

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Engelmann, Dirk; Friedrichsen, Jana; Kübler, Dorothea

Article — Published Version Fairness in markets and market experiments: insights from a field-plus-lab study and a failed replication

The Scandinavian Journal of Economics

Provided in Cooperation with: John Wiley & Sons

Suggested Citation: Engelmann, Dirk; Friedrichsen, Jana; Kübler, Dorothea (2024) : Fairness in markets and market experiments: insights from a field-plus-lab study and a failed replication, The Scandinavian Journal of Economics, ISSN 1467-9442, Wiley Periodicals, Inc., Hoboken, NJ, Vol. 126, Iss. 4, pp. 698-732, https://doi.org/10.1111/sjoe.12565

This Version is available at: https://hdl.handle.net/10419/313708

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.



http://creativecommons.org/licenses/by/4.0/

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU

The Scandinavian Journal of Economics

Scand. J. of Economics 126(4), 698–732, 2024 DOI: 10.1111/sjoe.12565

Fairness in markets and market experiments: insights from a field-plus-lab study and a failed replication

Dirk Engelmann

Humboldt-Universität zu Berlin, DE-10099 Berlin, Germany dirk.engelmann@hu-berlin.de

Jana Friedrichsen

Christian-Albrechts-Universität zu Kiel, DE-24098 Kiel, Germany friedrichsen@economics.uni-kiel.de

Dorothea Kübler

WZB Berlin Social Science Center and Technical University Berlin, DE-10785 Berlin, Germany kuebler@wzb.eu

Abstract

We investigate how fairness in a laboratory experiment framed as a market exchange relates to preferences for fair trade products elicited before and at the end of the experiment. We collected two samples, 10 years apart. In the original sample, fairness in the market experiment measured by the willingness to buy at a higher price when higher wages are paid to the worker correlates both with the choice of a fair trade product and with the willingness to pay a positive fair trade premium. These correlations are not significant in the replication experiment, which indicates at best a weak relationship.

Keywords: External validity; fair trade; social preferences *JEL classification*: C91; D01; D91

1. Introduction

Laboratory experiments have played a vital role in reintroducing social preferences and fairness into economics. Starting in the early 2000s, experiments on pro-sociality and fairness have been extended to market environments. They demonstrate that a market frame does not generally lead to the irrelevance of fairness concerns. In market experiments, a substantial share of consumers is willing to pay a higher price if firms pay workers a higher wage or reduce negative externalities (Rode et al., 2008; Bartling et al., 2015; Pigors and Rockenbach, 2016; Danz et al., 2022).¹

¹While Falk and Szech (2013) argue that markets crowd out concerns for fairness, a different interpretation of their data is provided by Breyer and Weimann (2015) and Sutter et al. (2020).

^{© 2024} The Authors. *The Scandinavian Journal of Economics* published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the *SJE*.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

For the external validity of pro-social behavior observed in market experiments, it is pertinent whether this behavior correlates with behavior in actual markets outside of the laboratory. Given the relevance and pervasiveness of market interactions, this constitutes an important case for the debate whether concerns for fairness that are frequently observed in the laboratory (Fehr and Schmidt, 2006) generalize to field behavior in a meaningful way (Levitt and List, 2007; Camerer, 2015). Levitt and List (2007) point to a number of important differences between laboratory experiments and behavior outside the laboratory in the realm of social preference experiments: participants in the laboratory know that they are being investigated by the researchers and their decisions may not remain anonymous; the context as well as the stakes matter for choices and often cannot be controlled perfectly by the experimenters; the participants are not the same in experiments and in the relevant natural contexts; the choices and time horizons are often restricted in an artificial manner in experiments. They argue that this limits the relevance of pro-social behavior found in the laboratory for field behavior.

In this paper, we provide a direct test of the relationship between market behavior in the laboratory, where externalities concern other participants, with choices that have externalities outside the laboratory. In particular, we elicit choices between a fair trade and a standard product and the willingness to pay (WTP) for fair trade, and then compare these choices with behavior in a market experiment. The market game is designed to closely resemble a situation that motivates the fair trade movement: consumers have market power, firms compete, and workers receive only a very small share of the surplus in the equilibrium without social preferences.² We implement a simplified version of the market experiment by Danz et al. (2022) that shares many features with the above-mentioned experiments on fairness in markets, such as Bertrand competition and externalities on workers who are participants in the experiment.

We investigate whether the willingness of consumers to pay a higher price if firms pay a higher wage in the experimental market correlates with two choices regarding actual fair trade products. First, participants are classified with respect to their choice between a fair trade chocolate bar and a larger conventional chocolate bar. This choice of chocolate takes place outside of

²We ignore complications that arise from market power being present at different stages of the supply chain of many fair trade products. For instance, traders, roasters, and retailers in the coffee supply chain have market power, and the overall effects on the distribution of surplus are unclear. At any rate, producers often face a monopsonist buyer, like the firm does in our experiment. Our design is not intended to, and is not well suited to, study fairness concerns of participants in the role of firms because firms act in a competitive environment. While consumers have strong market power, firms have only little. As a result, strategic concerns affect firm behavior. We therefore refrain from analyzing the behavior of subjects in the role of firms.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

the laboratory a few days before subjects participate in the market experiment. It involves a trade-off between the chocolate being produced and traded under fair trade standards and the size of the chocolate bar, thereby resembling everyday choices between fair trade and conventional products. This first measure allows us to evaluate the correlation between choices for actual fair trade products and behavior in a stylized market experiment.

For our second measure of fair trade preferences of consumers, we elicit the WTP for a bar of fair trade chocolate and a bar of conventional chocolate right after the market experiment. The difference between the two is the fair trade premium participants are willing to pay. While we elicit the fair trade premium in the laboratory, it is linked to an actual physical product, thereby affecting actual workers just like the purchase of a fair trade or conventional product in a store.

Our two measures of consumer fairness aim at alleviating the points of criticism raised by Levitt and List (2007). First, experimenter scrutiny is possibly felt less in the choice task before the laboratory experiment compared with the elicitation of the fair trade premium in the laboratory because it does not take place in the laboratory and may hence not be perceived as an experimental decision. Second, while our participants are not representative consumers, students are a relevant consumer group for fair trade products, just as other parts of the population. Also, we do not see any reason why a correlation detected for students should not also be present for other consumers. Third, while the stakes in our experiment are small, they are also small for consumers in many of the markets we are interested in, namely fair trade choices, or more generally fair consumer behavior. Because we use actual chocolate bars, the stakes are identical in the experiment and the field.³ Fourth, regarding the context, the advantage of the first task is that it is not influenced by the laboratory setting, whereas the second task involves money to express the preferences and it is easier to implement than the first, which could make it valuable for future studies. Finally and importantly, we compare choices between real consumption goods with choices in an experimental market that shares important features with the real market, rather than correlating non-experimental choices to abstract and context-free experimental games with a somewhat arbitrary connection between the two. Such exercises may be part of the explanation for the mixed results so far; see Galizzi and Navarro-Martínez (2018) for this argument

³Furthermore, we believe that the critique of using small stakes in the laboratory is partly misguided. Whereas it is indeed potentially problematic to draw inferences from laboratory choices for large-stake choices such as education choices, this does not imply that studying small-stake behavior is of minor interest. Many daily decisions are small-stakes choices for individuals but have a large impact in the aggregate. In particular, this is true for many consumption choices such as individual food purchases. These are the settings we have in mind.

^{© 2024} The Authors. *The Scandinavian Journal of Economics* published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the *SJE*.

and an overview of studies relating laboratory with field evidence on social preferences.

We conducted a first set of experiments in 2012 and a replication study in 2022. In the original experiment, we find that the likelihood of consumers buying from the firm that pays a higher wage and asks for a higher price is correlated at the individual level both with the likelihood to choose fair trade chocolate before the laboratory experiment and with the willingness to pay a positive fair trade premium elicited after the market experiment. It further correlates with positive attitudes toward fair trade and the stated frequency to buy fair trade products, as measured with a questionnaire at the end of the experiment. These results support the hypothesis that revealed fairness concerns in a stylized market experiment are indeed indicative of the WTP for fair behavior of firms in natural markets.⁴ The results of the original experiment are, therefore, in line with the idea that laboratory experiments assessing the determinants of fair behavior in experimental markets provide relevant insights into the determinants of fair behavior in the field but caution is needed as the study was very small.

In the follow-up study conducted 10 years after the original experiments, however, the results are generally much weaker and mostly insignificant even though the point estimates have the same sign as in the original study. Specifically, in the replication experiment, we find no significant correlation between any of our three indicators of fair market behavior in the market game and the choice of the fair trade chocolate over the conventional chocolate bar. Neither does the market behavior correlate significantly with the elicited fair trade premium in the replication data. We still see a significant positive correlation between the choice of fair trade chocolate and the willingness to pay a premium for fair trade. Further, attitudes toward fair trade that we infer from the questionnaire data correlate with the two measures of fair trade purchasing behavior but hardly relate to behavior in the abstract market game. When we analyze the pooled data from the original and the replication experiment, the results of the original experiment are substantially weakened but remain statistically significant. Together the two studies suggest that behavior in abstract market experiments may relate to fair purchasing behavior in real markets, but the effect is probably small in size.

Our paper is motivated by the controversy about the relevance of pro-social behavior in the laboratory for behavior in the field (Levitt and List, 2007; Camerer, 2015). Sports-card trading (List, 2006) is a prime example of the

⁴While we do not observe purchasing behavior in the field, which is difficult to link to data from laboratory experiments for data protection reasons, we do observe choices between products offered in actual markets that differ in their externalities on real workers. Purchasing behavior in the field might be more easily linked to data from lab-in-the-field experiments but our interest is explicitly to investigate the relevance of classical laboratory experiments.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

endeavor to link and compare the field and the laboratory. Other examples more closely related to issues of fairness include Fehr and Leibbrandt (2011) and Stoop et al. (2012) on cooperative behavior of fishermen when dealing with a common pool resource. A combination of field and lab-in-the-field experiments is used by Carlsson et al. (2014) to study whether cooperative behavior is stable at the level of the individual. They find strong evidence of such a correlation over time (i.e., a number of years) and over the different experimental formats for a non-student sample in Vietnam.

