University of Heidelberg

Department of Economics

Discussion Paper Series | No. 666

Norms in the Lab:
Inexperienced versus Experienced Participants

Robert J. Schmidt, Christiane Schwieren,
and Alec N. Sproten

September 2019

Norms in the Lab: Inexperienced versus Experienced Participants

Robert J. Schmidt, Christiane Schwieren & Alec N. Sproten

September 22, 2019

Abstract: Using coordination games, we study whether social norm perception differs between inexperienced and experienced participants in economic laboratory experiments. We find substantial differences between the two groups, both regarding injunctive and descriptive social norms in the context of participation in lab experiments. By contrast, social norm perception for the context of daily life does not differ between the two groups. We therefore conclude that learning through experience is more important than selection effects for understanding differences between the two groups. We also conduct exploratory analyses on the relation between lab and field norms and find that behaving unsocial in an experiment is considered substantially more appropriate than in daily life. This appears inconsistent with the hypothesis that social preferences measured in lab experiments are inflated and indicates a distinction between revealed social preferences as measured commonly and the elicitation of normatively appropriate behavior.

Highlights:

- Social norm perception is compared between inexperienced and experienced lab participants
- Substantial differences are observed in lab norms, but not in field norms
- The evidence suggests that learning is more important than selection-effects

Keywords: laboratory experiments, learning, selection effects, generalizability, methodology
JEL Classifications: B40, C90, C91

Corresponding author: Robert Johann Schmidt, Alfred-Weber-Institute for Economics, University of Heidelberg, Bergheimer Str. 58, 69115 Heidelberg, Germany. Phone: +49 152 34181300, email: rojoschmidt@gmail.com. We thank seminar participants at the UPF Barcelona, the GATE Lyon Saint Etienne, the Barcelona GSE Summer Forum, the 9th International Conference of the ASFEE in Nizza, the ESA-World Meeting in Berlin, and the ESA Asia-Pacific Meeting in Abu Dhabi for very valuable comments.

1. Introduction

Economic research makes extensive use of laboratory experiments for studying individual behavior in a controlled environment. Since the 1980s, the share of experimental studies published in general interest journals has risen continuously (Falk and Heckman, 2009). By now, lab experiments are an important source to inform economic theory and public policy (Nikiforakis and Slonim, 2015). However, methodological limitations of lab experiments, in particular the generalizability from the lab to the field, are regularly discussed (e.g., Dana et al., 2007; Galizzi and Navarro-Martínez, 2018; Levitt and List, 2007, 2008; Zizzo, 2010). Recently, more specific aspects of the recruitment process have been examined, such as the representativeness of registered students for the underlying student population (Abeler and Nosenzo, 2015; Cleave et al., 2013; Eckel and Grossman, 2000; Falk et al., 2013; Krawczyk, 2011; Slonim et al., 2013) or whether participants behave differently depending on the number of previous participations (Benndorf et al., 2017, Matthey and Regner, 2013).

In this study, we contribute to these literatures by examining two questions. First, we test whether social norm perception differs between inexperienced participants and experienced participants.[1] Second, by comparing differences between the two groups both for the lab and the field context, we attempt to investigate whether potential differences between the two groups are rather caused by learning through experience (when participating repeatedly) or by selection effects (through systematic differences in the probability to drop out from the pool depending on the participants´ characteristics).

To investigate these questions, we conduct a laboratory experiment and compare the two groups in a series of items that measure social norm perception. Precisely, we elicit social norm perception (i) regarding allocation behavior in the lab, (ii) regarding a series of unsocial behaviors in the lab and the field and (iii) regarding the evaluation of generalizability of behavior from the lab to the field. All questions are examined using the approach proposed by Krupka and Weber (2013) to elicit social norm perception via coordination games. In that approach, subjects are confronted with descriptions of behavior, and their task is to coordinate on appropriateness ratings. Assuming that social norms reflect shared perceptions about appropriate behaviors (Crawford and

[1] In our study we classify the degree of experience as follows: Inexperienced subjects did not yet participate in any economic (or psychological) experiment and experienced subjects participated at least ten times in an economic laboratory experiment.

Ostrom, 1995), the coordination outcome (i.e., the modal choice) reveals social norm perception within the participants´ population. We adopt that approach and sometimes adapt the set of available answers, such that the questions are suited to measure the kind of perceptions that we are interested in.[2]

A crucial assumption of the methodology applied in this experiment is that a coordination choice in the Krupka and Weber (2013) method is informative about a subject´s *own* perception about the question at hand. This view is justified if a respondent uses her own perception about a question when predicting the perception of others, in order to successfully anticipate the coordination outcome. Such behavior is implied by the false-consensus effect (Ross et al., 1977; Marks and Miller, 1987; Mullen et al., 1985), a well-documented phenomenon that follows Bayesian reasoning (Dawes, 1989, 1990; Engelman and Strobel, 2012; Schmidt, 2019b; Vanberg, 2019). Building on that assumption, we consider coordination choices as *informative signals* about the subjects´ actual *types*.[3] This allows us to draw conclusions on how the subjects themselves perceive the questions that they "answer" in the coordination games.

Our experiment yields three insights. First, social norm perception regarding behavior in the context of participation in a laboratory experiment differs significantly between inexperienced and experienced subjects. In a series of hypothetical dictator games, inexperienced subjects pronounce egalitarianism, while experienced subjects pronounce efficiency and the maximization of their earnings. Consistent with Matthey and Regner (2013), this indicates that behavior in experiments that involve allocation settings systematically depends on a subject's number of previous participations in lab experiments. Moreover, in the lab context, experienced subjects consider exploitation and deception as more appropriate than inexperienced participants, indicating that the two groups differ in experiments that involve these behaviors.

Second, by contrast to lab norms, neither field norms nor perceptions about the relation between the lab and the field differ between the two groups. This indicates that learning through experience is more important than selection effects for understanding the observed differences regarding lab norms.

[2] For example, in some items, we do not do ask whether a particular behavior is appropriate or not. Instead, we state that a particular behavior would be appropriate and the subjects´ task then is to coordinate on the degree of consent with that statement.

[3] For example, Schmidt (2019a) shows that injunctive and descriptive social norms elicited using coordination games are strongly related to revealed social preferences in a series of mini-dictator games on the subject level.

