Consumers' willingness to pay for plants protected by beneficial insects—Evidence from two stated-choice experiments with different subject pools

Mira Lehberger and Sven Grüner

Accepted Manuscript (Food Policy)

Mira Lehberger, Department of Fresh Produce Logistics, Horticultural Economics, Hochschule Geisenheim University, Geisenheim, Germany

Corresponding author, contact: mira.lehberger@hs-gm.de, +49 6722 502 89732

Postal Address: Von-Lade-Straße 1, 65366 Geisenheim

Sven Grüner, Institute of Agricultural Sciences, Agribusiness Management, Martin Luther University Halle-Wittenberg, Halle, Germany

Contract: sven.gruener@landw.uni-halle.de, +49 345 55 22 368

Postal Address: Karl-Freiherr-von-Fritsch-Str. 4, 06120 Halle (Saale)

Please cite as:

Lehberger, M., Grüner, S. (2021): Consumers' willingness to pay for plants protected by beneficial insects – Evidence from two stated-choice experiments with different subject pools. Food Policy, 102, 102100. <https://doi.org/10.1016/j.foodpol.2021.102100>

Consumers' willingness to pay for plants protected by beneficial insects – Evidence from two stated-choice experiments with different subject pools

Abstract

We investigate consumers' willingness to pay for edible and non-edible plants protected by beneficial insects and those protected by pesticides, respectively, as well as the influence of message framing on willingness to pay. We use basil, heather, and orchids as experimental objects in two stated-choice experiments (a lab-in-the-field experiment and an online experiment) conducted in Germany. In both experiments, we find strong evidence that consumers prefer plants protected by beneficial insects. Moreover, we find that message framing affects consumers' willingness to pay, where gain framing produces greater overall willingness to pay. To check for the robustness of our results, we implemented instruments to mitigate concerns for social desirability and hypothetical bias, finding that our results are robust in terms of our key findings. This is the first comprehensive experimental study specifically on consumer's willingness to pay for plants protected by beneficial insects. Our findings are valuable for practitioners, as the use of beneficial insects in greenhouse production is already rather common. Following our results, this information could be successfully used for marketing purposes.

Keywords

Biological control, framing experiments, willingness to pay, replication study, social desirability, hypothetical bias

1 Introduction

Biological control is commonly defined as the use of natural enemies for insect, weed, and disease management, one instance of which is beneficial insects (e.g., Barratt et al., 2018). A main aim of bio-based pest-management practices, including beneficial insects, is to reduce reliance on synthetic pesticides for plant protection (USDA, n.d.). Generally, the effectiveness of biological control is well established. For instance, a study on hot-spot application of beneficial insects in large-scale commercial rose farms in Kenya provides evidence that the use of insecticides (acaricides) can be drastically reduced without significantly reducing the number of harvested stems (Gacheri et al., 2015). This may be interesting information for consumers, as the public is very skeptical of the use of pesticides, as well as their potential residues in products. In Germany, Koch et al. (2017) find that consumers view pesticides as among the greatest threats to food quality and safety. Römer et al. (2019) find that the public's explicit (i.e., conscious) attitude, as well as the implicit (i.e., occurring without conscious awareness) attitude toward chemical crop protection, is negative. Several studies from outside of Germany further indicate that consumers prefer products that are free from pesticides (e.g., Edenbrandt et al., 2018; Khan et al., 2018). However, consumer attitudes and behaviors toward products protected by beneficial insects are less clear. Indeed, to the best of our knowledge, most practitioners in Germany do not inform consumers about this plant protection practice. First evidence from Germany shows that while consumers prefer beneficial insects over pesticides as a plant protection practice, they also have a low level of knowledge on both practices (Lehberger and Becker, 2020). Experimental results from North America suggests that U.S. consumers prefer floricultural crops labeled with 'use of beneficial insects' over those labeled with 'traditional insect control' (Wollaeger et al., 2015). Results from a Canadian survey by McNeil et al. (2010), regarding risk perception of biological control, indicate that participants consider biological control as a safer acceptable alternative to traditional pesticides for the production of food crops.

There are currently no registration requirements in Germany for the use of beneficial insects and, thus, official data on the application of beneficial insects are rather sparse. Following a survey by

the Federal Research Centre for Cultivated Plants (Koch et al., 2019), the greatest use of beneficial insects occurs in protected areas of vegetable and ornamental production, as well as in maize production. One of the most commonly used beneficial insects is the ichneumon wasp (*Encarsia Formosa*), which combats white flies on an estimated 67% of the glasshouse tomato production area and 16% of the glasshouse ornamental production area. In maize production, *Trichogramma brassicae* are used to fight the European corn borer. Relatively speaking, they are most often applied in the federal state of Baden-Wuerttemberg on approximately 20% of the maize production area. Here, the practice is subsidized (Ministry for Rural Areas and Consumer Protection Baden-Württemberg, 2017). Overall, the use of beneficial insects in Germany is highly dependent on the culture in focus, the specific pest infestation, the cultivation methods, and the governmental support in the region. An indication of the rising demand for beneficial insects is that the selection of commercially available beneficial insects in Germany has increased from below 10 in 1990 to more than 80 in 2016 (Koch et al., 2019).

While one may assume from the evidence above that informing consumers about the use of beneficial insects instead of synthetic-chemical insecticides is a promising marketing strategy, the exact information provided to consumers can be pivotal for its success. Findings on new technology suggest that the framing of information about the product can substantially affect perceptions, attitudes, and willingness to pay (WTP) (Britwum and Yiannaka, 2019; Thomas et al., 2019).¹ Kahneman and Tversky (1979) have popularized the concept of framing, an empirical violation of the invariance axiom. In contrast to rational choice predictions, different representation of an otherwise identical choice influences perceptions and/or behaviors (Tversky and Kahneman, 1986). In their famous *Asian Disease Problem*, Tversky and Kahneman (1981) describe a situation in which 600 people would be killed without any measures. Two programs are possible: program A is formulated in terms of ‘saved lives,’ whereas program B is framed as ‘people will die.’ In the literature, the two versions have since commonly been labeled as ‘gain and loss frames.’ Kahneman and Tversky (1979) define gains and losses relative to some neutral reference point. Reference points vary from person to person and are context-dependent (e.g., individuals’ expectations, status quo, goals). Researchers have developed several framing typologies. For example, Levin et al. (1998) distinguish between risky choices (i.e., comparison of choices for risky options), attributes (i.e., comparison of attractiveness ratings for the single item), and goals (i.e., comparison of rates of adoption of the behavior). Framing effects on behavior have been studied substantially in the field of health as well as in other economic contexts (Fiedler and Hillenbrand, 2020). Overall, reviews (Rothman et al., 2006; Wansink and Pope, 2015) as well as meta-analyses (Nabi et al., 2020; O’Keefe and Jensen, 2009) indicate that the effect of framing depends on a multitude of factors, such as person-specific, as well as outcome and behavior specific factors. Thus, it appears that neither message strategy can be viewed as dominant. However, in terms of consumer behavior, Britwum and Yiannaka (2019) show that a loss-framed message induced the highest level of WTP for two food-safety-enhancing technologies for ground beef. A tendency to weigh losses more heavily than gains has been coined by behavioral scientists as ‘loss aversion’ (Tversky and Kahneman, 1991). Contrary to predictions in accordance with loss aversion, Segev et al. (2015) find that in the context of green advertising, gain frames elicit more favorable consumer responses than loss frames.

In light of the described theoretical and empirical background this paper tackles two central research questions:

¹ While the use of beneficial insects is not a new technology, consumers are likely not very familiar with this plant protection practice. Indeed, Lehberger and Becker (2020) find that the vast majority of surveyed consumers have not heard of beneficial insects in context of food production.

