DIW BERLIN

1634

Discussion Papers

Deutsches Institut für Wirtschaftsforschung 2017

Who Cares about Social Image?

Jana Friedrichsen and Dirk Engelmann

IMPRESSUM

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
http://www.diw.de

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
http://www.diw.de/discussionpapers

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
http://ideas.repec.org/s/diw/diwwpp.html
http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html

Who cares about social image?*

Jana Friedrichsen† Dirk Engelmann‡

January 24, 2017

Abstract

This paper experimentally investigates how concerns for social approval relate to intrinsic motivations to purchase ethically. Participants state their willingness-to-pay for both a fair trade and a conventional chocolate bar in private or publicly. A standard model of social image predicts that all participants increase their fair trade premium when facing an audience. We find that the premium is indeed higher in public than in private. This effect, however, is driven by participants who preferred a conventional over a fair trade chocolate bar in a pre-lab choice. For those who chose the fair trade chocolate bar, public exposure does not change the fair trade premium. This is captured by a generalized model where intrinsic preferences and the concern for social approval are negatively correlated.

JEL-codes: D03, C91, D12
Keywords: image concerns, ethical consumption, fair trade, social approval, crowding out, experiments

*We thank Yves Breitmoser, Antonio Guarino, M'Hamed Helitim, Steffen Huck, Heiko Karle, Georg Kirchsteiger, Tobias König, Tatiana Kornienko, Johanna Möllerström, Rosemarie Nagel, Charlie Plott, Nora Szech, Mirko Tonin, Bertil Tungodden, Jean-Robert Tyran, Joël van der Weele, and Christian Waibel as well as seminar participants at the University of Cologne, Humboldt-University Berlin, Paris I–Sorbonne, the University of Marburg, the University of Birmingham, the University of Nottingham, University of Vienna, Texas A&M, NYU, CAU Kiel, DIW Berlin, the Berlin Behavioral Economics workshop, and at the International ESA conference 2012 in New York, the European ESA conference 2012 in Cologne, the annual meeting of the Gesellschaft für experimentelle Wirtschaftsforschung 2012 in Karlsruhe, the HeiMaX workshop 2012 in Heidelberg, the Spring Meeting of Young Economists 2013 in Aarhus, the Journées Louis-André Gérard-Varet 2013 in Aix-en-Provence, EEA 2013 in Gothenburg, EARIE 2013 in Evora, the annual meeting of the Verein für Socialpolitik 2013 in Düsseldorf, the CESifo Area Conference on Behavioral Economics 2013, and the Arne Ryde Workshop on Identity, Image and Economic Behavior, Image 2016 in Lund for helpful comments and suggestions. Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 is gratefully acknowledged. Jana Friedrichsen further acknowledges financial support by the Leibniz Competition through the project GlobalFood (SAW-2015-DIW-4).

†Corresponding author. DIW and Humboldt-Universität zu Berlin, Email: jfriedrichsen@diw.de or jana.friedrichsen@hu-berlin.de

‡Humboldt-Universität zu Berlin; CERGE-EI, Prague; CESifo, Munich, Email: dirk.engelmann@hu-berlin.de

1 Introduction

Substantial evidence exists that individuals desire to signal pro-social behavior in the lab (e.g. Andreoni and Bernheim, 2009; Ariely et al., 2009) and in the field (e.g. Carpenter and Myers, 2010; Soetevent, 2011). Specifically, empirical research also suggests that choices for products considered to be ethical are driven not only by intrinsic motivation but also by concerns for social approval (Griskevicius et al., 2010; Sexton and Sexton, 2014). How intrinsic motivation and image concerns interact is important for supply and pricing policy of these products (Friedrichsen, 2016). Furthermore, the interaction is relevant to assess potentially negative effects of incentives on behavior motivated by intrinsic motivation or image concerns (Bénabou and Tirole, 2006; Gneezy and Rustichini, 2000).[1] Little is known about this interaction between intrinsic motivation and concerns for social approval from empirical work. This paper starts filling this gap by assessing in a laboratory experiment, how intrinsic motivation and social image concerns interact.

In the experiment, we test whether individuals who are intrinsically motivated to support fair trade react more or less strongly to opportunities for image building than individuals who are not intrinsically motivated. Before subjects come to the experiment, we elicit a proxy for their intrinsic motivation for supporting fair trade by offering a choice between a fair trade and a larger conventional chocolate bar as an additional reward for taking part in the experiment. In the experiment itself, participants first engage in a market game unrelated to chocolate or fair trade products. Then we elicit in an incentive-compatible way their willingness to pay both for conventional and fair trade chocolate and derive for each individual subject the fair trade premium they are willing to pay, which is our main variable of interest. Our treatments vary whether the willingness to pay is kept private or has to be announced publicly. Thus, we vary whether participants can build an image among their fellow participants for being concerned with fair trade. In a questionnaire, we ask for knowledge about and attitudes towards

[1] Consider a supposedly ethical product (such as environmentally friendly, fair trade or those bundled with charitable donations) that is purchased by people who are intrinsically motivated to support the cause related to the product as well as those primarily concerned with the positive image derived from this purchase. Providing extrinsic incentives can dilute the image derived from the purchase and as a result crowd out purchases of image-motivated consumers. Crowding out of intrinsically motivated consumers may be problematic because they are arguably more willing to make an effort to observe whether the products satisfy the necessary standards. Intrinsically motivated agents are actually crowded out only if they are also concerned with their social image.

fair trade and confirm that the latter are positive, so that our variation of image building opportunities indeed allows participants to derive a positive social image.

Our theoretical framework follows Bénabou and Tirole (2006) in assuming that behaving pro-socially confers a positive social image or social esteem, so that image-concerned individuals should behave more pro-socially in the public sphere than in private.[2] Under the standard assumption that all individuals have identical image concerns, we derive the hypotheses that (1) participants who chose a conventional chocolate bar before the experiment will state on average a lower fair trade premium than those who chose a fair trade bar both in private and in public, and (2) that all participants increase their fair trade premium by the same amount if it is elicited publicly compared to the setting where it is elicited in private.

We find that in the private setting subjects who revealed to have no intrinsic motivation for fair trade by choosing the conventional chocolate bar before the experiment have significantly lower fair trade premiums than those who revealed an intrinsic fair trade preference by choosing the fair trade chocolate bar. In the public setting, however, on average both groups of participants stated fair trade premiums that were not statistically different. Only the participants who chose the conventional chocolate before the experiment exhibited a significantly larger fair trade premium in public than in private.

Thus, our experimental results confirm the first hypothesis for the private elicitation only but reject it for the public elicitation. Moreover, our results are not consistent with the second hypothesis as participants reacted differently to the image-building opportunity. Using a generalized model of image concerns, we show that all our experimental findings are rationalized if intrinsic motivation and image concerns are negatively correlated. Furthermore, we discuss alternative approaches and show that our data is inconsistent with models of fair trade purchases which would predict a positive correlation as, e.g., a model of expressive behavior.

These findings have three implications. First, incentives that encourage those not intrinsically motivated will not crowd out intrinsically motivated socially beneficial behavior in our sample, suggesting that incentivizing fair trade purchases may be a reasonable idea. Second, as a general insight, correlations between intrinsic motivation and image concerns are economically relevant, and theoretical models therefore need

[2] Alternative signaling explanations for charitable behavior argue that the size of the contribution reveals an individual's wealth (Glazer and Konrad, 1996; Harbaugh, 1998). This channel is absent from our design since purchases were made from the experimental endowment, but it may play a role in the field evidence on image concerns in giving.

to allow for significant correlations between types and their signaling concerns. Our stylized extended model suggests that this is feasible in a tractable way. Third, for this negative correlation an optimally designed product portfolio will induce partial pooling of different consumer types (cf. Friedrichsen, 2016).

Our experiment uses fair trade products, understood as bundles of a base product with a charitable contribution as in Reinstein and Song (2012).[3] By purchasing fair trade in the experiment, the participants can signal their pro-social attitudes. This is natural because consumer products are a frequently used signaling medium (e.g. Miller, 2009). Teyssier et al. (2015) find that WTPs for conventional chocolate decrease in public leading to higher premiums for fair trade in public but do not investigate intrinsic preferences or heterogeneity. In addition, a field experiment in German coffee shops finds that consumers donate larger amounts if these are bundled with a coffee purchase than if coffee and donations have to be bought separately (Koppel and Schulze, 2013). This is explained by social image concerns because the consumers have to communicate their choices to the staff if the donations are bundled with the product, whereas direct donations are just dropped into a box anonymously. As this paper focuses on consumer preferences and behavior, we do not discuss potential benefits or pitfalls of fair trade but refer to Dragusanu et al. (2014) and the literature surveyed therein for an economic analysis of it.[4]

We introduce our design in Section 2, develop a theoretical model to derive our hypotheses in Section 3, and present our results in Section 4. In Sections 5 and 6, we discuss our results and alternative modeling approaches. Section 7 concludes. Proofs are collected in Appendix A.

2 Experimental design and procedures

Our experimental design consists of two parts. First, after participants have registered for the experiment, but before they arrive at the laboratory, we derive a proxy for their preference for fair trade. Second, they participate in a laboratory experiment in which

[3]Such bundles need not be efficient, i.e., the price of the bundle may exceed the sum of the price of the private good and the donation as would be the case if retailers increase their markups for fair trade products. Consumers indeed also choose inefficient bundles (see Munro and Valente, 2016, for experimental evidence).

[4]A survey with a focus on fair trade's effects on the welfare of smallholders is provided by Dammert and Mohan (2015).

they take part in a market game before we elicit their willingness to pay a price premium for a product being fair trade.

