 2. For all screenshots of the task featuring an example article, please refer to the online Appendix. The experiment was programmed in oTree (Chen et al. 2016).

Figure 2: Example screenshots from the experiment.



Note: The participants were able to refer to the article at all times—it was visible on the left part of the screen.

As in typical real-world context, the crowdworkers were assigned different articles. Therefore, there were no objective benchmarks for their answers (e.g., in a context of an article describing a heatwave the answer to the question "Do you think this article relates to a disaster that would prompt some individuals to donate money in the form of disaster relief?" is necessarily subjective). In principle, there was no way to perfectly verify how well the task had been completed. Even the questions that seem more objective, like "Does the article report statistics about the number of people affected?" might still be ambiguous if the text reports, for example, that "families were displaced." For the purpose of evaluating the output, we assigned the same article to six crowdworkers, and, relying on the wisdom of crowds, treated the majority answer within that group as the benchmark. Participants were not informed about this aggregation rule, and it did not affect their payments. At the same time, this measure should be interpreted as a consensus-based proxy for quality rather than as objective ground truth. This approach appears

particularly reasonable given that we selected a relatively homogeneous group of high-performing crowdworkers with at least a bachelor's degree.

In addition, on the first three of the four pages with questions about the article, we included one attention check per page, which was placed among the other questions. This was intended to increase attention and, consequently, the quality of responses. Workers were informed ex-ante about attention checks and payment structure associated with them. However, these attention checks were unrelated to the output of interest to the employer and were identical for all participants. The participants had to answer those attention checks correctly in order to be paid an additional bonus of *£0.75*. This payment accounted for a large fraction of the final payment,[8] 42% on average. All attention checks were easy to answer and had an objectively true answer. Each time the participant failed an attention-check question, their bonus payment was reduced by *£0.25.* At this point, participants received a notification, and the mistake was explained (the question was repeated and the answer given by the participant and the correct answer were presented). More specifically, the first and third attention checks required the participant to move the slider until the question text turned green. The second attention check required the participant to select "yes" and subsequently fill in the word "cycling."

The task concluded with a series of questions about subjective perceptions of the task, focusing on job satisfaction and worker motivation.

3.1. Treatments

We followed a 3x2 between-subjects design. Each article was assigned to six participants such that all six treatments were implemented within the same article. The experiment manipulated two dimensions: nonmonetary incentives and payment framing.

The first treatment dimension concerned nonmonetary incentives. After each of the four pages with the questions about the article, participants were shown a short message screen for 1.7 seconds[9], depending on treatment condition. These messages either conveyed recognition, appreciation, or were absent in the control condition (see Table 1 for the exact phrasing). Messages were displayed only after task completion and were not randomized across pages. Because the messages appeared only after the first task screen, analyses of nonmonetary treatments use only outcomes measured after the first exposure to the treatment.

[8] The total payment was determined by the length of the article the participants worked on and the number of correctly answered attention checks.

[9] An average reader needs approximately 0.3 seconds for one word; that is 1.5 seconds for a phrase with five words. We increase the time a little bit to allow slower readers to still understand the phrase. Due to the preselection of participants with or in higher education, we judged this to be sufficient.

If a participant failed an attention check, the message screen was replaced with a notification about the incorrect answer. Such failures were rare: 96% of participants passed all three attention checks. Importantly, the attention checks differed from the primary task relevant to the employer, namely text analysis and classification.

The second treatment dimension involved framing payment as gains or losses. Participants received equivalent monetary incentives, framed either as bonuses for correct answers or deductions for incorrect answers. Payment information was provided at the outset and remained accessible throughout the experiment. Table 1 provides an overview of all treatment conditions. The exact wording of the instructions is reported in Appendix B.

Table 1: Experimental Treatments

Treatment	Description	Exact Phrases and Formulations
Panel A. Nonmonetary Incentives		
Control (C)	No recognition or appreciation shown; participants see blank screens.	—
Recognition (R)	Phrases expressing recognition of work, shown after a set of tasks is completed.	"Great work!", "You did a good job!", "Nice job!", "Well done!"
Work-Appreciation (A)	Phrases expressing appreciation for contribution, shown after a set of tasks is completed.	"Thank you!", "Your help makes a difference!", "We appreciate your support!", "Your work matters!"
Panel B. Payment Framing		
Gain (G)	Participants are informed that they will receive a baseline payment plus a bonus for each correctly solved attention check.	If you finish the task you will be paid [G: up to £1.65; L: at least £0.90]. Your payment depends on you meeting the quality requirements. The quality requirements are based on three simple questions that have a clear and objectively correct answer. These questions are positioned between other questions and tasks. Each time you [L: fail to] correctly answer the quality check question, your payment will be [G: raised; L: reduced] by £0.25. The maximum [G: additional payment is; L: reduction equals thus to] £0.75, in which case you will be [L: only] paid [G: £0.90].
Loss (L)	Participants are informed that they will receive a maximum payment, from which amounts are subtracted for each incorrectly solved attention check.	

Note. Gain and loss framing were payment-equivalent; the difference lay only in presentation.

3.2. Power

With a total of over 5,500 participants and a minimum of 1,800 participants per treatment arm (note that the two treatment dimensions were orthogonal), we had 0.8 power to detect a very small effect size of less than 0.1 in pairwise comparisons using a two-sided t-test. When

requiring power to be 0.95, we were still able to detect a very small effect size of 0.12. Compared to many other studies in this context, our study had a very high number of participants and thus much higher power, thereby allowing us to reevaluate previous findings more robustly.

3.3. Hypotheses

Based on the literature presented above, we preregistered (AEARTC registry, https://doi.org/10.1257/rct.6744) the following main hypothesis before any data collection:

Hypothesis 1 (H1): Recognition/work appreciation increases worker motivation/job satisfaction/performance.

We hypothesized that expressing recognition and work appreciation would lead to higher levels of work motivation, job satisfaction, and performance as presented in Figure 1. Specifically, in our experiment, we aimed to examine whether motivation, job satisfaction, or performance was higher in treatments A and R than in treatment C.

In addition, we formulated the following additional hypotheses:

Hypothesis 2 (H2): Performance is higher in A than in R.

According to our hypothesis, as shown in Figure 1, work appreciation (treatment A) would mainly affect motivation, which in turn would positively affect performance, while recognition was expected to mainly affect job satisfaction.[10]

Hypothesis 3 (H3): The probability of mistakes is lower in treatment L than in G for objectively verifiable tasks (attention checks) and higher for the less objective, employer-relevant tasks.

H3 focused on the concept of "work to rule," where adherence to strict procedural guidelines impacts task performance differently depending on the nature of the tasks. For objectively verifiable tasks (here: attention checks), we expected a lower probability of mistakes in treatment L than in treatment G. Conversely, for tasks that were less objectively assessable (i.e., performance relevant to the employer), we predicted a higher probability of mistakes in treatment L. This prediction was based on the notion that, in a loss domain scenario, participants

[10] This expectation was closely related to the one-off nature of the task. In repeated interactions, job satisfaction could influence motivation and performance.

would be inclined to strictly adhere to rules and guidelines for tasks that can be objectively validated, potentially sacrificing effort and creativity in tasks that cannot be objectively assessed.

Because the literature did not provide clear expectations regarding L and G with respect to subjective job satisfaction and motivation, we treated those comparisons as exploratory. Additionally, we preregistered two further hypotheses: H4 on heterogeneity by experience and H5 on the absence of interaction effects between nonmonetary and monetary treatments.

3.4. Dependent variables

The construction of the following dependent variables was prespecified in the preregistration:

3.4.1. Worker motivation

Worker motivation was assessed using both subjective and objective measures. For the subjective measure, we collected responses to questions adapted from the Multidimensional Work Motivation Scale (Gagné et al., 2015). Participants were asked to indicate their reasons for putting effort into the task using the following prompts: I put little effort into this task because I didn't think that it was worth it; I put effort into this task because others will only reward me financially if I put enough effort into this task; I put effort into this task because otherwise I would have felt bad about myself; Putting effort into this task aligns with my personal values. Responses were recorded using a slider on a 101-point scale ranging from 0 ("not at all") to 100 ("very much"). The scale was not visible to participants to avoid bunching on round numbers (using a slider of the type shown in Figure 2).

The final item relates to the literature on how a company's mission—as reflected in corporate social responsibility statements, declared social missions, or nonprofit orientation—affects workers' performance, reservation wages, the likelihood of applying for a job, the quality of the applicant pool, and retention (Bode et al., 2015; Burbano, 2016; Hedblom et al., 2019). Workers tend to exert more effort when an organization's mission aligns with their own preferences (Carpenter and Gong, 2016). However, most of this evidence comes from traditional employment settings, where workers face longer-term organizational relationships and stronger identification channels. In crowdsourcing markets, by contrast, tasks are typically short-term, anonymous, and low-commitment, so motivational content may operate differently or be less salient. We therefore examine whether participants perceive the task as aligned with their personal values and whether these perceptions vary across treatments.

For the objective measure of worker motivation, we analyzed the number of words or phrases selected in the disaster-word selection task. In this task, we asked the participants to

select ‘the most important words that refer to the natural disaster.’ The number of words was not specified, and participants were able to move to the next page after selecting one or more words. Of course, the potential number of relevant words differed by the article. In our analysis, we accounted for this by including article fixed effects, thus making comparisons within one article.

3.4.2. Performance

In our study, we evaluated performance using both objective and subjective measures. The objective measure involved a count of mistakes in attention checks, though these checks were not directly linked to outcomes relevant to the employer. This approach was similar to attendance-based pay (rather than productivity-linked pay) in traditional work settings. When studying the nonmonetary incentives, we only used the second and third attention check as an objective measure of performance, since recognition and appreciation phrases were only displayed to the participants after they finished the first set of tasks. For hypotheses related to the gain/loss framing, all attention checks served as objective performance measures, however, we note again that they do not reflect the relevant performance.

As for performance relevant for the employer, we had to rely on a less objective measure, based on the wisdom of crowds, because there were no clear correct answers. More specifically, we relied on answers to yes/no questions pertaining to the newspaper article. Individuals who answered ‘yes’ had to highlight the relevant text. We applied the following rule of thumb to assess accuracy: If the majority of the other participants in the same group of six gave the same answer to a question, we assumed those responses were correct. This procedure provides a practical benchmark for ambiguous annotation tasks, but it captures consensus-based quality rather than objective correctness. Participants were not informed about this aggregation rule, which made strategic coordination on what others might answer unlikely. Additional measures captured the internal consistency of individual responses. Workers who paid more attention to the task were expected to provide more consistent answers. Accordingly, for the emotion items, we expected responses to be more similar within the set of negative emotions and within the set of positive emotions than between the two sets. For each participant, we therefore calculated the following formula (i): -*((positive_emotions_sd*negative_emotions_sd))/(all_emotions_sd*all_emotions_sd)*, where sd denotes the standard deviation of individual answers. For the job-satisfaction items, we expected more attentive workers to give an answer to the question ‘fun’ that was close to the opposite of ‘boring.’ Accordingly, we defined the variable as (ii) -|*job_sat_fun - (100 -*

job_sat_boring)|, that is, the negative absolute difference between the two values. Note that the measures (i) and (ii) are negative by construction, with higher values (closer to zero) indicating greater internal consistency.

3.4.3. Job satisfaction

To assess job satisfaction, we selected specific questions based on relevant literature and tailored them to our context. For example, we partially drew on the work of Thompson and Phua (2012), Macdonald and Maclntyre (1997), and Lim (2008). However, these scales included questions not applicable to our context, such as those referring to the specifics of offline work, long-term aspects, or relationships with peers and supervisors. Consequently, we aimed to identify dimensions relevant to job satisfaction that were applicable to a one-off, online task. More specifically, at the end of our experiment, we asked participants to rate the level of interest, challenge, fun, boredom, and inspiration associated with the task using a slider analogous to those used in measuring work motivation. Additionally, respondents were prompted to rate their satisfaction with the payment scheme. These responses helped us to understand how participants perceived the task and their overall satisfaction with the payment following completion of the task.

4. Results

5,542 Prolific participants took part in the study between February 24 and September 30, 2022. In total, there were 12 sessions and participants were able to work at all times during the day or night. The recruiting announcement informed participants about the nature of the task—that they were expected to read a newspaper article thoroughly and answer the questions that followed. They were informed about the baseline payment and the existence of additional payments depending on their performance. The mean time to finish the tasks was 18 minutes and the median time was below 13 minutes. They earned, on average, £1.80. Thus, the payment was at the lower end of Prolific's recommended rate. At the same time, both pilot evidence and market practice for similar AI-training tasks suggest that such pay levels are not unusual in this setting.

The participants were, on average, 37.6 years old and 62% of them were women. By design, using the Prolific selection tool, we ensured all of them were native English speakers currently residing in the UK, holding at least a bachelor's degree, and that their acceptance rate in past studies was at least 99%. We considered these selection criteria important due to the complexity of the task. Before proceeding to analyze the results of our study, we tested whether the individual characteristics were well-balanced between treatments. Due to randomization

and a very large sample, we were not expecting any imbalances. In Appendix E, we present balancing tables. Out of 95 comparisons, there are three that are significant at $p<0.05$ and four that are significant at $p<0.1$, which is well in the expected range. However, due to those small imbalances, we used ordinary least squares (OLS) regression including a set of demographic control variables when studying the regressions outcomes. The final analysis sample was 5,507 because we lacked some information for a few participants.[11] Table 2 and 3 show some descriptive statistics for our outcome variables by treatment.

Table 2: Descriptive Statistics for the Loss and Gain Conditions

	Loss N=2774		**Gain** N=2768		
Variable	M	SE	M	SE	p
Panel A: Worker motivation					
I. I put effort into this task because I think that it was worth it.	78.63	0.40	77.72	0.41	0.116
II. I put effort into this task only because others will reward me financially.	51.72	0.58	52.15	0.58	0.597
III. I put effort into this task because otherwise I would have felt bad about myself	56.86	0.65	55.87	0.65	0.284
IV. Putting effort into this task aligns with my personal values.	71.39	0.53	69.98	0.54	0.063
V. Number of words entered in the task	18.41	0.25	18.80	0.28	0.295
Panel B: Performance					
I. Dummy attention checks 2 and 3 correct	0.99	0.00	0.99	0.00	0.352
II. Dummy all 3 attention checks correct	0.99	0.00	0.99	0.00	0.066
III. Share of questions answered in line with the majority	0.89	0.00	0.89	0.00	0.724
IV. Internal consistency regarding emotions	-0.88	0.00	-0.88	0.00	0.707
V. Internal consistency regarding job satisfaction	-26.95	0.42	-26.56	0.41	0.507
Panel C: Job satisfaction					
I. Did you find the task interesting?	67.27	0.50	67.88	0.49	0.386
II. Did you find the task challenging?	43.57	0.51	43.22	0.51	0.628
III. Did you find the task fun?	50.88	0.54	51.81	0.54	0.224
IV. Did you find the task boring?	24.07	0.48	23.53	0.48	0.429
V. Did you find the task inspiring?	27.62	0.45	28.10	0.46	0.456
VI. Are you satisfied with the payment scheme for the task?	75.44	0.43	74.10	0.45	0.032

Note. M = mean; SE = standard error.

[11] We used some information on individual characteristics provided by Prolific. However, participants sometimes revoked consent to share this information and sometimes the information was missing for other reasons.