Third, conducting exploratory analyses on whether norm perception in the lab corresponds to norm perception in the field, we find that norm perception between the lab and the field is correlated. However, using the same items to evaluate unsocial behaviors, once framed to the lab and once framed to the field context, shows that these contexts differ substantially. Specifically, behaving unsocially is considered significantly more appropriate in the lab than in the field. This appears inconsistent with the hypothesis that social preferences measured in economic experiments are inflated (e.g., Bardsley, 2008; Levitt and List, 2007; List, 2007) and indicates a distinction between revealed social preferences as measured commonly and the elicitation of normatively appropriate behavior using coordination games.

Our main conclusion is that, when conducting economic laboratory experiments, the degree of experience of participants needs to be taken care of. Since it is difficult to control explicitly for the exact number of previous participations by eliciting that information from subjects (as is done with gender or age), that characteristic needs to be properly randomized between treatments, when sessions are organized.[4] To ensure this, the recruitment bias identified by Benndorf et al. (2017), i.e., that the share of inexperienced subjects tends to be lower in early recruitment waves, needs to be considered.

The remainder of the paper is organized as follows. Section 2 provides a literature overview and derives hypotheses. Section 3 presents the experiment and section 4 the experimental results. Section 5 contains exploratory analyses on generalizability from the lab to the field and on socio-demographics. Section 6 summarizes and concludes.

2. Related Literature and Hypotheses

2.1. Related Literature

Our paper relates to studies which examine selection effects associated with recruitment to lab experiments. Two types of selection effects need to be distinguished in that regard: selection *into*

[4] Note that it is difficult to simply use the number of participations of a subject as recorded in the data base, since this would require to give up the anonymity of participants towards the experimenter. This is because it would be necessary to connect a participant's profile from the database with the data she produced in the experiment, which would require to identify which individual is sitting at which slot in the laboratory. Making that connection, however, is not in accordance with the usual policy to have subjects participate in economics experiments in a fully anonymous manner.

the subject pool and selection *out of* the subject pool. Selection into the subject pool results if subjects with specific characteristics have a higher probability of entering the subject pool through registration.[5] Selection out of the subject pool results if registered subjects with specific characteristics vary in the probability to drop-out after having participated once. In our study, we contribute to the topic of selection out of the subject pool by comparing first-time participants with those who participated many times, i.e., with the group of subjects that tends to remain in the pool.

For example, Casari et al. (2007) find that subjects are more likely to participate in a follow-up study, the more successful they were in monetary terms in a previous experiment. Similarly, Guillén and Veszteg (2012) find that earnings in previous experiments positively correlate with the probability of participating in future experiments. Thus, it has been hypothesized that more selfish subjects, which consequently earn more money in experiments, are more likely to regularly participate. As a result, it might be that common subject pools contain over-proportionally large shares of selfish individuals.

Another literature related to our study examines differences between inexperienced and experienced participants. Matthey and Regner (2013) use data about participants´ behavior in previously conducted dictator games, ultimatum games, and trust games and find that the number of participations is negatively correlated with sharing behavior in all three games. Based on post-experimental questionnaires, they conclude that repeated participation in experiments involving allocation decisions leads to learning effects through negative experiences. Benndorf et al. (2017) directly test for behavioral differences between participants with extensive lab experience and first-time participants across four one-shot two-player games (trust game, beauty contest, ultimatum game, and traveler's dilemma) and two individual decisions (lying task and risk preferences). In the trust game, experienced subjects trust less often, and they also behave significantly more selfish as second movers. In the risk elicitation tasks, experienced participants submit fewer non-monotonic strategies. The authors also document a recruitment bias as the share of inexperienced subjects was lower in early recruitment waves (i.e., in initial sessions of an experiment).

[5] The evidence on selection into the subject pool is mixed, although the majority of studies examining that kind of selection report null results. While Eckel and Grossman (2000) do identify differences in social preferences between registered and non-registered subjects, Abeler and Nosenzo (2015), Cleave et al. (2013), Falk et al. (2013), and Slonim et al. (2013) do not identify meaningful differences. Krawczyk (2011) examines optimal advertisement of participation in experiments and finds that recruitment is more effective when emphasizing pecuniary benefits of participation. Subjects that were recruited with advertisement of pecuniary benefits were less altruistic.

2.2. Hypotheses

We elicit social norm perception both for the lab and the field context. Based on the results from Benndorf et al. (2017) and Matthey and Regner (2013), which identify that behavior in the lab is related to experience, we hypothesize that this relationship is reflected in social norm perception.

Hypothesis 1. The perception of social norms for the context of lab experiments differs between inexperienced and experienced participants.

Regarding real-world norms, we again test the hypothesis that inexperienced and experienced participants differ. This hypothesis follows the idea that selection effects lead to an over-proportionally large share of selfish participants in the subject pool, as suggested by the literature on selection effects that result from drop-out (Casari et al., 2007; Guillén and Veszteg, 2012). Differences in field norms would thus indicate that selection also potentially explains differences in lab norms between inexperienced and experienced participants. By contrast, little or no differences in field norms would support the hypothesis that such differences mainly result from learning through repeated participation.

Hypothesis 2. The perception of social norms for the context of daily life differs between inexperienced and experienced participants.

3. Experimental Design and Procedures

3.1. Experimental Design

Conceptually, the experiment is divided into three parts and structured in five modules. In part 1, injunctive social norms and descriptive social norms are elicited in a series of hypothetical mini-dictator games. In part 2, injunctive social norms regarding unsocial behaviors in the lab and the field are measured. In part 3, subjects evaluate the generalizability from the lab to the field. Each module contains five items, and we use coordination games to measure social norm perception throughout the whole experiment (Krupka and Weber, 2013). In each item, subjects are asked a question, and they coordinate on one of four answer possibilities. At the end of the experiment, one

of the 25 items is selected at random. If a subject´s answer in that item matches the modal choice in the current session, the subject earns 10€ (and 0€ otherwise).[6]

Modules 1 and 2: Allocation decisions in the lab. We elicit *injunctive* social norms (module 1) and *descriptive* social norms (module 2) in a series of hypothetical allocation decisions. Injunctive norms indicate perceptions about normatively appropriate behavior in a specific context. They reflect what kind of behavior is approved or disapproved by the community and thereby motivate actions through the anticipation of social rewards or punishment. In contrast, descriptive social norms refer to what kind of behavior is assumed to be common or prevalent (Cialdini et al., 1990).[7]

Five mini-dictator games are used that allow differentiating between competing distributional motives: efficiency, egalitarianism, and profit maximization. At the beginning of the modules, subjects learn the rules of the classical dictator game paradigm[8] used in economic lab experiments and they are instructed to imagine that these allocation decisions would be used in an actual lab experiment.[9] Table 1 shows the five hypothetical mini-dictator games and how the choices correspond to the distributional motives.