1. To what extent do consumers prefer edible and non-edible (i.e., ornamental) plants that have been protected by beneficial insects over those protected by pesticides?
2. How does message framing influence consumers' WTP for these products?

We aim to answer these questions by carrying out two WTP experiments. We use edible and non-edible plants research objects, due to three main considerations: (i) In Germany, as elsewhere, these types of plants are rather often produced with beneficial insects and thus the research objects are realistic. (ii) While there is ample evidence on consumer behavior regarding certain types of plant protection regimes (such as organic or GM-free), this research typically focuses on edible-products (for a review on organic food consumption behavior, see e.g., Rana and Paul, 2017). While there is some evidence, that consumer also care about plant protection practices when it comes to non-edible ornamental plants (e.g., Getter et al., 2016; Wollaeger et al., 2015), consumer research in this field typically focuses on the visual or fragrance aspects of these products (e.g., Grygorczyk et al., 2016). (iii) Overall, the choice of research objects and overall setup allows for many insightful comparisons, i.e., by looking at plant types (edible versus non-edible), framing themes (loss versus gain) and plant protection practices (beneficial insects versus pesticides) relevant for consumer and behavioral researches and others who work in the field of food and non-food products.

We specifically used basil, heather, and orchid as experimental objects, as basil is currently the second most consumed herb in Germany (BVEO 2019), and consumers spend the most on heather, as an outdoor (bed and balcony) plant (AMI, 2019a), and orchids, as a blossoming indoor plant (AMI, 2019b). Inspired by experimental papers in psychology (e.g., Brainerd et al., 2017; Chu and Kita, 2011; Gray et al., 2014), we refrain from reporting only one single experiment (which is 'business as usual' in the majority of experimental papers in economics). There are at least two good reasons for this procedure. First, our initial experiment may generate research questions that we want to test with new data. Second, we check the robustness of our overall study results by comparing the findings of two similar experiments in completely different settings. The first experiment was carried out as a lab-in-the-field experiment, whereas the second one was an online experiment.² Due to the different recruiting procedures, the experimental subjects of both experiments vary considerably, which helps to increase the external validity of our overall findings if both experiments come to similar findings. Recent findings have shown insufficient replicability of experimental findings in the social sciences, sometimes referred to as 'replication crisis' (e.g., Camerer et al., 2016; Duvendack et al., 2015).

2 Experiment 1: Lab-in-the-field experiment with the interested public

2.1 Study design

Data collection took place at the biannual 'Open Campus' of the Hochschule Geisenheim University (Germany) in September 2019, where the public is welcome over two days to visit the campus and the institutes present their work. Visitors³ were invited to take part in the experiment, and if they agreed, they were asked to a separate room, where the experiment took place. Participants were randomly assigned to either experimental group 1 or 2. Both groups were shown the same three plants (basil, heather, orchid), but we placed a sign with either a gain or a loss message regarding plant protection practices in front of the plants. Following Tversky and Kahneman (1981) as well as

² Following Charness et al. (2013), we conduct two extra-laboratory experiments. Charness et al. (2013) use the terminology for experiments that deviates from standard experiments in the lab with students. In both experiments, we use nonstandard populations that deviate from the lab environment. Experiment 1 is carried out as a lab-in-the-field experiment and experiment 2 is Internet-based.

³ For reasons of practicability, couples and small visiting groups were invited together and were accordingly assigned to the same experimental group. Prior to each questioning, we randomized the set-up of plants.

many more recent works on framing in the realm of consumer behavior (Segev et al., 2015), we included no control group in our design.⁴ The signs read as follows:

Framing 1 (gain): "These plants were protected using beneficial insects. If you choose these plants, on average approx. 75% pesticides against pests are saved in production. To put it differently: Pesticide use for insect management will be lowered from 20kg/ha/year (without beneficial insects) to 5kg/ha/year (with beneficial insects)."

Framing 2 (loss): "These plants were protected using beneficial insects. If you do not choose these plants, on average approx. 300% more pesticides against pests will be used in production. To put it differently: Pesticide use for insect management will be raised from 5kg/ha/year (with beneficial insects) to 20kg/ha/year (without beneficial insects)."

While in general the potential of beneficial insects to reduce or even replace pesticides when fighting insect pests is well established (Gacheri et al., 2015; Opit et al., 2009), the exact numbers on the signs are not factual. This particular information is, to the best of our knowledge, not available.⁵ Following the typology by Levin et al. (1998) mentioned above, we use attribute labels to describe our gain and loss variants. Since Kahneman and Tversky (1979) stress that people think in terms of changes (instead of final assets), we provide the subjects with changes, both numerical and as percentages. We speak of increased and decreased use of pesticide (loss and gain framing, respectively) on the one hand and, on the other, give the percentage changes. The percentage changes are only a description of the numerical changes ($5 \rightarrow 20 = +300\%$ and $20 \rightarrow 5 = -75\%$).⁶ After reading the framing, participants were to answer a questionnaire, which consisted of three parts. The first part was to elicit consumers' WTP principally by these questions:

"I would pay € _____ for the basil [respectively heather and orchid] shown, where plant protection primarily relies on the use of beneficial organisms."

"I would pay € _____ for an identical basil [respectively heather and orchid] where plant protection does not rely on the use of beneficial organisms but only on pesticides."

The order of the WTP questions was randomized. We decided in favor of this type of open-ended contingent valuation, as we tried to keep the design as simple as possible because more complicated methods can lead to a decline in intrinsic motivation in the course of the experiment (e.g., Frey 2017). This holds especially for the relatively older participants in experiment 1. More complicated designs have proved to influence the willingness to participate (Asioli et al., 2020) and, as a consequence, may lead to selection-bias problems. We also decided in favor of the open-ended contingent valuation method because it is more practical and easier to understand (e.g., Hofstetter et al., 2020) and widely used in the field (e.g., Shi et al., 2014). Moreover, Shi et al. (2014) cite studies

⁴ While our main reason for not including a control group was to be consistent with the original literature, another reason was concern about the size of our sample. This was ex-ante difficult to predict, and an additional control group would call for a larger sample size to allow for meaningful group comparisons.

⁵ Pesticide use and reduction potential depend heavily on a multitude of factors (e.g., cultivated plant, pest infestation, production system, climate, regulations). For instance, Gacheri et al., (2015) found that the use of hot-spot applications of beneficial insects to fight spider mites in Kenyan rose production can reduce the number of treatments with acaricides substantially, but are also related to an increased application of other insecticides to fight other insect pests. Due to the lack of data regarding our products, it is impossible to determine whether the numbers on the signs are accurate approximations for either basil/heather/orchid production in general or, specifically, their production in Germany. However, a number of experts in plant protection viewed the numbers that our signs displayed and judged them as noncritical.

⁶ Within our study, we refer to information about the reduction in pesticide use as "gain" and information about the increase in pesticide use as "loss." The paper discusses whether this gain framed message/loss framed message dichotomy is already sufficient to alter individual's preferences. However, to be more precisely, we effectively have two reference points in our study design: 5kg/ha use of pesticide for loss framed messages (which can increase to 20kg/ha if beneficial insects are not used) and 20kg/ha use of pesticide for gain-framed messages (with a potential to reduce to 5kg/ha if beneficial insects are used).

that found similarities in the WTP resulting from close-ended and open-ended CVs in private goods. In addition, open-ended procedures do not suffer from anchoring effects (Kato and Hidaono, 2007).⁷ In the second part of the questionnaire, we elicited participants' knowledge and risk perception of pesticide and biological control, as well as their consumption behavior regarding plants and organic products. With the third part of the questionnaire we gathered socio-demographic information.

To increase the number of subjects, we incentivized the participants with a hedonic reward, in the form of sweets handed out after the experiment.