Proxy for intrinsic preference In order to derive a proxy for their intrinsic preference for fair trade, we offered subjects via email the choice between a bar of fair trade and a bar of conventional milk chocolate as an additional reward for coming to the experiment. This email was sent and had to be answered before they came to the laboratory but the chocolate bars were distributed only after the experiment. Since fair trade chocolate is on average more expensive, we offered a choice between a slightly larger (125g) bar of conventional chocolate and standard size (100g) bar of fair trade chocolate. As only few subjects chose the conventional chocolate in our first sessions, we offered a choice between two bars of conventional chocolate and one bar of fair trade chocolate in the following sessions. We balanced the design with respect to whether we offered one or two bars of conventional chocolate. The difference in trade-offs offered should have moved some subjects with positive willingness to pay to support fair trade between categories. This can only have weakened the comparison between the two classes of subjects.[5]

Market game The laboratory experiment itself consists of two parts. In the first part, the participants take part in a market game modified from Danz et al. (2012), with participants taking the roles of firms, consumers and workers. This market game serves two purposes. On the one hand, given that the second part is short, we used the opportunity to assess the generalizability of fair behavior observed in an experimental market by comparing behavior in this market with fair trade choices both before the experiment and in its second part. This analysis will be the focus of a companion paper. On the other hand, the market game serves to start the experiment in a relatively conventional fashion and thus removes the focus from the rather unusual chocolate purchase in the second part.

Willingness-to-pay In the second part, we use a random price mechanism (Becker et al., 1964) to elicit from each participant his or her willingness-to-pay (WTP) for both a fair trade and a conventional dark chocolate bar. Specifically, subjects enter a price between 0 and 2 Euros, where any multiple of €0.01 is permitted. Then we draw a

[5]The main results holds for either version of the trade-off. See also footnotes 16 and 17.

price from a uniform distribution of all integer multiples of €0.01. Subjects receive a bar of the chocolate type sold if their stated WTP for that type is at least as high as the randomly chosen price. Which type of chocolate is sold is determined randomly after the price has been chosen such that the mechanism is incentive compatible for both types of chocolate. We chose dark chocolate for this part of the experiment instead of milk chocolate. This ensures that subjects cannot end up with two identical chocolate bars, which could have reduced their willingness to pay for the type of chocolate of which they were already sure to receive one bar.[6] We also did not choose any well-known brands, in order to minimize the chance that subjects' willingness to pay was based on taste preferences due to personal experience or anchoring on market prices. From these two WTPs we infer individuals' willingness' to pay a premium for the fair trade chocolate as $\text{WTP}_{\text{fair}} - \text{WTP}_{\text{conv}}$. We call this the *fair trade permium*.

Treatments Our two treatments differ in whether the WTPs are elicited publicly or in private. In treatment *private*, individuals enter their WTPs privately at the computer. In treatment *public*, after they have entered their WTPs privately at the computer, all subjects stand up and announce their WTPs publicly among the group of participants. The difference in the fair trade permiums between the treatments serves as our measure for image concerns. A post-experiment questionnaire confirms that the participants consider supporting fair trade as a good deed.

We note that while the random price mechanism (Becker et al., 1964) is incentive compatible in theory, it has been pointed out that experimental subjects may misconceive this mechanism (Plott and Zeiler, 2005; Cason and Plott, 2014). Such misconceptions should be of much less concern in our experiment. Misconceptions appear to be more of an issue for elicitation of willingness to accept to forego an item rather than for willingness to pay to obtain an item. More importantly, we are only interested in the fair trade premium, and in particular in the question whether this differs significantly across groups or treatments. Tests of the related hypotheses are robust to any misconception that only leads to a bias that is monotone in the true WTP and is not systematically correlated with the treatment or subject group. One could imagine further misconceptions, but it appears that in order to test hypotheses relying on whether differences (and differences-in-differences and diffs-in-diffs-in-diffs) in WTP are significant, the random

[6] At the time of payment, most of our participants had actually already forgotten that for participating they would receive a milk chocolate bar of a type chosen beforehand.

Table 1: Prices drawn and number of chocolate bars paid out to participants.

	conventional				fair trade			
price in €	0.26	0.27	0.97	1.85	0.25	1.01	1.20	1.78
treatment	public	private	private	public	public	private	public	private
#participants	16	16	20	20	16	16	20	20
#bars sold	11	9	2	0	12	2	7	0

price mechanism is substantially more robust to misconceptions than when it is used to measure absolute WTP.

Procedures The experiment was computerized using zTree (Fischbacher, 2007) and 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). An English translation of the (German) instructions for the second part of the experiment is included in Appendix D. We have conducted 8 sessions with 16-20 participants each, and we had in total 144 participants. For the market game, each participant got a show-up fee of €5; for the second part of the experiment in which we elicit the WTPs, each participant got an additional endowment of €4. Average cash earnings were €18.63, including the show-up fee and the endowment in the second part and subtracting payments for chocolate if applicable.[7] In the second part, the payoff-relevant chocolate was conventional and fair trade in half of the sessions each. We handed out conventional chocolate to 22 subjects and fair trade chocolate to 21 subjects. Details about the (randomly chosen) prices at which chocolates were sold are collected in Table 1.

After entering their WTPs (but before they announce them in *public*), subjects fill in an extensive questionnaire regarding their attitudes towards and knowledge about fair trade. The answers to this questionnaire allows us to confirm the validity of our proxy for intrinsic motivation as those who chose fair trade chocolate are more frequent to report to buy fair trade products and less frequent to agree to negative statements regarding fair trade (see Figure 4 in Appendix C).

[7] In the market game in the first part of the experiment, participants in the role of firms earned €4.50 on average, those in the role of workers earned €6.31 on average, and those in the role of consumers earned €23.73 on average.

Our analysis evaluates decisions from 121 subjects who made their choice between fair trade and conventional chocolate via email as described above.[8] Among these 121 subjects, 32 chose conventional chocolate, while the remaining 89 chose fair trade before coming to the lab.

3 Theoretical background and hypotheses

In this section we derive hypotheses based on a standard model of image concerns where all consumers have identical concerns for social image and then consider the implications of heterogenous image concerns. Proofs are in Appendix A.

Preference Define an indicator $f \in \{0, 1\}$ for fair trade, i.e. $f = 0$ if the chocolate bar is a conventional one, and $f = 1$ if the chocolate bar is fair trade. Suppose utility of consuming chocolate of type f is

$$v(f, p(f), R) = uq(f) + fm + \lambda\rho fR - p(f) \tag{1}$$

where $p(f)$ is the price for a chocolate bar of type f, u is the marginal utility of chocolate, $q(f)$ is the (potentially quality-adjusted) size of the chocolate bar, and m is an individual's intrinsic motivation for supporting fair trade. The realized social image is denoted by R, ρ is the marginal utility of social image, and $\lambda \in \{0, 1\}$ indicates whether the purchasing decision is private (0) or public (1). We specify the social image below when we discuss the public elicitation of WTP.

Assume that u is normally distributed with mean $\bar{u}$ and variance σ_u.[9] Further, assume that preferences for fair trade, m, are normally distributed with mean $\bar{m}$ and variance σ_m, and that these are independent of the taste for chocolate u. Each individual privately knows her realizations of u and m.

[8] In addition, 23 newly recruited subjects participated in our experimental sessions but are not included in the analysis. For these subjects, the chocolate choice which we intended to use as a proxy for their intrinsic preference were taken in public during a recruitment day and not via email. Out of 222 new recruits, only 23 ever showed up in one of our experimental sessions and their chocolate choices are not balanced across treatments such that we cannot separately analyze this subgroup. A pooled analysis prevents itself due to the the public recruitment situation which biases the proxy.

[9] In the experimental design, negative prices are excluded. An individual can always state a WTP of zero in which case receiving a chocolate bar is a zero-probability event.

Classification by proxy Denote the size of the conventional chocolate by $q(0) = x$ and the size of the fair trade chocolate by $q(1) = x - d$, where $d > 0$. Individuals select the fair trade chocolate if

$$u(x - d) + m > ux \tag{2}$$

and choose conventional chocolate if the reverse inequality holds. An individual is indifferent between the fair trade bar and the larger conventional chocolate if her intrinsic motivation to support fair trade exactly compensates the reduction in quantity. Thus, the threshold valuation of fair trade that compensates the size difference at a given chocolate valuation u is defined as an increasing function $m(u) = du$. Given the independent distribution of m and u, those classified as fair trade choosers are expected to have a higher m and a lower u on average.

Willingness-to-pay If asked for her WTP, an individual may deviate from (1) by a random noise term. Moreover, for clarity of exposition, we abstract from the effect of the size of the chocolate bar on WTP and normalize the size to 1.[10] Specifically, we assume that the stated WTPs for both types of chocolate are given by

$$w_{\text{conv}}(u, m) = u + \eta_{\text{conv}} \text{ and } w_{\text{fair}}(u, m) = u + m + \lambda\rho R + \eta_{\text{fair}}, \tag{3}$$

where $\eta_{\text{conv}} \sim \mathcal{N}(0, \sigma_{\text{conv}})$ and $\eta_{\text{fair}} \sim \mathcal{N}(0, \sigma_{\text{fair}})$. Since heterogeneous tastes for chocolate u have a level effect on both WTPs, we focus on the individuals' willingness to pay a premium for fair trade chocolate as given by

$$a(m) := w_{\text{fair}}(u, m) - w_{\text{conv}}(u, m) = m + \lambda\rho R + \varepsilon \tag{4}$$

where $\varepsilon = \eta_{\text{conv}} + \eta_{\text{fair}}$ is a noise term that is normally distributed with $\varepsilon \sim \mathcal{N}(0, \sigma_R)$, and σ_R is the variance of the sum of both individual error terms.

Private As there is no scope for social signaling in the private case, only the individual's intrinsic valuation for fair trade and her utility from chocolate affect her WTP. The stated premium for fair trade is directly linked to the individual's valuation for fair trade: if $m' > m$, then $E[a|m'] = m' > m = E[a|m]$. Furthermore, upon know-

[10] In principle, the size difference may also affect stated WTP, and the model can be extended to cover this without affecting our hypotheses. Empirically, we can control for any size effect using the elicited WTP for conventional chocolate.

ing which type of chocolate an individual chose in the classification stage, we can update our expectation of her motivation to support fair trade because $E[m|\text{chose fair}] > E[m|\text{chose conv}]$. This yields the following hypothesis.