Table 1. Hypothetical Mini-Dictator Games used in Module 1 and 2

	Option 1		Option 2		Distributive Motives		
Game	Dictator	Recipient	Dictator	Recipient	Efficiency	Egalitarianism	Profit max.
1	15€	5€	11€	11€	Option 2	Option 2	Option 1
2	10€	10€	10€	15€	Option 2	Option 1	-
3	15€	5€	9€	9€	Option 1	Option 2	Option 1
4	10€	9€	9€	11€	Option 2	Option 1	Option 1
5	12€	8€	10€	10€	-	Option 2	Option 1

[6] We take care to make sure that subjects understand the coordination mechanism by reminding them before each item that their task is *not to state their own opinion*, but to *coordinate* with the remaining participants in the room. Only one item is paid in order to avoid hedging and ensure incentive compatibility (Azrieli et al., 2018).

[7] Cialdini et al. (1990) summarizes injunctive norms as "norms of ought" and descriptive norms as "norms of is". Experimental studies find that both types of norms explain behavior, but also that the two norms are conceptually different constructs that independently affect intentions and behavior (e.g., Cialdini and Kallgren, 1993; Kallgren et al., 2000; Rivis and Sheeran, 2003). This research also shows that behavior rates are typically highest when injunctive and descriptive norms are aligned.

[8] Each step of the classical dictator game paradigm (anonymity, randomization of roles, matching, decision rights, and payout function) is explained in detail. We use the term „Player A" for the dictator and „Player B" for the recipient.

[9] Precisely, they should imagine that the allocations would take place in an experiment like the one they are located in at that moment.

For simplicity, we always ask to evaluate option 1.[10] For injunctive norms in module 1, subjects indicate "how appropriate it would be, to choose option 1" in the role of the dictator by coordinating on these answer options: "very appropriate", "somewhat appropriate", "somewhat inappropriate" or "very inappropriate". For descriptive social norms in module 2 subjects indicate "which of the two options would be chosen more often" by coordinating on these answer options: "option 1 much more often", "option 1 somewhat more often", "option 2 somewhat more often" or "option 2 much more often".

Modules 3 and 4: Evaluation of unsocial behaviors. We study social norm perception in the lab and the field by eliciting injunctive social norms regarding a series of unsocial behaviors: selfishness, exploitation, spitefulness, deception, and willful ignorance. Subjects are confronted with the statement that the respective behavior would be appropriate and they then indicate the degree of consent with the respective statement by coordinating on: "fully agree", "somewhat agree", "somewhat disagree" or "fully disagree". We use the identical set of items and frame them once to the lab (module 3) and once to the field (module 4). Table 2 shows the statements we use in the two modules.

Table 2. Items used in Modules 3 and 4

As a participant in a laboratory experiment, it is appropriate to … / In daily life, it is appropriate to …
1) … mainly consider the own well-being.
2) … take advantage of other subjects, when this leads to a material advantage for oneself.
3) … harm other subjects, even when this does not lead to a material advantage for oneself.
4) … deceive other subjects, in order to materially gain from it.
5) … remain ignorant about the consequences that the own decisions have on other people.

Notes: The wording "As a participant in an experiment, it is appropriate to…" refers to module 3 and "In daily life, it is appropriate to…" refers to module 4.

Module 5: Generalizability of lab behavior. In module 5, we elicit perceptions about the generalizability of lab behavior. Again, we confront subjects with a set of statements and have them coordinate on the degree of consent. Table 3 contains the items used in module 5.

[10] Note that option 1 is always dominant for the dictator in terms of profit maximization.

Table 3. Items used in Module 5

1) As a participant in an experiment, I have **the same** moral standards regarding my own behavior as in daily life.
2) As a participant in an experiment, I have **the same** moral standards regarding the behavior of others as in daily life.
3) Selfishness in the lab is **not the same** as selfishness in daily life.
4) Social norms in the laboratory are **not the same** as social norms in daily life.
5) My behavior as a participant in an experiment is **representative** of my behavior in daily life.

Notes: In the experiment, none of the words were printed boldly. The bold print here is to illustrate which of the statements are affirmations (suggesting similarity between the two contexts) and which of the statements are negations (suggesting dissimilarity between the two contexts).

Order of modules. To mitigate order effects, we vary the order of modules as well as the order of items within modules. Moreover, we avoid that those modules which are subject to comparison (module 1 and 2; module 3 and 4) appear consecutively, in order to reduce spillover effects between modules. Also, we elicit norms of daily life always at the end to avoid priming field context before eliciting perceptions about the lab context. We test for order effects but do not find an interaction between the different order variants and the subjects´ choices.

3.2. Procedures

We conducted sessions either only with inexperienced subjects (no prior participation) or only with experienced subjects (at least 10 participations). Subjects were not informed about the fact that they were recruited as a specific subpart of the participant pool. In total, we recruited 82 inexperienced subjects and 68 experienced subjects. From the 82 inexperienced participants, 9 were excluded from the analysis because they stated in a post-experimental questionnaire that they already participated in at least one economic or psychological lab experiment before. Thus, 73 inexperienced and 68 experienced subjects remained in the analysis, leaving us with a total *N* of 141 observations and fairly balanced sample sizes for the two groups. The experiment was programmed in z-Tree (Fischbacher 2007), recruitment was done via hroot (Bock et al., 2014), and the sessions were conducted at the experimental laboratory of the University of Heidelberg, Germany, between November 2016 and May 2017. A typical session lasted about 35 minutes and subjects earned on average 10.30€ including a show-up fee of 3€.[11]

[11] A replication package, including data and data analysis files, is available at https://doi.org/10.11588/data/SS8CBF.

4. Results

4.1. Part 1: Allocation Decisions in the Laboratory

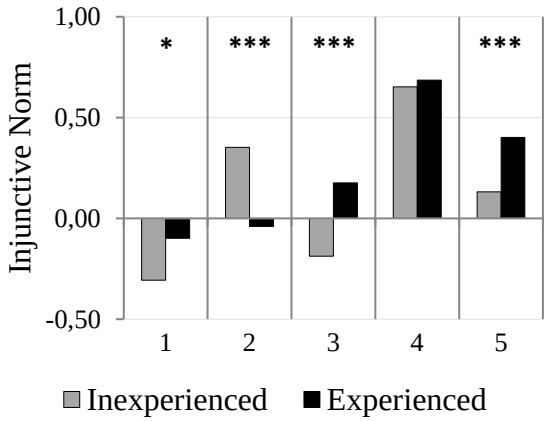
To analyze modules 1 and 2, we quantify the answers such that the resulting scores are normalized between -1 and 1. For injunctive norms in module 1, we quantify the answers as: 1 = ”very appropriate”, 1/3 = ”somewhat appropriate”, -1/3 = “somewhat inappropriate”, -1 = “very inappropriate”. Thus, the more positive (negative) the score in module 1, the more appropriate (inappropriate) it is considered to choose option 1 in the respective decision. For descriptive norms in module 2, we quantify the answers as: 1 = “option 1 much more often”, 1/3 = “option 1 somewhat more often”, -1/3 = “option 2 somewhat more often”, -1 = “option 2 much more often”. The more positive (negative) the score in module 2, the more common choosing option 1 (option 2) is considered in the respective decision. Figure 1 provides descriptive analysis to give an impression about coordination outcomes.