Table 3: Descriptive Statistics for the Nonmonetary-Dimension Treatments

Variable, see definitions in Table 2 above.	**Control** N=1848		**Recognition** N=1848		**Appreciation** N=1846		**Pairwise comparisons**		
	M	SE	M	SE	M	SE	p (C vs. R)	p (C vs. A)	p (R vs. A)
Panel A: Worker motivation									
I.	78.51	0.49	78.49	0.50	77.51	0.51	0.973	0.157	0.171
II.	51.55	0.70	52.55	0.70	51.71	0.71	0.316	0.870	0.403
III.	56.09	0.80	56.04	0.79	56.97	0.80	0.965	0.437	0.408
IV.	70.60	0.66	70.62	0.65	70.83	0.66	0.986	0.803	0.816
V.	18.42	0.32	18.79	0.33	18.60	0.31	0.425	0.693	0.677
Panel B: Performance									
I.	0.99	0.00	0.99	0.00	0.99	0.00	0.505	0.103	0.339
II.	0.99	0.00	0.99	0.00	0.99	0.00	0.701	0.556	0.332
III.	0.89	0.00	0.89	0.00	0.89	0.00	0.945	0.885	0.829
IV.	-0.88	0.00	-0.88	0.01	-0.88	0.01	0.904	0.957	0.948
V.	-27.59	0.52	-25.76	0.49	-26.91	0.51	0.010	0.351	0.104
Panel C: Job satisfaction									
I.	66.75	0.62	68.91	0.59	67.06	0.62	0.011	0.718	0.031
II.	45.17	0.64	42.86	0.61	42.16	0.62	0.009	0.001	0.424
III.	49.67	0.67	53.97	0.64	50.38	0.67	0.000	0.455	0.000
IV.	24.42	0.59	22.27	0.57	24.72	0.61	0.008	0.729	0.003
V.	27.22	0.55	27.56	0.55	28.79	0.58	0.661	0.048	0.124
VI.	74.96	0.54	75.26	0.53	74.08	0.55	0.697	0.254	0.124

Note. C = control, R = recognition, A = appreciation, M = mean, SE = standard error.

In what follows, we first describe the results with respect to the three different types of outcomes: worker motivation, performance, and job satisfaction, and then relate them to our hypotheses.

4.1. Worker motivation

The average response to the questions regarding the reasons for putting in effort was 78 out of 100 for 'I think that it was worth it' and 70 for 'aligns with my personal values.' The average ratings for the reasons 'because others will reward me financially' and 'because otherwise I would have felt bad about myself' were both around 52-57.

We show treatment effects in Table 4, Panel A. The table presents the results of OLS regressions with treatment dummies (no interactions), article fixed effects, and a set of baseline characteristics. The tables report the coefficients on the treatments and robust standard errors in brackets. Randomization-*t* p-values are presented in curled brackets (see Young 2019). We also report p-values for multiple hypothesis testing (MHT) using the Romano-Wolf method (Clarke 2021) in square brackets. They account for all regressions in Table 4 and all treatments in those regressions used to test our hypotheses 1–3, that is, 44 estimands. Note, however, that

they assume two-sided tests, even if our hypotheses are directional. We also compute the Westfall-Young randomization p-value for overall treatment significance across all regressions jointly (Young 2019). With a value of 0.021, we reject the null of no overall treatment effects across the full set of specifications.

Regarding subjective worker motivation regressions (Columns I–IV), the effects of the nonmonetary incentives were not significant. Regarding the objective measure—the number of words in the word-selection task (Column V)—we also did not find any significant results. Overall, we do not see any effect of either recognition or work appreciation on motivation.

Regarding the payment framing, we only found some (weak) negative effects of the gain framing on motivation in Columns I and IV. All other treatment differences were not significant.

4.2. Performance

The task-relevant performance was assessed based on the consistency of responses within groups of six participants (in a few cases, there were more). Participants were presented with 12 yes/no questions related to the content of the article. If they answered "yes," they also had to identify the relevant part of the article (see online Appendix for the list of questions). Approximately 5% of groups demonstrated perfect or near-perfect consistency, where all members answered all questions identically. Additionally, around 65% of groups achieved a consistency score of at least 90%, indicating high agreement within these groups, see Appendix D.

In the first set of regressions in Table 4, Panel B, we regressed the objective (first two columns) and less objective (last three columns) measures of performance on our treatment dummies. Again, the table reports the results of OLS regressions including treatment dummies, article fixed effects, and baseline individual characteristics. Note that the first two columns refer to attention checks—this performance is not relevant to the employer. Here, we found some weakly negative effects of the gain framing (Column II, all attention checks), suggesting that, in treatment L, individuals were more attentive to attention checks (or rather to the first attention check). Regarding the quality of responses to the questions on article characteristics (Column III), we found no differences between the treatments. This may partly be due to the fact that the answers within teams appeared to be quite consistent (see above). Finally, the last two columns pertain to the internal consistency of individual responses, specifically whether respondents gave comparable answers to semantically similar but reverse-worded items within the questionnaire. We found a positive and significant effect only for the recognition treatment in the last

column. Since this column relates to consistency in responses to the job satisfaction questions at the end of the experiment, the effect is likely driven by higher job satisfaction in this treatment (see below).

4.3. Job satisfaction

The average score for job interest was recorded at 68, indicating moderate engagement among participants. Conversely, the assessment of the job's challenging nature returned an average score of 43. Participants rated the job as moderately enjoyable, with a fun score of 51, while the job was rated as minimally boring, receiving a score of 23. However, it did not inspire significantly, as reflected by a score of 28. Notably, the satisfaction with payment was relatively high, achieving a commendable score of 75 out of 100.

In the last set of regressions (Table 4, Panel C), we tested how different treatments influenced job satisfaction. Again, the table presents the results of OLS regressions on treatment dummies, including article fixed effects and baseline individual characteristics. We found positive and highly significant effects of the recognition treatment on finding the task interesting and fun. In this treatment, participants also perceived the task as less boring and less challenging, with both effects being highly significant. Participants in the work-appreciation treatment found the task less challenging than those in the baseline condition while also considering it more inspiring. Importantly, the recognition and appreciation treatments used the same screens, timing, and placement; the difference lay only in wording. The observed increase in self-reported satisfaction in the recognition condition can therefore be attributed to treatment differences. Finally, we observed a negative effect of the gain framing on satisfaction with the payment scheme, although this result does not survive the most conservative multiple-testing adjustments.

4.4. Hypotheses tests

Regarding our main hypothesis H1, we find evidence that the recognition treatment improved some dimensions of job satisfaction, although only one coefficient remains significant after multiple hypothesis testing using the Romano-Wolf method. The work-appreciation treatment did not outperform the baseline on motivation or performance. The absence of detectable effects on performance may partly reflect the high consistency of responses among participants, suggesting either that the task was relatively easy for this screened sample or that simple automated praise is more effective at shifting reported task experience than measurable output quality.

We do not find evidence for H2 that work appreciation outperforms recognition in improving performance. This differs from the expectation that appreciation would be more closely tied to motivation and therefore more likely to affect output. One possible explanation is that in a short, one-off crowdwork setting, both forms of positive feedback are too weak to alter effort in a meaningful way. More broadly, this result suggests that the distinction between appreciation and recognition may matter less in highly standardized online work than in longer-term employment relationships.

Overall, we find some support for the first part of H3: the probability of mistakes is lower in the loss treatment L than in the gain treatment G for objectively verifiable tasks (Table 4, Panel B, Column II). However, for the less objective tasks, the coefficients are small and statistically insignificant (Table 4, Panel B, Columns III-V). Taken together with the other results, this suggests that gain framing does not dominate loss framing in this setting and that any advantage of loss framing is concentrated, at most, in simple and easily monitored subtasks.

Regarding H4, we present the results of the regressions in which we interact the treatment dummies with high experience of working on Prolific (median split regarding the number of jobs individuals have already performed on Prolific) in Appendix C, Table C1. While single interaction effects for job satisfaction were significant at $p<0.1$, we did not see any particular pattern in those results. The absence of robust heterogeneity by experience suggests that more experienced crowdworkers do not respond fundamentally differently to these interventions, thereby also closing off a path toward personalizing incentives by experience.

Regarding H5, Table C2 in Appendix C shows the results of interactions between monetary and nonmonetary treatments. As expected, we do not find interaction effects in the data, except for payment satisfaction. Given that treatments in general are mostly ineffective, this result is not surprising and suggests that combining these low-cost interventions does not generate additional gains beyond their separate effects. In this sense, the findings indicate that the treatments do not appear to amplify one another in any meaningful way in our setting. For the literature, this implies that simple recognition and payment framing may have limited scope for complementarity in short-term crowdwork tasks.

Table 4: Treatment effects

Panel A: Worker motivation					
	I put effort into this task because I think that it was worth it.	I put effort into this task only because others will reward me financially.	I put effort into this task because otherwise I would have felt bad about myself	Putting effort into this task aligns with my personal values.	Number of words entered in the task
	I	II	III	IV	V
Recognition	0.016	1.121	0.047	0.014	0.365
	(0.688)	(0.998)	(1.113)	(0.908)	(0.414)
	{0.983}	{0.283}	{0.962}	{0. 991}	{0.364}
	[1.000]	[0.999]	[1.000]	[1.000]	[1.000]
Work appreciation	-1.027	0.251	0.617	-0.017	0.123
	(0.708)	(0.997)	(1.122)	(0.924)	(0.410)
	{0.144}	{0.802}	{0.582}	{0. 987}	{0. 766}
	[0.986]	[1.000]	[1.000]	[1.000]	[1.000]
Gain Framing	-1.048*	0.558	-0.951	-1.639**	0.388
	(0.572)	(0.815)	(0.910)	(0.749)	(0.338)
	{0.082}	{0.477}	{0. 317}	{0.032}	{0. 221}
	[0.866]	[1.000]	[1.000]	[0.597]	[0.999]
Constant	79.182***	83.957***	58.310***	41.823**	10.109**
	(5.947)	(11.627)	(10.906)	(17.058)	(4.096)
Controls	yes	yes	yes	yes	yes
Article fixed effects	yes	yes	yes	yes	yes
Observations	5507	5507	5507	5507	5507
R^2	0.193	0.178	0.191	0.204	0.312
Panel B: Performance					
	Dummy attention checks 2 and 3 correct	Dummy all 3 attention checks correct	Share of questions answered in line with the majority	Internal consistency regarding emotions	Internal consistency regarding job satisfaction
	I	II	III	IV	V
Recognition	-0.002		0.000		2.016***
	(0.004)		(0.002)		(0.713)
	{0.527}		{0.785}		{0.0047}
	[1.000]		[1.000]		[0.166]
Work appreciation	-0.007		0.001		0.969
	(0.004)		(0.003)		(0.727)
	{0.118}		{0.937}		{0.182}
	[0.950]		[1.000]		[0.995]
Gain Framing	-0.002	-0.011**	-0.001	-0.002	0.432
	(0.003)	(0.005)	(0.002)	(0.005)	(0.585)
	{0.494}	{0.032}	{0.600}	{0.706}	{0.465}
	[1.000]	[0.597]	[1.000]	[1.000]	[1.000]
Constant	0.988***	0.959***	0.856***	-0.868***	-18.978*
	(0.020)	(0.030)	(0.032)	(0.073)	(9.798)
Controls	yes	yes	yes	yes	yes
Article fixed effects	yes	yes	yes	yes	yes
Observations	5507	5507	5507	5507	5507
R^2	0.168	0.167	0.513	0.429	0.183
Panel C: Job satisfaction					
	Did you find the task interesting?	Did you find the task challenging?	Did you find the task fun?	Did you find the task boring?	Did you find the task inspiring?	Are you satisfied with the payment scheme

						for the task?
	I	II	III	IV	V	VI
Recognition	2.174***	-2.357***	4.366***	-2.046***	0.368	0.438
	(0.829)	(0.836)	(0.915)	(0.793)	(0.774)	(0.750)
	{0.006}	{0.006}	{0.001}	{0.009}	{0.626}	{0.530}
	[0.259]	[0.171]	[0.001]	[0.279]	[1.000]	[1.000]
Work appreciation	0.250	-3.211***	0.802	0.453	1.541**	-0.638
	(0.864)	(0.845)	(0.942)	(0.830)	(0.785)	(0.779)
	{0.764}	{0.000}	{0.399}	{0.580}	{0.0496}	{0.410}
	[1.000]	[0.003]	[1.000]	[1.000]	[0.782]	[1.000]
Gain Framing	0.463	-0.403	0.811	-0.352	0.538	-1.548**
	(0.688)	(0.687)	(0.754)	(0.660)	(0.638)	(0.627)
	{0.497}	{0.535}	{0.278}	{0.595}	{0.397}	{0.020}
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[0.353]
Constant	66.417***	42.052***	59.639***	26.269***	38.225***	65.556***
	(8.816)	(14.304)	(10.560)	(7.837)	(10.943)	(13.570)
Controls	yes	yes	yes	yes	yes	yes
Article fixed effects	yes	yes	yes	yes	yes	yes
Observations	5507	5507	5507	5507	5507	5507
R^2	0.212	0.251	0.195	0.216	0.196	0.177

Note: Robust standard errors in parentheses; square brackets contain Romano-Wolf adjusted p-values for multiple testing (accounting for all regressions and all treatments specified in the regressions, i.e. 44 estimands); controls include article fixed effects, age, household income dummies, household size, female dummy, socioeconomic status dummies, student dummy, education level dummies, employment status dummy, nationality dummy, country of birth dummy; in Panel B coefficients for treatments on the nonmonetary dimension are estimated only for outcomes that were measured after at least one recognition or work appreciation phrase was visible; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5. Conclusions

Unlike employees in traditional firms, crowdworkers in online labor markets typically work independently without face-to-face interactions with employers or co-workers. Communication between crowdworkers and employers is predominantly one-sided, involving written instructions provided before the task commences. Furthermore, these relationships are typically short-term, as crowdworkers complete tasks for multiple employers within a brief timeframe. Additionally, the recruitment process is not extensive enough to allow for the selection of workers based on their presumed intrinsic motivation. This setting presents a unique environment for studying how incentives operate when work is routinized in procedure but still includes complex cognitive tasks, leaving room for interpretation and judgment.

This paper presents evidence from a preregistered large-scale field experiment on Prolific that examines whether automated recognition, appreciation, and payment framing affect crowdworkers' motivation, performance, and job satisfaction. We find that recognition—messages such as *Great work!*, *You did a good job!*, *Nice job!*, or *Well done!*—improves self-reported job satisfaction, but the evidence does not indicate comparable improvements in

motivation or output quality. Put differently, automated praise appears to change how workers experience the task more than how they execute it. Although numerous studies (e.g., Chadi et al., 2017; Chandler and Kapelner, 2013; Kosfeld et al., 2017) suggest that framing work as meaningful can enhance performance relative to neutral or meaningless framing, our appreciation treatment did not generate robust positive effects. One possible explanation is that brief automated phrases are simply too weak to alter behavior in a one-off crowdwork interaction; more concrete or personalized communication might matter more.

In line with some studies (e.g., Bulte et al. 2020; Fryer et al. 2022; Levitt et al. 2016), our findings confirm that framing rewards as losses rather than gains is somewhat more effective in influencing objective performance, but the effect is not very robust in our case. However, we also observe negative effects of the gain frame for motivation and job satisfaction.

We believe our study meets the four transparency conditions required for external validity as outlined by List (2020): We sampled a large proportion of the relevant population, attrition was not an issue, and the setting was natural. Scalability, of course, depends on how far we wish to extrapolate to other platforms, tasks, and crowdworker characteristics. Regarding other platforms and tasks, there is no reason to believe our setting is particularly unique, and we expect our findings to have external validity across similar platforms and one-off tasks. Regarding repeated interactions or interactions involving teamwork, we would expect the increased job satisfaction that we identified in our context to have a multiplier effect on performance. For example, Jalagat (2016) has suggested that increased job satisfaction may boost performance in contexts where employees work together—however, this aspect does not typically apply to crowdworkers, and repeated interactions are also less common.

We are fully aware that our results are likely influenced by the form of appreciation we used—an automated appreciation tool (or "bot manager"). While this may not be the same as receiving recognition from a real person, it appears to be the simplest and least expensive solution for this kind of task. In this sense, our intervention could be seen as a lower bound on what might be achieved with more personalized or involved recognition techniques. Our findings suggest that even a minimal intervention can positively impact job satisfaction, offering HR managers a low-cost solution to improve workers' mental well-being. Tasks that lack a clear or objectively correct solution might also pose a challenge for external validity. However, Allen et al. (2021) have shown that the "wisdom of crowds" principle is viable among crowdworkers.

When it comes to crowdworker characteristics, external validity is influenced by the difficulty in defining a typical crowdworker, as motivations and expectations vary widely. However, when it comes to complex or creative tasks, potential employers are likely to select

workers using an approach similar to ours. Whether our findings would hold with participants of different nationalities and educational levels remains an open question.