2.2 Randomization checks and data cleaning

As we were interested in the causal relationship between the framing variants and the subjects' WTP, we opted for randomized trials.⁸ Randomization means that subjects are randomly assigned to one of the groups to control for known and unknown confounders (Dunning, 2012). In general, the randomization process can be viewed as successful when the characteristics of the two framing variants are very similar to each other (e.g., the age or gender of the subjects in the two framing scenarios). However, the best measure of such success involves some controversy. Altman (1985: 126) argues that significance tests of baseline comparisons in randomized trials are at best tautological because: "performing a significance test to compare baseline variables is to assess the probability of something having occurred by chance when we know that it did occur by chance." Thus, we opt for standardized differences, as Schober et al. (2018) proposes. Following Cohen (1988), we distinguish between small (Effect Size Index=0.2), medium (Effect Size Index=0.5), and large effect sizes (Effect Size Index=0.8).

To ensure the quality of the data, we decided to do some plausibility checking of the answers. As individuals are boundedly rational (Simon, 1957), careless mistakes may happen. Thus, we had to consider some data cleaning to avoid extreme outliers distorting the results. This is not an easy task because there are rarely standardized ways to deal with a specific problem. In light of plausibility, we decided to clean the data in the following way:

1. As organic basil can be purchased from almost any supermarket in Germany for around €1 to €2, subjects who indicated a willingness to pay for basil at €10 or more were removed from the sample.
2. The type of heather used in the experiment can be purchased for around €3, however heather is only sometimes available at supermarkets and must typically be purchased in special outlets. Thus, consumers are assumingly less aware of the market value of heather. Thus, we allowed for more absolute variation in stated WTP. Subjects who indicated a willingness to pay at €30 or more were removed from the sample.

We decided against a cutoff WTP for orchids, as prices of these plants can vary immensely, depending on the variety and the outlet. Additionally, no participant left after considering criteria 1 and 2 indicated a WTP higher than €30 and thus no extreme outlier was detected.

Due to the cutoffs, 2 participants from the initial 91 participants, who filled out the questionnaire, were eliminated from the analysis regarding experiment 1, leaving a dataset of $n = 89$.

2.3 Results and discussion

Table 1 provides information about the characteristics of the participants overall and the two framing variants. On average, the experimental subjects were slightly older than 48 years. The age range of the subjects (18 to 82) illustrates their heterogeneity. Slightly more than half were female. None of the participants associated themselves with the third sex identifier ('other'). Over 80%

⁷ As one reviewer correctly points out, open-ended approaches are not free of criticism. Among the issues raised are incentive-compatibility and subjects' focus on costs instead of WTP (Schkade and Payne, 1994).

⁸ Note that there are other types of experiments throughout the literature. For example, the procedure of Holt and Laury (2002) to elicit the individuals' risk attitude does not require randomization. It is about measuring a parameter and not aimed at identifying causal relationships.

stated that they were not enrolled in a university degree program. Of those who had graduated or were enrolled in a degree program, over 40% indicated that their degree program was related to the agricultural, food, or horticultural sector. The vast majority (>80%) of the subjects were from the federal state (FS) of Hesse, where the experiment took place.

Table 1: Characteristics of the participants ^(a,b)

Characteristic	Specification	Frequency	%	Gain	Loss
Framing	Gain	46	51.69		
	Loss	43	48.31		
Gender	Female	51	57.95	54.35	61.90
	Male	37	42.05	45.65	38.10
Student	No	63	81.82	84.62	78.95
	Yes	14	18.18	15.38	21.05
Degree in agricultural, food or horticultural sector	No	68	76.40	78.16	74.42
	Yes	21	23.60	21.74	25.58
Residence in the federal state of Hesse	No	17	19.10	21.74	16.28
	Yes	72	80.90	78.26	83.72
Age	Mean years (Std. Dev.)	89	48.57 (16.48)	46.78 (16.36)	50.49 (16.58)

(a) A comparison of the framing variants with respect to the standardized differences (cf., Appendix 2 for more information) shows no large effect sizes, but medium effect sizes (age = 0.23; Degree program: Agricultural, food or horticultural sector: 0.09). All other variables can be associated with standardized differences that are smaller than 0.2. In general, the standardized differences indicate that the randomization did well, but not completely perfect, presumably due to the rather small sample size. As a robustness check, we also calculated p-values, with the null hypothesis that there are no differences between the framing variations. We cannot reject H0 for any of the variables of the Table, which indicates that randomization was successful. Thus, we can meaningfully compare mean values in our analysis. (b) Differences in n due to incomplete answers

Table 2 summarizes the experimental results. We distinguish between WTP for production with beneficial insects and with pesticides. We also differentiate between the two framing variants. A plausibility check of the mean WTPs suggests that the WTPs are very similar to market prices for these products, mitigating the concern that our open-end measure resulted in substantial overstatements. This mitigated the concern that our open-end measure and ‘hypothetical’ way to capture consumers’ WTP results in substantially higher WTP than ‘real’ WTP (e.g., Voelckner et al., 2006).

Table 2: Stated willingness to pay in experiment 1

Willingness to pay		Mean for plants protected by beneficial insects in € (Std. Dev.)	Mean for plants protected by pesticides in € (Std. Dev.)	Difference in means (Std. Dev.) between WTPs for plants protected by beneficial insects and pesticides
Basil	Overall	3.19 (1.06)	1.66 (.84)	1.53 (.22)
	Gain	3.31 (1.14)	1.77 (.79)	1.54 (.35)
	Loss	3.07 (.97)	1.52 (.89)	1.54 (.07)
Heather	Overall	3.54 (2.88)	1.89 (1.43)	1.65 (1.44)
	Gain	3.90 (3.73)	2.14 (1.61)	1.76 (2.12)
	Loss	3.15 (1.45)	1.61 (1.75)	1.54 (-.30)
Orchid	Overall	11.39 (5.46)	7.19 (4.10)	4.20 (1.36)
	Gain	12.89 (6.41)	8.24 (4.47)	4.65 (1.98)
	Loss	9.79 (3.66)	6.00 (3.35)	3.79 (.31)

The main results can be summarized as follows: (i) On average, the stated willingness to pay is higher for plants protected by beneficial insects than for those protected by pesticides. The differences of

the means are always substantially greater than 0 (last column of the Table 2). (ii) The differences in willingness to pay between plants protected by beneficial insects and pesticides are slightly higher in the gain frame. One exception is basil, where they are about equal. (iii) The gain frame also produces more extreme results. Beneficial insects in the gain frame have higher mean values than in the loss frame. Pesticides in the gain framing have a higher mean value than in the loss framing.

2.4 Starting point for experiment 2

The central goal of experiment 2 was to discover if the findings could be replicated when the subject pool is the 'general population.' Moreover, we aimed to check the robustness of the results, in light of social desirability bias (SDB) as well as hypothetical bias, commonly cited biases in contingent valuation experiments like ours (e.g., Börger, 2013; Murphy et al., 2010).

Regarding SDB, Fisher (1993: 303) explains this bias as follows: "Respondents are often unwilling or unable to report accurately on sensitive topics for ego-defensive or impression management reasons. The result is data that are systematically biased toward respondents' perception of what is 'correct' or socially acceptable." By carrying out a systematic literature review, Cerri et al. (2019) find that the vast majority of studies on sustainable food consumption and preferences neither discuss the thread that SDB stipulates to the validity of results nor offer any kind of remedy for it. While contingent valuation methods, such as the one applied here, are commonly used in surveys, their use is still somewhat controversial regarding the influence of hypothetical bias (Drichoutis et al., 2017). As Blumenschein et al. (2008: 114) indicate: "At the heart of this controversy is the extent to which hypothetical choices correspond to real economic choices," and as indicated above, ways to capture consumers 'hypothetical' WTPs typically result in greater WTP than 'real' WTP measurement instruments (e.g., Voelckner et al., 2006).