In order to draw statistical inferences on differences between inexperienced and experienced participants and to control for potential confounds, we conduct regression analyses (Table 4). Regression results suggest that inexperienced and experienced subjects differ both concerning injunctive norms and descriptive norms in most of the allocation decisions, with the latter diverging more strongly. In module 1, the two groups differ in three items before the correction procedure (items 2, 3, and 5), but two of these differences vanish after applying the Bonferroni correction.[12] The results regarding injunctive norms indicate that the two groups differ in particular regarding their evaluation of the competing motives captured in item 3. In that item, a dictator chooses between advantageous efficiency and egalitarianism. While opting for advantageous efficiency is considered rather inappropriate by inexperienced subjects (score for injunctive norm of -0.19), it is evaluated as rather appropriate by experienced subjects (score for injunctive norm of 0.18). In module 2, the differences are considerably stronger than in module 1. In particular, four items (1, 2, 3, and 5) differ between the two groups, and all of these significances survive the correction procedure.

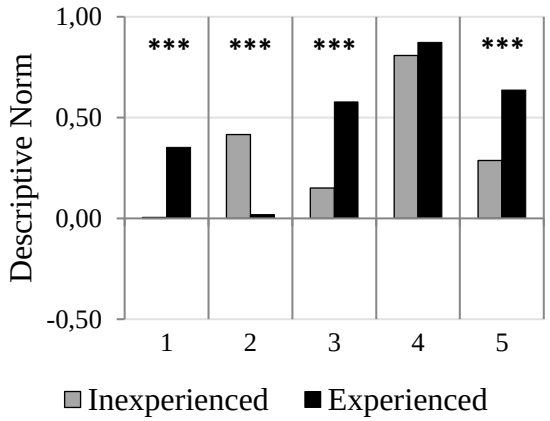
[12] We account for the fact that multiple items are used within modules to detect differences between inexperienced and experienced participants. In order to take care of the inflation of the overall type-I-error rate, we therefore multiply the *p*-values by the number of items within a module (i.e., by five).

Figure 1. Allocation Decisions in the Laboratory

Panel A. Injunctive Norms (Module 1)



Panel B. Descriptive Norms (Module 2)



Notes: *, **, *** indicates significance at the 10%, 5%, and 1% without correcting for multiple testing; Mann-Whitney-U tests. Positive (negative) values on Panel A represent that it is considered to be appropriate (inappropriate) to choose option 1 in the respective mini-dictator game. Positive (negative) values on Panel B represent that option 1 (option 2) is considered to be chosen more often in the respective mini-dictator game.

Table 4. Regression Analysis on Injunctive and Descriptive Social Norms

	Panel A. Injunctive Social Norms (Module 1)				
	1	2	3	4	5
Experienced	0.302 (0.194)	-0.404* (0.211)	0.608***/### (0.195)	0.123 (0.213)	0.447** (0.197)
# Obs.	141	141	141	141	141
	Panel B. Descriptive Social Norms (Module 2)				
	1	2	3	4	5
Experienced	0.555***/## (0.196)	-0.545***/## (0.203)	0.668***/### (0.202)	0.145 (0.254)	0.524***/## (0.205)
# Obs.	141	141	141	141	141

Notes: Ordered probit regressions. *, **, *** indicates levels of significance at the 10%, 5%, and 1% level before correcting for multiple testing. #,##, ### indicates levels of significance at the 10%, 5%, and 1% level after correcting for multiple testing according to the Bonferroni method. Standard errors are clustered on the individual level and reported in parentheses. Controls are gender, age, and field of study (economics or not).

We further analyze the distributive motives reflected in social norm perception. To this end, we calculate scores that reflect the relative importance of efficiency, egalitarianism, and profit

maximization on the individual level and run regression analyses (table 5).[13] The results show that experience is systematically related to these motives. Consistent with Benndorf et al. (2017) and Matthey and Regner (2013), inexperienced subjects pronounce egalitarianism, while experienced subjects pronounce efficiency and profit orientation. The results from modules 1 and 2 support hypothesis 1.

Result 1. Inexperienced and experienced participants differ both concerning injunctive and descriptive social norms in allocation decisions. Regarding both types of norms, egalitarianism is more pronounced in inexperienced subjects, while efficiency and profit orientation are more pronounced in experienced subjects.

Table 5. Regression Analysis on Distributive Motives

	Panel A. Injunctive Social Norms			**Panel B.** Descriptive Social Norms		
	Efficiency	Egalitarianism	Profit maximization	Efficiency	Egalitarianism	Profit maximization
Experienced	0.081**	-0.194***	0.152***	0.075**	-0.272***	0.210***
	(0.039)	(0.063)	(0.057)	(0.034)	(0.062)	(0.064)
Constant	-0.113	0.376**	-0.095	-0.137*	0.038	0.173
	(0.092)	(0.146)	(0.133)	(0.079)	(0.146)	(0.150)
# Obs.	141	141	141	141	141	141

Notes: OLS regressions. *, **, *** indicates significance at the 10%, 5%, and 1% level without correcting for multiple testing. Standard errors are clustered on the individual level and reported in parentheses. Controls are gender, age, and field of study (economics or not). As a robustness check, we conduct Tobit regressions which yield the same results.

4.2. Part 2: Evaluation of Unsocial Behaviors

To analyze module 3 (module 4), we again first present descriptive results in Figure 2 and then run regression analyses on the degree of consent that the described behaviors are considered to be appropriate in the laboratory (the field).[14] Comparing perceptions about the laboratory context, we

[13] Precisely, a score on the individual level for a particular distributive motive (efficiency/egalitarianism/profit) is calculated as follows. We take the choice of a particular subject in a particular item. We then multiply a dummy variable that indicates if option 1 is the efficient/equal/profit maximizing option (dummy = 1) or not (dummy = -1) with the choice score of the subject (i.e., 1, 1/3, -1/3 or -1) in that item. For each distributive motive, this is done for each item. The resulting scores are then averaged within the five items of a module regarding a specific distributive motive. This procedure is done for each subject, both for injunctive and for descriptive norms. The resulting scores then reflect the importance of a particular distributive motive in module 1 and module 2 on the subject level. In Appendix A.2 the reader finds an example of how we calculated the scores.