Previous research on the external validity of experimentally elicited fairness preferences focuses on aspects of giving and helping behavior, for example, considering dictator game choices and charitable giving (Benz and Meier, 2008; Franzen and Pointner, 2013; Winking and Mizer, 2013; Stoop, 2014; Galizzi and Navarro-Martínez, 2018).⁵ While charitable giving is an important activity, market interactions take up more of people's time and comprise a larger share of economic activity. Because we expect the external validity of laboratory experiments to depend on the exact context, studying the relevance of fairness in market experiments for the fairness in actual markets is necessary.

In addition to linking behavior in the laboratory to actions outside of the laboratory, our paper contributes to the literature on socially responsible or ethical purchases in experimental markets. Rode et al. (2008) was the first to establish that experimental consumers are willing to pay a higher price for the ethically differentiated product involving a donation. Danz et al. (2022) confirm that a significant fraction of consumers are willing to pay a higher price to support higher wages for dependent workers. Pigors and Rockenbach (2016) find that, in a monopoly market, socially responsible production (i.e., higher wages and higher prices) does not pay for the firm as it is not rewarded by the consumers. However, it becomes profitable in an oligopoly setting. Our set-up with a monopsonistic consumer and two competing firms resembles the latter set-up where it is profitable to firms to differentiate by engaging in corporate social responsibility (e.g., by paying better wages to the workers).

Bartling et al. (2015) designed an experiment to investigate the claim by Falk and Szech (2013) that markets erode moral behavior. In experiments conducted in Switzerland and in China, the study finds that consumers have a persistent preference for "clean" products (without negative externalities) and are willing to pay a higher price for them. Irlenbusch and Saxler (2019)

⁵More broadly, research on external validity also considers questions such as whether experimentally elicited time and risk preferences are good predictors of field behavior. For example, Sutter et al. (2013) find that time preferences elicited through choice lists can predict smoking and saving behavior, whereas risk and ambiguity preferences do not.

^{© 2024} The Authors. *The Scandinavian Journal of Economics* published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the *SJE*.

conducted an experiment to distinguish between three properties of markets: diffusion of responsibility, social information, and buyer/seller framing. They show that the latter two affect the fairness of subjects while they observe no effect of the diffusion of responsibility on socially responsible behavior. Extending the experimental design introduced by Falk and Szech (2013), Sutter et al. (2020) show that trading volume in a double auction market is lower if trading is associated with a negative externality on a third party than when it is not, indicating that moral concerns affect behavior also in these markets. Finally, moral behavior in individual tasks and in markets is affected in a similar way by institutional changes, such as the removal of anonymity, monetary incentives, etc. (Kirchler et al., 2016).

A number of studies couple products in laboratory markets with charitable donations outside of the laboratory to capture positive externalities. In Feicht et al. (2016), sellers in a Bertrand market can actively bundle their product with a charitable donation. Consumers are found to purchase from a firm with a higher credible donation only if price differences are negligible. In a related study by Soetevent et al. (2016) where the amount donated was less transparent (indicated as a percentage of the price of the good), considerable social behavior in markets is observed. If the charitable donation is high enough, participants are willing to pay higher prices than without the bundling, possibly because they overestimate the amount of money going to the charity. The role of information about corporate social responsibility for consumer choices is also the focus of Harrs et al. (2022). They study voluntary disclosure by firms who compete for consumers when the reliability of the information disclosed is varied by the experimenter. It turns out that high corporate social responsibility is chosen by the firms only when firms are able to disclose reliable information. Finally, Etilé and Teyssier (2016) investigate how different certification technologies affect market efficiency when firms choose charitable donations to be bundled with their product. They find that market efficiency is enhanced only if the certification is performed by a third party.

In a separate strand of the literature on ethical consumption, researchers have used surveys and field experiments to better understand the motivation of consumers. Consumers are generally found to report higher WTP for products that are produced under ethical conditions (e.g., fair trade) in survey studies and in studies with convenience samples; see, for example, De Pelsmacker et al. (2005) and Loureiro and Lotade (2005), and the overview by Andorfer and Liebe (2012). However, these intentions are not always reflected in purchasing behavior (e.g. Carrington et al., 2010), and actual market shares of fair trade products are still small in most markets.⁶ One possible reason is

⁶In 2017, the market shares of most fair trade products in Germany were in the one-digit range (TransFair e.V., 2018) even though fair trade is relatively big in Germany compared with other countries. See also footnote 20.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

that the price premium for fair trade is too high for many consumers. Indeed, in a field experiment where prices and labels of coffee were varied in the store, Hainmueller et al. (2015) find that demand increases with the presence of a fair trade label but depending on the initial price, consumer demand for the fair trade coffee may be highly price-elastic. Furthermore, they report substantial heterogeneity in the consumers' WTP for ethical sourcing, which they argue might be driven by differences in social preferences or in the perceived importance of ethical sourcing.

We present our experimental design in the next section and the results in Section 3. We conclude with a brief discussion.

2. Experimental design and procedures

The experiment consists of three parts. First, we derive a proxy for the participants' preference for fair trade chocolate after they have registered for the experiment, but before they come to the laboratory. Second, the subjects take part in a market game in the laboratory. Third, we elicit their WTP for fair trade and conventional chocolate with an incentive-compatible random-price mechanism in order to derive the premium they are willing to pay for fair trade. Our main interest concerns the relation between fairness in the market experiment and the fair trade choice in the first part as well as the fair trade premium elicited in the third part of the experiment.

2.1. The experiment

Part 1: choice between fair trade and conventional chocolate. In the first part of the experiment, we offer subjects the choice between fair trade and conventional milk chocolate as an additional reward for coming to the experiment. As fair trade chocolate is typically more expensive, we offered a choice between one (in half of the sessions two) slightly larger (125g) bar(s) of conventional chocolate and one standard size (100g) bar of fair trade chocolate.⁷ This was done via email. The email was sent out about two days before the laboratory experiment and had to be answered before the subjects came to the laboratory. The chocolate was distributed after the experiment. See Online Appendix A1.1 for an English translation of the recruitment email.

⁷In the first sessions, few subjects chose the conventional chocolate, and we therefore offered a choice between two bars of conventional and one bar of fair trade chocolate in the following sessions. In the replication study, the choice was always between one bar of fair trade and two bars of conventional chocolate.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

Part 2: market game in the laboratory. The second part of the experiment is based on a market game consisting of one consumer, two firms, and one worker. In the game, the consumer can buy up to 10 units of a fictitious homogeneous good, and for each unit that she buys she is redeemed with 20 points from the experimenter. Each firm is run by one manager and we refer to subjects in this role as firms. There is one worker who can produce up to 10 units of the good. The worker is an actual participant in the experiment, but has no choice to make. Having only one worker in each market who represents the workforce simplifies fair behavior for the consumer who can ignore horizontal equity concerns between workers.⁸ Firm $i \in \{1, 2\}$ chooses a uniform price per unit $p_i \in [0, 40]$ and a wage $w_i \in [0, 40]$ per unit. It can sell up to 10 units, $q_i \leq 10$. Wages are paid only for units that are actually sold. Thus, the worker can receive a wage for up to 10 units and has no costs.

If the consumer buys q_i units from firm *i* that has chosen the price p_i and wage w_i , she earns $20 - p_i$ for each unit, the firm makes a profit of $p_i - w_i$ and the worker earns w_i per unit bought at this firm. Total earnings of firm *i* are given by $q_i(p_i - w_i)$. Total earnings of the consumer are $q_1(20 - p_1) + q_2(20 - p_2)$. Total earnings of the worker are given by $q_1w_1 + q_2w_2$.

The timing of the game is as follows. After the two firms have made their choices, the consumer is informed about the price and the wage of each firm, (p_1, w_1) and (p_2, w_2) . The consumer can buy any combination of integer amounts from the two firms up to 10 units, $q_1 + q_2 \le 10$. At the end of each period, all market participants are informed about both firms' prices and wages as well as about the decision of the consumer and their earnings.

Let us consider the equilibria of the stage game when all agents maximize their payoffs. There are three subgame-perfect equilibria. In each of them, firm *i* sets $w_i = 0$. The equilibrium prices are $p_i = 0$, $p_i = 1$, or $p_i = 2$ for i = 1, 2. The consumer buys 10 units from the firm with the lower price as long as min $(p_1, p_2) < 20$, which always holds on the equilibrium path.⁹ If both firms choose the same price (< 20), the consumer buys 10 units, split in an arbitrary way between the firms. Note that, in equilibrium, almost the entire surplus goes to the consumer.¹⁰ In contrast, the payoffs are split equally

 $^{^{8}}$ Complex fairness concerns are studied by Danz et al. (2022) where each firm has its own worker.

⁹Off the equilibrium path, the consumer buys nothing if $\min(p_1, p_2) > 20$ for both firms and any number of units if $\min(p_1, p_2) = 20$.

¹⁰As the stage game has three equilibria with $p_i = 0$, $p_i = 1$, or $p_i = 2$, collusive equilibria of the repeated game exist due to the possibility to punish deviations. However, we do not find evidence of collusive firm behavior. In particular, such collusive equilibria exhibit a downward trend in prices, which we do not observe. Moreover, all equilibria, including collusive ones, involve wages equal to zero. A selfish consumer does not want to pay more for a higher wage, and thus a (selfish but collusive) firm has no reason to pay higher wages.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

among all four market participants if both firms choose $p_i = 15$ and $w_i = 5$ and the consumer buys five units from each of the firms. In this case, each participant earns 50 points. Therefore, we refer to these values as the "fair" price and wage, respectively.

The market game has the property that as long as the consumer buys 10 units, the total earnings in the market are constant. In particular, it does not matter for the total earnings from which firms, and at what prices, the consumer purchases the units. Thus, we can study fairness concerns of consumers toward workers that are not confounded with efficiency concerns.¹¹

The market game was repeated 20 times with fixed groups of four subjects, and subjects kept their role of firm, worker, or consumer throughout the entire market game. The payoffs in all 20 rounds were added to determine the total payment in this part of the experiment. The exchange rate was 100 points for 1 euro.

