

Burks, Stephen V. et al.

Working Paper

Lab measures of other-regarding preferences can predict some related on-the-job behavior: Evidence from a large scale field experiment

CeDEx Discussion Paper Series, No. 2015-21

Provided in Cooperation with:

The University of Nottingham, Centre for Decision Research and Experimental Economics (CeDEx)

Suggested Citation: Burks, Stephen V. et al. (2015) : Lab measures of other-regarding preferences can predict some related on-the-job behavior: Evidence from a large scale field experiment, CeDEx Discussion Paper Series, No. 2015-21, The University of Nottingham, Centre for Decision Research and Experimental Economics (CeDEx), Nottingham

This Version is available at:

<https://hdl.handle.net/10419/129838>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

Discussion Paper No. 2015-21

Stephen V. Burks, Daniele
Nosenzo, Jon Anderson,
Matthew Bombyk, Derek
Ganzhorn, Lorenz Götte and
Aldo Rustichini

December 2015

**Lab Measures of Other-
Regarding Preferences
Can Predict Some Related
on-the-Job Behavior:
Evidence from a Large
Scale Field Experiment**

CeDEX Discussion Paper Series

ISSN 1749 - 3293



CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of public economics, individual choice under risk and uncertainty, strategic interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit <http://www.nottingham.ac.uk/cedex> for more information about the Centre or contact

Suzanne Robey
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0)115 95 14763
Fax: +44 (0) 115 95 14159
suzanne.robey@nottingham.ac.uk

The full list of CeDEX Discussion Papers is available at

<http://www.nottingham.ac.uk/cedex/publications/discussion-papers/index.aspx>

Lab Measures of Other-Regarding Preferences Can Predict Some Related on-the-Job Behavior: Evidence from a Large Scale Field Experiment

Stephen V. Burks^{1,2,3}, Daniele Nosenzo^{3,4}, Jon Anderson⁵, Matthew Bombyk⁶,
Derek Ganzhorn⁷, Lorenz Götte^{2,8}, Aldo Rustichini^{7,9}

December 11, 2015

Abstract: We measure a specific form of other-regarding behavior, costly cooperation with an anonymous other, among 645 subjects at a trucker training program in the Midwestern US. Using a sequential, strategic form of the Prisoners' Dilemma, we categorize subjects as: Free Rider, Conditional Cooperator, and Unconditional Cooperator. We observe the subjects on the job for up to two years afterwards in two naturally-occurring choices—whether to send two types of satellite uplink messages from their trucks. The first identifies trailers requiring repair, which benefits fellow drivers, while the second benefits the experimenters by giving them some follow-up data. Because of the specific nature of the technology and job conditions (which we carefully review) each of these otherwise situationally similar field decisions represents an act of costly cooperation towards an anonymous other in a setting that does not admit of repeated-game or reputation-effect explanations. We find that individual differences in costly cooperation observed in the lab do predict individual differences in the field in the first choice but not the second. We suggest that this difference is linked to the difference in the social identities of the beneficiaries (fellow drivers versus experimenters), and we conjecture that whether or not individual variations in pro-sociality generalize across settings (whether in the lab or field) may depend in part on this specific contextual factor: whether the social identities, and the relevant prescriptions (or norms) linked to them that are salient for subjects (as in Akerlof and Kranton (2000); (2010)), are appropriately parallel.

Keywords: experiments; generalizability; external validity; parallelism; social identity; other-regarding behavior; costly cooperation, social preferences; prisoners' dilemma; trucker; truckload.

JEL: B4; C9; D03

¹ Division of Social Science, University of Minnesota, Morris (600 East 4th Street, Morris, MN 56267, USA)

² Institute for the Study of Labor (IZA) (P.O. Box 7240 D-53072, Bonn, Germany)

³ Center for Decision Research and Experimental Economics (CeDEx), University of Nottingham (University Park, Nottingham, NG7 2RD, UK)

⁴ School of Economics, University of Nottingham (University Park, Nottingham, NG7 2RD, UK)

⁵ Division of Science and Mathematics, University of Minnesota, Morris (600 East 4th Street, Morris, MN 56267, USA)

⁶ Innovations for Poverty Action (101 Whitney Ave, New Haven, CT 06510, USA)

⁷ Northwestern University School of Law (375 East Chicago Avenue Chicago, Illinois 60611-3069, USA)

⁸ University of Bonn, Institute for Applied Microeconomics (Adenauerallee 24-42, 53113 Bonn, DE)

⁹ Department of Economics, University of Minnesota, Twin Cities (4-101 Hanson Hall, 1925 4th Street S Minneapolis, MN 55455, USA)

1. Introduction¹

The existence of other-regarding preferences is well established in the laboratory: many participants in laboratory game experiments often choose to reduce their own material payoffs in order to increase (or decrease) the material payoff of another participant, even when there are no reputation-based reasons to do so. This suggests that many individuals are not exclusively motivated by the pursuit of material self-interest when engaging in social and economic interactions with others outside the laboratory. This conclusion, however, requires a crucial assumption—that the behaviors observed in a laboratory experiment are informative about behaviors outside the laboratory, where markedly different conditions often apply. The evidence for the range of applicability of this *assumption of parallelism* (Smith 1982) between the lab and field environments is relatively limited in the case of experiments that identify and measure other-regarding preferences.