3 Experiment 2: Online experiments with the general population

3.1 Study design

The data collection took place online in September 2019, via an online access panel. In contrast to interviewer-administered data-collection methods, such as in experiment 1, collecting self-administered online data already reduces social-desirability bias concerns (Krumpal, 2013). An external commercial panel provider invited participants to take part in the survey and monetarily incentivized voluntary and anonymous participation. The experimental design was similar to that in experiment 1. However, participants were individually assigned to an experimental group and, instead of the live representation of the plants and signs, pictures were shown. Also, the questionnaire was part of a larger survey on consumers' knowledge and risk perception related to plant protection measures. Thus, the questionnaire exceeded the one described in experiment 1. To obtain results comparable to those for the general German population, quotas were set for age, gender, place of residence, and income (cf., Table 3).

3.2 Extensions as a consequence of the results from experiment 1

In order to check the robustness of our results, we made three methodological additions. First, to determine whether participants read the sign carefully, we asked them to repeat the percentage of additional pesticide use and respective savings that the framings depicted. If they failed to do so, we dropped them from the analysis, mainly to ensure that the results regarding the framings could truly be attributed to the gain and loss messages⁹. Second, to explore the effect of social desirability in answers, the social desirability scale KSE-G (Kemper et al., 2012) was introduced into the

⁹ Attention checks such as ours can result in substantially smaller sample sizes (Oppenheimer et al., 2009) but are recommended for online surveys using access panels of a commercial research institute (Shamon and Berning, 2020), where participants are monetarily rewarded for their participation.

questionnaire. Third, to decrease concerns about hypothetical bias, we added an ex-post ‘certainty’ question. The following sections explain the latter two additions in more detail.

3.2.1 Social desirability scale

Research on the construct of social desirability derives mainly from psychology, and researchers have been interested in the subject for over 50 years (e.g., Crowne and Marlowe, 1960). While several different scales to capture the construct have emerged, the key idea behind them is principally the same: Researchers use psychometric scales that aim to capture social desirability and correlate these to the behavior of interest (or other survey answers) (Lusk and Norwood, 2011). The psychometric scales include a number of questions designed to elicit participants’ inclination to *say* they (do not) engage in activities that are socially (un)desirable, however frequent or infrequent, respectively (Stöber, 2001). For instance, a common question on social desirability scales includes the extent to which it is correct that one always stays factual and objective during an argument. Scholars calculate correlations between the scale and the answers to survey questions that are assumed susceptible to social desirability bias. To detect the effect of social desirability bias on this experiment, the last part of the questionnaire contained questions covering the short scale on social desirability (KSE-G) by Kemper et al. (2012). Here, two aspects of social desirability are captured separately: the tendency to minimize negative qualities (NQ-) and the tendency to exaggerate positive qualities (PQ+). Each aspect is measured using three items (see Appendix 1). We consequently calculated two independent scales (NQ-/ PQ+) by calculated the mean of the respective three items.¹⁰ We used the scales to identify whether the stated WTP substantially correlates with the two aspects of the scale and, consequently, to approximate the effect of social desirability bias.

3.2.2 Expressed certainty

The literature provides several methods to tackle hypothetical bias (HB). For example, ex-ante methods, such as cheap talk or consequentiality scripts, have been proposed and applied (e.g., Drichoutis et al., 2017; Britwum and Yiannaka, 2019). Also, Blumenschein et al. (2008) find evidence in a field experiment that certainty statements can mitigate the hypothetical bias. This so-called ‘expressed-certainty’ method traces back to Johannesson et al. (1998) and is a simple ex-post question, where study participants indicate in a follow-up question whether they are certain of their responses. A meta-analysis by Penn and Hu (2018: 1186) comes to the result that “[...] cheap talk, consequentiality as well as certainty follow-up all significantly contributed to explaining and mitigating the magnitude of HB.” In the study here, we use an ex-post certainty question to check whether the indicated WTP differs substantially between participants who indicate themselves as ‘rather certain’ and those who are ‘very certain.’ We do so, as (i) results by Penn and Hu (2018) suggest that ex-post methods work a little better in reducing the magnitude of HB than ex-ante methods, and (ii) an ex-post method allows us to compare the results regarding HB (i.e., those who are very and rather certain) without needing to add a control group.

3.3 Results and discussion

3.3.1 Main analysis

Results concerning the summary statistics of the sample (cf., Table 3) show that the quotas worked out well and the sample structure is similar to that of the general German population. Thirty participants were eliminated, as they were unable to repeat correctly the framing information (cf., section 3.2). Due to the cut-offs (cf., experiment 1), an additional 18 participants were eliminated. A total of $n = 187$ data sets from the initial 235 remained for the analysis.

¹⁰ To test for the reliability of the scales, we followed the suggestion by Hayes and Coutts (2020) and used McDonald’s omega rather than Cronbach’s alpha: $\omega(\text{NQ-}) = 0.664$; $\omega(\text{PQ+}) = 0.607$.

Table 3: Summary statistics of the sample (n= 187) ^(a)

Characteristic	Specification	Frequency	Percent (%)	German households in 2018 (%)
Age**	18-29	32	17.1	19.3
	30-39	30	16.0	17.8
	40-49	39	20.9	17.4
	50-59	43	23.0	22.5
	>60	43	23.0	23.2
Gender**	female	95	50.8	50.7
	male	92	49.2	49.3
Place of residence***	new FS (former East)	38	20.3	20.6
	old FS (former West)	149	79.7	79.4
Size of residence***	rural (<20,000)	65	34.8	37.9
	urban (>20,000)	122	65.2	62.1
	€0-€1500	33	17.6	16.5
Net household income***	€1501-€3000	70	37.4	36.3
	€3001-€4500	52	27.8	28.9
	>€4500	32	17.1	18.3

Note: +excluding people under 18; Sources: *Statistisches Bundesamt (2019a); **Statistisches Bundesamt (2019b), ***best for planning ^(a) According to the standardized differences, randomization worked out well. We use the sample variables as in Experiment 1, and all standardized values are smaller than 0.2: age (group): 0.12; female: 0.08; size of household: 0.16; number of children: 0.02; residence in the federal state of Hesse (= 1): 0.17. As a robustness check, we also calculated p-values, with the null hypothesis that there are no differences between the framing variations. We cannot reject H0 for any of the variables of the Table, which indicates that randomization was successful. Thus, we can meaningfully compare mean values in our analysis.

To be as similar to experiment 1 as possible, we use the same kind of presentation of the WTP results (cf., Table 4). We distinguish between the use of beneficial insects and the use of pesticides, as well as between the two framing variants. Comparing our results from experiment 1 with those of experiment 2, we find that:

- (i) Experiment 2 supports the finding (i.e., provides further evidence) that consumers' willingness to pay, on average, is greater for plants protected by beneficial insects than for those protected by pesticides.
- (ii) Experiment 2 partly supports the finding that the gain framing produces greater differences in willingness to pay between plants protected by beneficial insects and those protected by pesticides. The results for heather are in line with experiment 1. However, on average, loss framing induces marginal higher WTPs for basil and orchids (cf., last column of Table 4).
- (iii) Experiment 2 supports the finding that the gain framing produces more extreme willingness to pay for plants protected by beneficial insects and those protected by pesticides.