Part 3: elicitation of fair trade premium. The third part of the experiment yields a measure of the premium that participants are willing to pay for fair trade. From each participant, we elicit his or her WTP (between 0 and 2 euros) for both fair trade and conventional dark chocolate (WTP_{fair} and WTP_{conv}) by relying on a random-price mechanism (Becker et al., 1964). The participants were asked to state a price between 0 and 2 euros where any multiple of 0.01 euros was admissible. The random price was drawn from the uniform distribution of all integer multiples of 0.01 euros between 0 and 2 euros. Subjects bought a chocolate bar if their stated WTP for the bar was at least as high as the random price and if this bar was randomly chosen to be sold in the experiment (which was true only for one of the two bars). If a subject bought the chocolate, he or she paid the random price, not the stated WTP. The mechanism is incentive-compatible for both chocolate bars. The prices and relevant chocolate types that were drawn are given in Table A4 of Online Appendix A3.1.

Two treatments were conducted, one in which the WTPs were stated in private only and another where the participants stated their WTP publicly. The purpose of this variation was to study image concerns and the results are reported in Friedrichsen and Engelmann (2018). Their paper only uses the

¹¹Note that consumers have no incentive to signal that they care about fairness if in fact they do not, unless they want to preserve a positive (self-)image. This is in contrast to other experiments that try to assess the fairness concerns of players such as ultimatum, trust, and gift-exchange games. In these games, signaling typically increases the extent of fair behavior in early periods of repeated games, because the presence of a small share of fair players (or the mere possibility that they exist) makes it possible for selfish players to mimic them. In our experiment, selfish consumers want to signal that they do not care about the worker but only about low prices.

^{© 2024} The Authors. *The Scandinavian Journal of Economics* published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the *SJE*.

data from parts 1 and 3 but not from the market game in part 2. In this paper, we pool the two treatments of part 3 and check the robustness of our results by running separate tests and regressions per treatment.¹²

We chose dark chocolate for part 3 of the experiment instead of milk chocolate as in the first part. Thereby, we try to limit the potential effect that the WTP is reduced for the second bar of the same kind of chocolate. Moreover, we offered unknown brands to prevent subjects from choosing a chocolate only because of its brand and not because of its fair trade label. From the two WTPs that we elicit, we infer the premium that an individual is willing to pay for the chocolate being fair trade as WTP_{premium} = WTP_{fair} – WTP_{conv}.¹³

After entering their WTPs on the computer screen, subjects fill in an extensive questionnaire regarding their attitudes toward and knowledge about fair trade. The answers to this questionnaire allow us to confirm the validity of our proxies for a preference for fair trade.

2.2. Design features and dataset

Three remarks regarding the design are in order. First, if fair trade products are perceived to be of higher quality, those who choose fair trade chocolate might not only be concerned with the production methods, but might also expect a quality difference. This could weaken the correlation with pro-social choices in the market experiment, thus potentially leading us to underestimate the correlation between the behavior in the market game and the concerns for fair trade. Second, note that even if subjects choose fair trade chocolate or inflate their WTP for fair trade in order to impress the experimenters, this

¹²For the investigation of how behavior in the market game relates to the actual choices of one of the two chocolate bars in part 1, pooling the treatments is innocuous because the treatments only begin to differ after the market game. The main results of Friedrichsen and Engelmann (2018) are that the difference between the stated WTP for fair trade and for conventional chocolate was higher in the treatment with public choices than in the treatment with private choices, but this effect is driven exclusively by participants who chose the conventional chocolate in part 1.

¹³We note that while the random-price mechanism (Becker et al., 1964) is incentive-compatible in theory, experimental subjects may misconceive it (Plott and Zeiler, 2005; Cason and Plott, 2014). Such misconceptions arguably do not matter much in our experiment, because we are not concerned with measuring the WTP per se, but with checking whether it correlates with behavior in the experimental market. Hence, misconceptions would only be problematic if they were systematically related to characteristics that drive fair behavior in the market experiment. Moreover, we only analyze the fair trade premium, which is the differences cancel out. Nevertheless, there is noise in the WTPs, as measured by the Becker–DeGroot–Marschak mechanism, and, consequently, in the fair trade premium. To the extent that there is, we might underestimate the true correlation between the fair trade premium and the fairness preferences as exhibited in the market game. Partly due to this noise, though, we also focus on whether the fair trade premium is positive rather than on its absolute size.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

should not bias our main measure of interest upward, namely the relationship of these choices with behavior in a computerized market experiment, unless such experimenter demand is positively related to fair behavior in the market experiment.¹⁴ Using data from de Quidt et al. (2018) who estimate bounds on experimenter demand effects, we find no such correlation (i.e., that more pro-social individuals react more strongly to experimenter demand).¹⁵ Third, we switched from offering one bar of conventional chocolate to two bars for half of the sessions of the original experiment. Thus, our classification of those who chose the fair trade and the conventional product is noisy, which could bias any correlations with choices in the experimental market downwards in the original experiment. There is no such noise in the replication study.

The experiment was computerized using zTree (Fischbacher, 2007). The original experiment took place in the experimental economics laboratory mLab at the University of Mannheim in May, June, and October 2012. Participants were recruited using ORSEE (Greiner, 2015). A translation of the experimental instructions is given in Online Appendix A1.2. We conducted eight sessions with 16–20 participants each, with a total of 144 participants. Among these, 121 subjects chose between fair trade and conventional chocolate via email as described above, while there were 23 newly recruited subjects for whom the chocolate choice was collected in public during a recruitment day and not via email.¹⁶ The 144 participants are matched in groups of four so that

¹⁶We intended to recruit more subjects in this manner, but were not successful. Overall, 222 students received a chocolate bar on the recruitment day, but only 23 showed up to one of our

¹⁴We note that experimenter demand likely matters much less in the market experiment, because tracking an individual's behavior in the market experiment would require the experimenter to follow behavior during the experiment and link it to the individual before handing over the payment. Regarding the choices involving chocolate, the experimenter has to be informed of which type of chocolate to hand over or whether to hand over a chocolate, respectively, and thus has an informative signal about the participant's choices by the time of payment, the social interaction where experimenter demand should have the strongest effect. One could also argue that participants who choose fair trade products in our experiment also behave more fairly in the market experiment because they want to appear consistent. Then, however, our results show at the very least that our participants perceive the situation of the consumer in the market experiment as similar to a fair trade consumption choice, and thus interpret the market experiment as intended. Market experiments that investigate fair consumer behavior implicitly assume this. Hence our experiment at least supports the relevance of stylized market experiments in this respect.

¹⁵We use data from the positive demand frame in Experiment 7 of de Quidt et al. (2018) and only use observations with intermediate contribution levels in the first dictator game (contributions between 0.1 and 0.9, or more restrictively between 0.2 and 0.8) to avoid floor and ceiling effects. When regressing the change in dictator game giving on the original dictator game contribution in the remaining sample of 157 and 148 subjects, respectively, the coefficient of the original contribution is significantly negative. When we use the individual change as a proportion of the individually possible upward change as a dependent variable, the coefficient on the original contribution is positive but far from statistically significant.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

we have 36 independent groups of one consumer, one worker, and two firms each. Our main interest lies in the behavior of the 36 consumers, out of whom 13 had chosen the conventional chocolate bar before the laboratory experiment and 23 had chosen the fair trade chocolate bar. As each group plays the outlined market game for 20 periods, we have 20 observations per consumer or 720 observations at the market-period level. Among the consumers are six newly recruited subjects who made their chocolate choice on the recruitment day instead of via email. Two of these consumers had chosen the conventional chocolate and four the fair trade chocolate. The results are very similar if we restrict the analysis to those subjects who made their choice between chocolate bars via email, as shown in Online Appendix A3.7. For part 2 of the experiment, each participant received a show-up fee of 5 euros; for part 3, everyone received an additional endowment of 4 euros. Average cash earnings were 18.63 euros in total, including the show-up fee and the endowment, subtracting the payments for the chocolate if applicable. In the market game, participants in the role of firms earned 4.50 euros on average, those in the role of workers earned 6.31 euros on average, and those in the role of consumers earned 23.73 euros on average. In contrast to the equilibrium with common knowledge of rationality and selfish players where the consumers earn at least 36 euros, the workers earn 0 euros, and the firms earn at most 2 euros, we observe that both the workers and the firms are better off than predicted while consumers do worse.¹⁷

We conducted a replication of the original experiment in the TU-WZB lab in Berlin between September 2021 and March 2022. This replication study was preregistered at AsPredicted.¹⁸ We collected data for 58 markets (i.e., from 232 participants) in sessions of 8–12 participants, due to COVID-19 restrictions. Out of these participants, 66 had chosen the conventional chocolate bar and 166 had chosen the fair trade one. We are mainly interested in the 58 consumers, 15 of whom had chosen the conventional chocolate bar and 43 the fair trade one. With each group playing 20 rounds of the market game, we collected a total of 1,160 observations for consumers at the market-period level. For part 2 of the experiment, each participant received a show-up fee of 7 euros; for part 3, they received an additional endowment of 4 euros. Average earnings in the market game were 9.50 euros, average total cash earnings were

¹⁸See https://aspredicted.org/tv33u.pdf.

experimental sessions. This explains the low number of subjects in this group. It also indicates that handing out chocolate to motivate students to sign up for experiments is not a very effective recruitment mechanism.

¹⁷In fact, workers on average earn even more than firms. This difference, however, disappears if we exclude profits from rounds where firms erroneously chose wages that exceeded prices (p > 0.9 in a two-sided *t*-test with unequal variances). In Danz et al. (2022), where there is one worker per firm, the workers are worse off than the firms on average.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

20.38 euros, which includes the show-up fee and the endowment, subtracting the payments for the chocolate if applicable. In the market game, average earnings of participants in the role of firms were 5.40 euros, in the role of workers 6.20 euros, and in the role of consumers 21.02 euros. As in the original experiment, both the workers and the firms are better off than predicted by the equilibrium with common knowledge of rationality and selfish players, while consumers do worse.

3. Experimental results

Our first finding for the original dataset is that the two measures of fair trade preferences, the chocolate choice made before the actual laboratory experiment and the willingness to pay a premium for fair trade as compared to conventional chocolate, are highly correlated with each other. The stated fair trade premium in part 3 of the experiment is significantly higher on average for those consumers who chose a fair trade chocolate (average of 30.83 euro cents, SD = 44.96 euro cents, N = 23) than for those who chose a conventional chocolate (average of 2.31 euro cents, SD = 7.25 euro cents, N = 13). This difference is highly significant in a two-sided *t*-test with unequal variances (p = 0.0066) and the distributions are also significantly different according to a Wilcoxon rank-sum test (p = 0.0048) with (*ex post* calculated) power of the mean comparison of 0.8143. Thus, the probability of not rejecting the null when it is false is below 20 percent.