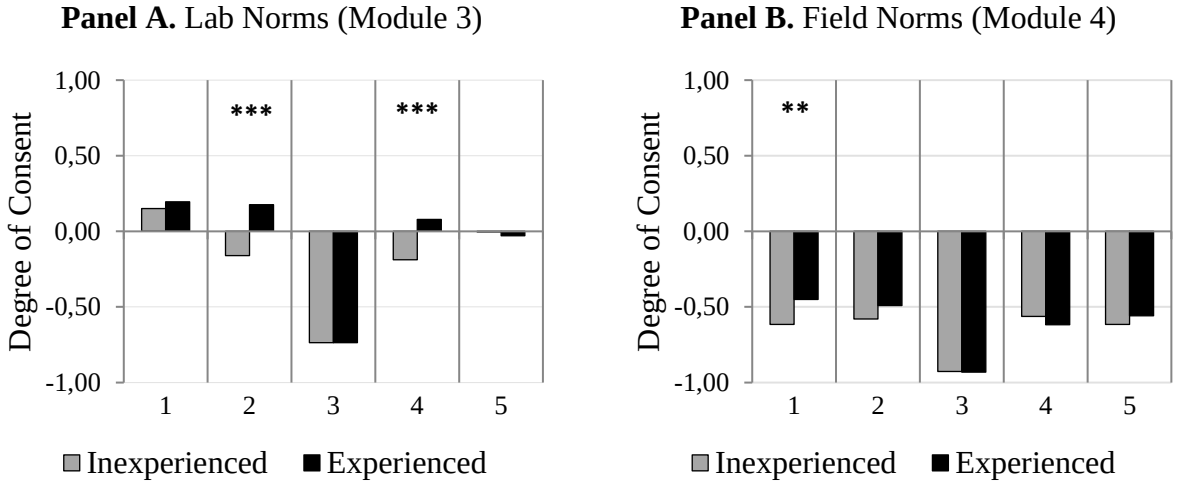
[14] Figure 2 shows that behaving unsocial in the lab is considered significantly more appropriate than behaving unsocial in the context of daily life (we elaborate on that finding in section 5.1).

find that for experienced subjects, it is significantly more appropriate to exploit and deceive other participants within the lab context (see the regressions in Panel A of Table 6). The results are highly significant and robust to the correction procedure. The observed differences further support hypothesis 1.

By contrast to lab norms, real-world norms are homogenous with respect to experience (see panel B of Table 6). Item 1, which refers to selfishness, is marginally significantly different between the groups, indicating that selfishness in daily life is considered more appropriate by experienced participants. This difference, however, is not robust to the correction procedure. The results thus do not support hypothesis 2.

Result 2. Lab norms differ between the two groups, as it is significantly more appropriate to exploit and deceive other participants within the lab context for experienced subjects. By contrast, real-world norms are homogenous with respect to experience.

Figure 2. Unsocial Behaviors in the Lab and the Field



Notes: *, **, *** indicates significance at the 10%, 5%, and 1% without correcting for multiple testing; Mann-Whitney-U tests. Positive (negative) values indicate that the behavior described in an item is considered more (less) appropriate.

Table 6. Regression Analysis on Appropriateness of Unsocial Behaviors

	Panel A. Unsocial Behavior in the Lab (Module 3)				
	Selfishness	Exploitation	Spitefulness	Deception	Ignorance
Experienced	0.134	0.639***/###	-0.019	0.640***/###	-0.041
	(0.193)	(0.200)	(0.231)	(0.198)	(0.193)
# Obs.	141	141	141	141	141
	Panel B. Unsocial Behavior in the Field (Module 4)				
	Selfishness	Exploitation	Spitefulness	Deception	Ignorance
Experienced	0.384*	0.216	0.109	-0.166	0.222
	(0.204)	(0.204)	(0.327)	(0.209)	(0.210)
# Obs.	141	141	141	141	141

Notes: Ordered probit regressions. *, **, *** indicates levels of significance at the 10%, 5%, and 1% level before correcting for multiple testing. #,##, ### indicates levels of significance at the 10%, 5%, and 1% level after correcting for multiple testing according to the Bonferroni method. Regressions in Panel A (Panel B) are ran on the degree of consent that the respective behavior is considered to be an appropriate behavior in the lab (field) context. Standard errors are clustered on the individual level and reported in parentheses. Controls are gender, age, and field of study (economics or not).

4.3. Part 3: Perceptions about Generalizability

In module 5, subjects are asked to evaluate the generalizability of lab behavior. Table 7 shows regression analyses on the degree of consent with the items used in module 5. Item 1 indicates that experienced subjects more strongly agree to the statement that they do not have the same moral standards regarding their own behavior as in daily life. That is, experienced subjects more strongly agree to a statement that suggests that the two contexts are different. However, the difference is not robust to the correction procedure. We therefore do not find that the two groups differ in how they evaluate the generalizability of lab behavior.

Result 3. Inexperienced and experienced subjects do not differ in how they perceive the relation between behavior in the lab and behavior in the field.

Table 7. Regression Analysis on Perceptions about Generalizability

	Item 1 (+)	Item 2 (+)	Item 3 (-)	Item 4 (-)	Item 5 (+)
Experienced	-0.398**	-0.022	-0.285	0.002	-0.191
	(0.200)	(0.196)	(0.198)	(0.197)	(0.199)
# Obs.	141	141	141	141	141

Notes: Ordered probit regressions. *, **, *** indicates uncorrected levels of significance at the 10%, 5%, and 1% level. #,##, ### indicates levels of significance at the 10%, 5%, and 1% level after correcting for multiple testing according to the Bonferroni method. Regressions are ran on the degree of consent with the respective item. The (+) and (-) indicate whether the statements represent affirmations (implying similarity between the lab and field) or negations (implying dissimilarity between the lab and field). Standard errors are clustered on the individual level and reported in parentheses. Controls are gender, age, and field of study (economics or not).