While much remains to be explored regarding motivation, satisfaction, and performance among crowdworkers, our study suggests some practical implications and recommendations. First, providing recognition to crowdworkers is crucial. Future research could investigate whether personalized recognition, such as a prerecorded thank-you message from the company's CEO or showcasing a beneficiary of the work upon task completion, would have a greater impact. Second, the concept of recognition could extend beyond crowdworkers to any scenario involving indirect interactions between parties. For instance, in online prerecorded classes—whether for teaching in online courses or motivating in online fitness sessions—similar recognition strategies could be employed to enhance engagement and satisfaction. Moreover, our study highlights the dual nature of many crowdwork tasks, which blend routine components with creative judgment. Understanding this interplay is essential, as motivational interventions may have different effects depending on whether a task leans more toward routinization or creativity.

References:

Abdullah, A. A., & Wan, H. L. (2013). Relationships of non-monetary incentives, job satisfaction and employee job performance. *International Review of Management and Business Research*, 2(4), 1085. https://www.irmbrjournal.com/download.php?file=papers/1384881024.pdf

Allen, J., Arechar, A. A., Pennycook, G., and Rand, D. G. (2021). Scaling up fact-checking using the wisdom of crowds. *Science advances*, 7(36). https://doi.org/10.1126/sciadv.abf4393.

Amabile, T.M. (1996). Creativity In Context: Update To The Social Psychology Of Creativity (1st ed.). Routledge. https://doi.org/10.4324/9780429501234

Ariely, D., Gnezy, U., Loewenstein, G., and Mazar, N. (2009). Large Stakes and Big Mistakes. *Review of Economic Studies*, *76*(2), 451–469. https://doi.org/10.1111/j.1467-937X.2009.00534.x

Armantier, O., and Boly, A. (2015). Framing of Incentives and Effort Provision. *International Economic Review*, *56*(3), 917–938. https://doi.org/10.1111/iere.12126

Attanasi, G., Chessa, M., Gil-Gallen, S. and Llerena, P. (2021). A Survey on Experimental Elicitation of Creativity in Economics. *Revue d'économie industrielle*, 174(2), 273-324. https://doi.org/10.4000/rei.10448.

Barrero, J. M., Bloom, N., and Davis, S. J. (2023). The Evolution of Work from Home. *Journal of Economic Perspectives*, 37 (4), 23–49. https://doi.org/10.1257/jep.37.4.23

Bénabou, R., and Tirole, J. (2003). Intrinsic and Extrinsic Motivation. *Review of Economic Studies*, *70*(3), 489–520. https://doi.org/10.1111/1467-937X.00253

Bénabou, R., and Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review*, *96*(5), 1652–1678. https://doi.org/10.1257/aer.96.5.1652

Bode, C., Singh, J., and Rogan, M. (2015). Corporate Social Initiatives and Employee Retention. *Organization Science*, *26*(6), 1702–1720. https://doi.org/10.1287/orsc.2015.1006

Bowling, N. A. (2013). Job Satisfaction, Motivation and Performance. In M. C. W. Peeters, J. D. Junge, and T. W. Taris (Eds.), *An introduction to contemporary work psychology* (pp. 321–341). West Sussex: John Wiley and Sons. https://doi.org/10.1002/9781394259564.ch13

Bradler, C., Dur, R., Neckermann, S., and Non, A. (2016). Employee Recognition and Performance: A Field Experiment. *Management Science*, *62*(11), 3085–3099. https://doi.org/10.1287/mnsc.2015.2291

Bradler, C., Neckermann, S., & Warnke, A. J. (2019). Incentivizing Creativity: A Large-Scale Experiment with Performance Bonuses and Gifts. *Journal of Labor Economics*, 37(3), 793–851. https://doi.org/10.1086/702649

Brancati, C. U., Pesole, A., and Fernández-Macías, E. (2019). *Digital Labour Platforms in Europe: Numbers, Profiles, and Employment Status of Platform Workers* EUR 29810 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-08955-1, https://doi.org/10.2760/16653

Bulte, E., List, J. A., and van Soest, D. (2020). Toward an Understanding of the Welfare Effects of Nudges: Evidence from a Field Experiment in the Workplace. *The Economic Journal*, *130*(632), 2329–2353. https://doi.org/10.1093/ej/ueaa054

Burbano, V. C. (2016). Social Responsibility Messages and Worker Wage Requirements: Field Experimental Evidence from Online Labor Marketplaces. *Organization Science*, *27*(4), 1010–1028. https://doi.org/10.1287/orsc.2016.1066

Carpenter, J., and Gong, E. (2016). Motivating Agents: How Much Does the Mission Matter? *Journal of Labor Economics*, *34*(1), 211–236. https://doi.org/10.1086/682345

Cassar, L., and Meier, S. (2018). Nonmonetary Incentives and the Implications of Work as a Source of Meaning. *Journal of Economic Perspectives*, *32*(3), 215–238. https://doi.org/10.1257/jep.32.3.215

Cerasoli, C. P., Nicklin, J. M., and Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, *140*(4), 980–1008. https://doi.org/10.1037/a0035661

Chadi, A., Jeworrek, S., and Mertins, V. (2017). When the Meaning of Work Has Disappeared: Experimental Evidence on Employees' Performance and Emotions. *Management Science*, *63*(6), 1696–1707. https://doi.org/10.1287/mnsc.2016.2426

Chandler, D., and Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior and Organization*, *90*, 123–133. https://doi.org/10.1016/j.jebo.2013.03.003

Charness, G., Cooper, M., & Reddinger, J. L. (2020). Wage policies, incentive schemes, and motivation. *Handbook of labor, human resources and population economics*, 1-33. https://doi.org/10.1007/978-3-319-57365-6_125-1

Charness, G., Cooper, M., & Reddinger, J. L. (2023). Non-financial motivation in the workplace. *Elgar Encyclopedia of Labour Studies*, 134–140. https://doi.org/10.4337/9781800377547.ch32

Charness, G., & Grieco, D. (2019). Creativity and Incentives. *Journal of the European Economic Association*, 17(2), 454–496. https://doi.org/10.1093/jeea/jvx055

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*, vol 9: 88-97. https://doi.org/10.1016/j.jbef.2015.12.001

Clarke, D., 2021. rwolf2 Implementation and Flexible Syntax. http://www.damianclarke.net/computation/rwolf2.pdf

Czibor, E., Hsu, D., Jimenez-Gomez, D., Neckermann, S., and Subasi, B. (2022). Loss-Framed Incentives and Employee (Mis-)Behavior. *Management Science*, *68*(10), 7518–7537. https://doi.org/10.1287/mnsc.2021.4280

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–668. https://doi.org/10.1037/0033-2909.125.6.627

DellaVigna, S., and Pope, D. (2018). What Motivates Effort? Evidence and Expert Forecasts. *The Review of Economic Studies*, *85*(2), 1029–1069. https://doi.org/10.1093/restud/rdx033

de Quidt, J., Fallucchi, F., Kölle, F., Nosenzo, D., and Quercia, S. (2017). Bonus versus penalty: How robust are the effects of contract framing? *Journal of the Economic Science Association*, *3*(2), 174–182. https://doi.org/10.1007/s40881-017-0039-9

Ellingsen, T., and Johannesson, M. (2008). Pride and Prejudice: The Human Side of Incentive Theory. *American Economic Review*, *98*(3), 990–1008. https://doi.org/10.1257/aer.98.3.990

Fehr, E., and Goette, L. (2007). Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment. *American Economic Review*, *97*(1), 298–317. https://doi.org/10.1257/aer.97.1.298

Fryer Jr, R. G., Levitt, S. D., List, J., and Sadoff, S. (2022). Enhancing the efficacy of teacher incentives through framing: A field experiment. American Economic Journal: Economic Policy, 14(4), 269-299. https://doi.org/10.1257/pol.20190287

Gagné, M., Forest, J., Vansteenkiste, M., Crevier-Braud, L., Van den Broeck, A., Aspeli, A. K., Bellerose, J., Benabou, C., Chemolli, E., Güntert, S. T., Halvari, H., Indiyastuti, D. L., Johnson, P. A., Molstad, M. H., Naudin, M., Ndao, A., Olafsen, A. H., Roussel, P., Wang, Z., & Westbye, C. (2015). *Multidimensional Work Motivation Scale (MWMS)* [Database record]. APA PsycTests. https://doi.org/10.1037/t45942-000

Gneezy, U., and Rustichini, A. (2000). A Fine is a Price. *The Journal of Legal Studies*, *29*(1), 1–17. https://doi.org/10.1086/468061

Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, 25(4), 191–210. https://doi.org/10.1257/jep.25.4.191

Gonzalez-Cabello, M., Siddiq, A., Corbett, C. J., & Hu, C. (2024). Fairness in crowdwork: Making the human AI supply chain more humane. Business Horizons. https://doi.org/10.1016/j.bushor.2024.09.003

Gross, D. P. (2020). Creativity Under Fire: The Effects of Competition on Creative Production. *The Review of Economics and Statistics*, 102(3), 583–599. https://doi.org/10.1162/rest_a_00831

Hedblom, D., Hickman, B., and List, J. (2019). Toward an Understanding of Corporate Social Responsibility: Theory and Field Experimental Evidence National Bureau of Economic Research Working Paper No. w26222. https://doi.org/10.3386/w26222

Herzberg, F., Mausner, B., and Snyderman, B. (1959). *The Motivation to Work*. John Wiley and Sons Inc.

Hossain, M., and Kauranen, I. (2015). Crowdsourcing: a comprehensive literature review. *Strategic Outsourcing: An International Journal*, *8*(1), 2–22. https://doi.org/10.1108/SO-12-2014-0029

Howe, J. (2006). The Rise of Crowdsourcing. *Wired Magazine*, *14*(6), 176–183.

Jalagat, R. (2016). Job Performance, Job Satisfaction, and Motivation: A Critical Review of their Relationship. *International Journal of Advances in Management and Economics*, *5*(6), 36–42. https://www.managementjournal.info/index.php/IJAME/article/view/64.

Katz, L., and Krueger, A. (2018). The Rise and Nature of Alternative Work Arrangements in the United States, 1995-2015. *ILR review*, 72(2), 382-416. https://doi.org/10.1177/0019793918820008

Kosfeld, M., and Neckermann, S. (2011). Getting More Work for Nothing? Symbolic Awards and Worker Performance. *American Economic Journal: Microeconomics*, *3*(3), 86–99. https://doi.org/10.1257/mic.3.3.86

Kosfeld, M., Neckermann, S., and Yang, X. (2017). The Effects of Financial and Recognition Incentives across Work Contexts: The Role of Meaning. *Economic Inquiry*, *55*(1), 237–247. https://doi.org/10.1111/ecin.12350

Kvaløy, O., Nieken, P., and Schöttner, A. (2015). Hidden benefits of reward: A field experiment on motivation and monetary incentives. *European Economic Review*, *76*, 188–199. https://doi.org/10.1016/j.euroecorev.2015.03.003

Levitt, S. D., List, J. A., Neckermann, S., and Sadoff, S. (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. American Economic Journal: Economic Policy, 8(4), 183-219. https://doi.org/10.1257/pol.20130358

Lim, S. (2008). Job satisfaction of information technology workers in academic libraries. Library and Information Science Research, 30(2), 115-121. http://dx.doi.org/10.1016/j.lisr.2007.10.002

List, J. (2020). Non est Disputandum de Generalizability? A Glimpse into The External Validity Trial. Cambridge, MA. https://doi.org/10.3386/w27535

Locke, E. A. (1970). The supervisor as 'motivator': his influence on employee performance and satisfaction. In B. Bass, R. Cooper, and J. A. Haas (Eds.), *Managing for Accomplishment* (pp. 57–67).

Macdonald, S., and Maclntyre, P. (1997). The generic job satisfaction scale: Scale development and its correlates. Employee Assistance Quarterly, 13(2), 1-16. https://doi.org/10.1300/J022v13n02_01

Ockenfels, A., Sliwka, D., and Werner, P. (2015). Bonus Payments and Reference Point Violations. *Management Science*, *61*(7), 1496–1513. https://doi.org/10.1287/mnsc.2014.1949

Scherer, K. R. (2005). What are emotions? And how can they be measured? Social Science Information, 44(4), 693-727. https://doi.org/10.1177/0539018405058216

Tamam, M. B., and Sopiah. (2022). Effect of Work Motivation and Employee Performance. *International Journal of Law Policy and Governance*, *1*(2), 103–110. https://doi.org/10.54099/ijlpg.v1i2.437

Thompson, E. R., and Phua, F. T. (2012). A brief index of affective job satisfaction. Group and Organization Management, 37(3), 275-307. https://doi.org/10.1177/1059601111434201

Wu, H., Li, S., Zheng, J., and Guo, J. (2020). Medical students' motivation and academic performance: the mediating roles of self-efficacy and learning engagement. *Medical Education Online*, *25*(1), 1–9. https://doi.org/10.1080/10872981.2020.1742964

Young, Alwyn (2019). Channelling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results. *Quarterly Journal of Economics*, 134 (2): 557-598. doi.org/10.1093/qje/qjy029

Appendix for "Motivating Crowdworkers with Nonmonetary Incentives and Payment Framing—Evidence from a Large-Scale Experiment"

Appendix A: Pre-screening criteria and the available pool of participants

Table A1: Pre-screening criteria and the available pool of participants on Prolific

Variable	description	criterium	Available sample (participants active in the past 90 days as of 19.11.2020)
Geographic variables/ Current country of residence		UK	49,819
Languages/first language		English	39,132
Education/Highest education completed		Undergraduate degree (BA/BSc/other); Graduate degree (MA/MSc/MPhil/other); Doctorate degree (PhD/other)	18,965
Basic demographic variables/Age		18-100= all categories	18,915
Basic demographic variables/sex		male or female	18,834
Socioeconomic variables/ Socioeconomic Status	Participants were asked the following question: **Think of a ladder (see image) as representing where people stand in society. At the top of the ladder are the people who are best off - those who have the most money, most education and the best jobs. At the bottom are the people who are worst off - who have the least money, least education and the worst jobs or no job. The higher up you are on this ladder, the closer you are to people at the very top and the lower you are, the closer you are to the bottom. Where would you put yourself on the ladder? Choose the number whose position best represents where you would be on this ladder.**	1-10, all categories	18,547

Socioeconomic variables/ Household Size		all categories	18,467
Socioeconomic variables/ Household Income (GBP)	**What is your total household income per year, including all earners in your household (after tax) in GBP?**	all categories	18,253
Geographic variables/ Country of Birth		all categories	**18,223**

In addition, we restricted the sample to individuals with approval rates between 99 and 100. This did not significantly reduce the available subject pool.

Appendix B: Treatment variation

First dimension:
C Control

[Blank pages] x 4

R recognition

1. *Great work!*
2. *You did a good job!*
3. *Nice job!*
4. *Well done!*

A Appreciation

1. *Thank you!*
2. *Your help makes a difference!*
3. *We appreciate your support!*
4. *Your work matters!*

Second dimension:
Treatment G (gain domain)

Payment

If you finish the task you will be paid at least ***£0.90****.*[1] *Your final payment depends on you meeting the quality requirements. The quality requirements are based on three simple questions that have a clear and objectively correct answer. Those questions are positioned between other questions and tasks. Each time you correctly answer the quality check question, your payment will be raised by* ***£0.25****. The maximum additional payment equals thus to* ***£0.75*** *in which case you will be paid* ***£1.65****.*

After the first {second}[third] failed attention check:
You did not answer the quality check question correctly. Your final payment of £0.90 {1.15}[1.40] will not be raised by £0.25.
Treatment L (loss domain)

Payment

If you finish the task you will be paid up to ***£1.65****.*[2] *Your final payment depends on you meeting the quality requirements. The quality requirements are based on three simple questions that have a clear and objectively correct answer. These questions are positioned between other questions and tasks. Each time you fail to answer the quality check question correctly, your payment will be reduced by* ***£0.25****. The maximum reduction equals thus to* ***£0.75****, in which case you will be only paid* ***£0.90****.*

After the first [second] {third} failed attention check:

You did not answer the quality check question correctly. Your final payment of £1.65 [1.40] {1.15} will be reduced by £0.25.

[1] The exact baseline payment depends on the length of the article. This is an example for the shortest ones.
[2] This is an example payment, see above.

Appendix C: Additional hypotheses specified in the preregistration (not stated in the main paper)

Table C1: Regressions for H4: [Experience] More experienced participants react less to recognition/work meaning (we use a median split).