Hypothesis 1. *(Private elicitation) If elicited in private, the average fair trade premium is higher for individuals classified as fair trade choosers than for those classified as conventional choosers,* $E[a|\textit{chose fair}] > E[a|\textit{chose conv}]$.

Hypothesis 1 follows if, conditional on their basic utility from chocolate, the WTP for fair trade is higher for fair trade choosers because they have the stronger fair trade preference m which made them choose the fair trade chocolate bar over the conventional one. Alternatively, Hypothesis 1 would also be implied if the fair trade choosers simply care less about chocolate, i.e., have a smaller u rather than a higher m. Specifically, the observed fair trade premium should decrease in the chocolate valuation because $a = u(x-d) + m - ux + \varepsilon = m - du + \varepsilon$. We therefore consider the following hypothesis, which is derived from the type classification.

Hypothesis 2. *(Size effects) Individuals who chose the fair trade chocolate have on average lower valuations for chocolate as such, i.e.* $E[w_{\textit{conv}}(u,m)|\textit{chose fair}] < E[w_{\textit{conv}}(u,m)|\textit{chose conv}]$.

Public Individuals may care about their social image as a pro-social person. In the private setting, there is a direct link between an individual's pro-sociality as given by her preference for fair trade, m, and her fair trade premium a. This suggests that the WTP difference can be used to signal pro-sociality.

Assume that the **social image** function $R(a,u) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ attached to an observed fair trade premium a is the public's belief about an individual's attitude m toward fair trade controlling for her taste for chocolate u. This inference function maps each pair of observed fair trade premium and WTP for the conventional chocolate bar, (a, w_{conv}), to an expected fair trade preference, which in equilibrium must be consistent with the conditional expectation of an individual's type upon observing her stated premium a, i.e., $R(a,u) = E[m|a,u]$ in equilibrium. Note that we need the willingness to pay for both types of chocolate to compute the social image. When only the WTP for fair trade chocolate is observed, the inference on m is confounded by heterogeneity in the base utility from chocolate u. If we also observe the WTP for conventional chocolate, we

learn u so that clean inference on m is possible.[11] For ease of notation, we drop the argument u of the image in the following.

Since preferences as well as the error terms in stated WTP are normally distributed, all WTP differences on the real line can occur with positive probability, and in particular all differences that can technically be observed in our experiment, i.e., all $a \in [-2, 2]$ are expected to occur with positive probability by individuals in an equilibrium. Then, the inference function, and so the anticipated social image, is well-defined on the entire interval. In line with the literature (e.g., Bénabou and Tirole, 2006), we concentrate on the case where the image function is differentiable.

Assumption 1. *Assume that the inference function $R(a)$ is differentiable in a.*

To obtain a higher observed premium and improve her social image, an individual may increase her WTP for fair trade. Stating a WTP for fair trade is costly though. Specifically, we assume that an increase in the WTP for fair trade and thereby the fair trade premium from the privately optimal level $a = m$ to $a = m + \Delta$, is associated with a cost $c(\Delta) = k\frac{\Delta^2}{2}$ for some scalar $k > 0$.[12] We expect individuals to increase their stated price premium in response to social image concerns as long as the marginal expected benefit exceeds the marginal expected costs, $c'(\Delta) = k\Delta$. Thus, in the public context, we expect an individual with preferences (u, m) to state a fair trade premium $a = m + \Delta + \varepsilon$ such that

$$\rho \frac{\partial R(m + \Delta)}{\partial \Delta} = k\Delta \tag{5}$$

Knowing this, an observer expects that this agent has a fair trade preference

$$E[m|a] = a - \frac{\rho}{k} \frac{\partial R(m + \Delta)}{\partial \Delta} \tag{6}$$

Solving the resulting first-order differential equation, we obtain the following result.

Proposition 1. *Let all agents have the same image concern ρ. There is a unique (differentiable-image) equilibrium, in which an agent with preferences (m, u) states a fair trade premium $a = m + \rho/k$. The marginal social image return equals ρ/k, and it is constant across types.*

[11] We assume that w_{conv} is a reliable measure of u also in the public setting. This simplifies the theoretical exposition and is supported by our data. See discussion below.

[12] The random price mechanism in our experiment induces a quadratic cost function that is scaled by the upper bound of the price domain. When an individual increases her WTP for fair trade chocolate, she faces the following additional net cost for an increase by Δ above the privately optimal level, which we here denote by w: $C(\Delta) = \int_w^{w+\Delta} \frac{p}{\bar{p}} dp - \frac{\Delta}{\bar{p}} w = \frac{1}{2}\frac{\Delta^2}{\bar{p}}$, where $\bar{p}$ is the upper bound on the price interval.

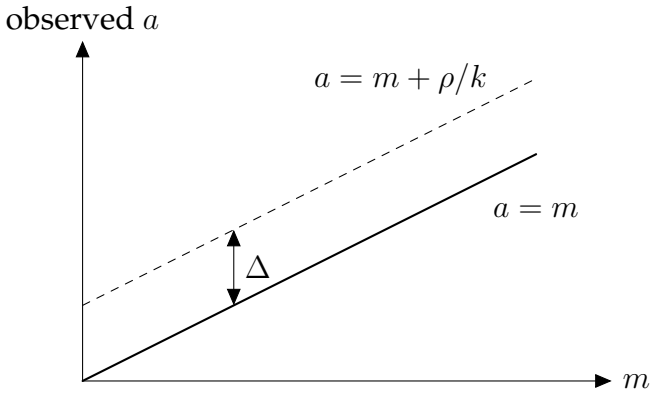

Figure 1: Expected upward shift in a due to public exposure is homogeneous (parallel) for equilibrium or naive inference according to Proposition 1 and Corollary 1.

Individuals of all preference types inflate their fair trade premium by increasing their WTP for fair trade by the same amount (see Figure 1). They do so in an attempt to obtain a better social image. In equilibrium, however, everyone is just "running to keep in the same place" (Hopkins and Kornienko, 2004), and despite the increase in her fair trade premium, each type's preference for fair trade is correctly inferred. Even if individuals are unable to anticipate the equilibrium inferences but make naive inferences instead, the same incentives to inflate one's fair trade premium are at work, even though a positive image premium is obtained in that case.

Corollary 1. *If the social image of a fair trade premium a is naively inferred as $R(a) = a$, an agent with preferences (m, u) states a fair trade premium $a = m + \rho/k$.*

The two previous results imply, that public exposure increases stated WTP of both motivated and unmotivated individuals in the two benchmark cases of equilibrium inference and naive inference so that our treatment should be effective.

Hypothesis 3. *(Difference public vs. private) The average stated fair trade premium is higher if elicited publicly than privately both for conventional choosers and for fair trade choosers.*

More specifically, because all individuals are expected to increase their WTP by the same amount, the stated fair trade premiums of conventional choosers and fair trade choosers should differ not only in the private but also in the public treatment.

Hypothesis 4. *(Public elicitation) If elicited in public, the average stated fair trade premium is higher for fair trade choosers than for conventional choosers.*

The previous results assumed that individuals value image to the same extent. We further note that given our quadratic cost function, (5) shows that as long as the inference function is increasing, anyone with a positive concern for image would inflate her fair trade premium in the public treatment. But if individuals systematically differ in how much the value their social image, the reaction to public exposure will be heterogeneous.

Corollary 2. *If conventional choosers have systematically lower (higher) concern for social image than fair trade choosers, they increase their fair trade premium by a smaller (larger) amount in response to public exposure than the fair trade choosers.*

The intuition is the following: By (5) an individual who cares more about social image, i.e., has a larger ρ, would choose to inflate her fairtrade premium even more. Therefore, if those who are truly fair-minded cared more about their social image as suggested, e.g., by theories of expressive preferences (see, e.g., Hillman, 2010), we should see a larger treatment effect for those who are more concerned with fair trade.

Hypothesis 5. *(Expressive behavior) The treatment effect from public exposure is larger for fair trade choosers than for conventional choosers.*

4 Experimental results

In line with the results from previous studies (e.g. De Pelsmacker et al., 2005; Loureiro and Lotade, 2005), we find significant heterogeneity in the intrinsic preferences for fair trade. In our sample, the minimum fair trade premium amounted to -0.49 euros whereas the maximum fair trade premium amounted to 1.79 euros. In the following, we investigate to what extent the variation in fair trade premiums can be organized within our theoretical framework.

We first analyze whether the classification based on the chocolate choice before the experiment indeed translates into higher fair trade premiums in the private treatment. In line with Hypothesis 1, for the conventional choosers we observe an average fair trade premium $a =$ €-0.06, while for the fair trade choosers we observe an average fair trade premium $a =$ €0.26. The fair trade premiums differ across the two groups according to a Mann-Whitney test ($p < 0.001$).

In contrast to Hypothesis 2, the difference in fair trade premiums is not in part driven by lower willingness to pay for conventional chocolate by the fair trade choosers. Re-

gressing the chocolate choice before the experiment on the stated WTP for conventional chocolate, the latter is insignificant (p-value$= 0.895$ in private, $= 0.979$ in public, and $= 0.919$ pooled) indicating that the difference in sizes of the two chocolate bars did not play a role for the observed sorting. Also non-parametric tests yield no evidence that those with lower valuation for chocolate are more likely to choose fair trade chocolate, but instead the sorting is driven by fair trade preferences. In the private treatment, the WTP for the conventional chocolate bar, w_{conv}, is nearly identical across the two groups ($= 0.45$ for fair trade choosers and $= 0.46$ for conventional choosers, Mann-Whitney test $p = 0.467$), whereas the WTP for the fair trade chocolate bar, w_{fair}, is substantially higher for fair trade choosers ($= 0.71$) than for conventional choosers ($= 0.40$), (Mann-Whitney test, $p = 0.051$; $p = 0.022$ if we exclude participants with WTP below 2 cents for both bars of chocolate).[13]

Concerning the predicted treatment effect on the fair trade premium, we see a clear treatment effect in line with Hypothesis 3 on the conventional choosers. Making choices public increases the average stated fair trade premium of a conventional chooser from €-0.06 in private to €0.10 in public. These two fair trade premiums are significantly different (Mann-Whitney test, $p = 0.005$). The treatment difference is driven by an increase in the willingness to pay for fairtrade chocolate rather than a decrease in the willingness to pay for conventional chocolate, w_{fair} increases from €0.40 to €0.58, but w_{conv} even marginally increases (€0.46 in private and €0.48 in public).[14]

In contrast, there is no significant treatment effect on the fair trade premium of fair trade choosers. The average stated fair trade premium of a fair trade chooser even decreases from €0.26 in private to €0.15 in public, but this decrease is not statistically significant (Mann-Whitney test, $p = 0.122$). Hence Hypothesis 3 is not supported for the fair trade choosers. This result is inconsistent with the assumption of identical image concerns of all participants and instead suggests that the conventional choosers are concerned with their image, while fair trade choosers are not.