The difference in stated fair trade premiums depending on the chocolate choice is still present in the replication data but smaller and with mixed findings regarding the significance (p = 0.047 in a two-sided *t*-test with unequal variances and p = 0.066 in a Wilcoxon rank-sum test). In this dataset, those who choose conventional chocolate state a substantial fair trade premium of 15.27 euro cents on average, whereas the stated fair trade premium by fair trade choosers of 34.05 euro cents on average is similar to the original data. When we consider the discretized dummy *premium*, which takes on a value of 1 if the premium is strictly positive, we observe a significant positive correlation with the chocolate choice in both the original and the replication data (Spearman's $\rho = 0.4914$, p = 0.0023 in the original data and Spearman's $\rho = 0.2589$, p = 0.0497 in the replication data). See Table 1 for details.¹⁹

We also check whether the WTP for fair trade in the experiment is consistent with such figures in markets outside the laboratory. In the original

¹⁹These results hold up in the entire sample, that is, also if we include subjects who were in the roles of firms or workers in the market game. Details are contained in Table A5 in Online Appendix A3.2.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

	Mean premium (1)	$\begin{array}{l} \text{Premium} \leq 0\\ (2) \end{array}$	$\begin{array}{l} \text{Premium} > 0\\ (3) \end{array}$	Total (4)
Panel A. Original	l experiment			
Conventional	2.31	10	3	13
Fair trade	30.83	6	17	23
Total	20.53	16	20	36
Panel B. Replicat	ion experiment			
Conventional	15.27	8	7	15
Fair trade	34.05	11	32	43
Total	29.19	19	39	58

Table 1. Choices of chocolate type and fair trade premiums by consumers

Notes: Column 1 displays mean values of the individual fair trade *premiums* in euro cents. Columns 2 and 3 state absolute numbers of participants.

experiment, 14 percent of participants who chose the fair trade chocolate and 2.3 percent of those who chose conventional before the laboratory experiment have a WTP that weakly exceeds the market price. In the replication data, a share of 21 percent of fair trade choosers and 13.6 percent of the conventional choosers have a WTP weakly exceeding the market price of the fair trade chocolate. These shares are consistent with the actual market shares of fair trade products. While these are typically in the one-digit range globally, fair trade cocoa's market share has increased to 16 percent in Germany in 2021 (Lebensmittelzeitung, 2022).²⁰

According to the post-experimental questionnaire, participants in the role of consumers on average consider fair trade products to be of slightly higher quality than conventional products with a perceived mean quality difference of 0.42 on a scale from -3 to 3. One might therefore hypothesize that participants who chose fair trade chocolate before the laboratory experiment believe it to be of higher quality and therefore state a higher WTP. This conjecture is not supported in our original data. Quality perceptions as stated in the post-experimental questionnaire are not statistically different between participants who chose fair trade or conventional chocolate (p > 0.6 in a two-sided *t*-test) and are also not statistically different between participants with a positive or a weakly negative fair trade premium (p > 0.3 in a two-sided *t*-test, both with unequal variances). The findings are somewhat different in

²⁰In Germany in 2017, fair trade cocoa achieved a market share of 8 percent, and fair trade coffee one of 4.1 percent making coffee the fair trade product with the highest revenue; the most successful fair trade products in terms of market share were roses (28 percent) and bananas (12 percent) (TransFair e.V., 2018). Globally, cocoa sold with a fair trade label accounts for less than 1 percent of the market (see the webpage of the International Cocoa Organization, https://www.icco.org/chocolate-industry/).

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

the replication data, where consumers who chose fair trade chocolate believe it to be of significantly higher quality than those who chose the conventional chocolate (p = 0.0052 in a two-sided *t*-test). At the same time, we again find that the probability to state a positive premium does not relate to quality perceptions (p > 0.5 in a two-sided *t*-test). If we regress individual fair trade premiums on quality perceptions and the chocolate choice, the coefficient of quality perceptions remains insignificant in both the original and the replication data. We are therefore confident that quality perceptions only play a minor role.

Let us now consider consumer behavior in the market game that subjects played in the laboratory. Given that the worker has no bargaining power in the market game, we call a consumer fair if she buys from the firm with the higher wage and higher price. Note that such behavior leads to a more equal distribution of payoffs on average, as the average wage offered was 3.33, which is below the fair wage of 5, and the average posted price was 8.98 as compared to the fair price of 15. The average realized wage (i.e., taking into account the quantities actually sold by both firms) is 3.25 and thereby significantly higher than the firms' average realized profit margin (price - wage) of 2.32.²¹ We use the same two indicator measures of fair market behavior by consumers that were used by Danz et al. (2022): BuySomeBW (with BW standing for "by wage") is an indicator of the consumer purchasing at least one unit from the firm offering a higher wage and asking for a higher price, i.e., $w_i > w_i$, $p_i > p_i$, and $q_i > 0$, conditional on such a high-wage/high-price offer being available. BuyMoreBW is an indicator of the consumer purchasing more units at the high-wage/high-price firm (if existing), $w_i > w_j$, $p_i > p_j$, and $q_i > q_j$. We also employ a third, continuous measure where we calculate the fraction of units that a consumer bought at the high-wage/high-price firm, denoted by FairShare.²² These three measures allow us to focus on the consumers' choices and to abstract from the absolute levels of wages and prices. Note that alternative measures, such as the Gini coefficient to measure equality

²¹These numbers refer to the original experiment. In the replication data, a similar pattern emerges. The average wage offered was 3.55, the average posted price was 10.20, the average realized wage is 3.30, and the average realized profit margin 2.86.

²²In Online Appendix A3.4, we also consider two additional indicator measures for the sake of the robustness of our analysis, namely BuyMin50BW and BuyAllBW. BuyMin50BW is an indicator of the consumer purchasing at least as many units at the high-wage/high-price firm (if existing) as at the competitor, $w_i > w_j$, $p_i > p_j$, and $q_i \ge q_j$. The results are very similar to those with BuyMoreBW. BuyAllBW is an indicator that takes the value 1 if a consumer buys all units from the firm with the higher wage and price. Although not statistically significant, the results are in line with those presented in the main text. We attribute the absence of a significant effect to the overall low number of cases where a consumer buys all units at the high-wage/high-price firm.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

of payoffs, are affected by firm behavior, and thus less suited to isolate consumers' fairness concerns.

3.1. Fair trade choices and consumer decisions in experimental markets

To address whether fair consumer choices in the abstract market experiment are related to preferences for fair trade products, we begin with relating the behavior in part 2 with the decisions for fair trade chocolate that we elicited in part 1 of the experiment. In each of our following analyses, we start with the original data and then report how the replication data differ. We restrict attention to those instances in which fair behavior is possible (i.e., situations where one of the two firms offers a higher wage and asks for a higher price than its competitor). This is the case in slightly less than half of all interactions in the original dataset. The restriction leaves us with 319 observations, 110 with a consumer who chose conventional chocolate and 209 with a consumer who chose the fair trade chocolate in part 1. Consumers who had chosen the conventional chocolate faced a situation where a high-wage/high-price offer was available on average 8.46 times (SD = 3.13, min = 4, max = 15), and those who had chosen fair trade faced such a situation on average 9.09 times $(SD = 3.60, \min = 3, \max = 16)$. The equality of the two averages cannot be rejected (two-sided *t*-test with unequal variances, p = 0.5900; also, the distributions are not significantly different according to a Wilcoxon rank-sum test, p = 0.6197).²³ On average, the high-wage/high-price offer achieves a market share of 24.41 percent if we consider only observations where such an offer was available, with a similar market share in the replication data. Thus, fair choices matter for the outcome of markets.

As shown in the left-most parts of the three panels of Figure 1 for the original experiment, consumers who chose the fair trade chocolate in the first part were more likely to make a fair choice in the market experiment according to all three of our measures (averages and standard deviations are collected in Table A6 in Online Appendix A3.3). While the number of observations is relatively small when we aggregate the market data to one observation per consumer (i.e., 36 observations), the power of our tests is sufficient for our main measure BuyMoreBW and is high enough for all measures, making a sign error unlikely; see Gelman and Carlin (2014) for a discussion of errors

²³This is similar in the replication data. We observe a high-wage/high-price offer in 137 out of 300 interactions of consumers who chose the conventional chocolate in part 1, and in 458 out of 860 interactions of fair trade choosers. Those who chose conventional chocolate faced a high-wage/high-price situation on average 9.13 times (SD = 3.54, min = 4, max = 16) and those who had chosen the fair trade chocolate faced on average 10.65 such situations (SD = 3.78, min = 1, max = 16). Again, we find no statistically significant differences.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.



Figure 1. Differences of fair purchasing behavior by chocolate choice

Notes: Original experiment: N = 36, 13 chose conventional chocolate, 23 fair trade. Replication experiment: N = 58, 15 chose conventional chocolate, 43 fair trade. The figures show box plots of the consumer-level average of the respective variable, taking into account all situations with a high-wage/high-price firm. The boxes illustrate the 25th to 75th percentile, with the median given by a white line; whiskers mark the lower and upper adjacent values and circles outside observations. If no box is visible, the 75th percentile is zero.

© 2024 The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

in sign and magnitude of an effect.²⁴ In the replication data, the difference in behavior between those choosing fair trade and conventional chocolate measured by BuySomeBW has the same sign but is smaller; see Figure 1(a). The reason for this is that participants who chose conventional chocolate purchase more often from the high-wage/high-price firm in the replication experiment than in the original experiment (see Figure 1 and Table A6 in Online Appendix A3.3).

In the original dataset, those participants who chose fair trade chocolate in part 1 are almost twice as likely to buy some units at the high-wage/high-price firm, BuySomeBW, than those who chose conventional chocolate (54 percent compared to 30 percent). However, the variation is relatively large and this difference is only marginally significant in a two-sided *t*-test with unequal variances (p = 0.0789) and the difference between the distributions is insignificant according to a Wilcoxon rank-sum test (p = 0.1105). The replication data reveal a similar pattern. Those participants who chose the fair trade chocolate in part 1 were more likely to purchase some units at the high-wage/high-price firm than those choosing conventional chocolate but the difference fails to reach significance (54 percent compared to 43 percent, p > 0.33 in both a *t*-test and a Wilcoxon rank-sum test).