5. Exploratory Analyses

5.1. Generalizability of Lab Norms

Although it is not the focus of this study, the data allows us to study generalizability by comparing modules 3 and 4 and testing whether lab and field norms correspond. By holding the content of the items fixed and varying the framing (lab vs. field), we can isolate the effect of the context. For that sake, we again quantify the answers such that the resulting scores are normalized between -1 and 1. The more positive (negative) the score, the stronger subjects agree (disagree) with the statement that a particular behavior is considered to be appropriate in the respective context.

We first conduct correlation analyses between lab norms and field norms. Table 8 shows that these are positive and mainly significant. We next compare the absolute values of the degree of consent between the two modules. Table 9 shows the same data as presented in Figure 2, but now the results regarding the two contexts are compared. The results indicate that perceptions about the appropriateness of the unsocial behaviors described in the items differ substantially between the two contexts, as each behavior is considered significantly *less appropriate in the field than in the lab*. This applies independently from the degree of experience. All differences remain highly statistically significant after the correction procedure. The finding stands in contrast with the hypothesis, that social preferences measured in the lab are inflated (e.g., Levitt and List 2007) and the results thus might indicate a distinction between revealed social preferences as elicited commonly and the elicitation of normatively appropriate behavior using coordination games.

Result 3. Unsocial behavior in the lab context is considered substantially more appropriate than in the context of daily life.

Table 8. Correlations between Laboratory and Field Norms

	Selfishness	Exploitation	Spitefulness	Deception	Ignorance
Inexperienced	0.204*	0.302***/##	0.291**/#	0.204*	0.260**
Experienced	0.103	0.091	0.526***/###	0.268**	0.041
All subjects	0.157*	0.237***/##	0.404***/###	0.220***/##	0.157*

Notes: Spearman rank correlation. *, **, *** indicates significance at the 10%, 5%, and 1% level before correcting for multiple testing. #,##, ### indicates levels of significance at the 10%, 5%, and 1% level after correcting for multiple testing according to the Bonferroni method.

Table 9. Comparison Between Laboratory and Field Norms

		Selfishness	Exploitation	Spitefulness	Deception	Ignorance
Inexperienced participants	Lab context	0.15	-0.16	-0.73	-0.19	0.00
	Field context	-0.62	-0.58	-0.93	-0.56	-0.61
	Difference	-0.77***/###	-0.42***/###	-0.20***/###	-0.37***/###	-0.61***/###
Experienced participants	Lab context	0.20	0.18	-0.73	0.08	-0.03
	Field context	-0.45	-0.49	-0.93	-0.62	-0.56
	Difference	-0.65***/###	-0.67***/###	-0.20***/###	-0.70***/###	-0.53***/###

Notes: Two-sided t-tests. *, **, *** indicates significance at the 10%, 5%, and 1% level before correcting for multiple testing. #,##, ### indicates levels of significance at the 10%, 5%, and 1% level after correcting for multiple testing according to the Bonferroni method. The numbers represent the degree of consent that the behavior stated in an item is considered to be an appropriate behavior in the respective context.

5.2. Socio-Demographics: Age, Gender and Field of Study

Throughout all analyses, we control for age, gender and economic study. In Appendix A.1, we report complete regression analyses including coefficients of control variables (age, gender, and field of study). In this section, we report those findings that are significant at the 5%-level without correction for multiple testing.

We find that all differences between inexperienced and experienced participants are fully independent of differences in age after controlling for experience. However, we identify some interesting patterns regarding gender and field of study. First, consistent with previous studies on gender effects in dictator game giving (e.g., Andreoni and Vesterlund, 2001; Eckel and Grossman,

1998), female subjects are significantly more guided by egalitarianism and significantly less guided by efficiency in experimental allocation tasks than male subjects (see Table 11). However, females consider it to be more appropriate to ignore the consequences that their own decisions have on other people in a lab experiment. Second, economics students consider several unsocial behaviors (exploitation, spitefulness, and deception) in the laboratory context as more appropriate than non-economics students (see Panel A of Table 12). These differences, however, only refer to norm perception in the laboratory context. Regarding norm perception in the field context, all sub-group differences vanish.

Result 4. Injunctive and descriptive social norms of female subjects are guided more strongly by egalitarianism and less strongly by efficiency than social norms of male subjects.

Result 5. In the lab context, economics students consider exploitation, spitefulness, and deception as more appropriate than non-economics students.

6. Summary and Conclusion

We compare social norm perception of inexperienced and experienced participants in economic laboratory experiments using the Krupka and Weber (2013) approach. We find that the two groups differ both concerning injunctive norms and descriptive norms in allocation decisions in the lab, with the latter diverging more strongly. Consistent with Benndorf et al. (2017) and Matthey and Regner (2013), egalitarianism is more pronounced in norm perception of inexperienced subjects, while efficiency and profit maximization dominate in experienced subjects. We complement these results with the finding that experienced subjects consider exploitation and deception of other participants in the lab as more appropriate than inexperienced subjects. The results demonstrate that not only revealed social preferences (Matthey and Regner, 2013) are related to the number of participations, but that also social norm perception, which potentially mediates the differences in behavior, is different between subjects with varying degrees of experience.

We also compare norm perception for the context of daily life and find that these do not differ between the two groups. We thus do not find support for the hypothesis that selection effects through drop-out lead to an over proportionally large share of selfish individuals in the subject pool (Casari et al. 2007, Guillén and Veszteg 2012). We therefore conclude that learning is more important than selection effects for explaining differences that are linked to experience. For a

conclusive analysis of the relative importance of learning and selection, however, further research is necessary. In particular, it might be interesting to compare field behavior more comprehensively, as the scope of analyses of that context was smaller than the scope on the analysis of the lab context in our study.

Finally, we conduct exploratory analyses in order to contribute to the generalizability debate by comparing social norm perception between the lab and the field. We find that norm perception between the two contexts is correlated. However, independent from the degree of experience, behaving unsocially in the lab is considered significantly more appropriate than in the real-world. This finding stands in contrast with the hypothesis that social preferences measured in the lab are inflated (e.g., Levitt and List 2007) and indicates a distinction between revealed social preferences and the elicitation of normatively appropriate behavior using coordination games.

Our results corroborate the idea that, when conducting economic laboratory experiments, the degree of "lab experience" of participants needs to be taken care of. We therefore conclude that it is essential to make sure that this characteristic is properly randomized between treatments, and that this should be monitored in the invitation phase of the recruitment process. To ensure this, the recruitment bias identified by Benndorf et al. (2017), i.e., that the share of inexperienced subjects tends to be lower in early recruitment waves, needs to be considered.