As a result of the heterogeneous treatment effect, the difference between the fair trade premiums between the conventional choosers and the fair trade choosers nearly

[13] One reason why the size difference between the two bars of chocolate does not appear to affect the willingness to pay may be that fair trade chocolate is typically perceived to be of higher quality, which is also supported by our questionnaire data. A perceived quality difference may then compensate for the size difference.

[14] The fair trade premium could in principle also be inflated by decreasing the WTP for conventional chocolate. This does not appear to happen, possibly because individuals bracket narrowly (Read et al., 1999) and consider the image of buying fair trade only when stating WTP for fair trade chocolate.

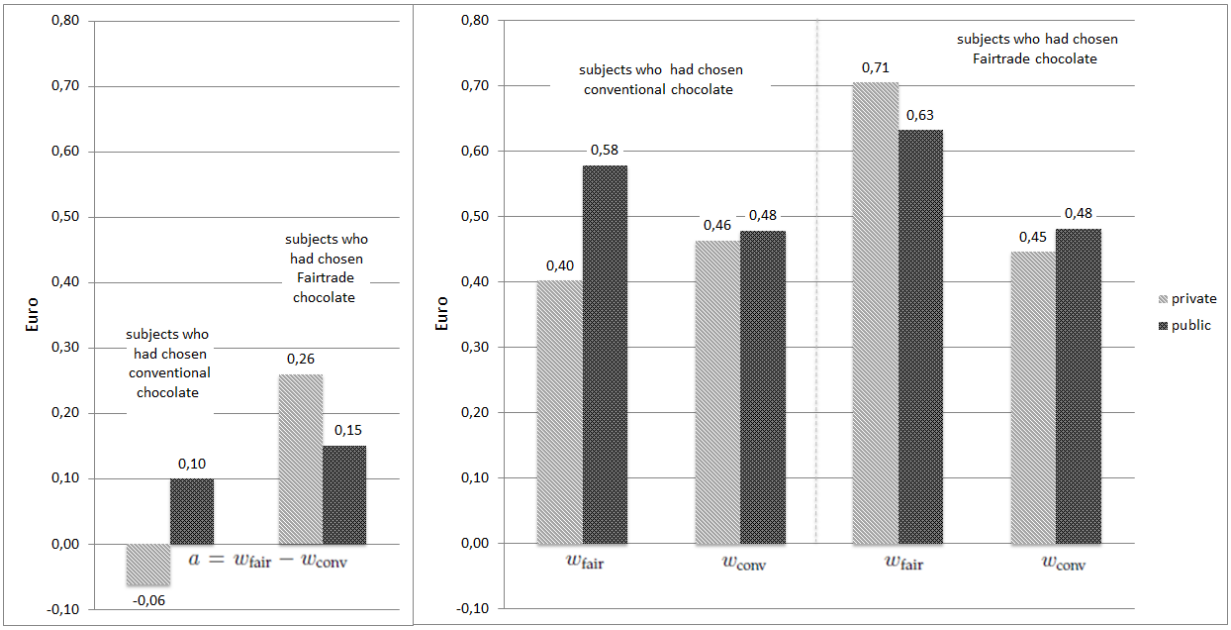

Figure 2: Averages of stated WTP by treatment and by chocolate choice. Left panel: averages of fair trade premiums a, right panel: averages of WTPs for the conventional chocolate bar, w_{conv}, and of the WTPs the fair trade chocolate bar, w_{fair}.

disappears in the public treatment and is not significant anymore (Mann-Whitney test, $p = 0.123$). Thus, while behavior in the private treatment differs in line with intrinsic motivation, the two groups become indistinguishable in the public treatment, and Hypothesis 4 is not supported. Moreover, our results contradict Hypothesis 5 because we find that individuals who chose conventional chocolate react more strongly to the treatment than the fair trade choosers. Instead, using Corollary 2, our results would be consistent with a negative correlation between intrinsic preferences for fair trade and image concerns. Figure 2 illustrates the average fair trade premiums and willingness to pay for both types of chocolate, both groups of participants and both treatments.

The change in the distribution of fair trade premiums for the two different groups is illustrated in Figure 3. On the left, we see that for the conventional choosers substantial mass is shifted from the left of 0 to the right of 0 when moving from the private to the public treatment. In contrast, there is no clear shift for the fair trade choosers, as shown on the right. There is no robust effect of the treatment on the variance. The variance of fair trade premiums for the conventional choosers is higher in the public than the private treatment (0.057 vs. 0.02), while we find the opposite effect for the fair trade choosers (0.069 vs. 0.11). Furthermore, these differences are reversed if we exclude two outliers, namely the fair trade chooser with a very large fair trade premium in the private treatment (the variance falls from 0.11 to 0.063) and the conventional chooser with a fair trade premium of 1 in the public treatment (the variance decreases from 0.057 to 0.008). While there is no ex-ante reason to exclude these outliers, this exercise shows that for both groups the differences in variances across treatments can be traced to a single individual and are hence not conclusive.

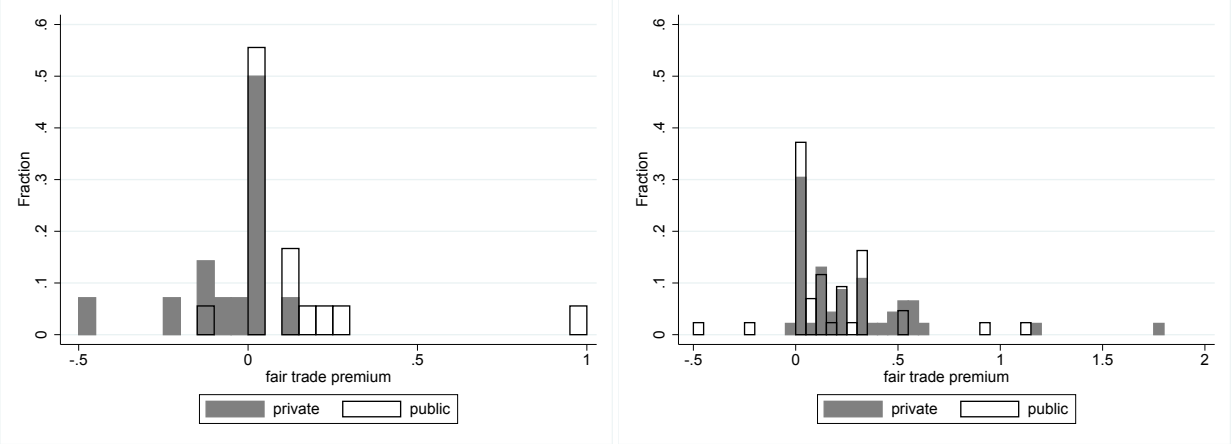

Figure 3: Histograms of distribution of fair trade premiums a, bin size: 5 cents. Left panel: conventional choosers, right panel: fair trade choosers.

On the aggregate level, for the fair trade choosers we observe a slight decrease of the fair trade premium in the public treatment. This could be seen as suggesting that these subjects choose fair trade to support their self-image and that the expected pooling of those only driven by social image in the public treatment leads to a decrease of the self-image derived from the stated fair trade premium and hence their fair trade premium actually decreases if self-image is derived as if in the eye of a neutral observer as in the model by Benabou and Tirole (2003).[15] Such motivations do not seem to play a role here because the treatment effect on the fair trade choosers completely disappears once we control for the profits from the first part of the experiment. In an OLS regression of the fair trade premium restricted to the subjects who chose fair trade before the experiment, the coefficient on the dummy for the public treatment is actually positive, but very small and far from being significantly different from zero ($p > 0.8$).[16]

We confirm the heterogeneity of the treatment effect with OLS regressions of the fair trade premium on dummies for the *public* treatment and whether the subject had chosen fair trade chocolate (*FTchoice*) before coming to the experiment as well as the subject's earnings from the market game in the first part of the experiment (*marketprofit*). We also

[15] In Appendix B, we provide an extended model that would be consistent with a negative treatment effect for fair trade choosers.

[16] Table 3 in Appendix C shows results from regressions on the two subsamples of individuals who chose conventional and fair trade chocolate, respectively. Only for those who chose conventional chocolate the treatment effect is significant. Also if we split the fair trade choosers further into those who chose a bar of fair trade chocolate over one, respectively two, bars of conventional chocolate, the treatment effect is insignificant for both subgroups.

Table 2: Regression of the stated fair trade premium $a = w_{\text{fair}} - w_{\text{conv}}$ on (ex ante) fair trade choice, profits from the first part of the experiment, treatment (public) and interaction terms. Column 1 includes all subjects. In column 2, we exclude subjects who bid less than 2 cents for each type of chocolate. Standard errors in parentheses.