When we consider the more stringent measure of fair purchasing, BuyMoreBW, we find larger differences in the original data. Consumers who chose the fair trade chocolate are, on average, five times as likely to purchase more units from the high-wage/high-price firm than consumers who chose the conventional chocolate (about 19 percent compared to about 4 percent). This difference is statistically significant in a two-sided *t*-test with unequal variances (p = 0.0028) and, additionally, the distributions differ according to a Wilcoxon rank-sum test (p = 0.0153). Finally, if we consider the average number of units a consumer bought at the high-wage/high-price firm (FairShare) in the original experiment, we find that, on average, those who chose fair trade in the role of consumers bought 27 percent of their basket at the high-wage/high-price firm, whereas those who chose conventional

²⁴When we split the sample according to the *ex ante* choice of a chocolate bar, the power is highest for the comparison of means of BuyMoreBW with 0.88, indicating that the probability of making a type-2 error is 12 percent. For the other measures, it is lower with 0.42 for BuySomeBW, 0.67 for BuyMin50BW, and 0.66 for FairShare. In light of the discussions about the possible exaggeration bias in statistically significant findings (e.g., Gelman and Carlin, 2014; Ioannidis et al., 2017), we acknowledge that our point estimates might overstate the true effects. Because of the small sample, our original study is not powered to find small effects. The minimal effect size that we are able to detect at a power of 80 percent, and at $\alpha = 0.05$, is d = 1.03 for comparisons between consumers with a weakly negative fair trade premium and those with a positive premium.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

chocolate bought only 11.3 percent at the high-wage/high-price firm in the original data. This difference is also statistically significant (two-sided *t*-test with unequal variances, p = 0.0203; the distributions differ according to a Wilcoxon rank-sum test, p = 0.0193). Surprisingly, moving to the stricter measures of fair purchases, the difference between those choosing fair trade and conventional chocolate vanishes in the replication data, see Figures 1(b) and (c). Both for those choosing fair trade and those opting for the conventional chocolate bar, BuyMoreBW takes a value of close to 14 percent on average. Similarly, for the replication data, the average number of units a consumer bought at the high-wage/high-price firm (FairShare) is only slightly larger for those who chose the fair trade chocolate with 24 percent on average as compared to 19.7 percent for those who chose conventional chocolate, and this difference fails to reach significance (p > 0.47 in either test).

We estimate linear probability models with random effects and standard errors clustered at the subject level, reported in Table 2, to investigate the relationship between fair consumer choices in the market experiment and the choice of a chocolate bar in part 1. These analyses take into account the fact that the data contain multiple observations per consumer and allow us to control for the costs and benefits of purchasing at the more expensive firm by including price and wage differences as controls. We construct a dataset that contains all market interactions in which one of the two offers comprised the higher price and the higher wage.

In the original experiment (see Panel A in Table 2), the propensity to buy some units from the high-wage/high-price firm, BuySomeBW, is 25 percentage points higher if the subject chose fair trade chocolate rather than conventional chocolate in part 1. The effect becomes only slightly smaller when we control for wage and price differences. Similarly, a subject is about 17 percentage points more likely to buy more units from the high-wage/high-price firm, BuyMoreBW, if the subject chose fair trade chocolate, and the effect is again only slightly smaller when we include controls.²⁵

These results do not replicate in our second study as can be seen in Panel B of Table 2. The dummy variable FT, indicating whether a participant chose the fair trade chocolate bar in the first part of the experiment does not significantly relate to our indicators of fair purchasing behavior in the replication study. We note also that the overall R^2 is much lower in the regressions using the

²⁵We find similar results if we include interaction effects or estimate a linear probability model without random effects (see Tables A10 and A11 in Online Appendices A3.5 and A3.6). If we run a fixed-effects regression instead of the random-effects specification, and omit the dummy for the subjects' chocolate choice, we find that the decision in favor of the fair trade chocolate bar correlates significantly positively with the estimated individual fixed effect for both dependent variables, BuySomeBW and BuyMoreBW, among our 36 participants (p < 0.01; results not reported).

^{© 2024} The Authors. *The Scandinavian Journal of Economics* published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the *SJE*.

	BuySon	neBW	BuySom	eBW	BuyMo	reBW	BuyMo	oreBW
	Coef.	<i>p</i> -value	Coef.	<i>p</i> -value	Coef.	<i>p</i> -value	Coef.	<i>p</i> -value
Panel A	. Original e	xperiment						
FT	0.2503*	0.0461	0.2409*	0.0484	0.1724***	0.0003	0.1649***	0.0004
Δp			-0.0260**	0.0076			-0.0158^{*}	0.0136
Δw			0.0022	0.8253			0.0159*	0.0288
Const.	0.3020**	0.0030	0.4018***	0.0001	0.0362	0.0555	0.0536*	0.0302
Ν	319		319		319		319	
R^2 overall	0.0970		0.1484		0.0577		0.0827	
Panel B	. Replication	n experime	ent					
FT	0.1130	0.3341	0.1241	0.2792	0.0032	0.9542	0.0080	0.8831
Δp			-0.0243^{***}	0.0002			-0.0193***	0.0001
Δw			0.0177^{+}	0.0581			0.0206*	0.0103
Const.	0.4322***	0.0000	0.4722***	0.0000	0.1399**	0.0022	0.1553***	0.0009
Ν	595		595		595		595	
R^2 overall	0.0080		0.0790		0.0000		0.0420	
Panel C	. Original a	nd replicat	tion experimer	nt pooled				
FT	0.1745*	0.0427	0.1797*	0.0321	0.0720^{+}	0.0589	0.0720^{+}	0.0537
Δp			-0.0242^{***}	0.0000			-0.0182***	0.0000
Δw			0.0101	0.1315			0.0188***	0.0006
Const.	0.3722***	0.0000	0.4351***	0.0000	0.0926**	0.0011	0.1097***	0.0003
Ν	914		914		914		914	
R^2 overall	0.0301		0.0968		0.0084		0.0458	

 Table 2.
 Probability of fair consumer behavior measured by BuySomeBW and BuyMoreBW regressed on market characteristics and chocolate choice in part 1

Notes: Linear probability model with individual-level random effects and robust standard errors clustered on subject. FT denotes choice of fair trade chocolate in part 1 of the experiment. Δw and Δp denote the difference in wages and prices between the two firms. $^+p < 0.1$; $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

replication data than in the original experiment. Panel C of Table 2 provides results from analyses using the pooled data from both experiments. Subjects choosing the fair trade chocolate are significantly more likely to buy some or the majority of their units at the high-wage/high-price firm but the effect is smaller than in the original experiment, in particular for BuyMoreBW.

As the dummy variables do not take into account all information available about the relative weight a consumer puts on purchases at the high-wage/high-price firm, we also analyze purchasing behavior based on the variable FairShare. Note that this variable is missing if a consumer does not purchase anything, which happens three times in our original dataset and four times in the replication. Panel A of Table 3 shows that those who chose the fair trade chocolate bar in the first part of the original experiment purchased a significantly larger fraction of their basket (about 16 percentage points more) at the high-wage/high-price firm. Again, these results do not replicate in our

© 2024 The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

	FairSh	FairShare		FairShare		
	Coef.	<i>p</i> -value	Coef.	<i>p</i> -value		
Panel A. Or	iginal experiment					
FT	0.1633**	0.0088	0.1575**	0.0092		
Δp			-0.0161**	0.0057		
Δw			0.0090	0.1332		
Const.	0.1147*	0.0101	0.1531***	0.0008		
Ν	316		316			
R^2	0.0881		0.1398			
Panel B. Rej	olication experiment					
FT	0.0382	0.5187	0.0455	0.4274		
Δp			-0.0195***	0.0000		
Δw			0.0146*	0.0157		
Const.	0.2006***	0.0001	0.2318***	0.0000		
Ν	591		591			
R^2 overall	0.0019		0.0848			
Panel C. Or	iginal and replication	experiment pool	ed			
FT	0.0913*	0.0341	0.0932*	0.0257		
Δp			-0.0182^{***}	0.0000		
Δw			0.0121**	0.0049		
Const.	0.1609***	0.0000	0.1964***	0.0000		
Ν	907		907			
R^2 overall	0.0190		0.0925			

Table 3. Fair consumer behavior measured by FairShare regressed on market characteristics and chocolate choice in part 1

Notes: Random-effects regression with robust standard errors clustered on subject. FT denotes choice of fair trade chocolate in part 1 of the experiment. Δw and Δp denote the difference in wages and prices between the two firms. ${}^{+}p < 0.1$; ${}^{*}p < 0.05$; ${}^{**}p < 0.01$; ${}^{**}p < 0.001$.

second experiment, as shown in Panel B of Table 3 but in the pooled analyses of both studies the effect is significant with a smaller point estimate (and weaker significance) than in the original study.²⁶

In all regressions, we find that consumers react to the costs of being fair because all our measures of fair behavior decrease significantly in the price difference between the firms. They react less systematically to the benefits of their actions. While the coefficient for the wage difference between firms is positive in all regressions, only BuyMoreBW increases significantly with the wage difference (and FairShare in the replication data).

 $^{^{26}}$ In Table A12 in Online Appendix A3.6, we show that these results are essentially the same when we use an ordinary least-squares regression without random effects instead of the random-effects specification.

^{© 2024} The Authors. *The Scandinavian Journal of Economics* published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the *SJE*.

3.2. Fair trade premium and consumer decisions in experimental markets

Next, we investigate the relationship between the choices in the experimental market and the fair trade premium elicited in part 3 of the experiment. Do participants who are willing to pay more for a fair trade product buy (more units) from the high-wage/high-price firm? We dichotomize the stated fair trade premium to the indicator *premium*, because the continuous values are noisy measures of fair trade preferences due to the treatment effect in the public and private condition.²⁷ Again, we restrict attention to those instances in which one of the two firms offered a higher wage and asked for a higher price than its competitor. The number of such instances does not differ significantly for situations where consumers display a positive or a zero/negative fair trade premium.²⁸

As shown in Figure 2, consumers who stated a positive fair trade premium are more likely to make fair choices in the market experiment according to all three of our measures in the original data. The figure displays the market choices of the consumers, differentiated by their stated fair trade premium in part 3 (the relevant averages and standard deviations are shown in Table A7 in Online Appendix A3.3). Participants with a positive premium are more than twice as likely to buy some units at a high-wage/high-price firm (61 percent compared to 26 percent). This difference is significant in a two-sided *t*-test with unequal variances (p = 0.0041) and the distributions differ significantly according to a Wilcoxon rank-sum test (p = 0.0057). The indicator BuyMoreBW shows even larger differences. Consumers with a positive fair trade premium are substantially more likely to have purchased

 $^{^{27}}$ Friedrichsen and Engelmann (2018) show that the fair trade premiums of those who chose conventional chocolate are higher in the public than in the private treatment. In effect, the average fair trade premium in public does not differ between individuals who chose the fair trade chocolate and those who chose the conventional chocolate in part 1 of the experiment. In Online Appendix A3.9, we analyze the continuous fair trade premiums in more detail. The results are in the same direction, but of weaker significance due to the noise in the continuous measure. In the replication experiment, behavior appears to be noisier, which might explain why the continuous premium again performs no better than the dichotomized premium (see discussion in Section 3.4).