Table 4: Stated willingness to pay in experiment 2

Willingness to pay		Mean for plants protected by beneficial insects in € (Std. Dev.)		Mean for plants protected by pesticides in € (Std. Dev.)		Difference in means (Std. Dev.) between WTPs for plants protected by beneficial insects and pesticides	
Basil	Overall	2.80	(1.33)	1.61	(1.15)	1.19	(.19)
	Gain	2.94	(1.25)	1.76	(1.15)	1.18	(.09)
	Loss	2.61	(1.42)	1.41	(1.12)	1.20	(.30)
Heather	Overall	2.97	(2.10)	1.65	(1.44)	1.32	(.67)
	Gain	3.21	(2.28)	1.75	(1.44)	1.46	(.85)
	Loss	2.65	(1.81)	1.52	(1.44)	1.14	(.37)
Orchid	Overall	8.18	(5.45)	5.12	(4.35)	3.06	(1.10)
	Gain	8.46	(5.46)	5.42	(4.33)	3.04	(1.13)
	Loss	7.81	(5.46)	4.73	(3.38)	3.08	(1.08)

Overall, we conclude from experiment 2 (consistent with experiment 1) that consumers prefer plants protected by beneficial insects. This conclusion is strengthened by the very low p-values for both experiments (see Appendix 3). This finding is in line with prior evidence suggesting that consumers prefer the use of beneficial insects over the use of pesticides (Lehberger and Becker, 2020; Wollaegeer et al., 2015). It is particularly relevant for producers who already use beneficial insects for plant protection, as it indicates that marking their plant protection practice can result in a competitive advantage over other producers, a potential economic incentive for producers to start using beneficial insects and marking their products accordingly. Also, our evidence indicates that the gain framing produces higher levels of WTP for both types of plant protection practice.¹¹ Thus, loss aversion, which should lead to higher levels of WTP induced by the loss framed message, appears not to be a main driver of stated preferences in our samples. However, our findings are similar to those contending that gain frames are more effective in green advertising (Segev et al., 2015). Drawing from results by Wansink and Pope (2015) and Lehberger and Becker (2020) it may be argued that, in our study, the gain frame works better, as people have little knowledge about the behavior in question and are not highly involved. However, it is interesting that the gain frame produces higher WTP for plants protected with pesticides as well, compared to the loss frame for plants produced with pesticides.

3.3.2 Robustness checks

(i) Social desirability scale and correlations with WTP

An inflated stated willingness to pay for beneficial insects and a reduced willingness to pay for pesticides would be consistent with the social desirability bias. However, Table 5 shows that the correlations between the two aspects of the social desirability scale (PQ+/NQ-) and willingness to pay are both small in magnitude overall and mostly not statistically significant at the 5% level. The highest (and negative) correlation is found between NQ- and the WTP regarding beneficial insects. This indicates that the more participants *are inclined to* minimize their negative qualities, the *higher* is their WTP for beneficial insects. However, even the ‘highest’ statistically significant correlations are low (-0.14, -0.17) to moderate (-0.35). Thus, we may conclude that preferences rather than social desirability bias leads to the participants’ indicated WTP.

¹¹ We also checked the robustness of our findings by carrying out regression analysis. The negative association between the loss framing and the willingness to pay was prevalent in both experiments, even after experimenting with several controls, such as age and gender.

Table 5: Correlation between social desirability bias and willingness to pay

	PQ(+) and beneficial insects rho's (sig. level)	PQ(+) and pesticides rho's (sig. level)	NQ(-) and pesticides rho's (sig. level)	NQ(-) and beneficial insects rho's (sig. level)
Basil	.05 (.53)	-.02 (.78)	.04 (.56)	-.13 (.09)
Heather	.04 (.56)	.01 (.91)	.07 (.35)	-.14 (.05)
Orchid	.08 (.30)	.04 (.60)	-.17 (.02)	-.35 (.00)

(ii) Expressed certainty

Only 25.14% of our participants answered that they were 'very certain' of their indicated WTP. Looking at the mean values, a pattern can be identified in the data (cf., Table 6). With beneficial insects, the mean of those who are 'very certain' exceeds those who are 'rather certain'; with pesticides, it is the reverse. This effect seems to be robust over all three cultures, but the effect size is low. Following the interpretation by Blumenschein et al. (2008), these results indicate that the 'real' difference between the WTP for plants protected by beneficial insects and for those protected by pesticides is even higher than our initial results suggested. Following Blumenschein et al. (2008), this is because the participants who indicate they are 'very certain' are the ones for whom the hypothetical bias is mitigated. Thus, results again support our first finding, namely, that participants' willingness to pay, on average, is higher for plants protected by beneficial insects than for those protected by pesticides.

Table 6: Expressed certainty (1= rather certain, 2= very certain)

	Basil (Beneficial insects)		Basil (Pesticides)		Heather (Bene- ficial insects)		Heather (Pesticides)		Orchids (Bene- ficial insects)		Orchids (Pesticides)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Rather certain (74.86%)	2.78	1.30	1.67	1.18	2.93	2.21	1.73	1.50	7.97	5.65	5.29	4.64
Very certain (25.14%)	2.96	1.39	1.46	1.07	3.20	1.79	1.45	1.29	9.21	4.72	4.77	3.44
H0: S1=S2 ^(a)	0.43		0.53		0.10		0.45		0.04		0.92	

^(a) P-values according to rank sum tests

3.4 Policy implications

From an ecological point of view, biocontrol measures are particularly environmentally friendly, having no or minimal effects on nontarget organisms (Koch et al., 2019). Beneficial insects, specifically, are effective in fighting insect pests while reducing the need for synthetic-chemical insecticides, especially in greenhouse production (Gacheri et al., 2015). Thus, policies to further encourage the use of beneficial insects have been and can be defended from an environmental-sustainability viewpoint. This can be an important consideration for policy makers mandating the use of beneficial insects for certain pests in greenhouse production systems. The voluntary adoption of biocontrol measures is highly dependent on contextual, as well as person-specific factors (Goldberger and Lehrer, 2016; Sharifzadeh et al., 2017) and leads to very different application rates in the federal states in Germany (Koch et al., 2019). Our evidence suggests that consumers prefer plants protected by beneficial insects. In light of these findings, a potential price premium that consumers would pay may even offset the additional cost to domestic producers. A price premium may be especially important, as the majority of ornamental plants sold in Germany are imported (AMI, 2020) and, thus, would coexist with German products. An alternative or addition to mandating

the use of beneficial insects may be the provision of further economic incentives to encourage the adoption of beneficial insects, e.g., governmental subsidies. This may be a promising measure to increase the use of beneficial insects, especially in outside production in Germany, where regional subsidies have already succeeded in producing higher adoption rates (Koch et al., 2019). Economic incentives could also aim toward developing a niche market for plants protected with beneficial insects, which coexist and compete with conventionally produced plants. Our research indicates that nudges (Thaler and Sunstein, 2009), such as different framings analyzed within this paper, can effectively steer consumer behavior toward more environmentally friendly actions. However, especially in outside production, where the efficiency of beneficial insects is less predictable, increasing governmentally funded research may be an effective strategy for developing more effective and safe ways to fight insect pests with beneficial insects, without disrupting existing ecosystems. Finally, knowledge transfer and consulting services specifically for the application of beneficial insects may be a fruitful strategy for increasing the adoption of this plant protection practice. This is especially relevant because an effective application of beneficial insects is knowledge-intensive, and a lack of knowledge is one of the key barriers to adoption (Goldberger and Lehrer, 2016).

4 Conclusion

Our two experimental studies investigated consumers' WTP for plants protected by beneficial insects and by pesticides, respectively, as well as the influence of gain and loss framing on indicated WTP. Concerning the WTP for different plant protection practices, our research results clearly indicate that consumers have a greater WTP for plants protected by beneficial insects than for those protected by pesticides. This holds true for all three plants included in the study (basil, heather, and orchid). With regard to the influence of message framing on the differences in WTP between plants protected by beneficial insects and those protected by pesticides, we find mixed evidence: While both gain and loss framing largely lead to similar results, the comparative impact of the gain framed message was stronger in the lab-in-the-field experimental setting than in the online setting. A third main finding of our research is that both experiments indicate that the gain frame produced more extreme WTP.