Panel A: Worker motivation

	I put effort into this task because I think that it was worth it.	I put effort into this task only because others will reward me financially.	I put effort into this task because otherwise I would have felt bad about myself	Putting effort into this task aligns with my personal values.	Amount of words entered in the Task
	I	II	III	IV	V
Recognition	0.533	1.291	-0.876	0.677	0.152
	(1.046)	(1.464)	(1.624)	(1.377)	(0.613)
Work meaning	-0.195	0.176	-0.429	-0.786	0.584
	(1.070)	(1.444)	(1.658)	(1.439)	(0.615)
Gain Framing	-1.144	1.981*	-2.183	-2.032*	0.286
	(0.863)	(1.194)	(1.333)	(1.131)	(0.489)
Recognition * Experience over median	-0.962	-0.427	1.950	-1.264	0.435
	(1.490)	(2.176)	(2.391)	(1.980)	(0.914)
Work meaning * Experience over median	-1.543	0.202	2.173	1.635	-0.938
	(1.559)	(2.152)	(2.443)	(2.025)	(0.904)
Gain Framing * Experience over median	0.189	-2.830	2.517	0.747	0.220
	(1.238)	(1.789)	(1.976)	(1.628)	(0.733)
Experience over median	4.398***	5.010***	-1.402	1.206	-0.181
	(1.227)	(1.773)	(2.016)	(1.611)	(0.740)
Constant	77.156***	81.671***	59.945***	41.567**	10.114**
	(5.939)	(11.714)	(10.912)	(17.207)	(4.100)
Observations	5507	5507	5507	5507	5507
R^2	0.199	0.181	0.192	0.205	0.313

Panel B: Performance

	Dummy attention checks 2 and 3 correct	Dummy all 3 attention checks correct	Share of questions answered in line with the majority	Internal consistency regarding emotions	Internal consistency regarding job satisfaction
	I	II	III	IV	V
Gain Framing	0.000	-0.010	-0.000	-0.004	1.141
	(0.005)	(0.007)	(0.003)	(0.007)	(0.832)
Recognition	0.002		-0.000		1.501
	(0.006)		(0.004)		(1.022)
Work meaning	-0.003		-0.002		0.520
	(0.006)		(0.004)		(1.034)
Recognition * Experience over median	-0.008		0.001		0.934
	(0.009)		(0.005)		(1.559)
Work meaning *	-0.008		0.005		0.821

Experience over median	(0.009)		(0.005)		(1.574)
Gain Framing * Experience over median	-0.005 (0.007)	-0.000 (0.011)	-0.002 (0.004)	0.011 (0.011)	-1.396 (1.283)
Experience over median	0.012* (0.007)	0.003 (0.007)	0.000 (0.005)	-0.007 (0.008)	-1.179 (1.305)
Constant	0.980*** (0.021)	0.958*** (0.030)	0.857*** (0.032)	0.873*** (0.073)	-18.440* (9.978)
Observations	5507	5507	5507	5507	5507
R^2	0.169	0.167	0.513	0.429	0.183

Panel C: Job satisfaction

	Did you find the task interesting?	Did you find the task challenging?	Did you find the task fun?	Did you find the task boring?	Did you find the task inspiring?	Are you satisfied with the payment scheme for the task?
	I	II	III	IV	V	VI
Recognition	1.353	-2.862**	2.425*	-0.370	0.012	0.897
	(1.254)	(1.229)	(1.359)	(1.202)	(1.138)	(1.128)
Work meaning	1.062	-4.747***	0.653	0.585	1.434	0.044
	(1.280)	(1.246)	(1.374)	(1.223)	(1.135)	(1.144)
Gain Framing	0.214	-1.244	1.313	-0.100	-0.061	-1.125
	(1.028)	(1.013)	(1.114)	(0.996)	(0.939)	(0.922)
Recognition * Experience over median	1.662	1.068	3.792*	-3.364*	0.760	-0.906
	(1.818)	(1.826)	(2.001)	(1.739)	(1.679)	(1.645)
Work meaning * Experience over median	-1.640	3.122*	0.202	-0.270	0.233	-1.285
	(1.893)	(1.832)	(2.051)	(1.799)	(1.693)	(1.695)
Gain Framing * Experience over median	0.560	1.699	-0.885	-0.618	1.218	-0.849
	(1.498)	(1.492)	(1.653)	(1.454)	(1.397)	(1.352)
Experience over median	-0.057	-2.013	-1.687	0.248	-0.857	4.174***
	(1.548)	(1.488)	(1.688)	(1.470)	(1.377)	(1.375)
Constant	66.521***	43.802***	60.839***	25.431***	38.899***	63.513***
	(8.738)	(14.537)	(10.801)	(7.799)	(10.960)	(13.546)
Observations	5507	5507	5507	5507	5507	5507
R^2	0.212	0.251	0.196	0.217	0.196	0.180

Note: Controls include article fixed effects, age, household income dummies, household size, female dummy, socioeconomic status dummies, student dummy, education level dummies, employment status dummy, nationality dummy, country of birth dummy; In Panel B coefficients for treatments on the nonmonetary dimension are estimated only for outcomes that were measured after at least one recognition or work appreciation phrase was visible; robust standard errors in paratheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C2: Regression results for H5: [Interaction effect] There is no interaction effect.

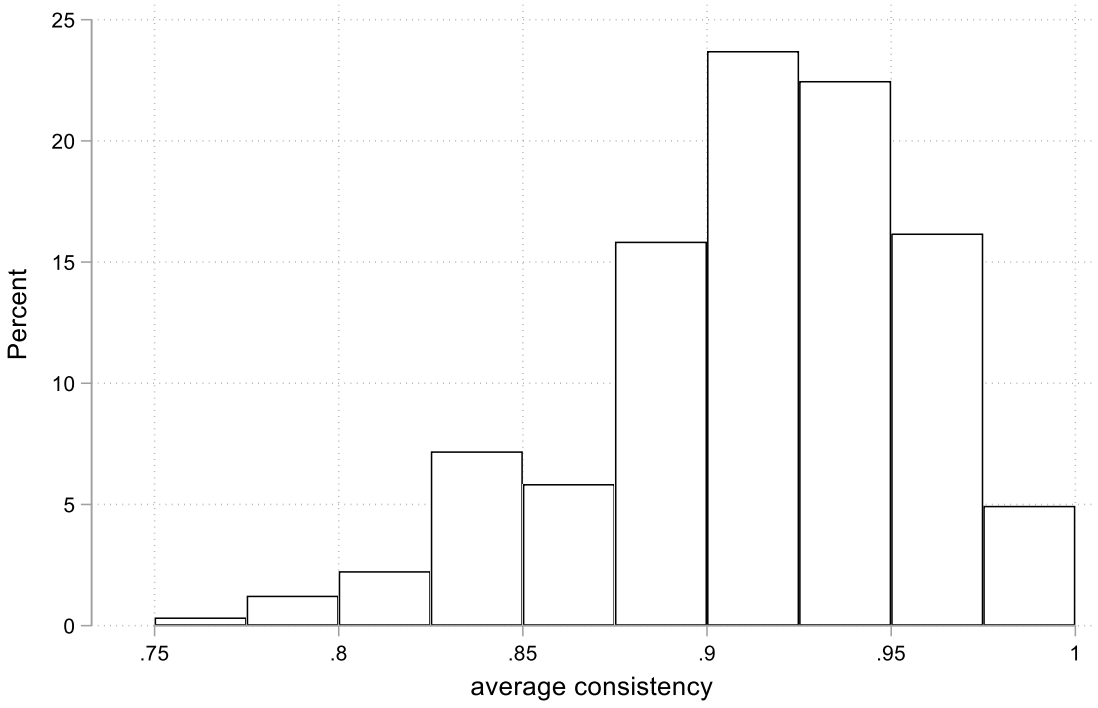
Panel A: Worker motivation					
	I put effort into this task because I think that it was worth it.	I put effort into this task only because others will reward me financially.	I put effort into this task because otherwise I would have felt bad about myself	Putting effort into this task aligns with my personal values.	Amount of words entered in the Task
	I	II	III	IV	V
Recognition	-0.566	1.631	0.542	-0.933	-0.052
	(0.955)	(1.401)	(1.587)	(1.285)	(0.549)
Work meaning	-1.544	0.835	1.516	-0.743	0.220
	(1.002)	(1.410)	(1.611)	(1.297)	(0.581)
Gain Framing	-1.782*	1.289	-0.021	-2.756**	0.173
	(0.974)	(1.413)	(1.585)	(1.296)	(0.579)
Recognition*Gain	1.169	-1.025	-0.996	1.900	0.836
	(1.376)	(1.999)	(2.226)	(1.821)	(0.837)
Work meaning*Gain	1.038	-1.171	-1.803	1.454	-0.195
	(1.412)	(2.000)	(2.254)	(1.853)	(0.816)
Constant	79.630***	83.504***	57.717***	42.499**	10.215**
	(5.980)	(11.566)	(10.853)	(17.078)	(4.065)
Observations	5507	5507	5507	5507	5507
R^2	0.193	0.178	0.191	0.204	0.313
Panel B: Performance					
	Dummy attention checks 2 and 3 correct	Dummy all 3 attention checks correct	Share of questions answered in line with the majority	Internal consistency regarding emotions	Internal consistency regarding job satisfaction
	I	II	III	IV	V
Gain Framing	0.003	-0.011**	-0.004	0.002	0.280
	(0.006)	(0.005)	(0.004)	(0.005)	(1.031)
Recognition	0.001		-0.001		2.249**
	(0.006)		(0.003)		(1.012)
Work meaning	-0.002		-0.003		0.506
	(0.006)		(0.003)		(1.023)
Recognition*Gain	-0.007		0.003		-0.466
	(0.008)		(0.005)		(1.430)
Work meaning*Gain	-0.010		0.006		0.930
	(0.009)		(0.005)		(1.452)
Constant	0.985***	0.959***	0.858***	0.868***	-18.849*
	(0.020)	(0.030)	(0.033)	(0.073)	(9.837)
Observations	5507	5507	5507	5507	5507
R^2	0.168	0.167	0.513	0.429	0.183
Panel C: Job satisfaction					
	Did you find the task interesting?	Did you find the task challenging ?	Did you find the task fun?	Did you find the task boring?	Did you find the task inspiring?	Are you satisfied with the payment scheme for the

						task?
	I	II	III	IV	V	VI
Recognition	2.504**	-2.729**	4.279***	-1.919*	0.920	2.353**
	(1.172)	(1.177)	(1.305)	(1.122)	(1.083)	(1.050)
Work	0.478	-3.876***	-0.202	0.745	2.556**	0.982
meaning	(1.232)	(1.199)	(1.341)	(1.182)	(1.098)	(1.084)
Gain Framing	0.837	-1.095	0.085	-0.073	1.583	0.812
	(1.211)	(1.190)	(1.329)	(1.147)	(1.090)	(1.076)
Recognition*	-0.663	0.749	0.177	-0.256	-1.111	-3.843**
Gain	(1.660)	(1.677)	(1.836)	(1.587)	(1.546)	(1.501)
Work	-0.458	1.334	2.015	-0.585	-2.037	-3.248**
meaning*	(1.733)	(1.691)	(1.889)	(1.658)	(1.566)	(1.561)
Gain						
Constant	66.192***	42.493***	60.132***	26.089***	37.559***	64.119***
	(8.823)	(14.214)	(10.592)	(7.861)	(11.053)	(13.470)
Observations	5507	5507	5507	5507	5507	5507
R^2	0.212	0.251	0.195	0.216	0.197	0.178

Note: Controls include article fixed effects, age, household income dummies, household size, female dummy, socioeconomic status dummies, student dummy, education level dummies, employment status dummy, nationality dummy, country of birth dummy; In Panel B coefficients for treatments on the nonmonetary dimension are estimated only for outcomes that were measured after at least one recognition or work appreciation phrase was visible; robust standard errors in paratheses; * p < 0.10, ** p < 0.05, *** p < 0.01

Appendix D: Additional results

Figure D1: Consistency per yes/no question. The score ranges between 0.5 (half of the group chose 'yes,' half chose 'no' as an answer) and 1 (all group members chose the same answer).



Appendix E: Balancing tables

Table E1: Individual characteristics by control treatment and treatments: recognition and appreciation

	Control treatment (1)			Recognition treatment (2)			Work-appreciation treatment (3)			1=2	1=3	2=3
	Mean	Standard error	Number of observations	Mean	Standard error	Number of observations	Mean	Standard error	Number of observations	T-test p-value	T-test p-value	T-test p-value
Participant's age	37,61	0,29	1847	37,73	0,29	1848	37,74	0,29	1842	0,76	0,76	0,99
Household income	6,00	0,07	1678	5,97	0,07	1647	6,06	0,07	1639	0,73	0,53	0,33
Household size	2,89	0,03	1840	2,79	0,03	1840	2,86	0,03	1832	0,03	0,55	0,13
Pariticpant's gender (female)	0,62	0,01	1848	0,62	0,01	1848	0,64	0,01	1846	0,97	0,24	0,26
Socioeconomic Status	5,57	0,03	1848	5,52	0,03	1848	5,56	0,03	1846	0,37	0,83	0,50
Student dummy	0,14	0,01	1848	0,13	0,01	1848	0,13	0,01	1846	0,38	0,39	0,99
Undergraduate degree (BA/BSc/other)	0,68	0,01	1848	0,67	0,01	1848	0,65	0,01	1846	0,73	0,11	0,22
Graduate degree (MA/MSc/MPhil/other)	0,27	0,01	1848	0,28	0,01	1848	0,31	0,01	1846	0,74	0,02	0,05
Doctor degree (PhD/other)	0,05	0,00	1848	0,05	0,01	1848	0,04	0,00	1846	0,76	0,23	0,13
Unemployed (and job seeking)	0,05	0,01	1848	0,04	0,00	1848	0,05	0,01	1846	0,25	0,71	0,43
Not in paid work (e.g. homemaker', 'retired or disabled)	0,10	0,01	1848	0,08	0,01	1848	0,09	0,01	1846	0,11	0,40	0,44
Due to start a new job within the next month	0,02	0,00	1848	0,02	0,00	1848	0,01	0,00	1846	0,21	0,90	0,17
Part-Time	0,17	0,01	1848	0,17	0,01	1848	0,19	0,01	1846	0,66	0,14	0,06
Full-Time	0,53	0,01	1848	0,56	0,01	1848	0,51	0,01	1846	0,15	0,28	0,01
Other	0,03	0,00	1848	0,03	0,00	1848	0,03	0,00	1846	0,85	0,71	0,85
Nationality (UK dummy)	0,95	0,00	1848	0,95	0,01	1848	0,94	0,01	1846	0,41	0,09	0,39
Country of birth (UK dummy)	0,92	0,01	1848	0,91	0,01	1848	0,91	0,01	1846	0,48	0,22	0,60

Table E2: Individual characteristics by loss and gain treatments

	Loss frame (1)			Gain frame (2)			1=2
	Mean	Standard error	Number of observations	Mean	Standard error	Number of observations	T-test p-value
Participant's age	37,65	0,24	2773	37,74	0,24	2764	0,77
Household income	6,09	0,05	2494	5,93	0,06	2470	0,04
Household size	2,86	0,03	2759	2,84	0,03	2753	0,59
Participant's sex	0,62	0,01	2774	0,63	0,01	2768	0,36
Socioeconomic status	5,56	0,03	2774	5,54	0,03	2768	0,60
Student (dummy)	0,13	0,01	2774	0,13	0,01	2768	0,55
Undergraduate degree (BA/BSc/other)	0,65	0,01	2774	0,68	0,01	2768	0,08
Graduate degree (MA/MSc/MPhil/other)	0,30	0,01	2774	0,28	0,01	2768	0,06
Doctor degree (PhD/other)	0,04	0,00	2774	0,05	0,00	2768	0,78
Unemployed (and job seeking)	0,05	0,00	2774	0,05	0,00	2768	0,58
Not in paid work (e.g. homemaker', 'retired or disabled)	0,09	0,01	2774	0,09	0,01	2768	0,31
Due to start a new job within the next month	0,02	0,00	2774	0,02	0,00	2768	0,35
Part-Time	0,18	0,01	2774	0,17	0,01	2768	0,42
Full-Time	0,53	0,01	2774	0,54	0,01	2768	0,32
Other	0,03	0,00	2774	0,03	0,00	2768	0,77
Nationality (UK dummy)	0,94	0,00	2774	0,95	0,00	2768	0,22
Country of birth (UK dummy)	0,91	0,01	2774	0,91	0,01	2768	0,76