²⁸In the original experiment, consumers whose fair trade premium is positive faced a high-wage/high-price offer 9.6 times on average, SD = 3.44, min = 5, max = 16 (in the replication experiment: mean= 10.62, SD = 3.39, min = 1, max = 16), those whose stated fair trade premium was zero or negative faced such an offer on average 7.94 times, SD = 3.24, min = 3, max = 15 (in the replication experiment: mean = 9.53, SD = 3.39, min = 1, max = 16). The two averages are not statistically different at conventional levels (two-sided *t*-test with unequal variances, p = 0.1458 (original) and p = 0.3501 (replication); also the distributions do not differ significantly according to a Wilcoxon rank-sum test, p = 0.1456 (original) and p = 0.2863 (replication)).

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.



Figure 2. Differences in fair purchasing behavior by premium

Notes: Original experiment: N = 36, 20 with a fair trade premium > 0, 16 with a fair trade premium ≤ 0 . Replication experiment: N = 58, 39 with a fair trade premium > 0, 19 with a fair trade premium ≤ 0 . The figures show box plots of the consumer-level average of the respective variable, taking into account all situations with a high-wage/high-price firm. The boxes illustrate the 25th to 75th percentile, with the median given by a white line; whiskers mark the lower and upper adjacent values and circles outside observations. If no box is visible, the 75th percentile is zero.

© 2024 The Authors. *The Scandinavian Journal of Economics* published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the *SJE*.

more units from the high-wage/high-price firm than those with a zero or negative premium (about 22 percent compared to about 3 percent). This difference is highly statistically significant in a two-sided *t*-test with unequal variances (p = 0.0014) as is the difference between distributions according to a Wilcoxon rank-sum test (p = 0.0006). Finally, if we take into account how many units a consumer bought at the high-wage/high-price firm and average this over all available instances (FairShare), we find that consumers with a positive premium bought on average 30 percent of their basket at the high-wage/high-price firm, whereas those with a zero or negative premium bought only 10 percent at the high-wage/high-price firm. This difference is also highly statistically significant (two-sided *t*-test with unequal variances, p = 0.0027; also the distributions differ significantly according to a Wilcoxon rank-sum test, p = 0.0021).²⁹

In the replication experiment, the correlation between the premium dummy and the indicators for fair market behavior has the same sign but is much smaller. This can be taken from Figure 2. Averages and standard deviations are collected in Table A7 in Online Appendix A3.3). Participants with a positive premium are more likely to buy some units at the high-wage/high-price firm (58 percent compared to 38 percent) and the difference is marginally significant at the 10 percent level (p = 0.0699 in a two-sided *t*-test with unequal variances and p = 0.0705 in a Wilcoxon rank-sum test). The indicator BuyMoreBW, however, hardly differs between the two groups; consumers with a positive fair trade premium are only slightly more likely to purchase more units from the high-wage/high-price firm than those with a zero or negative premium (15 percent compared to 13 percent), and we cannot reject that they are equal (p = 0.7347) in a two-sided *t*-test with unequal variances and p = 0.5583 in a Wilcoxon rank-sum test). For FairShare, the difference is slightly larger; consumers with a positive premium bought on average 26 percent of their basket at the high-wage/high-price firm, whereas those with a zero or negative premium bought only 17 percent at the high-wage/high-price firm. But this difference is again smaller than in the original data and fails to reach statistical significance (two-sided *t*-test with unequal variances, p = 0.1509; Wilcoxon rank-sum test, p = 0.1105).

We again estimate a linear probability model with random effects and standard errors clustered at the level of subjects. This allows us to study the relationship between the consumer choices in the market experiment and the fair trade premium in more detail. Using a dataset that contains all market

²⁹When we split the sample according to the incidence of a positive fair trade premium, the power is highest for the comparison of means of BuyMoreBW with 0.93. It is 0.74 for BuySomeBW, and 0.88 for FairShare.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

interactions in which one of the two offers comprised the higher price and the higher wage, we regress the measures BuySomeBW and BuyMoreBW on a dummy for a positive fair trade premium (*premium*) and market characteristics. From Panel A of Table 4, it can be taken that the coefficient of the *premium* dummy is significantly different from zero and economically relevant in size for both measures, BuySomeBW and BuyMoreBW, in the original experiment. The incidence of BuySomeBW is about 34 percentage points higher on average for a consumer with a positive fair trade premium and the incidence of BuyMoreBW is about 17 percentage points higher than for a consumer with a weakly negative fair trade premium.

Overall, we find that in the original experiment, a positive premium is associated with a higher propensity to buy from the high-wage/high-price firm. This relationship has the same direction but is not significant in the replication

	BuySo	meBW	BuySon	neBW	BuyMo	reBW	BuyMo	reBW
	Coef.	<i>p</i> -value	Coef.	<i>p</i> -value	Coef.	p-value	Coef.	<i>p</i> -value
Panel A.	Original ex	periment						
Premium	0.3491**	0.0022	0.3369**	0.0025	0.1839***	0.0007	0.1739**	0.0011
Δp			-0.0258^{**}	0.0099			-0.0156^{*}	0.0224
Δw			0.0019	0.8534			0.0153*	0.0469
Const.	0.2670**	0.0019	0.3677***	0.0001	0.0415	0.1615	0.0609^{+}	0.0713
Ν	319		319		319		319	
R^2 overall	0.1153		0.1667		0.0600		0.0844	
Panel B.	Replication	experime	ent					
Premium	0.1849+	0.0840	0.1643	0.1193	0.0163	0.7866	0.0062	0.9169
Δp			-0.0241***	0.0002			-0.0192***	0.0000
Δw			0.0179	0.0549			0.0206*	0.0105
Const.	0.3908***	0.0000	0.4516***	0.0000	0.1311*	0.0130	0.1568**	0.0054
Ν	595		595		595		595	
R^2 overall	0.0236		0.0860		0.0015		0.0423	
Panel C.	Original an	d replicat	tion experim	ent pooled	l			
Premium	0.2546**	0.0012	0.2362**	0.0022	0.0827*	0.0446	0.0733+	0.0693
Δp			-0.0239***	0.0000			-0.0179***	0.0000
Δw			0.0102	0.1295			0.0189***	0.0006
Const.	0.3340***	0.0000	0.4106***	0.0000	0.0905**	0.0058	0.1121**	0.0016
Ν	914		914		914		914	
R^2 overall	0.0492		0.1050		0.0128		0.0481	

Table 4. Probability of fair consumer behavior measured by BuySomeBW and

 BuyMoreBW regressed on market characteristics and an indicator of a positive fair trade

 premium in part 3

Notes: Linear probability model with individual-level random effects and robust standard errors clustered on subject. Δw and Δp denote the difference in wages and prices between the two firms. p < 0.1; p < 0.05; p < 0.01; p < 0.001.

© 2024 The Authors. *The Scandinavian Journal of Economics* published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the *SJE*.

experiment, for which we present the results in Panel B of Table $4.^{30}$ As can be seen in Panel C of Table 4, if we pool the original experiment and the replication, the correlations remain significant (for BuyMoreBW only at the 10 percent level if we include controls for price and wage differences).

As the WTP in the third part of the original experiment was announced privately or in public, depending on the treatment, we also conduct the analysis for both treatments separately. We find that our findings hold in both the private and the public treatment and, moreover, the estimated coefficients for the fair trade premium are nearly identical in both treatments (see Tables A25 and A26 in Online Appendix A3.10). The replication experiment did not include this treatment variation.

We also analyze whether purchasing behavior based on the variable FairShare correlates with the stated fair trade premium in part 3 of the experiment. Panel A of Table 5 shows that in the original experiment those with a positive fair trade premium purchase a significantly larger fraction of their basket (about 18 percentage points more) at the high-wage/high-price firm than those with a weakly negative fair trade premium. As with the indicators discussed above, the correlations are weaker and not significant in the replication experiment, as shown in Panel B of Table 5.³¹ Again, if we pool the original experiment and the replication, the correlation remains significant (see Panel C of Table 5).

3.3. Attitudes toward fair trade

After the experiment, we administered a questionnaire to learn about individual attitudes, knowledge, and motivations with respect to fair trade. The questionnaire contained 25 statements for which the participants had to indicate their level of agreement.³² The questionnaire is designed to cover the most important arguments in favor of and against supporting fair trade, important facts about the fair trade principles, and potential purchasing motivations.

³⁰The results are very similar and remain significant in the original experiment whereas they remain insignificant in the replication experiment, when we include interaction effects or use a linear probability model without random effects (see Tables A10 and A13 in Online Appendices A3.5 and A3.6). We also run a fixed-effects regression and omit the dummy for a positive premium. We find that having a positive fair trade premium correlates significantly positively with the estimated individual fixed effect both for BuySomeBW and for BuyMoreBW in our original sample of 36 consumers (p = 0.001; results not reported).

³¹In Table A14 in Online Appendix A3.6, we show that these results are essentially the same when we use ordinary least-squares regressions instead of the random-effects specification but the point estimate for the original experiment is slightly smaller in the OLS. Furthermore, in Table A27 we conduct the random-effects analysis separately for each treatment of the original experiment. Again, our results hold for both treatments.