A major advantage of our study is that we carried out several measures to mitigate concerns about social desirability and hypothetical biases and found that our results are robust in terms of our key findings. Some of these methods to test the robustness of the results are common in other research disciplines but novel in research on sustainable food. For instance, while social desirability scales are commonly used in psychological research, Cerri et al. (2019) find that only 1 out of 388 reviewed studies on sustainable food consumption applies such a scale. Moreover, we argue that we improved the credibility of our results by conducting a replication study. Still, our research constitutes a starting point; therefore, we would welcome further replications of our work. Another advantage of our study is that our experiments are carried out with all types of members of the general public. Experiments in psychology and economics are still often carried out with students as subjects, although many researchers are shifting away from this practice. While there are many advantages to experiments with students (e.g., low recruitment cost), these are offset by several disadvantages, in terms of external validity (i.e., generalizability) (Frchette, 2015a, b). Recruiting non-students helps to increase the external validity of findings.

Our study also suffers from limitations that may serve as a starting point for further research. The decisions made by the participants in our experiments are hypothetical. We cannot exclude the possibility that people make different decisions when they are at the point of sale and must choose between several products. Just think of the complexity they would face if they deliberately thought

of the pros and cons of each product. Thus, further research must verify the real WTP in a point-of-sale situation. Field experiments could help to clarify this point. Another limitation is that our samples may be judged as rather small. Further study with more participants could consider conducting subgroup analyses for groups of interest, e.g., comparisons between young and old people or, possibly, gender differences, which could include a control group. Furthermore, replication studies are required to verify the robustness of our results with respect to different bidding mechanisms. We used open-ended questions to measure the subjects' WTP. Other studies could focus, for example, on multiple price list designs, random lottery pair designs, ordered lottery selection designs, the Becker–DeGroot–Marschak design, and/or experimental auctions (Harrison and Rutström, 2007; Corrigan et al., 2009). Moreover, in our paper, we define information about the reduction in pesticide use as “gain” and information about the increase in pesticide use as “loss.” This procedure is associated with two reference points. Further research might instead consider using one neutral reference point (e.g., 20 kg/ha) from which gains and losses would be varied. Finally, while the work of Tversky and Kahneman (1981) and their concept of loss aversion dominate the behavioral economic literature, our results are more in line with findings indicating that the influence of message framing is very context-specific (Wansink and Pope, 2015). Further research should elaborate on the behavioral drivers of framing in the sustainable food context.

For horticultural producers and retailers, our overall findings are very valuable. Consumers evaluate plant protection with beneficial insects very positively. The use of beneficial insects in horticultural production, especially in the ornamental and herb-production sectors, is already rather common. Thus, our results indicate that producers who already use beneficial insects may advertise this, in order to gain a competitive advantage. Moreover, our research findings indicate that adequate framing could further increase consumers' WTP. Given that neutral framings do not exist in reality, policy makers or marketing representatives (e.g., for grocery stores) can use our findings to nudge people in the desired direction. Consequently, information provided should be carefully selected and its effect should be tested when advertisements are developed.

Acknowledgment

The authors declare that we have no relevant or material financial interests that relate to the research described in this paper. The funding for the data collection was provided by the Hochschule Geisenheim University. Sven Grüner is a recipient of financial support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation – 388911356).

References

- Altman, D.G. (1985). Comparability of Randomised Groups. *The Statistician*, 34(1), 125–136.
- AMI (2019a): Starkes Jahr für Beet und Balkonpflanzen. <https://www.bundesverband-zierpflanzen.de/top-10-zierpflanzenmarkt.html> Accessed 28 January 2020.
- AMI (2019b): Markt zeigt sich weitestgehend stabil. <https://www.bundesverband-zierpflanzen.de/top-10-zierpflanzenmarkt.html> Accessed 28 January 2020.
- AIM (2020). Warenstromanalyse 2018, Blumen, Zierpflanzen & Gehölz. Bundesministerium für Ernährung and Landwirtschaft. <https://www.bmel.de/SharedDocs/Downloads/DE/Landwirtschaft/Pflanzenbau/warenstromanalyse-2018.html> Accessed 18 January 2021.
- Asioli, D., Mignani, A., & Alfnes, F. (2020). Quick and easy? Respondent evaluations of the Becker–DeGroot–Marschak and multiple price list valuation mechanisms. *Agribusiness*, early view. <https://doi.org/10.1002/agr.21668>

- Barratt, B.I.P., Moran, V.C., Bigler, F., & van Lenteren, J.-C. (2018). The status of biological control and recommendations for improving uptake for the future. *BioControl*, 63, 155-167.
- Blumenschein, K., Blomquist, G.C., Johannesson, M., Horn, N., & Freeman, P.R. (2008). Eliciting Willingness to Pay in the Without Bias: Evidence from a Field Experiment. *The Economic Journal*, 118(525), 114-137.
- Börger, T. (2013). Keeping up appearances: Motivations for socially desirable responding in contingent valuation interviews. *Ecological Economics*, 87, 155-165.
- Brainerd, C.J., Nakamura, K., Reyna, V.F., & Holliday, R.E. (2017). Overdistribution Illusions: Categorical Judgments Produce Them, Confidence Ratings Reduce Them. *Journal of Experimental Psychology*, General 146(1), 20-40.
- Britwum, K., & Yiannaka, A. (2019). Consumer willingness to pay for food safety interventions: The role of message framing and issue involvement. *Food Policy*, 86, 101726.
- BVEO (2019). *Top-10 der Kräuter in deutschen Privathaushalten*. Press release from 05/02/2019, Berlin: Pressebüro deutsches Obst und Gemüse.
- Camerer, C.F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., & Johannesson, M., et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433-1436.
- Cerri, J., Thogerson, J., & Testa, F. (2019). Social desirability and sustainable food research - A systematic literature review. *Food Quality and Preferences*, 71, 136-140.
- Charness, G., Gneezy, U., & Kuhn, M.A. (2013). Experimental methods: Extra-laboratory experiments-extending the reach of experimental economics. *Journal of Economic Behavior & Organization*, 91, 93-100.
- Chu, M., & Kita, S. (2011). The Nature of Gestures' Beneficial Role in Spatial Problem Solving. *Journal of Experimental Psychology: General*, 140(1), 102-116.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed). Lawrence Erlbaum Associates Publishers: Hillsdale, NJ.
- Corrigan, J.R., Depositario, D.P.T., Nayga, R.M., Wu, X., & Laude, T.P. (2009). Comparing Open-Ended Choice Experiments and Experimental Auctions: An Application to Golden Rice. *American Journal of Agricultural Economics*, 91(3), 837-853.
- Crowne, D.P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349-354.
- Drichoutis, A.C., Vassilopoulos, A., Lusk, J.L., & Nayga, R.M. (2017). Consumer preferences for fair labour certification. *European Review of Agricultural Economics*, 44(3), 455-474.
- Dunning, T. (2012). *Natural Experiments in the Social Sciences: A Design-based Approach*. Cambridge University Press: Cambridge.
- Duvendack, M., R.W. Palmer-Jones, & W.R. Reed, (2015). Replications in Economics: A Progress Report. *Econ Journal Watch*, 12(2), 164-191.
- Edenbrandt, A.K., Gamborg, C., & Thorsen, B.J. (2018). Consumers' Preferences for Bread: Transgenic, Cisgenic, Organic or Pesticide-free? *Journal of Agricultural Economics*, 69(1), 121-141.
- Fiedler, S., & Hillenbrand, A. (2020). Gain-loss framing in interdependent choice. *Games and Economic Behavior*, 121, 232-251.
- Fisher, R.J. (1993). Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research*, 20(2), 303-315.
- Fréchette, G.R. (2015a). Laboratory Experiments: Professionals versus Students. Fréchette, G.R. & Schotter, A. (eds.): *Handbook of Experimental Economic Methodology*. Oxford University Press, Oxford: 360-390.
- Fréchette, G.R. (2015b). Experimental Economics across Subject Populations. Kagel, J.H., & Roth, A.E. (eds.): *The Handbook of Experimental Economics*, Vol. 2. Princeton University Press, Princeton: 435-480.