Appendix F: Treatment effects

Table F: Alternative Specification to Table 1 (without control variables, with article fixed effects)

Panel A: Worker motivation					
	I put effort into this task because I think that it was worth it.	I put effort into this task only because others will reward me financially.	I put effort into this task because otherwise I would have felt bad about myself	Putting effort into this task aligns with my personal values.	Number of words entered in the Task
	I	II	III	IV	V
Recognition	-0.035	1.015	-0.081	0.012	0.378
	(0.695)	(1.001)	(1.117)	(0.926)	(0.419)
Work appreciation	-1.013	0.167	0.838	0.218	0.173
	(0.708)	(0.999)	(1.126)	(0.935)	(0.411)
Gain Framing	-0.912	0.418	-0.980	-1.403*	0.389
	(0.574)	(0.816)	(0.915)	(0.759)	(0.339)
Constant	78.977***	51.334***	56.602***	71.307***	18.224***
	(0.565)	(0.816)	(0.926)	(0.757)	(0.335)
Observations	5542	5542	5542	5542	5542
R^2	0.168	0.159	0.166	0.162	0.294
Panel B: Performance					
	Dummy attention checks 2 and 3 correct	Dummy all 3 attention checks correct	Share of questions answered in line with the majority	Internal consistency regarding emotions	Internal consistency regarding job satisfaction
	I	II	III	IV	V
Recognition	-0.002		-0.000		1.835**
	(0.004)		(0.002)		(0.714)
Work appreciation	-0.007		0.001		0.691
	(0.004)		(0.002)		(0.728)
Gain Framing	-0.003	-0.011**	-0.001	0.002	0.388
	(0.003)	(0.005)	(0.002)	(0.005)	(0.586)
Constant	0.988***	0.972***	0.890***	0.878***	-27.791***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.592)

Observations	5542	5542	5542	5542	5542
R^2	0.158	0.152	0.507	0.427	0.160

Panel C: Job satisfaction

	Did you find the task interesting?	Did you find the task challenging?	Did you find the task fun?	Did you find the task boring?	Did you find the task inspiring?	Are you satisfied with the payment scheme for the task?
	I	II	III	IV	V	VI
Recognition	2.146**	-2.312***	4.285***	-2.135***	0.313	0.294
	(0.841)	(0.859)	(0.918)	(0.803)	(0.777)	(0.754)
Work appreciation	0.308	-3.001***	0.697	0.313	1.558**	-0.875
	(0.868)	(0.870)	(0.942)	(0.836)	(0.784)	(0.778)
Gain Framing	0.615	-0.350	0.924	-0.531	0.478	-1.339**
	(0.695)	(0.704)	(0.756)	(0.669)	(0.639)	(0.626)
Constant	66.446***	45.342***	49.219***	24.676***	26.995***	75.630***
	(0.706)	(0.701)	(0.766)	(0.679)	(0.626)	(0.626)
Observations	5542	5542	5542	5542	5542	5542
R^2	0.180	0.194	0.177	0.183	0.176	0.159

Note: No controls, with article fixed effects, We also run regressions without article fixed effects, this does not affect the results; robust standard errors in paratheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Online Appendix for Motivating Crowdworkers with Nonmonetary Incentives and Payment Framing— Evidence from a Large-Scale Experiment

May 21, 2025

1 Screenshots from the experiment

1.1 Treatment version: Recognition x Loss

Figure 1: Introduction page

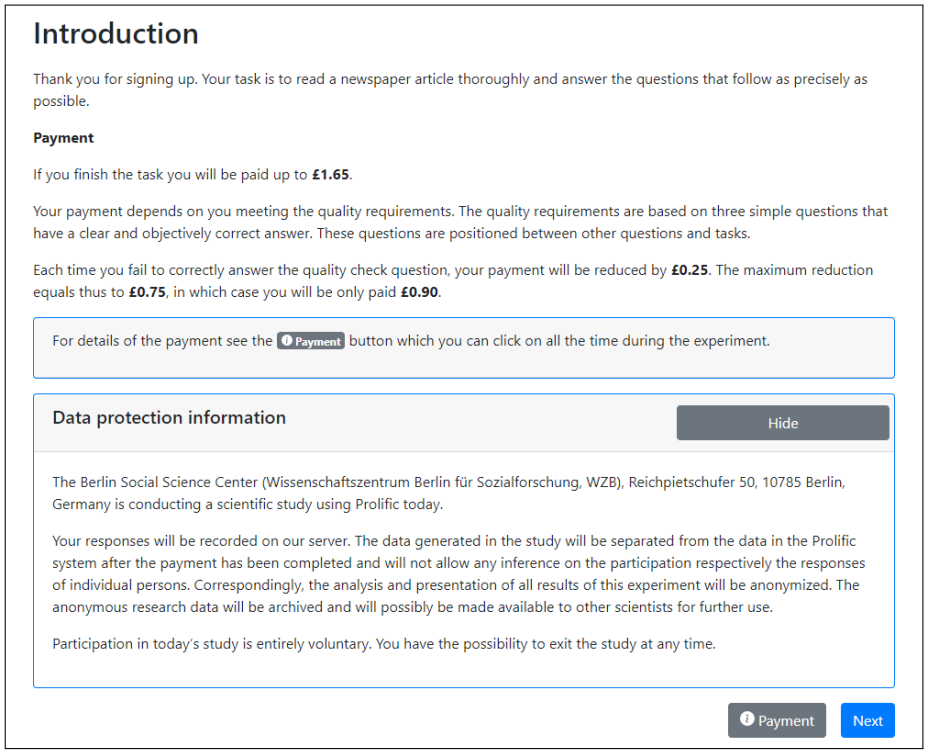


Figure 2: Article page

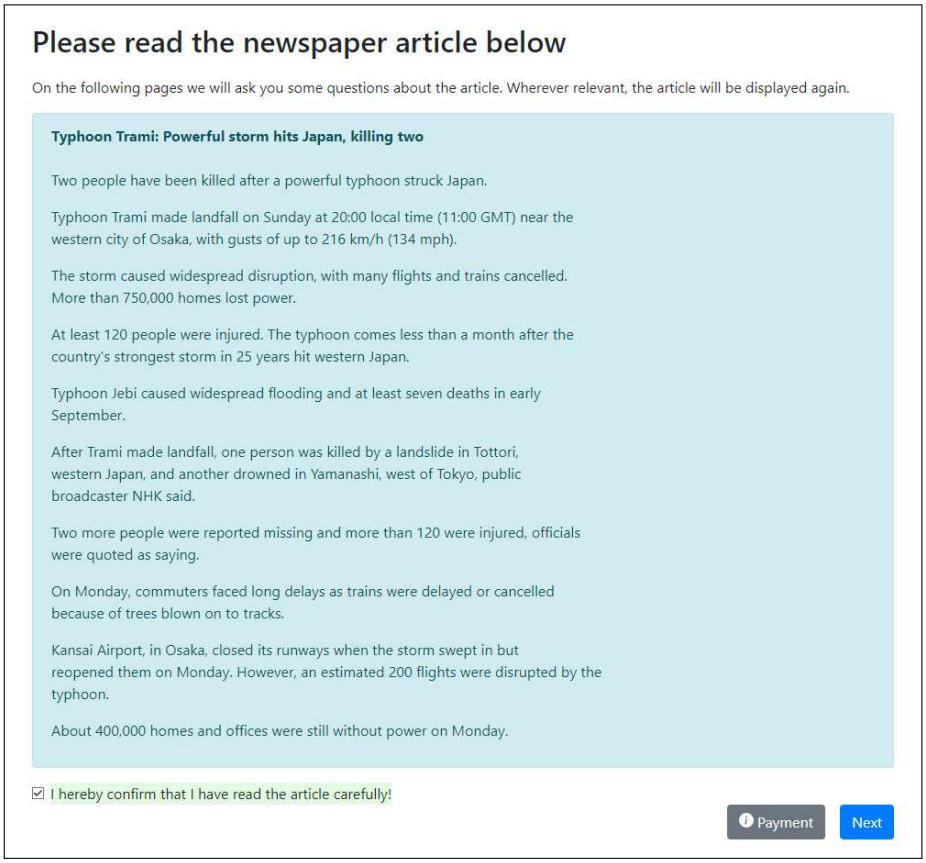


Figure 3

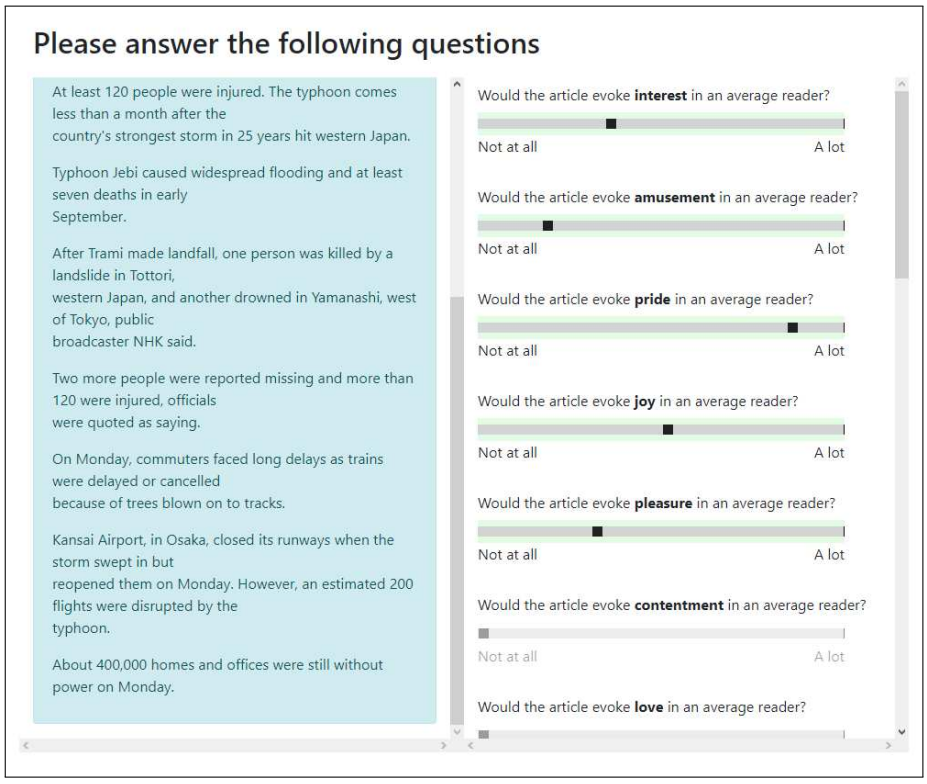

Please answer the following questions
At least 120 people were injured. The typhoon comes less than a month after the country's strongest storm in 25 years hit western Japan.
Typhoon Jebi caused widespread flooding and at least seven deaths in early September.
After Trami made landfall, one person was killed by a landslide in Tottori, western Japan, and another drowned in Yamanashi, west of Tokyo, public broadcaster NHK said.
Two more people were reported missing and more than 120 were injured, officials were quoted as saying.
On Monday, commuters faced long delays as trains were delayed or cancelled because of trees blown on to tracks.
Kansai Airport, in Osaka, closed its runways when the storm swept in but reopened them on Monday. However, an estimated 200 flights were disrupted by the typhoon.
About 400,000 homes and offices were still without power on Monday.
Would the article evoke interest in an average reader?
Not at all
A lot
Would the article evoke amusement in an average reader?
Not at all
A lot
Would the article evoke pride in an average reader?
Not at all
A lot
Would the article evoke joy in an average reader?
Not at all
A lot
Would the article evoke pleasure in an average reader?
Not at all
A lot
Would the article evoke contentment in an average reader?
Not at all
A lot
Would the article evoke love in an average reader?