³²A translation of the questionnaire is contained in Online Appendix A1.3.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

	FairShare		FairS	FairShare	
	Coef.	<i>p</i> -value	Coef.	<i>p</i> -value	
Panel A. Orig	inal experiment				
Premium	0.1939**	0.0016	0.1849**	0.0021	
Δp			-0.0159**	0.0084	
Δw			0.0087	0.1624	
Const.	0.1103**	0.0094	0.1504***	0.0004	
Ν	316		316		
R^2 overall	0.0887		0.1383		
Panel B. Repli	ication experiment				
Premium	0.0721	0.2036	0.0563	0.3124	
Δp			-0.0194***	0.0000	
Δw			0.0147^{*}	0.0153	
Const.	0.1800***	0.0001	0.2266***	0.0000	
Ν	591		591		
R^2 overall	0.0134		0.0913		
Panel C. Orig	inal and replication	experiment pooled			
Premium	0.1214**	0.0034	0.1088**	0.0070	
Δp			-0.0180^{***}	0.0000	
Δw			0.0121**	0.0051	
Const.	0.1481***	0.0000	0.1918***	0.0000	
Ν	907		907		
R^2 overall	0.0328		0.0988		

Table 5. Fair consumer behavior measured by FairShare regressed on market characteristics and an indicator of a positive fair trade premium in part 3

Notes: Random-effects regression with robust standard errors clustered on subject. *Premium* is an indicator of a positive fair trade premium in part 3. Δw and Δp denote the difference in wages and prices between the two firms. $^+p < 0.1$; $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

We find that choices observed in part 1 of the experiment reflect the stated purchasing behavior and attitudes. Those participants who chose fair trade chocolate report to buy fair trade products more frequently and in the aggregate reveal a more positive attitude regarding fair trade (see Figure A2 in Online Appendix A2.1).

Next, we use an exploratory factor analysis to determine linear combinations of the original questionnaire statements that can be used to summarize the response behavior with fewer variables without losing much information. We conduct the factor analyses separately for the original and the replication experiment. For both, it turns out that the questionnaire data can be explained by three latent factors.³³

³³The factor analysis finds a set of q common factors such that linear combinations of the q factors reconstruct the p original variables. The coefficients of the factors in the linear

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

Factor 1 loads on attitudes toward fair trade and its effect on farmers and the market. In our interpretation, it measures the attitude that fair trade is an effective tool to do good and to transfer money to the poor. We see differences in that the ban on chemicals and GMO played a role for this factor in the original but not in the replication experiment, whereas in the latter prohibitively high prices of fair trade appeared more important. We label factor 1 "FT helps". A second factor (Factor 2 in the original experiment and Factor 3 in the replication experiment) is called "social pressure", because it characterizes individuals who report purchasing fair trade products not out of intrinsic interest but because of image concerns or in order to conform with the wishes of family or friends. In the replication experiment, this factor also loads on stated purchasing behavior, which suggests that these social motives have gained in importance. The third factor (Factor 3 in the original experiment and Factor 2 in the replication experiment) mostly loads on questions that do not represent value statements but agreement with a relatively narrow interpretation of fair trade as a system that is focused on above-market level prices and pre-financing of harvests. We call this factor "fair trade knowledge".³⁴ Even though world market prices for cocoa were similar in 2022 and 2012, the idea that producer prices are generally too low has lost importance in the replication experiment.

We investigate how these factors correlate with the behavior in the market game, the chocolate choice, and the stated fair trade premium. The detailed results from the regressions with the set of attitudes are collected in Online Appendix A3.8. These analyses help us to understand the results from the original experiment as well as the differences in the replication experiment.

The attitudes measured by our questionnaire support the interpretation that both the fair trade choice and the stated premium relate to real fair trade preferences in the original experiment. There, the decision in favor of the fair trade chocolate bar correlates positively with the belief that purchasing fair trade products is a good thing (FT helps) although this correlation only becomes marginally significant if we exclude six newly recruited consumer participants (see Tables A18 and A19). The fair trade premium also correlates

combination are called factor loadings. We employed a maximum-likelihood factor analysis and selected the model with three factors because it had a lower value on Schwarz's BIC than alternative models with fewer or more factors. Our interpretation of the factors relies on a varimax rotation. Details are provided in Online Appendix A2.3.

 $^{^{34}}$ Specifically, this factor loads positively on Item 12, "I consider the following criteria of fair trade [not important at all (-3) to particularly important (+3)]: (b) improvement of working conditions, (c) fair trade minimum price, (d) fair trade premium on top of minimum price, and (e) prefinancing of harvest", which are all important factual aspects of the fair trade system. The full questionnaire is included in Online Appendix A1.3 and details on the factor loadings in both experiments can be found in Table A3.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

positively and significantly with FT helps (see Tables A19 and A20). In contrast, being influenced by image concerns and peers (social pressure) does not significantly correlate with the fair trade choice or the stated fair trade premium, and neither does the factor related to fair trade knowledge. This is consistent with the observation that factor 2 (social pressure) mostly reflects social motivations to purchase fair trade and factor 3 (fair trade knowledge) captures knowledge about fair trade. Neither of the two must necessarily be correlated with intrinsic preferences for or against fair trade, which are rather associated with factor 1 (FT helps). In the replication experiment, we observe a stronger correlation between the choice of the fair trade chocolate bar and the factor FT helps but a weaker relation of the same factor with the willingness to pay a premium for fair trade chocolate than in the original experiment. None of the other two factors is significantly related to the chocolate choice or the stated premium. Details are contained in Tables A18, A19, and A20.

Regarding decisions in the experimental market, the analysis reveals that those consumers who believe in fair trade as an effective tool to help farmers (high values of FT helps) are more likely to purchase more than 50 percent of their basket at a high-wage/high-price firm in the market game of the original experiment. We do not find this relationship in the replication experiment, where we instead see that FT helps tends to relate to the likelihood to purchase at least some units at the high-wage/high-price firm (p = 0.06)(see Table A21). In addition, we find that the factor fair trade knowledge, which captures the association of fair trade with above market prices and not with judgements regarding its function to help farmers for example, correlates negatively with the probability to buy at least some units from the high-wage/high-price firm (see Table A21) and negatively with the share purchased at the high-wage/high-price firm (FairShare; see Table A22) in the original experiment but not in the replication experiment. These findings support the notion that behavior in the abstract market experiment relates to attitudes and behavior regarding fair trade in markets outside of the laboratory in the original and the replication experiment, though less so in the latter.

3.4. Differences between the original and the replication experiment

While the results of the replication mostly go in the same direction as in the original experiment, they are weaker and often not statistically significant. One possible reason for this lies in differences between the two experiments. We attempted to make them as similar as possible but some changes were unavoidable, and we cannot rule out that they affected the results. First, we employ only organic chocolate in the replication experiment (one is fair trade, the other conventional), which was not the case in the original study.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

The reason is that it has become virtually impossible in Germany to buy chocolate that is neither organic nor fair trade.³⁵ This change in the market is also reflected in people's stated purchasing behavior of fair trade products. According to consumer surveys, only 8 percent of consumers regularly purchased fair trade products in 2013 but 25 percent did so in 2020; the share of those purchasing fair trade products at least sometimes rose from 44 percent to 70 percent in the same period (Forum Fairer Handel, 2022). Differences in the answers to our fair trade questionnaire are consistent with the idea that fair trade was perceived differently in 2012 and 2022.³⁶ Moreover, the fact that the chocolate bars are labeled as organic may influence the willingness to pay a premium for the chocolate being fair trade. Due to moral licensing as observed by Engel and Szech (2020) for the WTP for organic production and social certification of towels, it is possible that the fair trade premium is lower for organic chocolate. However, as illustrated by Figure 3, the premiums in the replication experiment are not lower than in the original experiment. Among participants in the role of consumers, the difference between the original and replication experiment is particularly strong for those choosing conventional chocolate: whereas only about a fifth of them were willing to pay a positive but very small premium for fair trade in the original experiment, about a quarter to a third state a substantial premium in the replication experiment. We also note that while a zero premium is common, very few participants state a negative premium. As there is substantial variation in the positive premium, we focus on whether participants state a non-positive (typically zero) or positive premium, as captured by our premium indicator.

We further note that the replication experiment was conducted at a different location from the original experiment and at a time when the lab procedures were still affected by the COVID-19 pandemic, such that sessions had to be smaller. We cannot rule out that participants in the replication experiment are different from those in the original experiment, for example, due to selection effects; participants in the role of consumers displayed more irrational purchases and more self-interested purchasing decisions in the replication experiment than in the original experiment, with the consequence that consumer and worker profits are lower in the replication than in the original data. A final difference is that among the relatively large group of those choosing fair trade in the replication experiment, 12 percent apparently were not interested in obtaining a bar of chocolate or had a rather high marginal valuation of money as they stated a WTP of less than 2 euro cents for

³⁵While not every non-organic chocolate carries the fair trade label, alternative labels that promise varying extents of social benefits to the cocoa farmers have become ubiquitous.

 $^{^{36}}$ Online Appendices A2.2 and A2.3 contain details on how questionnaire answers and the inferred attitudes differ between the original and the replication experiment.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

Figure 3. Cumulative distribution functions of stated fair trade premiums by participants in the consumer role in original and replication experiments



(b) Premiums of fair trade choosers

Notes: In the original (replication) experiment, 13 (15) consumers chose the conventional chocolate bar and 23 (43) the fair trade one.

each type of chocolate (in the following called "no demand"). In the original experiment, only 4 percent of those choosing fair trade were no demand types. For those choosing conventional chocolate, we observe the opposite with 23 percent being classified as no demand in the original experiment versus no one in the replication experiment.

While gender is often discussed as a potential correlate of pro-social behavior, we confirm that the results of the original and the replication

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

experiment are not driven by one gender alone or by differences between genders. We checked this by running our main analyses separately for male and female consumers (see Tables A28 and A29 in the Online Appendix for details).

4. Discussion

Our results yield mixed evidence on the question whether a stylized market experiment is suitable for measuring preferences regarding consumption goods with externalities on third parties. The original experiment suggested a significant positive correlation between lab behavior and actual consumption choices captured by fair trade choices before the experiment and a positive willingness to pay for a fair trade over a conventional product. In the replication study, these correlations are much weaker and mostly not significant although the signs are identical to the original study. One possible reason is that fair trade has become more mainstream over time, which might weaken the correlation between pro-sociality in the market game and the purchase of fair trade products. Nevertheless, in both the original and the replication data, the subjects' answers to a post-experimental questionnaire support the hypothesis that all three choices reflect fair trade preferences in the field.