- Frey, B. (2017). Policy consequences of pay-for-performance and crowding-out. *Journal of Behavioral Economics for Policy*, 1(1), 55-59.
- Gacheri, C., Kigen, T., & Sigsgaard, L. (2015). Hot-spot application of biocontrol agents to replace pesticides in large scale commercial rose farms in Kenya. *BioControl*, 60, 795-803.
- Getter, K.L., Behe, B.K., & Wollaeger, H.M. (2016). Comparative Consumer Perspectives on Eco-friendly and Insect Management Practices on Floriculture Crops. *HortScience*, 26(1), 46-53.
- Goldberger, J.R., & Lehrer, N. (2016). Biological control adoption in western U.S. orchard systems: Results from grower surveys. *Biological Control*, 102, 101-111.
- Gray, K., Ward, A.F., & Norton, M.I. (2014). Paying It Forward: Generalized Reciprocity and the Limits of Generosity. *Journal of Experimental Psychology: General*, 143(1), 247-254.
- Grygorczyk, A., Mhlanga, S., & Lesschaeve, I. (2016). The most valuable player may not be on the winning team: Uncovering consumer tolerance for color shades in roses. *Food Quality and Preference*, 47(A), 23-28.
- Håkansson, C. (2008). A new valuation question: analysis of and insights from interval open-ended data in contingent valuation. *Environmental and Resource Economics*, 39, 175-188.
- Harrison, G.W., & Rutström, E.E. (2007). Risk aversion in the laboratory. *Research in Experimental Economics*, 12, 41-196.
- Hayes, A.F., & Coutts, J.J. (2020). Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But... *Communication Methods and Measures*, 14(1), 1-24.
- Hofstetter, R., Miller, K.M., Krohmer, H., & Zhang, Z.H. (2020). A de-biased direct question approach to measuring consumers' willingness to pay. *International Journal of Research in Marketing*, in Press. <https://doi.org/10.1016/j.ijresmar.2020.04.006>.
- Holt, C.A., & Laury, S.K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, 92(5), 1644-1655.
- Johannesson, M., Liljas, B., & Johansson, P-O. (1998). An Experimental Comparison of Dichotomous Choice Contingent Valuation Questions and Real Purchase Decisions. *Applied Economics*, 30, 643-647.
- Kahan, D.M., Kysar, D., Braman, D., Slovic, P., Cohen, G.L., & Gastil, J. (2009). *Cultural cognition and nanotechnology risk perceptions: An experimental investigation of message framing*. Cultural Cognition Project at Yale Law School Report.
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-292.
- Kato, T., & Hidano, N. (2007). Anchoring Effects, Survey Conditions, and Respondents' Characteristics: Contingent Valuation of Uncertain Environmental Changes. *Journal of Risk Research*, 10(6), 773-792.
- Kemper, C.J., Beierlein, C., Bensch, D., Kovaleva, A. & Rammstedt, B. (2012). *Eine Kurzskala zur Erfassung des Gamma-Faktors sozial erwünschten Antwortverhaltens: Die Kurzskala Soziale Erwünschtheit-Gamma (KSE-G)*. Working Paper 25. Mannheim: Gesis – Leibniz-Institut für Sozialwissenschaften.
- Khan, J., Khanal, A-R., Ho Lim, K., Jan, A.U., & Shah, S.A. (2018). Willingness to Pay for Pesticide Free Fruits: Evidence from Pakistan. *Journal of International Food & Agribusiness Marketing*, 30(4), 392-408.
- Koch, S., Epp, A., Lohman, M., & Böhl, G- F. (2017). Pesticide Residues in Food: Attitudes, Beliefs, and Misconceptions among Conventional and Organic Consumers. *Journal of Food Protection*, 80(12), 2083-2089.
- Koch, E., Herz, A., Kleespies, R.G., Schmitt, A., Stephan, D., & Jehle, J.A. (2019). *Statusbericht Biologischer Pflanzenschutz 2018*. Berichte aus dem Julius Kühn-Institut 203. Braunschweig: Julius Kühn-Institut.

- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality and Quantity*, 47, 2025-2047.
- Lehberger, M., & Becker, C. (2020). Plant protection practices: how do risk perception, subjective and objective knowledge influence the preference of German consumers. *British Food Journal*, 123(4), 1465-1477.
- Levin, I. P., Schneider, S.L., & Gaeth, G. J. (1998). All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects. *Organizational Behavior and Human Decision Processes*, 76(2), 149-188.
- Lusk, J., & Norwood, B. (2011). Social Desirability Bias in Real, Hypothetical, and Inferred Valuation Experiments. *American Journal of Agricultural Economics*, 93, 528-534.
- McNeil, J.N., Cotnoir, P.-A., Leroux, T., Laprade, R., & Schwartz, J.-L. (2010). A Canadian national survey on the public perception of biological control. *BioControl*, 55(4), 445-454.
- Ministry for Rural Areas and Consumer Protection Baden-Württemberg (2017). Förderprogramm für Agrarumwelt, Klimaschutz und Tierwohl Baden-Württemberg [Support program for agri-environment, climate protection and animal welfare Baden-Württemberg], https://mlr.baden-wuerttemberg.de/fileadmin/redaktion/m-mlr/intern/dateien/publikationen/Landwirtschaft/2017_Fakt_Foerderprogramm_Agrarumwelt_Klimaschutz_Tierwohl.pdf, accessed 27 January 2021. Accessed 27 January 2021.
- Murphy, J.J., Stevens, T.H., & Yadav, L. (2010). A Comparison of Induced Value and Home-Grown Value Experiments to Test for Hypothetical Bias in Contingent Valuation. *Environmental and Resource Economics*, 47(1), 111-123.
- Nabi, R.L., Walter, N., Oshidary, N., Endacott, C.G., Love-Nichols, Z., Lew, Z.J., & Aune, A., (2020). Can Emotions Capture the Elusive Gain-Loss Framing Effect? A Meta-Analysis. *Communication Research*, 47(8), 1107-1130.
- O'Keefe, D.J., & Jensen, J.D. (2009). The relative persuasiveness of gain-framed and loss-framed messages for encouraging disease detection behaviors: a meta-analytic review. *Journal of Communication*, 59, 296-316.
- Opit, G., Perrett, J., Holt, K., Nechols, J., Margolies, D., & Williams, K. (2009). Comparing Chemical and Biological Control Strategies for Twospotted Spider Mites (Acari: Tetranychidae) in Commercial Greenhouse Production of Bedding Plants. *Journal of economic entomology*, 102(1), :336-46. 336-46.
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4), 867-872.
- Penn, J.M., Hu, W. (2018). Understanding Hypothetical Bias: An Enhanced Meta-Analysis. *American Journal of Agricultural Economics*, 100(4)4, 1186-1206.
- Rana, J., & Paul, J. (2017). Consumer behavior and purchase intention for organic food: A review and research agenda, *Journal of Retailing and Consumer Services*, 38, 157-165.
- Römer, U., Schaak, H., & Mußhoff, O. (2019). The perception of crop protection: Explicit vs. implicit association of the public and in agriculture. *Journal of Environmental Psychology*, 66, 101346.
- Rothman, A.J., Bartels, R.D., Wlaschin, J., & Salovey, P. (2006). The Strategic Use of Gain- and Loss-Framed Messages to Promote Healthy Behavior: How Theory Can Inform Practice. *Journal of Communication*, 56(1), S202-S220.
- Schkade, D.A. & Payne, J.W. (1994). How people respond to contingent valuation questions: a verbal protocol analysis of willingness to pay for an environmental regulation. *Journal of Environmental Economics and Management*, 26(1), 88-109.
- Schober, P., Bossers, S.M., & Schwarte, L.A. (2018). Statistical Significance Versus Clinical Importance of Observed Effect Sizes: What Do P Values and Confidence Intervals Really Represent? *Anesthesia & Analgesia*, 126(3), 1068-1072.