Figure 4

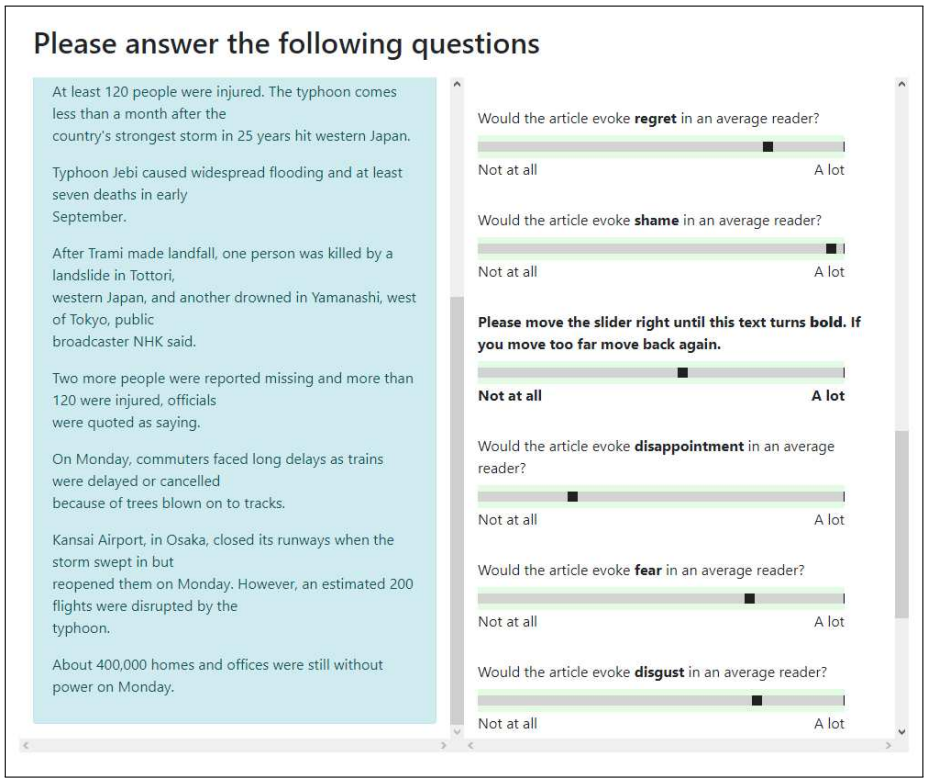


Figure 5: Treatment page I: Recognition

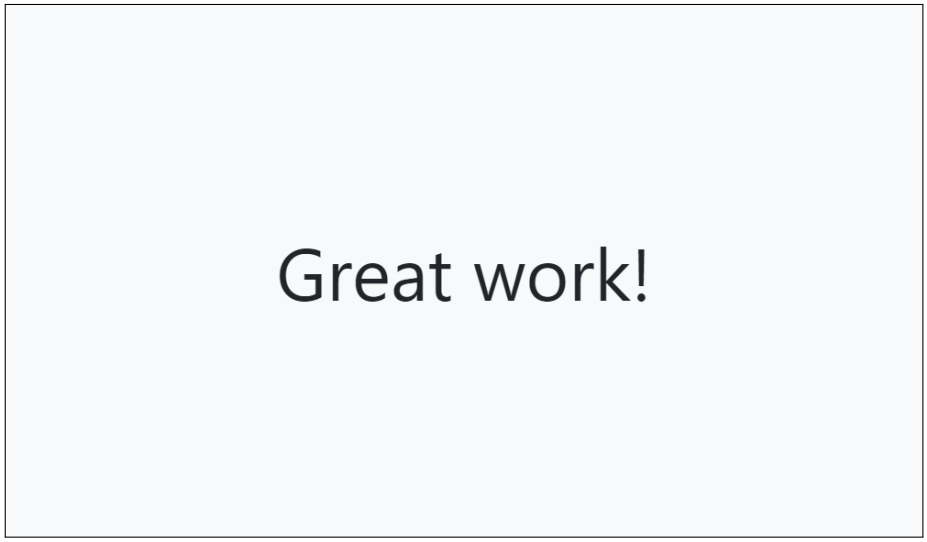


Figure 6

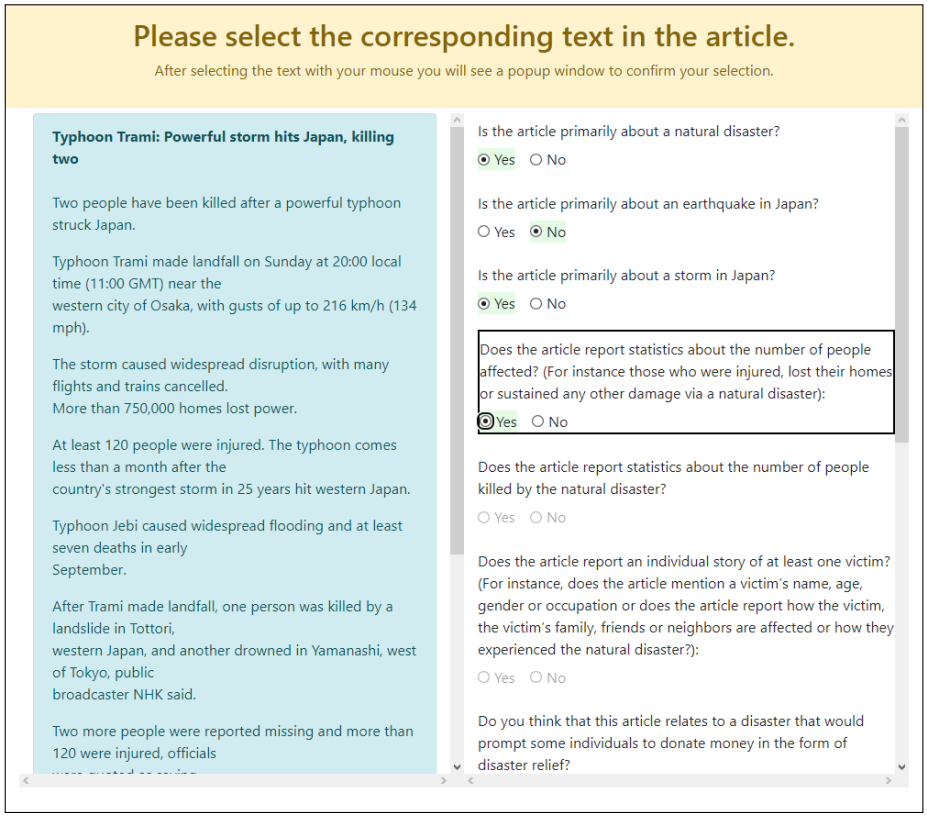

Please select the corresponding text in the article.
After selecting the text with your mouse you will see a popup window to confirm your selection.
Typhoon Trami: Powerful storm hits Japan, killing two
Two people have been killed after a powerful typhoon struck Japan.
Typhoon Trami made landfall on Sunday at 20:00 local time (11:00 GMT) near the western city of Osaka, with gusts of up to 216 km/h (134 mph).
The storm caused widespread disruption, with many flights and trains cancelled.
More than 750,000 homes lost power.
At least 120 people were injured. The typhoon comes less than a month after the country's strongest storm in 25 years hit western Japan.
Typhoon Jebi caused widespread flooding and at least seven deaths in early September.
After Trami made landfall, one person was killed by a landslide in Tottori, western Japan, and another drowned in Yamanashi, west of Tokyo, public broadcaster NHK said.
Two more people were reported missing and more than 120 were injured, officials
Is the article primarily about a natural disaster?
Yes No
Is the article primarily about an earthquake in Japan?
Yes No
Is the article primarily about a storm in Japan?
Yes No
Does the article report statistics about the number of people affected? (For instance those who were injured, lost their homes or sustained any other damage via a natural disaster):
Yes No
Does the article report statistics about the number of people killed by the natural disaster?
Yes No
Does the article report an individual story of at least one victim? (For instance, does the article mention a victim's name, age, gender or occupation or does the article report how the victim, the victim's family, friends or neighbors are affected or how they experienced the natural disaster?):
Yes No
Do you think that this article relates to a disaster that would prompt some individuals to donate money in the form of disaster relief?

Figure 7

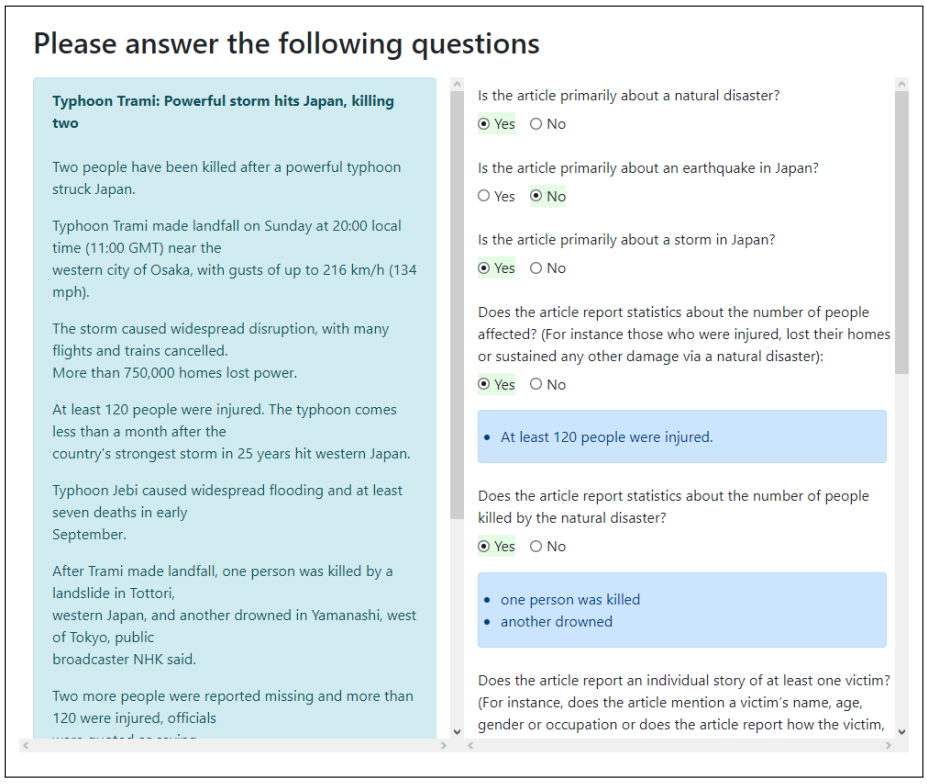

Please answer the following questions
Typhoon Trami: Powerful storm hits Japan, killing two
Two people have been killed after a powerful typhoon struck Japan.
Typhoon Trami made landfall on Sunday at 20:00 local time (11:00 GMT) near the western city of Osaka, with gusts of up to 216 km/h (134 mph).
The storm caused widespread disruption, with many flights and trains cancelled.
More than 750,000 homes lost power.
At least 120 people were injured. The typhoon comes less than a month after the country's strongest storm in 25 years hit western Japan.
Typhoon Jebi caused widespread flooding and at least seven deaths in early September.
After Trami made landfall, one person was killed by a landslide in Tottori, western Japan, and another drowned in Yamanashi, west of Tokyo, public broadcaster NHK said.
Two more people were reported missing and more than 120 were injured, officials
Is the article primarily about a natural disaster?
Yes
No
Is the article primarily about an earthquake in Japan?
Yes
No
Is the article primarily about a storm in Japan?
Yes
No
Does the article report statistics about the number of people affected? (For instance those who were injured, lost their homes or sustained any other damage via a natural disaster):
Yes
No
At least 120 people were injured.
Does the article report statistics about the number of people killed by the natural disaster?
Yes
No
one person was killed
another drowned
Does the article report an individual story of at least one victim? (For instance, does the article mention a victim's name, age, gender or occupation or does the article report how the victim,

Figure 8

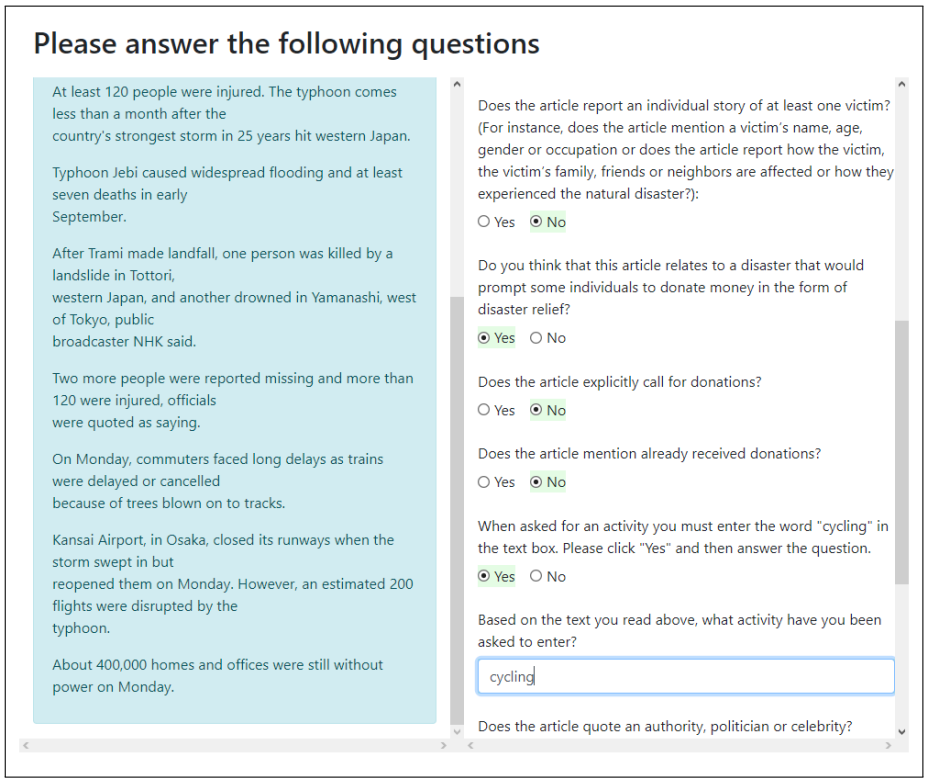

Please answer the following questions
At least 120 people were injured. The typhoon comes less than a month after the country's strongest storm in 25 years hit western Japan.
Typhoon Jebi caused widespread flooding and at least seven deaths in early September.
After Trami made landfall, one person was killed by a landslide in Tottori, western Japan, and another drowned in Yamanashi, west of Tokyo, public broadcaster NHK said.
Two more people were reported missing and more than 120 were injured, officials were quoted as saying.
On Monday, commuters faced long delays as trains were delayed or cancelled because of trees blown on to tracks.
Kansai Airport, in Osaka, closed its runways when the storm swept in but reopened them on Monday. However, an estimated 200 flights were disrupted by the typhoon.
About 400,000 homes and offices were still without power on Monday.
Does the article report an individual story of at least one victim? (For instance, does the article mention a victim's name, age, gender or occupation or does the article report how the victim, the victim's family, friends or neighbors are affected or how they experienced the natural disaster?):
Yes
No
Do you think that this article relates to a disaster that would prompt some individuals to donate money in the form of disaster relief?
Yes
No
Does the article explicitly call for donations?
Yes
No
Does the article mention already received donations?
Yes
No
When asked for an activity you must enter the word "cycling" in the text box. Please click "Yes" and then answer the question.
Yes
No
Based on the text you read above, what activity have you been asked to enter?
cycling
Does the article quote an authority, politician or celebrity?

Figure 9

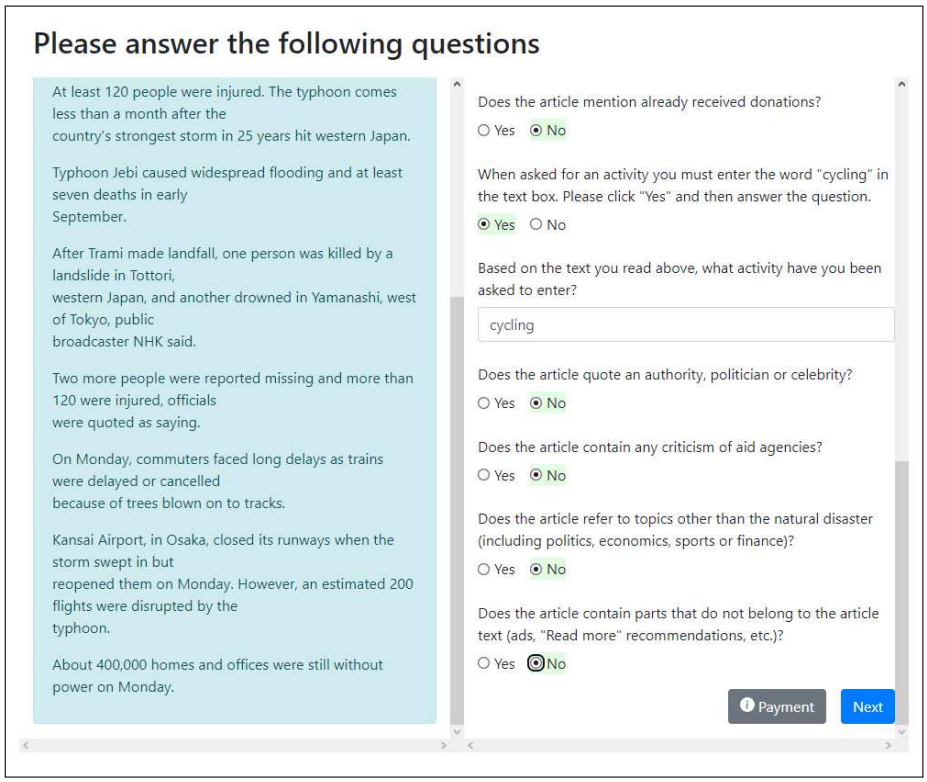


Figure 10: Treatment page II: Recognition

Figure 11

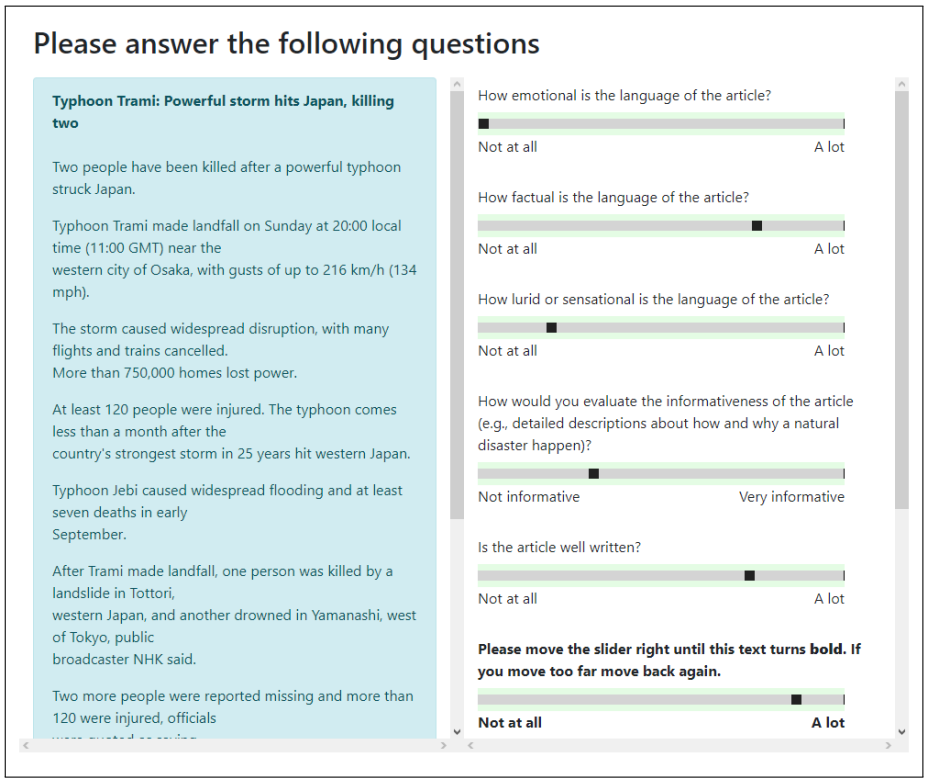

Please answer the following questions
Typhoon Trami: Powerful storm hits Japan, killing two
Two people have been killed after a powerful typhoon struck Japan.
Typhoon Trami made landfall on Sunday at 20:00 local time (11:00 GMT) near the western city of Osaka, with gusts of up to 216 km/h (134 mph).
The storm caused widespread disruption, with many flights and trains cancelled. More than 750,000 homes lost power.
At least 120 people were injured. The typhoon comes less than a month after the country's strongest storm in 25 years hit western Japan.
Typhoon Jebi caused widespread flooding and at least seven deaths in early September.
After Trami made landfall, one person was killed by a landslide in Tottori, western Japan, and another drowned in Yamanashi, west of Tokyo, public broadcaster NHK said.
Two more people were reported missing and more than 120 were injured, officials
How emotional is the language of the article?
Not at all
A lot
How factual is the language of the article?
Not at all
A lot
How lurid or sensational is the language of the article?
Not at all
A lot
How would you evaluate the informativeness of the article (e.g., detailed descriptions about how and why a natural disaster happen)?
Not informative
Very informative
Is the article well written?
Not at all
A lot
Please move the slider right until this text turns bold. If you move too far move back again.
Not at all
A lot