	(1)	(2)
FTchoice	0.313***	0.341***
	(0.085)	(0.093)
public	0.292**	0.306**
	(0.122)	(0.133)
FTchoice*public	-0.290**	-0.289**
	(0.119)	(0.131)
marketprofit	0.007*	0.006
	(0.004)	(0.004)
marketprofit*public	-0.012**	-0.012*
	(0.006)	(0.007)
Observations	121	104
adj. R^2	0.113	0.113
Prob> F	0.002	0.005

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

include the interaction effects between the two dummies and between (*marketprofit*) and the treatment dummy.

Looking at all 121 individuals, we find the following results (see Table 2, column 1). Our treatment dummy *public* is significant ($+29.2$ Cents, $p = 0.018$) implying that making choices public increases individuals' willingness to pay a premium for fair trade chocolate. Moreover, the heterogeneity of the treatment effect is confirmed by the negative interaction effect between having chosen fair trade chocolate and *public* ($p = 0.016$). The size of coefficient on the interaction is similar in size to the aggregate treatment effect. Thus, in our experiment, a treatment effect indicative of image concerns is significantly less pronounced and virtually absent for those individuals who had chosen the fair trade chocolate. [17]

In line with Hypothesis 1, we further find that having chosen fair trade chocolate before the experiment is associated with a significantly higher fair trade premium (+31.3

[17] This result remains qualitatively unchanged if we run the analysis separately for individuals facing the choice between one fair trade chocolate bar and one, respectively two, bar(s) of conventional chocolate (see Tables 5 and 6). We therefore conclude that the change in the classification trade-off is not problematic.

Cents, $p < 0.001$). Higher earnings in the first stage also increase the stated fair trade premium; the effect is marginally significant (coefficient of 0.007, i.e. +0.7 Cents per €1 higher income, $p = 0.056$). The interaction between first stage profits and *public* is significantly negative with a coefficient of -0.012 ($p = 0.047$). This implies that in the public treatment, first stage profits have no significant effect. We confirm this in a separate regression conditioning on the treatment being public, where the coefficient of first stage profits is insignificant ($p > 0.1$).

Whereas the significance of income is not surprising as it illustrates a simple income effect,[18] at a first glance the irrelevance of income in the public treatment is. A possible explanation is that since income from the market game was higher if an individual behaved more unfairly, individuals with high profits shy away from stating very large willingness to pay in the public treatment so that they do not reveal their unfair behavior.

Our main results remain qualitatively unchanged if we exclude 17 individuals with "no demand", i.e., individuals who state a willingness to pay of less than 2 cents for each of the two types of chocolate (see Table 2, column 2). The main difference is that the profit from the first part has no significant impact anymore (and the interaction effect with the treatment dummy is significant only at the 10% level).[19]

To control for conceivable (though arguably unlikely) spill-over effects from the interaction in the market game, we run robustness checks clustering standard errors on the group level (with 36 independent groups, see Table 4, columns 3 and 4, in Appendix C) and with group-level random effects (see Table 4, columns 5 and 6, in Appendix C). The main results are unchanged.

[18] Indeed in the private treatment both w_{fair} and w_{conv} increase with the income from the first part of the experiment.

[19] One participant, a fair trade chooser in the public treatment, stated after the experiment to have accidentally swapped w_{fair} and w_{conv} (this is one of the two participants with a negative fair trade premium shown in Figure 3). The aggregate data and statistical analysis reported in this paper use the original data as he entered them, because some participants always make mistakes and it seems somewhat arbitrary to correct those that some participants report later to be mistakes. Nevertheless, we also performed robustness checks with the WTPs as he claimed he had wanted to enter them. The only difference we observe is that the significance of the impact of the profits from the first part is weaker in some of the regressions, but the impact of the profit is not our concern.

5 Explaining the heterogeneous treatment effect

As discussed in the previous section, our results that on average fair trade choosers do not react to the treatment whereas conventional choosers do, would be consistent with a negative correlation between individuals' intrinsic preferences for fair trade and their desire for a positive social image (compare Corollary 2). In this section, we discuss how this negative correlation between intrinsic motivation and image concerns can arise in a psychologically plausible way, and provide a tractable model. We then discuss alternative explanations and argue how they are inconsistent with our data.

5.1 An extended model with endogenous image concerns

A negative correlation between image concerns and intrinsic motivation is psychologically plausible because social-psychological research has shown that people are more prone to actively manage the impressions that others have about them if they, correctly or incorrectly, perceive that the image that they send differs from the image that they would like to send or that they consider socially acceptable (Leary and Kowalski, 1990). Specifically, "people with low self-esteem may respond to social pressure (and act less consistently with their inner compass) than do people with high self-esteem" (MacDonald and Leary, 2012, p.363). We translate this into an economic model where the concern for one's social image decreases in the individual's self-image.

Formally, the intuition that individuals may care more about their social image if they perceive a discrepancy between their own type and social ideal can be expressed by the following utility function:

$$v(f, p(f), R(a)) = uq(f) + fm + s(m) + \lambda(S - s(m))R(a) - p(f), \tag{7}$$

where $s(m)$ is the self-image (normalized to lie between 0 and S), which increases in own type.[20,21] Utility increases in self-image and in social image. However, the marginal utility of social image is decreasing in self-image. The optimal exaggeration Δ of the fair trade premium in the public treatment is then determined by the first-order condition

[20] In Appendix B, we show that the predictions are very similar if we assume that the person is not entirely sure about her own type, and the realized self-image depends on the stated fair trade premium.

[21] Self-image has been discussed in a broader literature of non-signaling models of morally motivated behaviors, e.g. Brekke et al. (2003).

$$\lambda(S - s(m))\frac{\partial R(m+\Delta)}{\partial \Delta} = c'(\Delta). \tag{8}$$

It is easy to see that the utility-maximizing Δ is falling in m and thus those who are more strongly intrinsically motivated will react less to the possibility of social image building, unless R' is (strongly) increasing. However, the quadratic cost function still implies that the optimal Δ is positive as long as $R' > 0$ which will hold in any separating equilibrium. Furthermore, the distribution of fair trade premiums should become more compressed as willingness-to-pay are publicized. This should also hold if we look separately at those who chose the fair trade chocolate bar and those who chose the conventional one. Our evidence regarding such a compression effect is mixed as discussed in Section 4 (see also Figure 3).

5.2 Alternative explanations

There are several perceivable alternative explanations for our results. If a positive image can be realized by revealing a positive fair trade premium above a certain threshold but this image does not further improve in the size of the premium, the fact that fair trade choosers do not increase their fair trade premium is not informative about their image concerns. However, if this was the case the fair trade premium should cluster at a specific positive level, which it does not (see Figure 3).

Another possible argument is that for fair trade choosers the increase in image from separating compared to pooling with the conventional choosers could be small, such that the fair trade choosers would not find it worthwhile to increase their fair trade premium to achieve separation (for example, because relatively few conventional choosers would pool with them, which would hence not dilute the image much). This argument requires a perception of a dichotomous distribution of fair trade and conventional choosers, and thus the variance in observed positive fair trade premiums speaks against this hypothesis.

Finally, some participants might see a public good character in fair trade but decide to free-ride in private. In public, however, they decide to contribute and thereby give an example to others who might follow them and contribute in the future because then their costs of contribution would have an effect beyond the one of their own consumption choice. Note that our experiment is one shot so that such future contributions would have to happen outside the lab.

6 Discussion and implications

Our experimental results demonstrate that indeed individuals may systematically differ in their image concerns, in contrast to the approach taken traditionally in the literature on image concerns and in particular conspicuous consumption. Therefore, it is important to take this heterogeneity into account in economic modeling and in policy recommendations. These results are in line with indirect evidence from several studies in other domains. For instance, Filippin et al. (2013) analyze tax morale in Italian cities and find that those intrinsically motivated are less affected by the possibility of having a negative social reputation (for withholding taxes). This finding is corroborated by field experimental evidence from Germany (Boyer et al., 2016; Dwenger et al., 2016). Results by Riedl and Smeets (2015) indicate that also among professional financial investors intrinsic motivation and social image concerns are negatively correlated. They show that "selfish" investors choose socially responsible mutual funds if these are not associated with tax benefits because they care about their image. Similar signaling motivations are not found among those who are classified as pro-social. Also in the domain of conspicuous consumption, empirical evidence in Charles et al. (2009) is consistent with a negative correlation between wealth and the desire to signal wealth. In contrast, Grossman (2015) in a laboratory experiment finds more compelling evidence for social signaling concerns if he excludes "selfish-types" and "money-maximizers", indicating a positive relation between image motivation and intrinsic giving in his sample, opposite to what we find.

While the accumulated evidence from the field is in line with the negative correlation we find, the absence of image concerns for the fair trade choosers in our experiment does not rule out that the same individuals exhibit concerns for their social image in other circumstances. But we expect our findings to generalize in ethical consumption settings, where we would expect that the more intrinsically motivated consumers are less affected by image concerns than those consumers who care little intrinsically.

Our results have implications for the optimal design of policy interventions that intend to direct consumers toward more ethical purchasing behavior, and also for the optimal design of marketing campaigns of private companies. Based on our findings, we expect that addressing the signaling desire of consumers who are not sufficiently intrinsically motivated to buy ethically can increase ethical consumption without having to fear an image-based crowding out of intrinsically motivated buyers. Since the intrinsically motivated subjects in our experiment are not influenced by social image

building opportunities, they would not be affected if the derived image is diluted because those not intrinsically motivated are encouraged by extrinsic incentives (such as image building opportunities but also, e.g., material rewards) to buy the same products. Hence extrinsic incentives are not likely to crowd out intrinsic motivation in our setting. Indeed, crowding-out effects of pro-social behavior have, to the best of our knowledge, not been observed in the context of ethical consumption, though they have been observed in non-market settings such as blood donations (Mellström and Johannesson, 2008; Lacetera and Macis, 2010). This suggests that the market for fair trade products can be enlarged by appealing to consumers' image concerns. But the long run effect of an publicity campaign for fair trade would have to take into account possible market responses. The negative correlation between intrinsic motivation and image concerns would suggest that the profit maximizing strategy of a monopolist will attempt to pool consumers who intrinsically value fair trade with those who only care about their image as may in fact be a result of fair trade products being served in specialty stores as well as discounters. In this case, increasing the value of the associated image through public campaigns may only increase prices without positive effects on farmer or consumer welfare (see Friedrichsen (2016) for the underlying theoretical results and further discussion).