There are several possible mechanisms why the correlations between choices in the abstract market and actual product choices are weak in our experiments. First, the social distance among the decision-makers and the affected parties differ substantially in the market experiment and the fair trade choices. In the market experiment, the affected party (the worker) is another participant in the laboratory and thus physically proximate as well as likely comparable across several demographic variables. In the fair trade choices, the affected parties (actual workers in the coccoa production) are both physically and socially remote. Second, whether wages are fair or not is much more transparent in the market experiment, because participants are informed about them. Third, interaction in the market experiment is repeated, which is not the case for the other two tasks. In repeated games, strategic concerns can play a larger role. However, we do not find a significant time trend in the willingness to purchase from the high-wage/high-price firm in our market experiment.

Based on our findings, we cannot resolve the controversy regarding the relevance of pro-social behavior in the laboratory for behavior in the field (Levitt and List, 2007; Camerer, 2015). At the same time, the results do not undermine the relevance of the lively debate on markets and morals fueled by experimental evidence (Falk and Szech, 2013; Bartling et al., 2015; Breyer and Weimann, 2015; Pigors and Rockenbach, 2016; Sutter et al., 2020). Overall, our failure to replicate the findings of the original study demonstrates the need to employ a variety of experiments with many designs as well as replications

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

to investigate research questions such as the relationship between fairness in the lab and in the field.

Acknowledgments

We thank Björn Bartling, Alexander Cappelen, Johanna Möllerström, Lisa Spantig, Bertil Tungodden, and especially Roel van Veldhuizen for helpful comments and suggestions. Jana Hamdan and Andrés Alberto Pérez Asiain provided excellent research assistance, and Adam Lederer language editing. Support by the German Science Foundation (DFG) through CRC TRR 190 (project number 280092119) is gratefully acknowledged. D. Kübler thanks the Cluster of Excellence SCRIPTS. J. Friedrichsen acknowledges financial support from the Leibniz Association through the project SAW-2015-DIW-4. Open Access funding enabled and organized by Projekt DEAL.

Supporting information

Additional supporting information can be found online in the supporting information section at the end of the article.

Online appendix Replication package

References

- Andorfer, V. A. and Liebe, U. (2012), Research on fair trade consumption—a review, *Journal* of Business Ethics 106, 415–435.
- Bartling, B., Weber, R. A., and Yao, L. (2015), Do markets erode social responsibility?, *Quarterly Journal of Economics 130*, 219–266.
- Becker, G., DeGroot, M., and Marschak, J. (1964), Measuring utility by a single-response sequential method, *Behavioral Science* 9, 226–232.
- Benz, M. and Meier, S. (2008), Do people behave in experiments as in the field? Evidence from donations, *Experimental Economics* 11, 286–281.
- Breyer, F. and Weimann, J. (2015), Of morals, markets and mice: be careful drawing policy conclusions from experimental findings!, *European Journal of Political Economy* 40, 387–390.
- Camerer, C. F. (2015), Chapter 14 The promise and success of lab–field generalizability in experimental economics: a critical reply to Levitt and List, in G. R. Fréchette and A. Schotter (eds), *Handbook of Experimental Economic Methodology*, Oxford University Press, Oxford, 249–295.
- Carlsson, F., Johansson-Stenman, O., and Nam, P. K. (2014), Social preferences are stable over long periods of time, *Journal of Public Economics 117*, 104–114.
- Carrington, M. J., Neville, B. A., and Whitwell, G. J. (2010), Why ethical consumers don't walk their talk: towards a framework for understanding the gap between the ethical purchase intentions and actual buying behaviour of ethically minded consumers, *Journal of Business Ethics* 97, 139–158.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

- Cason, T. N. and Plott, C. R. (2014), Misconceptions and game form recognition: challenges to theories of revealed preference and framing, *Journal of Political Economy* 122, 1235–1270.
- Danz, D., Engelmann, D., and Kübler, D. (2022), Do legal standards affect ethical concerns of consumers?, *European Economic Review 144*, 104044.
- De Pelsmacker, P., Driesen, L., and Rayp, G. (2005), Do consumers care about ethics? Willingness to pay for fair-trade coffee, *Journal of Consumer Affairs* 39, 363–385.
- de Quidt, J., Haushofer, J., and Roth, C. (2018), Measuring and bounding experimenter demand, American Economic Review 108 (11), 3266–3302.
- Engel, J. and Szech, N. (2020), A little good is good enough: ethical consumption, cheap excuses, and moral self-licensing, *PloS One 15*, e0227036.
- Etilé, F. and Teyssier, S. (2016), Signaling corporate social responsibility: third-party certification versus brands, *Scandinavian Journal of Economics 118*, 397–432.
- Falk, A. and Szech, N. (2013), Morals and markets, Science 340, 707-711.
- Fehr, E. and Leibbrandt, A. (2011), A field study on cooperativeness and impatience in the tragedy of the commons, *Journal of Public Economics 95*, 1144–1155.
- Fehr, E. and Schmidt, K. M. (2006), Chapter 8 The economics of fairness, reciprocity and altruism–experimental evidence and new theories, in S.-C. Kolm and J. M. Ythier (eds), *Handbook of the Economics of Giving, Altruism and Reciprocity*. Vol. 1, North-Holland, Amsterdam, 615–691.
- Feicht, R., Grimm, V., and Seebauer, M. (2016), An experimental study of corporate social responsibility through charitable giving in Bertrand markets, *Journal of Economic Behavior* & Organization 124, 88–101.
- Fischbacher, U. (2007), Zurich toolbox for ready-made economic experimental *Economics 10*, 171–178.
- Forum Fairer Handel (2022), Zahlen und Fakten zum Fairen Handel in Deutschland, https://www.forum-fairer-handel.de/fairer-handel/zahlen-und-fakten-zum-fairen-handel-indeutschland/verbraucherbefragung.
- Franzen, A. and Pointner, S. (2013), The external validity of giving in the dictator game, *Experimental Economics* 16, 155–169.
- Friedrichsen, J. and Engelmann, D. (2018), Who cares about social image?, European Economic Review 110, 61–77.
- Galizzi, M. M. and Navarro-Martínez, D. (2018), On the external validity of social preference games: a systematic lab-field study, *Management Science* 65, 976–1002.
- Gelman, A. and Carlin, J. (2014), Beyond power calculations: assessing type s (sign) and type m (magnitude) errors, *Perspectives on Psychological Science* 9, 641–651.
- Greiner, B. (2015), Subject pool recruitment procedures: organizing experiments with ORSEE, *Journal of the Economic Science Association 1*, 114–125.
- Hainmueller, J., Hiscox, M. J., and Sequeira, S. (2015), Consumer demand for fair trade: evidence from a multistore experiment, *Review of Economics and Statistics* 96, 402–417.
- Harrs, S., Rockenbach, B., and Wenner, L. M. (2022), Revealing good deeds: disclosure of social responsibility in competitive markets, *Experimental Economics* 25, 1349–1373.
- Ioannidis, J. P. A., Stanley, T. D., and Doucouliagos, H. (2017), The power of bias in economics research, *Economic Journal* 127, F236–F265.
- Irlenbusch, B. and Saxler, D. J. (2019), The role of social information, market framing, and diffusion of responsibility as determinants of socially responsible behavior, *Journal of Behavioral and Experimental Economics* 80, 141–161.
- Kirchler, M., Huber, J., Stefan, M., and Sutter, M. (2016), Market design and moral behavior, Management Science 62, 2615–2625.
- Lebensmittelzeitung (2022), Fairer Handel: Kakaoprodukte mit dem Fairtrade-Siegel legen deutlich zu, https://www.lebensmittelzeitung.net/handel/nachrichten/fairer-handel -kakaoprodukte-mit-dem-fairtrade-siegel-legen-deutlich-zu-163555.

^{© 2024} The Authors. The Scandinavian Journal of Economics published by John Wiley & Sons Ltd on behalf of Föreningen för utgivande av the SJE.

- Levitt, S. D. and List, J. A. (2007), What do laboratory experiments measuring social preferences reveal about the real world?, *Journal of Economic Perspectives 21* (2), 153–174.
- List, J. A. (2006), The behavioralist meets the market: measuring social preferences and reputation effects in actual transactions, *Journal of Political Economy* 114, 1–37.
- Loureiro, M. L. and Lotade, J. (2005), Do fair trade and eco-labels in coffee wake up the consumer conscience?, *Ecological Economics* 53, 129–138.
- Pigors, M. and Rockenbach, B. (2016), Consumer social responsibility, *Management Science* 62, 3123–3137.
- Plott, C. R. and Zeiler, K. (2005), The willingness to pay/willingness to accept gap, the "endowment effect", subject misconceptions and experimental procedures for eliciting valuations, *American Economic Review 95* (3), 530–545.
- Rode, J., Hogarth, R. M., and Le Menestrel, M. (2008), Ethical differentiation and market behavior: an experimental approach, *Journal of Economic Behavior & Organization 66*, 265–280.
- Soetevent, A. R., Bao, T., and Schippers, A. L. (2016), A commercial gift for charity, Tinbergen Institute Discussion Paper No. 16-009/VII, https://papers.tinbergen.nl/16009.pdf.
- Stoop, J. (2014), From the lab to the field: envelopes, dictators and manners, *Experimental Economics* 17, 304–313.
- Stoop, J. T., Noussair, C. N., and van Soest, D. P. (2012), From the lab to the field: cooperation among fishermen, *Journal of Political Economy* 102, 1027–1056.
- Sutter, M., Huber, J., Kirchler, M., Stefan, M., and Walzl, M. (2020), Where to look for the morals in markets?, *Experimental Economics* 23, 30–52.
- Sutter, M., Kocher, M. G., Glätzle-Rützler, D., and Trautmann, S. T. (2013), Impatience and uncertainty: experimental decisions predict adolescents' field behavior, *American Economic Review 103* (1), 510–531.
- TransFair e.V. (2018), TransFair e.V. Jahres- und Wirkungsbericht 2017, TransFair e.V., https://www.fairtrade-deutschland.de/fileadmin/DE/mediathek/pdf/fairtrade_transfair _jahresbericht_2017_web.pdf.
- Winking, J. and Mizer, N. (2013), Natural-field dictator game shows no altruistic giving, *Evolution and Human Behavior 34*, 288–293.

First version submitted December 2019; final version received February 2024.