- Segev, S., Fernandes, J., & Wang, W. (2015). The Effects of Gain Versus Loss Message Framing and Point of Reference on Consumer Responses to Green Advertising. *Journal of Current Issues & Research in Advertising*, 36(1), 35-51.
- Shamon, H., & Berning, C.C. (2020). Attention Check Items and Instructions in Online Surveys with Incentivized and Non-Incentivized Samples: Boon or Bane for Data Quality? *Survey Research Methods*, 14(1), 55-77.
- Sharifzadeh, M.S., Damalas, C.A., Abdollahzadeh, G., & Ahmadi-Gorgi, H., (2017). Predicting adoption of biological control among Iranian rice farmers: An application of the extended technology acceptance model (TAM2). *Crop Protection*, 96, 88-96.
- Shi, L., Gao, Z., & Chen, X. (2014). The cross-price effect on willingness-to-pay estimates in open-ended contingent valuation. *Food Policy*, 46, 13-21.
- Simon, H.A. (1957). *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. John Wiley and Sons: New York.
- Statistisches Bundesamt (2019a). Bevölkerung: Deutschland, Stichtag, Altersjahre (Code „12411-0005“). <https://www-genesis.destatis.de/genesis/online> Accessed 10 December 2019.
- Statistisches Bundesamt (2019b). Bevölkerung: Deutschland, Stichtag, Geschlecht (Code „12411-0003“). <https://www-genesis.destatis.de/genesis/online> Accessed 10 December 2019.
- Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment*, 17, 222-232.
- Thaler, R.H., & Sunstein, C.R. (2009). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale: Yale University Press.
- Thomas, F., Midler, E., Lefebvre, M., & Engel, S. (2019). Greening the common agricultural policy: a behavioural perspective and lab-in-the-field experiment in Germany. *European Review of Agricultural Economics*, 46(3), 367-392.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453-458.
- Tversky, A., & Kahneman, D. (1986). Rational Choice and the Framing of Decisions. *The Journal of Business*, 59(4), 251-278.
- Tversky, A., & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *The Quarterly Journal of Economics*, 106(4), 1039-1061.
- USDA (United States Department of Agriculture) (not dated). Biological Control Program. <https://nifa.usda.gov/program/biological-control-program> Accessed 30 September 2020.
- Voelckner, F. (2006). An empirical comparison of methods for measuring consumers' willingness to pay. *Marketing Letters*, 17, 137-149.
- Wansink, B., & Pope, L. (2015). When do gain-framed health messages work better than fear appeals? *Nutrition Review*, 73(1), 4-11.
- Wollaeger, H.M., Getter, K.L., & Behe, B.K. (2015). Consumer Preferences for Traditional, Neonicotinoid-free, Bee-friendly, or Biological Control Pest Management Practices on Floriculture Crops. *HortScience*, 5, 721-732.

Appendix 1

Social Desirability Scale after Kemper et al. (2012): Answers ranged on a five-point scale from 'doesn't apply at all = 1' to 'applies completely = 5'. The PQ+ covers question 2, 4 and, 6. NQ- covers questions 1, 3 and, 5. Scales were calculated by building means. Questions go as follows:

- (1) It has happened that I have taken advantage of someone in the past.
- (2) Even if I am feeling stressed, I am always friendly and polite to others.
- (3) Sometimes I only help people if I expect to get something in return.
- (4) In an argument, I always remain objective and stick to the facts.

- (5) I have occasionally thrown litter away in the countryside or on to the road.
(6) When talking to someone I always listen carefully to what the other person says.

Appendix 2

Standardized differences are similar to Cohen's d; they address the effect size as a measure of mean (or, if the target variable is dichotomous, prevalence) and variation of the sample. No claim to generalizability (as is the case with p-values) is required. Two sources (Altman, 1985, and Schober et al., 2018) illustrate the measure as a recommended instrument to deal with baseline comparisons.

Appendix Table A1: Calculating standard differences (continuous and categorical baseline)

Continuous variables		Dichotomous variables	
	$d = \frac{(\bar{x}_T - \bar{x}_C)}{\sqrt{\frac{s_T^2 + s_C^2}{2}}}$		$d = \frac{(\hat{p}_T - \hat{p}_C)}{\sqrt{\frac{\hat{p}_T(1 - \hat{p}_T) + \hat{p}_C(1 - \hat{p}_C)}{2}}}$
$\bar{x}_T(\bar{x}_C)$	Sample mean of covariate, treated subjects (untreated subjects)	$\hat{p}_T(\hat{p}_C)$	Prevalence/mean of the covariate, treated subjects (untreated subjects)
$s_T^2(s_C^2)$	Sample variance of covariate, treated subjects (untreated subjects)		

Since randomization worked well in our study, mean differences can be interpreted as effect size. Regarding expectation, randomization balances confounders across treatment conditions. This holds not only for known but also for unknown confounders. Therefore, randomization helps to provide more valid findings than would simple regressions of observational studies (e.g., Dunning, 2012). Larger mean differences indicate larger effects.

Appendix 3

In the following, we provide p-values (of two sided rank sum tests) to better understand the WTP of the differences between the two framing variants and differences between plant protection through beneficial insects and pesticides. However, one should keep in mind that per definition p-values shrink with the sample size. Thus, the p-values are in experiment 2 ceteris paribus lower than in experiment 1.

Appendix Table A2: Stated Willingness to pay in experiment 1: P-Values

Plant	Subject of interest	Comparison (H0: both are identical)	Prob > z
Basil	Beneficial insects	Gain vs Loss Framing	0.2679
Basil	Pesticides	Gain vs Loss Framing	0.2505
Heather	Beneficial insects	Gain vs Loss Framing	0.6395
Heather	Pesticides	Gain vs Loss Framing	0.1159
Orchids	Beneficial insects	Gain vs Loss Framing	0.0270
Orchids	Pesticides	Gain vs Loss Framing	0.0327
Basil	Gain	Beneficial insects vs Pesticides	0.0000
Basil	Loss	Beneficial insects vs Pesticides	0.0000
Heather	Gain	Beneficial insects vs Pesticides	0.0000
Heather	Loss	Beneficial insects vs Pesticides	0.0000
Orchids	Gain	Beneficial insects vs Pesticides	0.0002
Orchids	Loss	Beneficial insects vs Pesticides	0.0000

Appendix Table A3: Stated Willingness to pay in experiment 2: P-Values

Plant	Subject of interest	Comparison (H0: both are identical)	Prob > z
Basil	Beneficial insects	Gain vs Loss Framing	0.0548
Basil	Pesticides	Gain vs Loss Framing	0.0291
Heather	Beneficial insects	Gain vs Loss Framing	0.1284
Heather	Pesticides	Gain vs Loss Framing	0.1933
Orchids	Beneficial insects	Gain vs Loss Framing	0.3585
Orchids	Pesticides	Gain vs Loss Framing	0.2221
Basil	Gain	Beneficial insects vs Pesticides	0.0000
Basil	Loss	Beneficial insects vs Pesticides	0.0000
Heather	Gain	Beneficial insects vs Pesticides	0.0000
Heather	Loss	Beneficial insects vs Pesticides	0.0000
Orchids	Gain	Beneficial insects vs Pesticides	0.0000
Orchids	Loss	Beneficial insects vs Pesticides	0.0001