Appendix

A.1. Complete Regression Analyses reporting Coefficients of Control Variables

A.1.1. Complete Regression Analysis on Injunctive and Descriptive Social Norms

Table 10. Regression Analysis on Injunctive and Descriptive Social Norms

	Panel A. Injunctive Social Norms (Module 1)				
	1	2	3	4	5
Experienced	0.302 (0.194)	-0.404* (0.211)	0.608*** (0.195)	0.123 (0.213)	0.447** (0.197)
Age	0.024 (0.018)	0.000 (0.018)	0.020 (0.018)	0.003 (0.019)	0.023 (0.019)
Female	0.141 (0.191)	0.749*** (0.208)	0.035 (0.191)	0.119 (0.211)	-0.131 (0.192)
Economics	0.223 (0.191)	-0.133 (0.207)	0.175 (0.191)	-0.274 (0.209)	0.056 (0.193)
# Obs.	141	141	141	141	141
	Panel B. Descriptive Social Norms (Module 2)				
	1	2	3	4	5
Experienced	0.555*** (0.196)	-0.545*** (0.203)	0.668*** (0.202)	0.145 (0.254)	0.524*** (0.205)
Age	-0.011 (0.017)	0.020 (0.019)	0.022 (0.020)	0.077* (0.043)	0.045* (0.027)
Female	-0.244 (0.192)	0.565*** (0.201)	-0.154 (0.197)	0.256 (0.249)	-0.208 (0.202)
Economics	-0.009 (0.191)	-0.074 (0.201)	0.136 (0.197)	-0.085 (0.248)	0.093 (0.203)
# Obs.	141	141	141	141	141

Notes: Ordered probit regressions. *, **, *** indicates significance at the 10%, 5%, and 1% level without correcting for multiple testing. Standard errors are clustered on the individual level and reported in parentheses.

A.1.2. Complete Regression Analysis on Distributive Motives

Table 11. Regression Analysis on Distributive Motives

	Injunctive Social Norms			Descriptive Social Norms		
	Efficiency	Egalitarianism	Profit maximization	Efficiency	Egalitarianism	Profit maximization
Experienced	0.081**	-0.194***	0.152***	0.075**	-0.272***	0.210***
	(0.039)	(0.063)	(0.057)	(0.034)	(0.062)	(0.064)
Age	0.000	-0.007	0.006	-0.002	0.001	0.005
	(0.004)	(0.006)	(0.005)	(0.003)	(0.006)	(0.006)
Female	-0.127***	0.112*	0.010	-0.078**	0.156**	-0.058
	(0.039)	(0.062)	(0.056)	(0.034)	(0.062)	(0.064)
Economics	0.028	-0.082	0.028	0.026	-0.034	0.018
	(0.039)	(0.062)	(0.056)	(0.034)	(0.062)	(0.064)
Constant	-0.113	0.376**	-0.095	-0.137*	0.038	0.173
	(0.092)	(0.146)	(0.133)	(0.079)	(0.146)	(0.150)
# Obs.	141	141	141	141	141	141

Notes: OLS regressions. *, **, *** indicates significance at the 10%, 5%, and 1% level; standard errors are clustered on the individual level and reported in parentheses. As a robustness check, we conduct Tobit regressions which yield the same results.

A.1.3. Complete Regression Analysis on Evaluation of Unsocial Behaviors

Table 12. Regression Analysis on Appropriateness of Unsocial Behaviors

	Panel A. Unsocial Behavior in the Lab (Module 3)				
	Selfishness	Exploitation	Spitefulness	Deception	Ignorance
Experienced	0.134 (0.193)	0.639*** (0.200)	-0.019 (0.231)	0.640*** (0.198)	-0.041 (0.193)
Age	-0.014 (0.017)	0.012 (0.018)	-0.001 (0.024)	-0.028 (0.019)	0.018 (0.018)
Female	-0.014 (0.191)	0.230 (0.195)	0.367 (0.230)	0.249 (0.193)	0.561*** (0.195)
Economics	0.037 (0.191)	0.453** (0.196)	0.636*** (0.221)	0.628*** (0.197)	0.154 (0.192)
# Obs.	141	141	141	141	141
	Panel B. Unsocial Behavior in the Field (Module 4)				
	Selfishness	Exploitation	Spitefulness	Deception	Ignorance
Experienced	0.384* (0.204)	0.216 (0.204)	0.109 (0.327)	-0.166 (0.209)	0.222 (0.210)
Age	-0.012 (0.019)	0.000 (0.018)	-0.008 (0.036)	-0.002 (0.019)	-0.003 (0.019)
Female	-0.248 (0.200)	-0.221 (0.203)	0.416 (0.336)	-0.082 (0.206)	0.260 (0.208)
Economics	0.175 (0.200)	-0.074 (0.203)	-0.021 (0.320)	-0.087 (0.207)	-0.066 (0.207)
# Obs.	141	141	141	141	141

Notes: Ordered probit regressions. *, **, *** indicates significance at the 10%, 5%, and 1% level without correcting for multiple testing. Standard errors are clustered on the individual level and reported in parentheses. Regressions in Panel A (Panel B) are ran on the degree of consent that the respective behavior is considered to be an appropriate behavior in the lab (field) context.

A.1.4. Complete Regression Analysis on Perceptions Generalizability

Table 13. Regression Analysis on Perceptions about Generalizability

	Agreement with Statements				
	Item 1 (+)	Item 2 (+)	Item 3 (-)	Item 4 (-)	Item 5 (+)
Experienced	-0.398**	-0.022	-0.285	0.002	-0.191
	(0.200)	(0.196)	(0.198)	(0.197)	(0.199)
Age	-0.003	-0.001	0.007	-0.001	0.016
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Female	-0.115	0.098	0.123	0.001	0.065
	(0.196)	(0.194)	(0.196)	(0.195)	(0.196)
Economics	0.133	0.133	0.048	0.313	0.302
	(0.196)	(0.195)	(0.196)	(0.197)	(0.198)
# Obs.	141	141	141	141	141

Notes: Ordered probit regressions. *, **, *** indicates significance at the 10%, 5%, and 1% level without correcting for multiple testing. Standard errors are clustered on the individual level and reported in parentheses. Regressions are ran on the degree of consent with the respective item. The (+) and (-) indicate whether the statements represent affirmations (implying similarity between the lab and field) or negations (implying dissimilarity between the lab and field).

A.2. Example Calculation of Scores for Distributive Motives

We illustrate how we calculate the scores that reflect the relative importance of efficiency, egalitarianism, and profit maximization on the individual level. The following table depicts choices of an exemplary subject in module 1 or module 2. The columns "Efficiency-Dummy" / "Egalitarianism-Dummy" / "Profit-Dummy" indicate if Option 1 in the respective item is the efficient / egalitarian / profit-maximizing choice.