Figure 12

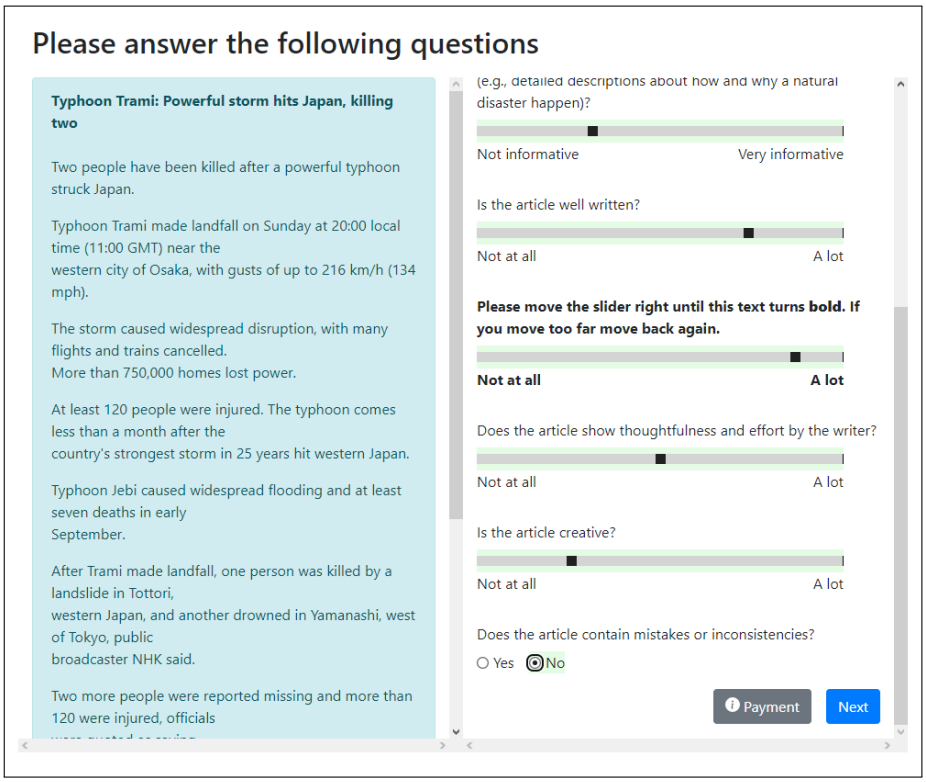


Figure 13: Treatment page III: Recognition

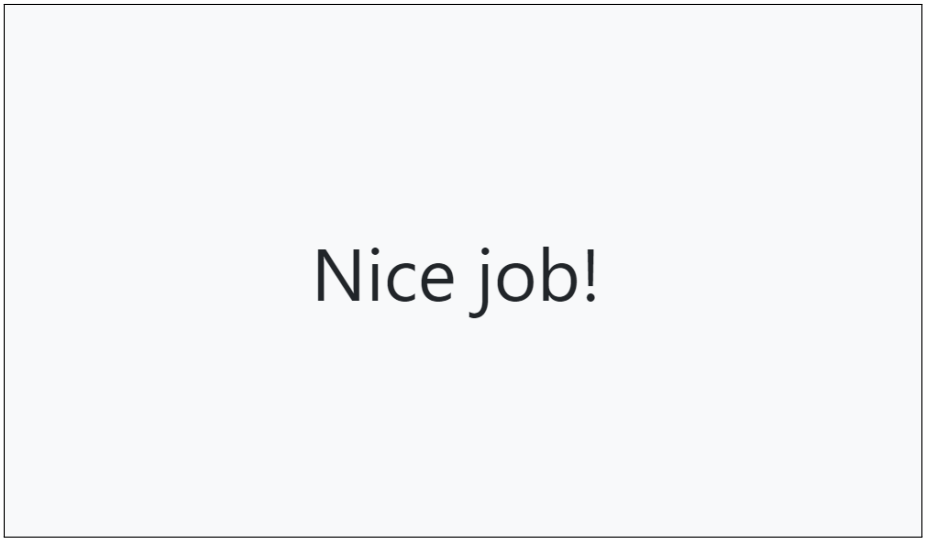


Figure 14

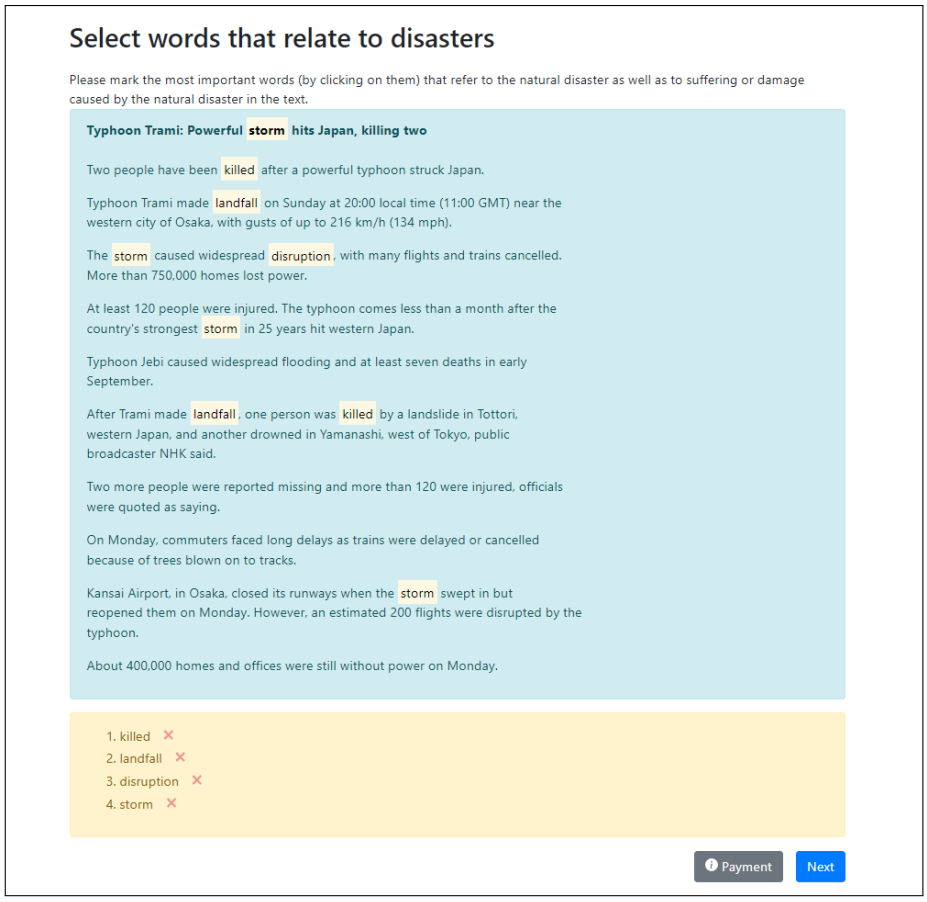

Select words that relate to disasters
Please mark the most important words (by clicking on them) that refer to the natural disaster as well as to suffering or damage caused by the natural disaster in the text.
Typhoon Trami: Powerful storm hits Japan, killing two
Two people have been killed after a powerful typhoon struck Japan.
Typhoon Trami made landfall on Sunday at 20:00 local time (11:00 GMT) near the western city of Osaka, with gusts of up to 216 km/h (134 mph).
The storm caused widespread disruption, with many flights and trains cancelled. More than 750,000 homes lost power.
At least 120 people were injured. The typhoon comes less than a month after the country's strongest storm in 25 years hit western Japan.
Typhoon Jebi caused widespread flooding and at least seven deaths in early September.
After Trami made landfall, one person was killed by a landslide in Tottori, western Japan, and another drowned in Yamanashi, west of Tokyo, public broadcaster NHK said.
Two more people were reported missing and more than 120 were injured, officials were quoted as saying.
On Monday, commuters faced long delays as trains were delayed or cancelled because of trees blown on to tracks.
Kansai Airport, in Osaka, closed its runways when the storm swept in but reopened them on Monday. However, an estimated 200 flights were disrupted by the typhoon.
About 400,000 homes and offices were still without power on Monday.
1. killed
2. landfall
3. disruption
4. storm
Payment
Next

Figure 15: Treatment page IV: Recognition

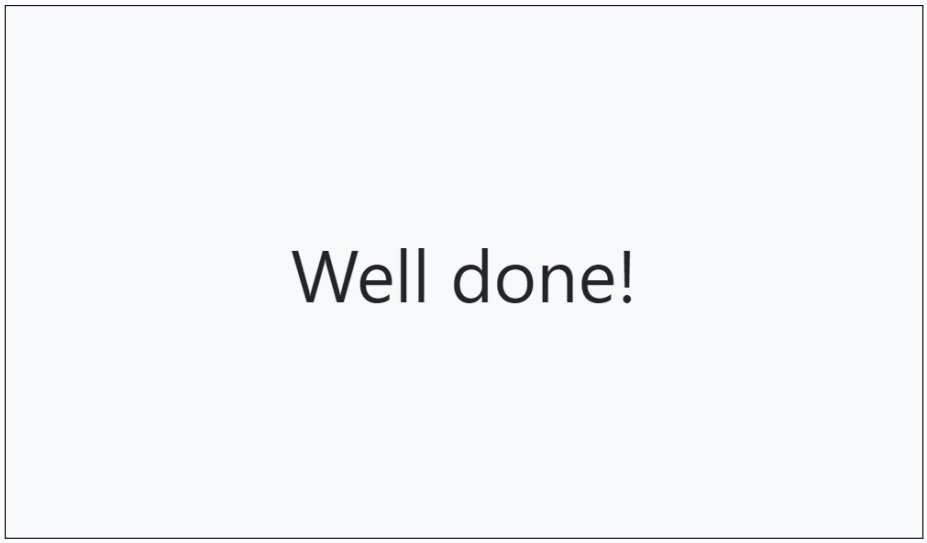


Figure 16

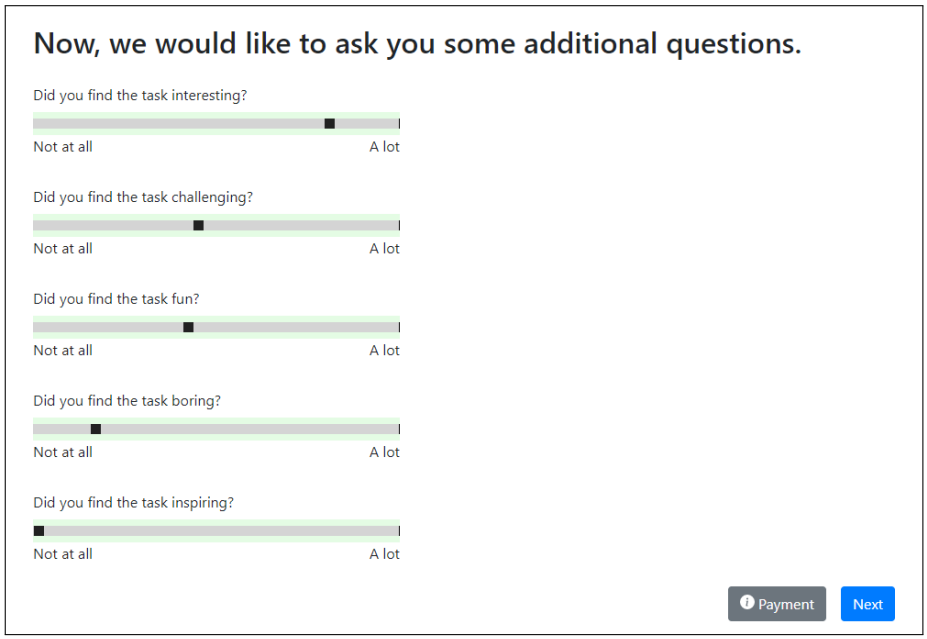


Figure 17

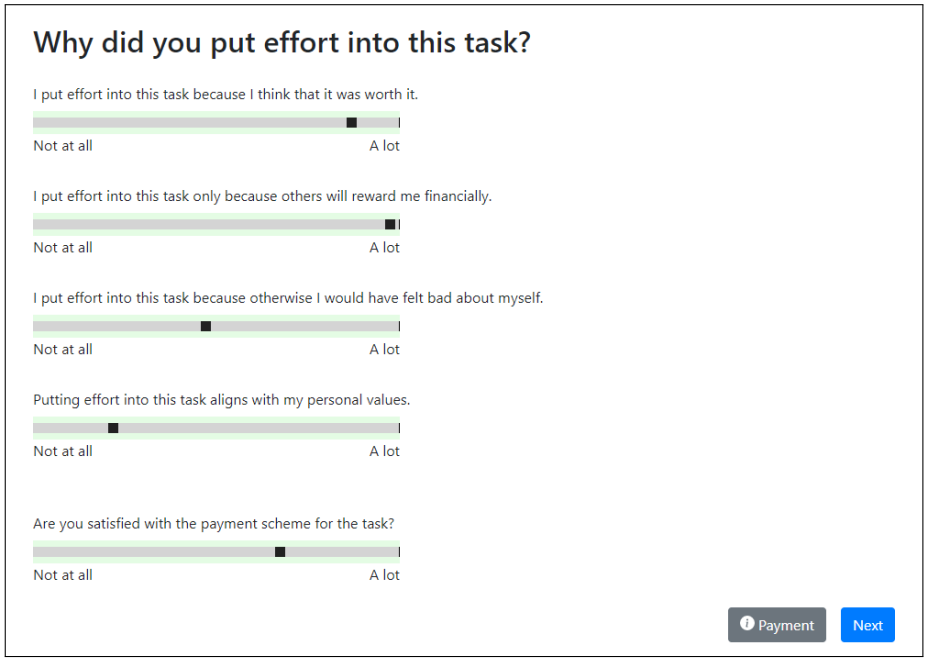

Why did you put effort into this task?
I put effort into this task because I think that it was worth it.
Not at all
A lot
I put effort into this task only because others will reward me financially.
Not at all
A lot
I put effort into this task because otherwise I would have felt bad about myself.
Not at all
A lot
Putting effort into this task aligns with my personal values.
Not at all
A lot
Are you satisfied with the payment scheme for the task?
Not at all
A lot
Payment
Next

Figure 18


Thank you for rating the article!
Payment
You answered 3 out of 3 quality checks correctly. You will receive £1.65 for this task on your Prolific account None.
Due to Prolific's payment scheme you will get £0.90 once we approve your submission. The remaining £0.75 will be paid as a bonus.
Comments
This is the pilot version of our study and your feedback is greatly appreciated.
Was the estimated time to finish the study:
way too short
somewhat too short
just right
somewhat too high
much
too high
Do you have any comments about this survey:
No
Payment
Next