We find that the elicited fair trade premiums and thereby total expected revenue for fair trade products increases with public exposure of individual decisions. The effect of increased public scrutiny on consumer welfare, however, depends on whether a positive act yields prestige or negative acts are stigmatized. In our study, the positive effect on the fair trade premium of the conventional choosers is mostly driven by an increase in their WTP for fair trade chocolate in line with our model and the assumption that social prestige is derived from support for fair trade. This is in contrast to findings by Teyssier et al. (2015) who find an increase in the fair trade premium driven by a decrease in the WTP for conventional chocolate. One possible explanation for this difference is how the framing of the choice situation is perceived by the individuals. If it is perceived as attaching a positive image to fair trade, individuals will expect an image gain when purchasing the fair trade product and increase their WTP for the fair trade option. If it is perceived as attaching a negative image to conventional products, individuals will expect a utility loss from purchasing conventional chocolate and reduce their WTP for conventional chocolate accordingly. Cappelen et al. (2016) find evidence for both social esteem and social pressure in a dictator game study.

7 Conclusion

We have addressed the heterogeneity in image concerns by studying the effect of opportunities for image building on fair trade premiums for experimental participants with different intrinsic motivation. We find that participants with low intrinsic motivation to buy fair trade react positively to image building opportunities, whereas those with high intrinsic motivation do not. This is inconsistent with a standard model of image concerns that assumes identical image concerns for all participants. Our results suggest instead that in our setting intrinsic motivation and image concerns are negatively correlated. We have developed a model that can capture this effect in a tractable way based on a psychologically plausible interaction between self image and social image. We have argued that taking this correlation into account is important for consumer policy and firm behavior.

References

Andreoni, J. and B. D. Bernheim (2009). Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects. *Econometrica 77*(5), 1607–1636.

Ariely, D., A. Bracha, and S. Meier (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review 99*(1), 544–555.

Becker, G., M. DeGroot, and J. Marschak (1964). Measuring utility by a single-response sequential method. *Behavioral Science 9*(3), 226–232.

Benabou, R. and J. Tirole (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies 70*(3), 489–520.

Bénabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *American Economic Review 96*(5), 1652–1678.

Boyer, P., N. Dwenger, and J. Rincke (2016). Do taxes crowd out intrinsic motivation? Field-experimental evidence from Germany. *Journal of Public Economics 144*, 140–153.

Brekke, K. A., S. Kverndokk, and K. Nyborg (2003). An economic model of moral motivation. *Journal of Public Economics 87*(9), 1967–1983.

Cappelen, A. W., T. Halvorsen, E. Ø. Sørensen, and B. Tungodden (2016). Face-saving or fair-minded: What motivates moral behavior? *Journal of the European Economic Essociation*. Forthcoming.

Carpenter, J. and C. K. Myers (2010). Why Volunteer? Evidence on the Role of Altruism, Image, and Incentives. *Journal of Public Economics 94*(11-12), 911–920.

Cason, T. N. and C. R. Plott (2014). Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy 122*(6), 1235–1270.

Charles, K. K., E. Hurst, and N. Roussanov (2009). Conspicuous consumption and race. *Quarterly Journal of Economics 124*(2), 425–467.

Dammert, A. C. and S. Mohan (2015). A survey of the economics of fair trade. *Journal of Economic Surveys 29*(5), 855–868.

Danz, D., D. Engelmann, and D. Kübler (2012). Do legal standards affect ethical concerns of consumers? An experiment on minimum wages. University of Mannheim Working Paper Series, No. 12-3.

De Pelsmacker, P., L. Driesen, and G. Rayp (2005). Do consumers care about ethics? willingness to pay for fair-trade coffee. *Journal of Consumer Affairs 39*(2), 363–385.

Dragusanu, R., D. Giovannucci, and N. Nunn (2014). The economics of fair trade. *Journal of Economic Perspectives 28*(3), 217–236.

Dwenger, N., H. Kleven, I. Rasul, and J. Rincke (2016). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany. *American Economic Journal: Economic Policy 8*(3), 203–232.

Filippin, A., C. V. Fiorio, and E. Viviano (2013). The effect of tax enforcement on tax morale. *European Journal of Political Economy 32*, 320–331.

Fischbacher, U. (2007). Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Friedrichsen, J. (2016). Signals sell: Designing a product line when consumers have social image concerns. WZB Discussion Paper SP II 2016–202, Berlin: WZB.

Glazer, A. and K. A. Konrad (1996). A signaling explanation for charity. *American Economic Review 86*(4), 1019–1028.

Gneezy, U. and A. Rustichini (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics 115*(3), 791–810.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with orsee. *Journal of the Economic Science Association 1*(1), 114–125.

Griskevicius, V., J. M. Tybur, and B. Van den Bergh (2010). Going green to be seen: Status, reputation, and conspicuous conservation. *Journal of Personality and Social Psychology 98*(3), 392–404.

Grossman, Z. (2015). Self-signaling and social-signaling in giving. *Journal of Economic Behavior & Organization 117*, 26–39.

Harbaugh, W. T. (1998). What do donations buy? A model of philanthropy based on prestige and warm glow. *Journal of Public Economics 67*(2), 269–284.

Hillman, A. L. (2010). Expressive behavior in economics and politics. *European Journal of Political Economy 26*(4), 403–418.

Hopkins, E. and T. Kornienko (2004). Running to keep in the same place: Consumer choice as a game of status. *American Economic Review 94*(4), 1085–1107.

Koppel, H. and G. G. Schulze (2013). The importance of the indirect transfer mechanism for consumer willingness to pay for fair trade products: Evidence from a natural field experiment. *Journal of Consumer Policy 36*(4), 369–387.

Lacetera, N. and M. Macis (2010). Do all material incentives for pro-social activities backfire? The response to cash and non-cash incentives for blood donations. *Journal of Economic Psychology 31*(4), 738 – 748.

Leary, M. R. and R. M. Kowalski (1990). Impression management: A literature review and two-component model. *Psychological Bulletin 107*(1), 34.

Loureiro, M. L. and J. Lotade (2005). Do fair trade and eco-labels in coffee wake up the consumer conscience? *Ecological Economics 53*(1), 129–138.

MacDonald, G. and M. R. Leary (2012). Individual differences in self-esteem. In M. R. Leary and J. P. Tangney (Eds.), *Handbook of Self and Identity*, Chapter 17, pp. 354–377. Guilford Press.

Mellström, C. and M. Johannesson (2008). Crowding out in blood donation: Was Titmuss right? *Journal of the European Economic Association 6*(4), 845–863.

Miller, G. F. (2009). *Spent: Sex, Evolution, and Consumer Behavior*. Viking Adult.

Munro, A. and M. Valente (2016). Green goods: Are they good or bad news for the environment? evidence from a laboratory experiment on impure public goods. *Environmental and Resource Economics 65*(2), 317–335.

Plott, C. R. and K. Zeiler (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.

Read, D., G. Loewenstein, and M. Rabin (1999). Choice bracketing. *Journal of Risk and Uncertainty 19*(1-3), 171–197.

Reinstein, D. and J. Song (2012). Efficient consumer altruism and fair trade products. *Journal of Economics & Management Strategy 21*(1), 213–241.

Riedl, A. and P. Smeets (2015). Why do investors hold socially responsible mutual funds? Mimeo, Maastricht University.

Sexton, S. E. and A. L. Sexton (2014). Conspicuous conservation: The prius halo and willingness to pay for environmental bona fides. *Journal of Environmental Economics and Management 67*(3), 303–317.

Soetevent, A. R. (2011). Payment choice, image motivation and contributions to charity: Evidence from a field experiment. *American Economic Journal: Economic Policy 3*(1), 180–205.

Teyssier, S., F. Etilé, and P. Combris (2015). Social-and self-image concerns in fair-trade consumption. *European Review of Agricultural Economics 42*(4), 579–606.

A Proofs

Proof of Proposition 1. By choosing her WTP such that the observed fair trade premium is a and the WTP for the conventional bar of chocolate is u, an observer infers that for this agent the following holds:

$$a = m + \frac{\rho}{k}\frac{\partial R(m+\Delta)}{\partial \Delta} + \varepsilon \tag{9}$$

Knowing that stated values may differ from true ones only by a mean-zero normally distributed noise term, the observer will expect that this agent has a fair trade preference equal to

$$E[m|a] = a - \frac{\rho}{k}\frac{\partial R(m+\Delta)}{\partial \Delta} \tag{10}$$

We use the observation that $\frac{\partial R(m+\Delta)}{\partial \Delta} = \frac{\partial R(a)}{\partial a}$, and that the social image is defined as the conditional expectation of an individual's fair trade preference m upon observation of her fair trade premium and WTP for the conventional chocolate, $R(a) = E[m|a, u]$. Then, the public inference constitutes a first-order differential equation

$$R(a) = a - \frac{\rho}{k}R'(a) \tag{11}$$

for which the generic solution is $R(a) = a - \frac{\rho}{k} + e^{-ak/\rho}C$. The agent's optimization problem is globally concave only for $C = 0$ so that the unique solution for our context is

$$R(a) = a - \frac{\rho}{k} \tag{12}$$

Hence $R'(a) = 1$ and $a = m + \frac{\rho}{k}$ by 9.