Item	Choices in Module 1	Choices in Module 2	Efficiency-Dummy	Egalitarianism-Dummy	Profit-Dummy
1	Somewhat inappropriate (-1/3)	Option 2 somewhat more often (-1/3)	-1	-1	1
2	Very inappropriate (-1)	Option 2 much more often (-1)	-1	1	-
3	Somewhat inappropriate (-1/3)	Option 1 somewhat more often (+1/3)	1	-1	1
4	Very appropriate (+1)	Option 1 much more often (+1)	-1	1	1
5	Somewhat appropriate (+1/3)	Option 1 much more often (+1)	-	-1	1

The score representing the importance of the distributive motives in the two modules are then calculated as follows:

- Efficiency (injunctive): [(-1/3)*(-1)+(-1)*(-1)+(-1/3)*1+1*(-1)+(1/3)*0]/5=**0.00**
- Egalitarianism (injunctive): [(-1/3)*(-1)+(-1)*1+(-1/3)*(-1)+1*1+(1/3)*(-1)]/5=**0.07**
- Profit max. (injunctive): [(-1/3)*1+(-1)*0+(-1/3)*1+1*1+(1/3)*1]/5=**0.13**

- Efficiency (descriptive): [(-1/3)*(-1)+(-1)*(-1)+(1/3)*1+1*(-1)+1*0]/5=**0.13**
- Egalitarianism (descriptive): [(-1/3)*(-1)+(-1)*1+(1/3)*(-1)+1*1+1*(-1)]/5=**-0.20**
- Profit max. (descriptive): [(-1/3)*1+(-1)*0+(1/3)*1+1*1+1*1]/5=**0.40**

In Table 5, the level of experience is then regressed on these values.

References

Abeler, J., & Nosenzo, D. (2015). Self-selection into laboratory experiments: pro-social motives versus monetary incentives. *Experimental Economics*, *18*(2), 195-214.

Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *The Quarterly Journal of Economics*, *116*(1), 293-312.

Azrieli, Y., Chambers, C. P., & Healy, P. J. (2018). Incentives in experiments: A theoretical analysis. *Journal of Political Economy*, *126*(4), 1472-1503.

Bardsley, N. (2008). Dictator game giving: altruism or artefact?. *Experimental Economics*, *11*(2), 122-133.

Benndorf, V., Moellers, C., & Normann, H. T. (2017). Experienced vs. inexperienced participants in the lab: do they behave differently?. *Journal of the Economic Science Association*, *3*(1), 12-25.

Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, *71*, 117-120.

Casari, M., Ham, J. C., & Kagel, J. H. (2007). Selection bias, demographic effects, and ability effects in common value auction experiments. *American Economic Review*, *97*(4), 1278-1304.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015.

Cleave, B. L., Nikiforakis, N., & Slonim, R. (2013). Is there selection bias in laboratory experiments? The case of social and risk preferences. *Experimental Economics*, *16*(3), 372-382.

Crawford, S. E., & Ostrom, E. (1995). A grammar of institutions. *American Political Science Review*, *89*(3), 582-600.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67-80.

Dawes, R. M. (1989). Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology*, *25*(1), 1-17.

Dawes, R. M. (1990). The potential nonfalsity of the false consensus effect.

Eckel, C. C., & Grossman, P. J. (2000). Volunteers and pseudo-volunteers: The effect of recruitment method in dictator experiments. *Experimental Economics*, *3*(2), 107-120.

Engelmann, D., & Strobel, M. (2012). Deconstruction and reconstruction of an anomaly. *Games and Economic Behavior*, *76*(2), 678-689.

Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. S*cience*, *326*(5952), 535-538.

Falk, A., Meier, S., & Zehnder, C. (2013). Do lab experiments misrepresent social preferences? The case of self-selected student samples. *Journal of the European Economic Association*, *11*(4), 839-852.

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

Galizzi, M. M., & Navarro-Martínez, D. (2018). On the external validity of social preference games: a systematic lab-field study. *Management Science*.

Guillén, P., & Veszteg, R. F. (2012). On "lab rats". *The Journal of Socio-Economics*, *41*(5), 714-720.

Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A focus theory of normative conduct: When norms do and do not affect behavior. *Personality and Social Psychology Bulletin*, *26*(8), 1002-1012.

Krawczyk, M. (2011). What brings your subjects to the lab? A field experiment. *Experimental Economics*, *14*(4), 482-489.

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary?. *Journal of the European Economic Association*, *11*(3), 495-524.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world?. *Journal of Economic perspectives*, *21*(2), 153-174.

Levitt, S. D., & List, J. A. (2008). Homo economicus evolves. *Science*, *319*(5865), 909-910.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, *115*(3), 482-493.

Marks, G., & Miller, N. (1987). Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological Bulletin*, *102*(1), 72.

Matthey, A., & Regner, T. (2013). On the independence of history: experience spill-overs between experiments. *Theory and Decision*, *75*(3), 403-419.

Mullen, B., Atkins, J. L., Champion, D. S., Edwards, C., Hardy, D., Story, J. E., & Vanderklok, M. (1985). The false consensus effect: A meta-analysis of 115 hypothesis tests. *Journal of Experimental Social Psychology*, *21*(3), 262-283.

Nikiforakis, N., & Slonim, R. (2015). Editors' preface: introducing JESA. *Journal of the Economic Science Association*, *1*(1), 1-7.

Reno, R. R., Cialdini, R. B., & Kallgren, C. A. (1993). The transsituational influence of social norms. *Journal of Personality and Social Psychology*, *64*(1), 104.

Rivis, A., & Sheeran, P. (2003). Descriptive norms as an additional predictor in the theory of planned behaviour: A meta-analysis. *Current Psychology*, *22*(3), 218-233.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*(3), 279-301.

Schmidt, R. J. (2019a). Do Injunctive or Descriptive Social Norms Elicited Using Coordination Games Better Explain Social Preferences?, University of Heidelberg: AWI Discussion Paper Series.

Schmidt, R. J. (2019b). Capitalizing on the (False) Consensus Effect: Two Tractable Methods to Elicit Private Information, University of Heidelberg: AWI Discussion Paper Series.

Slonim, R., Wang, C., Garbarino, E., & Merrett, D. (2013). Opting-in: Participation bias in economic experiments. *Journal of Economic Behavior & Organization*, *90*, 43-70.

Vanberg, C. (2019). A short note on the rationality of the false consensus effect, University of Heidelberg: AWI Discussion Paper Series No. 662.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75-98.