1.2 Remaining treatment versions

Figure 19: Variation in first dimension: Neutral

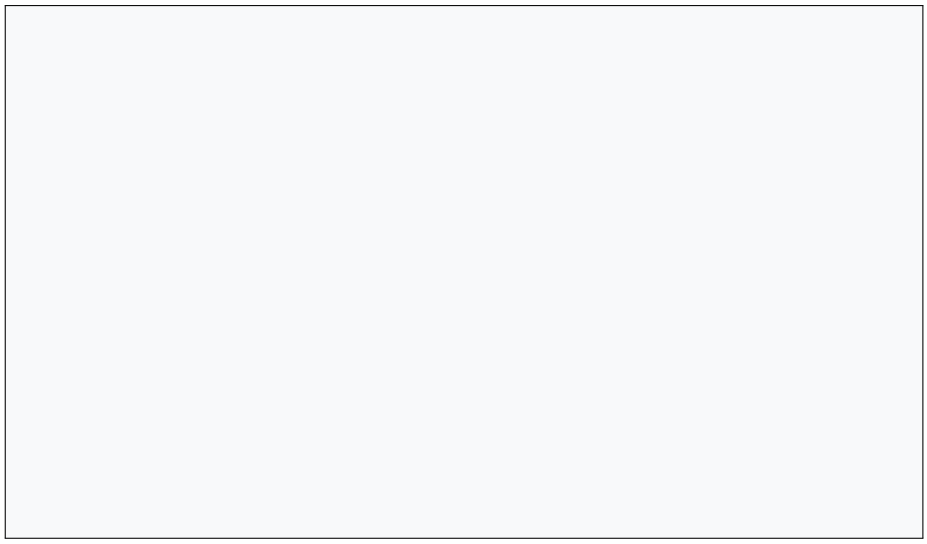


Figure 20: Variation in first dimension: Appreciation I

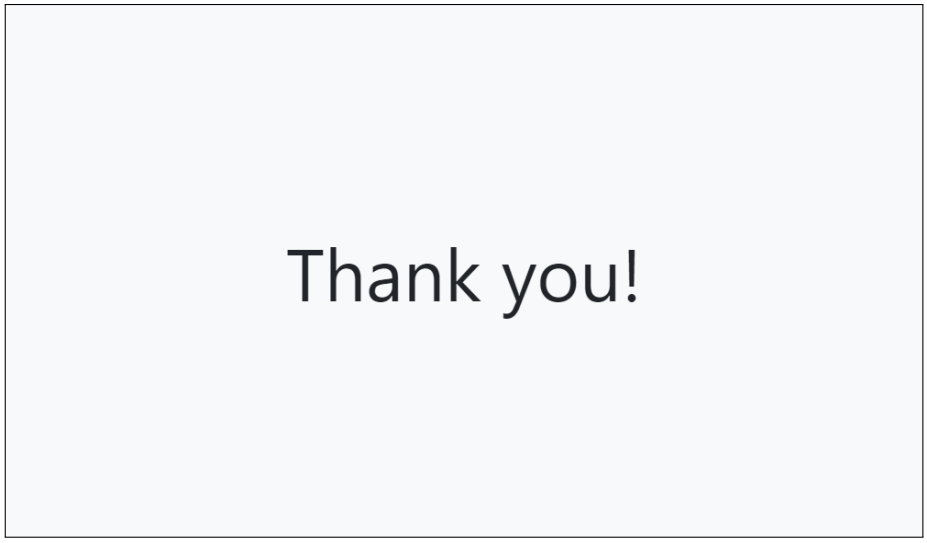


Figure 21: Variation in first dimension: Appreciation II

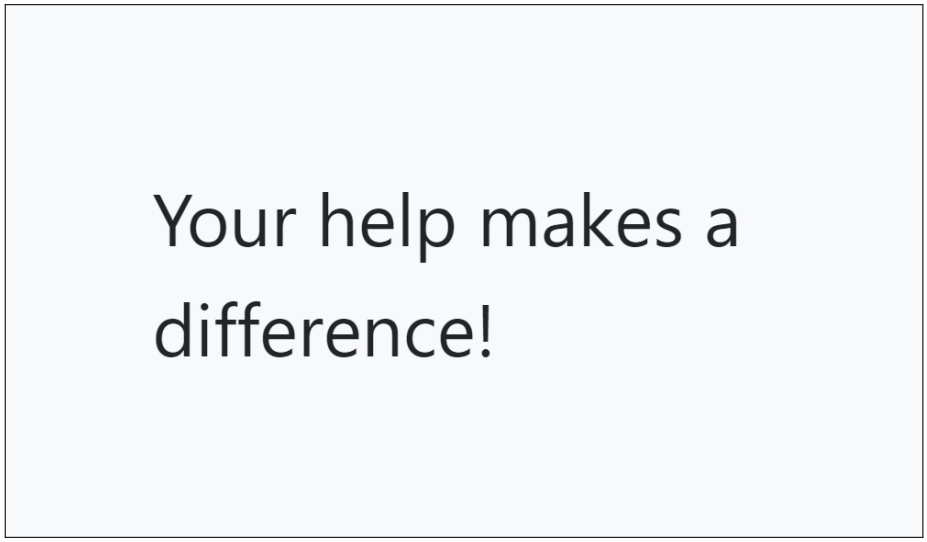


Figure 22: Variation in first dimension: Appreciation III

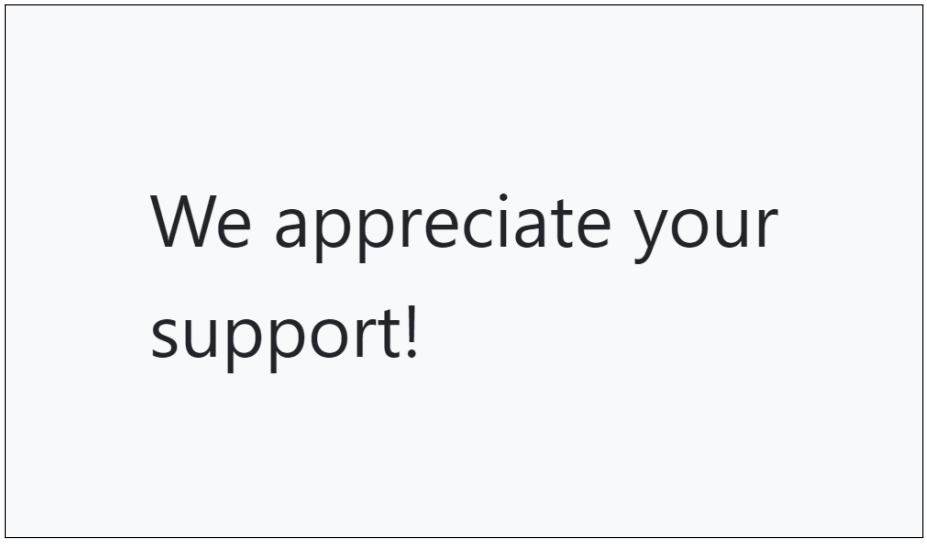


Figure 23: Variation in first dimension: Appreciation IV

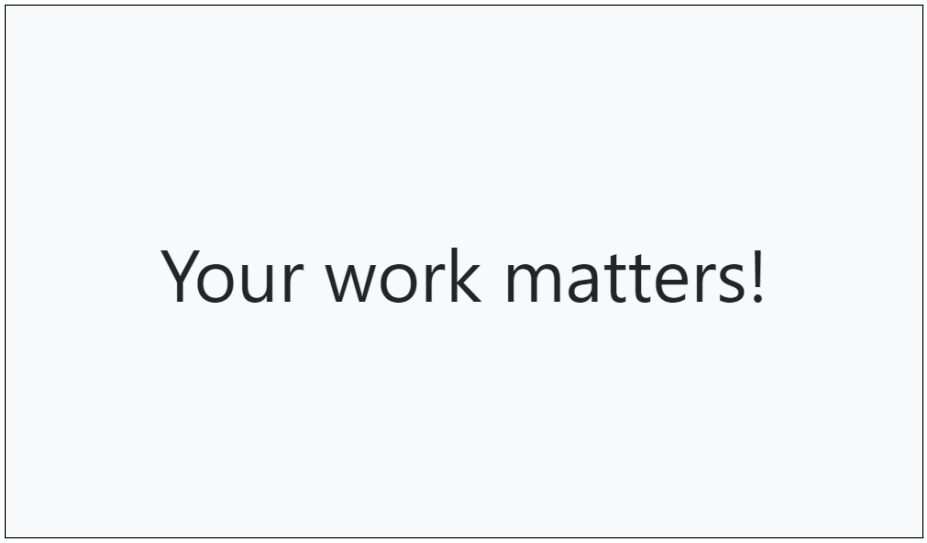


Figure 24: Variation in second dimension: Loss

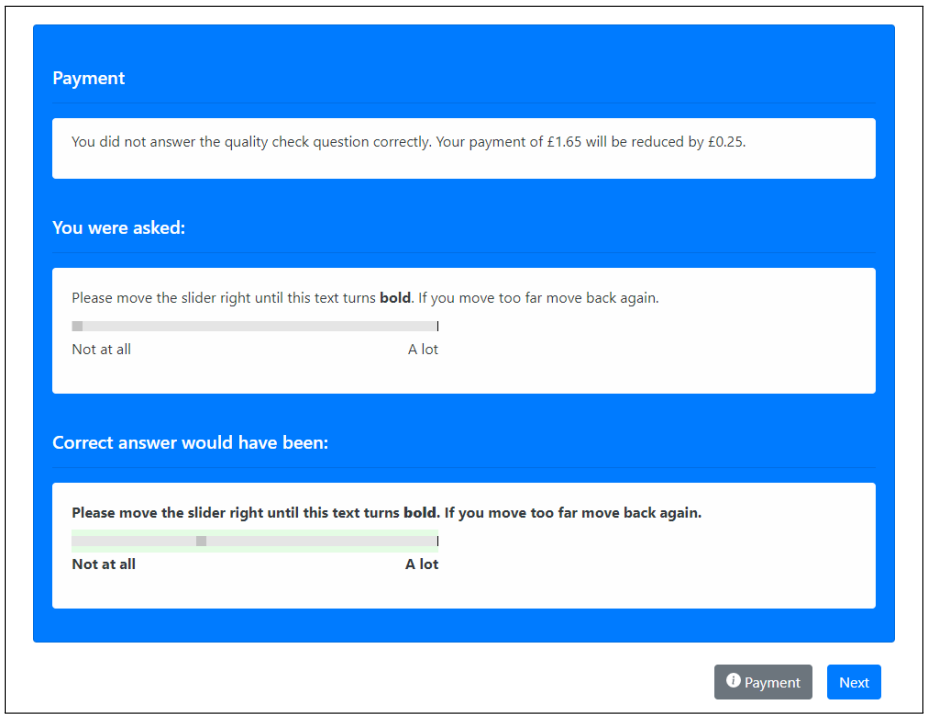


Figure 25: Variation in second dimension: Introduction page

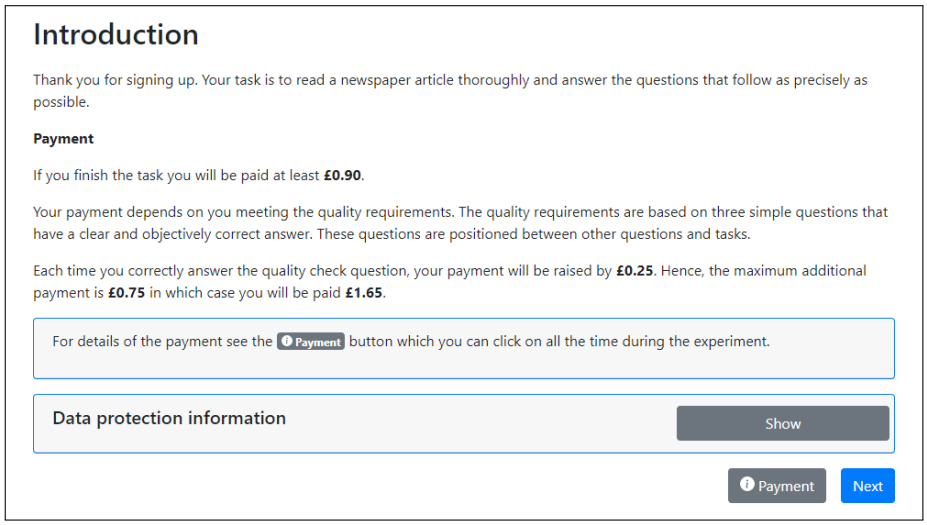


Figure 26: Variation in second dimension: Gain I

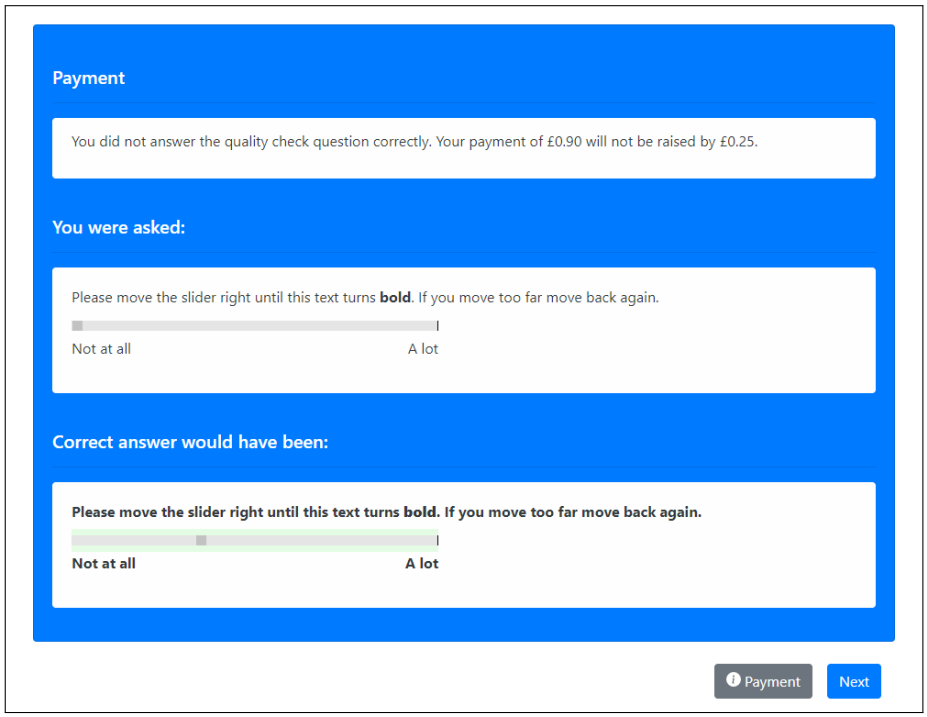


Figure 27: Variation in second dimension: Gain II

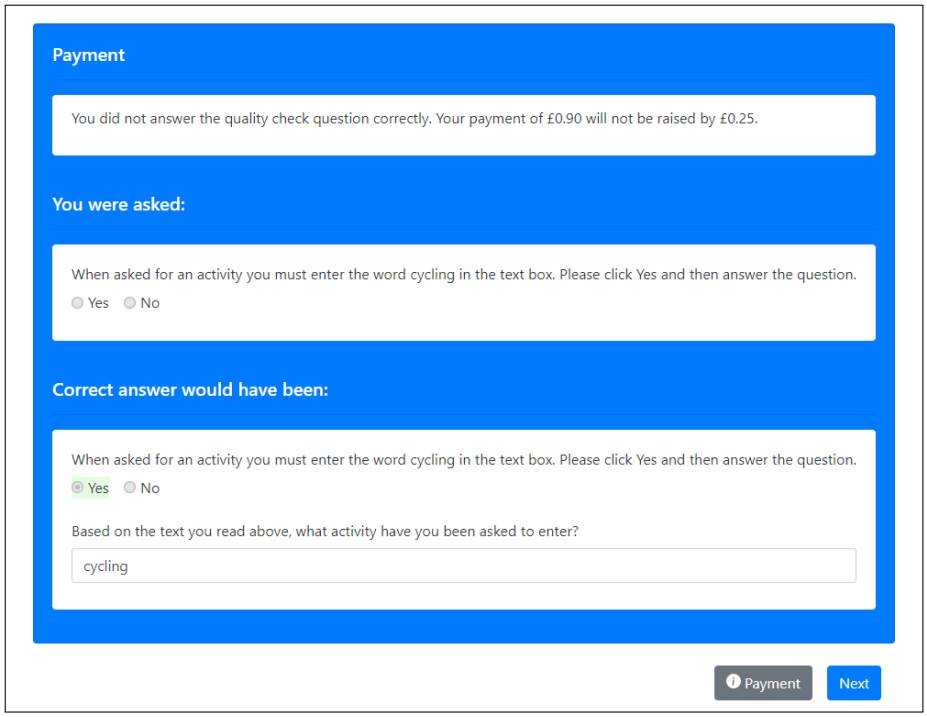


Figure 28: Variation in second dimension: Gain III