□

The argument in the proof of Proposition 1 is using that the base utility for chocolate u is observed and the inference is calculated as if it is done separately for any possible u, because in principle, in equilibrium the stated fair trade premium could depend on u. The result, however, is independent of u and hence the same inflation of image is obtained for any base utility of chocolate u and any level of intrinsic motivation m.[22]

Proof of Corollary 1. Assume that $R(\hat{a}) = \hat{a}$, i.e., upon observing a stated price premium the public infers this to be the individual's attitude toward fair trade. Such naive inferences would be accurate if applied to WTP that were stated in private or if individuals' WTP where unaffected by public exposure. The term on the left-hand side of (5) is positive for every $a = m + \Delta$, if $\rho > 0$. Thus, we would expect individuals of all types who care about social image to increase their stated WTP for fair trade in response to public exposure. So the distribution of WTP shifts upwards. Moreover, the marginal gain in reputation is constant for naive inferences so that all WTP are shifted upward by the same amount. □

Proof of Corollary 2. Consider first the case that the conventional choosers (those with lower m) have systematically lower concerns for social image ρ_{conv} than that of the fair trade choosers ρ_{fair}. Assume initially that the derivative of the inference function R' is again constant. This implies that the fair trade choosers will inflate their fair trade premium by Δ_{fair} such that $\rho_{\text{fair}} R'(m_{\text{fair}} + \Delta_{\text{fair}}) = k\Delta_{\text{fair}}$ and conventional choosers by Δ_{conv} such that $\rho_{\text{conv}} R'(m_{\text{conv}} + \Delta_{\text{conv}}) = k\Delta_{\text{conv}}$. For R' constant this would imply that $\Delta_{\text{fair}} > \Delta_{\text{conv}}$. Then, however, because fair trade choosers have a higher m, the equilibrium inference would have to discount high fair trade premiums more, such that R is concave and not linear as assumed. However, the concavity of R cannot be so extreme that $\Delta_{\text{fair}} \leq \Delta_{\text{conv}}$ because then in equilibrium inference R would have to take this into account and not be concave. Thus in equilibrium if fair trade choosers care more about social image than conventional choosers, they will inflate their fair trade premium more than conventional choosers, but the difference will be smaller than if R' was constant. The same logic applies to the case that conventional choosers have systematically higher concerns for social image. Then equilibrium inference R has to be convex, but not to a degree such that $\Delta_{\text{conv}} \leq \Delta_{\text{fair}}$ and as a result conventional choosers will inflate their fair trade premium more than fair trade choosers but again the difference will be smaller than it would be if equilibrium inference R' was constant. □

[22] Note that in the experiment, participants could not influence which type of chocolate they would obtain, but only whether they would obtain the randomly selected type. Hence the trade-off they make is between the price they would pay for the fair trade chocolate and the image they might gain as a result. This trade-off is not influenced by how much the size difference matters for them.

B Extension to imperfect self-knowledge

Here we show that the conclusion from the self-knowledge case presented in Section 5.1 remains true if we assume that individuals are unsure of their type. In particular, we allow the self-image to positively depend on the stated fair trade premium.

Then, utility is given by the following equation

$$v(f, p(f), R(a)) = uq(f) + fm + s(m, a) + \lambda(S - s(m, a))R(a) - p(f),$$

where $s(m, a)$ is the self-image (normalized to lie between 0 and S), which increases in own type and also increases in the stated fair trade premium.

Utility increases in self-image and in social image. However, the marginal utility of social image is decreasing in self-image. The optimal exaggeration of the fair trade premium in the public treatment is then determined by the first-order condition

$$\frac{\partial s(m, m+\Delta)}{\partial \Delta} - \lambda \frac{\partial s(m, m+\Delta)}{\partial \Delta} R(m+\Delta) + \lambda(S - s(m, m+\Delta)) \frac{\partial R(m+\Delta)}{\partial \Delta} = c'(\Delta). \quad (13)$$

As the self-image is imperfect, an increase in the stated fair trade premium improves the individual's self image, $\partial s(m, m+\Delta)/\partial \Delta \geq 0$. Then for any given level of motivation m, the first term on the left-hand side of (13) is positive and represents the direct effect of an improved self-image on utility. This effect is present also in the private elicitation though, so that it would not count into our treatment effect. The second term is negative and represents the decrease in the utility from social image due to increases in self-image. The third term is positive, but decreasing in m and captures that marginal utility from social image is smaller for larger self-image. The total effect of an increase in the stated fair trade premium on utility depends on the relative importance of self- and social signaling motivations.

In the light of our experimental design, we are mostly interested in the effect of making choices public. To that respect, the extension to include uncertainty about one's own type does not qualitatively affect our predictions. Assuming that both s and R have constant slopes, it becomes apparent that individuals with a higher level of intrinsic motivation will react less strongly to the publicity treatment for two reasons. Like in the case with self-knowledge, individuals with higher levels of intrinsic motivation have a lower marginal utility from the social image. In addition, increasing the stated fair trade premium in response to publicity improves an individual's self-image and thus lowers the utility from the social image. This marginal loss in utility is higher for individuals with higher intrinsic motivation because they expect (and realize) a higher social image.

If the loss in marginal utility of social image that is induced by inflating the stated WTP differential and the associated increase in the self-image is larger than the marginal increase in utility from social image (i.e., if the second term in (13) is absolutely larger than the third term), then an individual may even state a lower fair trade premium in public than in private. This is more likely to be the case the higher is the intrinsic prefer-

ence for fair trade because then both the self-image and the social image are high. Such an effect would, however, be inconsistent with a separating equilibrium and the standard assumption of a differentiable image function for a continuous type distribution.

C Additional results

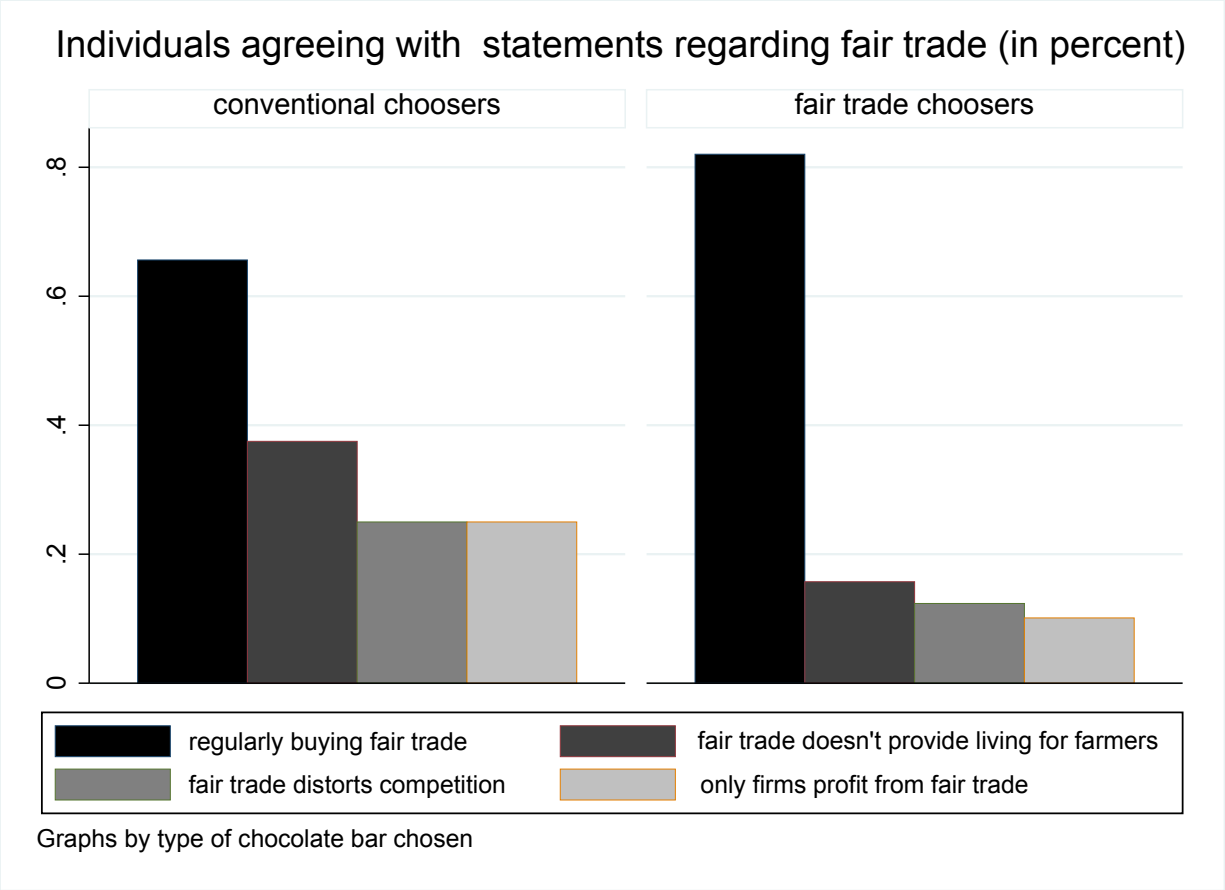

Figure 4: The choice of fair trade chocolate is associated with a higher frequency of buying fair trade and a more positive attitudes towards fair trade.

Table 3: Regression of the stated fair trade premium $a = w_{\text{fair}} - w_{\text{conv}}$ on (ex ante) fair trade choice, profits from the first part of the experiment, treatment (public) and interaction terms. The sample is split according to chocolate choice. Column 2 and 4 exclude subjects who bid less than 2 cents for each type of chocolate. Standard errors in parentheses.

	fair trade		conventional	
	all	pos. demand	all	pos. demand
	(1)	(2)	(3)	(4)
public	0.014	0.043	0.258**	0.267**
	(0.100)	(0.117)	(0.111)	(0.123)
marketprofit	0.008*	0.006	0.004	0.004
	(0.005)	(0.005)	(0.006)	(0.007)
marketprofit*public	-0.013	-0.015	-0.008	-0.008
	(0.009)	(0.010)	(0.008)	(0.009)
Observations	89	76	32	28
R^2	0.071	0.054	0.182	0.194
Prob > F	0.098	0.262	0.126	0.152

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Regression of the stated fair trade premium $a = w_{\text{fair}} - w_{\text{conv}}$ on (ex ante) fair trade choice, profits from the first part of the experiment, treatment (public) and interaction terms. Columns 1 and 2 are the benchmark specifications (also reported in the main text). Columns 3 and 4 account for standard errors clustering at the group level, columns 5 and 6 include group-level random effects. In columns 2, 4, and 6 we exclude subjects who bid less than 2 cents for each type of chocolate. Standard errors in parentheses. (a) R^2: within$= 0.1391$, between $= 0.1217$, overall $= 0.1503$. (b) R^2: within $= 0.1983$, between $= 0.0075$, overall $= 0.1556$

	standard		clustered SE		group-level RE	
	all	pos. demand	all	pos. demand	all	pos. demand
FTchoice	0.313***	0.341***	0.313***	0.341***	0.313***	0.336***
	(0.085)	(0.093)	(0.060)	(0.060)	(0.085)	(0.093)
public	0.292**	0.306**	0.292**	0.306**	0.292**	0.288**
	(0.122)	(0.133)	(0.109)	(0.114)	(0.122)	(0.137)
FTchoice*public	-0.290**	-0.289**	-0.290***	-0.289**	-0.290**	-0.260*
	(0.119)	(0.131)	(0.106)	(0.111)	(0.119)	(0.133)
marketprofit	0.007*	0.006	0.007	0.006	0.007*	0.006
	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)
marketprofit*public	-0.012**	-0.012*	-0.012*	-0.012*	-0.012**	-0.011*
	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)
Observations	121	104	121	104	121	104
R^2	0.150	0.156	0.150	0.156	(a)	(b)
Prob> F	0.002	0.005	0.000	0.000	0.001	0.003

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Regression of the stated fair trade premium $a = w_{\text{fair}} - w_{\text{conv}}$ on (ex ante) fair trade choice, profits from the first part of the experiment, treatment (public) and interaction terms. Columns 1, 3, and 5: subsample of individuals who chose between one fair trade chocolate bar and one conventional one. Columns 2, 4, and 6: subsample of individuals who chose between one fair trade chocolate bar and two conventional ones. Columns 3 and 4 account for standard errors clustering at the group level, columns 5 and 6 include group-level random effects. Standard errors in parentheses. (a) R^2: within$= 0.1985$, between $= 0.1757$, overall $= 0.2076$. (b) R^2: within $= 0.1642$, between $= 0.1103$, overall $= 0.1711$, (c): Prob $> \chi^2$.

	standard		clustered SE		group-level RE	
	1–to-1	2-to-1	1-to-1	2-to-1	1-to-1	2-to-1
FTchoice	0.293	0.251***	0.293**	0.251***	0.293	0.251***
	(0.245)	(0.082)	(0.138)	(0.060)	(0.245)	(0.082)
public	0.308	0.198*	0.308**	0.198	0.308	0.198*
	(0.299)	(0.117)	(0.116)	(0.138)	(0.299)	(0.117)
FTchoice*public	-0.259	-0.202*	-0.259	-0.202	-0.259	-0.202*
	(0.297)	(0.118)	(0.166)	(0.138)	(0.297)	(0.118)
marketprofit	0.015**	-0.001	0.015*	-0.001	0.015**	-0.001
	(0.006)	(0.004)	(0.008)	(0.003)	(0.006)	(0.004)
marketprofit*public	-0.023**	-0.001	-0.023**	-0.001	-0.023**	-0.001
	(0.011)	(0.006)	(0.010)	(0.005)	(0.011)	(0.006)
Observations	57	64	57	64	57	64
R^2	0.208	0.171	0.208	0.171	(a)	(b)
Prob> F	0.032	0.048	0.030	0.006	0.0202(c)	0.0351(c)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Regression of the stated fair trade premium $a = w_{\text{fair}} - w_{\text{conv}}$ on (ex ante) fair trade choice, profits from the first part of the experiment, treatment (public) and interaction terms, excluding subjects who bid less than 2 cents for each type of chocolate. Columns 1, 3, and 5: subsample of individuals who chose between one fair trade chocolate bar and one conventional one. Columns 2, 4, and 6: subsample of individuals who chose between one fair trade chocolate bar and two conventional ones. Columns 3 and 4 account for standard errors clustering at the group level, columns 5 and 6 include group-level random effects. Standard errors in parentheses. (a) R^2: within$= 0.1937$, between $= 0.1706$, overall $= 0.2002$. (b) R^2: within $= 0.3828$, between $= 0.0002$, overall $= 0.2265$

	standard		clustered SE		group-level RE	
	1–to-1	2-to-1	1-to-1	2-to-1	1-to-1	2-to-1
FTchoice	0.347	0.255***	0.347**	0.255***	0.347	0.257***
	(0.265)	(0.086)	(0.136)	(0.063)	(0.264)	(0.081)
public	0.299	0.240*	0.300**	0.240	0.300	0.224
	(0.319)	(0.125)	(0.114)	(0.149)	(0.319)	(0.139)
FTchoice*public	-0.298	-0.152	-0.298*	-0.152	-0.298	-0.096
	(0.320)	(0.128)	(0.170)	(0.138)	(0.320)	(0.128)
marketprofit	0.012	-0.001	0.012	-0.001	0.012*	-0.002
	(0.007)	(0.004)	(0.009)	(0.003)	(0.007)	(0.004)
marketprofit*public	-0.021	-0.002	-0.021*	-0.002	-0.021*	-0.003
	(0.013)	(0.007)	(0.010)	(0.006)	(0.013)	(0.007)
Observations	50	54	50	54	50	54
R^2	0.2000	0.232	0.200	0.232	(a)	(b)
Prob> F	0.071	0.023	0.0170	0.002	0.0512(c)	0.0055(c)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D Instructions for the second part of the experiment

Below is the English translation of original instructions for the second part of the experiment, which were in German. The instructions for the market game (first part of the experiment) are omitted as we do not analyze that part of the experiment here. The instructions for the two treatments only differed in the last paragraph as indicated.

Instructions for the second part of the experiment

In the second part of the experiment, you make two simple decisions and answer a brief questionnaire.

For the second part of the experiment, you get €4 in addition to your earnings from the first part. You can spend part of these €4 to purchase a bar of chocolate.

Purchase of chocolate

The potential purchasing of chocolate takes place according to the following mechanism:

- There are two types of chocolate, one is fair trade, the other one conventional. Both will be shown to you before you make your decision.
- Please state your maximal willingness-to-pay for each type of chocolate on the screen. Your willingness-to-pay must lie between €0 and €2 and you can choose any amount in Cents in this interval.
- Only your willingness-to-pay for one of the two types of chocolate will be payoff-relevant in the end. Thus, you will get at most one bar of chocolate. The relevant type of chocolate is determined randomly and you will learn only after you have made your decisions which one it is. The same type of chocolate is payoff-relevant for all participants.
- Before it has been determined which type of chocolate is payoff-relevant, the price for the chocolate bar is drawn at random. This price is the same for all participants and is independent of the type of chocolate. The price can be between €0 and €2 and any amount in Cents is equally likely.
- If your stated maximal willingness-to-pay for the relevant type of chocolate is at least as high as the randomly drawn price, you obtain one bar of this type of chocolate and the price is subtracted from the €4 that you were endowed with for the second part of the experiment. If your stated maximal willingness-to-pay is lower than the randomly drawn price, you will not get a bar of chocolate and you do not pay anything; thus you keep your €4.

Please note that with your stated willingness-to-pay you cannot influence the price of the chocolate but only whether you will get a bar. Therefore, you should indeed state how much you would pay at most for the respective type of chocolate. Then you get the chocolate when-ever you do not have to pay more for it than what you are at most

willing to pay and you do not get a chocolate whenever you would have to pay more than your maximal willingness-to-pay.

Example 1: You state a maximal WTP of €0.13 for fair trade chocolate and of €1.93 for conventional chocolate. Suppose the randomly determined price is €0.78. If fair trade is drawn to be payoff-relevant, you obtain €4 but no chocolate bar because your stated maximal willingness-to-pay of €0.13 is lower than the price of €0.78. If the conventional chocolate is chosen to be payoff-relevant instead, you get a bar of conventional chocolate and you pay €0.78. In this case, you will get €4 - €0.78 = €3.22 and a bar of conventional chocolate for this part of the experiment.

Example 2: You state a maximal WTP of €1.34 for fair trade and of €0.62 for conventional chocolate. Suppose the randomly determined price is €0.44. If fair trade is chosen to be payoff-relevant, you get a bar of fair trade chocolate and you pay €0.44. In this case, you obtain €4 - €0.44 = €3.56 and a bar of fair trade chocolate for this part of the experiment. If the conventional chocolate is chosen to be payoff-relevant instead, you get a bar of conventional chocolate and you pay €0.44. In this case, you obtain €4 - €0.44 = €3.56 and a bar of conventional chocolate for this part of the experiment.

From the examples you see that you cannot influence the price through your stated maxi-mal willingness-to-pay (in example 2, you pay the same price in both cases even though the maximal willingness-to-pay is different) but only whether you obtain a bar of chocolate (example 1).

Questionnaire

After all participants have entered their willingness-to-pay for both types of chocolate, we ask you to fill in a brief questionnaire on the screen.

Distribution of chocolate (private treatment)

After filling in the questionnaire you will first be informed about the randomly drawn price and be reminded of your stated willingness-to-pay for both types. You will then be informed which type of chocolate is payoff-relevant and whether you will obtain a bar. At the end of the experiment you receive your chocolate, if you get one, in the adjoining room together with the money that you earned in the experiment. None of the other participants will learn whether you receive a bar of chocolate, your willingness-to-pay for it or how much money you obtain.

Distribution of chocolate (public treatment)

After filling in the questionnaire you will first be informed about the randomly drawn price and be reminded of your stated willingness-to-pay for both types. Each of you will then be asked individually to announce your stated maximal willingness-to-pay for the both types of chocolate.

You will then be informed which type of chocolate is payoff-relevant and whether you will obtain a bar.

If your stated willingness-to-pay for the payoff-relevant type is at least as high as the price, you will be asked to come forward to pick up your bar of chocolate.

You will receive the money that you earned in the experiment in the adjoining room. None of the other participants will learn how much money you obtain.