A key question when parallelism is at issue is whether and to what extent “the social situation of being in an experiment changes the actions under observation” (Bardsley, Cubitt et al., 2010 , p. 233). In the context of experiments that measure other-regarding preferences, the concern is that the prevalence of other-regarding behaviors observed in the lab might be the result of forces that are germane to the social situation of being in a laboratory experiment, but not to the social situations in the field to which we ultimately wish to generalize the insights from the experiment. Levitt and List (2007), for instance, argue that participants’ behavior in a lab experiment might be systematically influenced by a number of factors such as participants’ awareness that their actions are being scrutinized by the investigators, the context in which the decision situation is embedded, the stakes of the game, and the possibility for subjects who have other-regarding preferences to differentially self-select into the study. To the extent that these factors differ systematically between the lab and the field, the evidence of the importance of

¹ Acknowledgements. We received helpful comments from seminars audiences at the London Experimental Workshop, the University of Paris 1, University of Nottingham, and participants at the Economic Science Association session at the Allied Social Science Associations meeting in Denver (CO), the Wharton School of Business workshop on the Advances in Field Experiments, Carnegie Mellon University workshop on the Behavioral Decision-Making Research in Management 2010, the Asia/Pacific Economic Science Association regional meeting in Melbourne. The Truckers and Turnover Project acknowledges financial and in-kind support from the cooperating firm, and financial support from the MacArthur Foundation’s Research Network on the Nature and Origin of Norms and Preferences, the Sloan Foundation’s Industry Studies Program, the University of Nottingham, and the University of Minnesota, Morris. Nosenzo acknowledges support from the Leverhulme Trust (ECF/2010/0636). The views expressed are those of the authors, and do not necessarily reflect those of the supporting entities

other-regarding preferences in the lab may not be readily generalized to draw conclusions about the importance of such preferences in the field.²

In this paper, following the precept that “which kinds of behavior exhibit parallelism and which do not can only be determined empirically” (Smith 1982, p. 936), we combine laboratory experiments with data on the behavior of the *same subjects* in the field in order to examine the extent to which other-regarding behaviors observed in a laboratory social situation correlate with the behaviors that the same individuals exhibit in comparable, naturally-occurring, social situations outside the laboratory. A key feature of our study, which distinguishes it from previous studies in this area (e.g., Barr and Zeitlin (2010), Fehr and Leibbrandt (2011), Kosfeld and Rustagi (2015), Rustagi, Engel et al. (2010) - see Section 5 for a detailed discussion of this literature), is that we are able to correlate our subjects’ laboratory behavior with their behavior in two closely related, but distinct, field situations. One of the crucial differences between the two field situations is that in one situation subjects interact with individuals from the same “social category” in both the lab and field settings, whereas in the second situation subjects interact with individuals from different social categories in the lab versus in the field. We find that our measures of other-regarding preferences generated in lab experiments can be successfully extrapolated to the first, but not the second, field situation. This provides the first empirical evidence we know of in which variations in the projection of lab-measured pro-sociality to a field setting are associated with differences in the social identities of the beneficiaries in the field.

As we discuss in detail in Section 2, our subject pool is a sample of truck drivers employed by a large motor carrier headquartered in the Midwest of the United States. The drivers were first recruited to take part in a laboratory experiment during the middle of a two-week residential training program provided to inexperienced new employees, where they participated in a sequential version of the prisoner’s dilemma. We use choices in this setting to classify drivers in three main categories: Free-Riders, who always choose actions that maximize own-payoff, Conditional Cooperators, who deviate from own-payoff maximization by behaving cooperatively

² A number of experimental studies examine the extent to which each of these factors may pose a challenge to parallelism. For example, Anderson, Burks et al. (2013), Cleave, Nikiforakis et al. (2013), Falk, Meier et al. (2013), and Abeler and Nosenzo (2015) study whether allowing subjects to self-select into experiments distorts the measurement of other-regarding preferences. Similarly, numerous studies vary experimentally the stakes of the game, or the extent to which subjects’ actions in an experiment are observable (by the experimenter and/or other subjects). For reviews see, e.g., Levitt and List (2007) and Camerer (2015).

if others do, and Unconditional Cooperators, who behave cooperatively independently of others' behavior.

We combine this laboratory data with field data about drivers' on-the-job behavior in the months following the experiment. In particular, our data include information about drivers' satellite communications with the trucking company and, specifically, records of the transmission of "Trailer Maintenance Reports" (TMRs). These reports, which typically take a few minutes to send, are used by drivers to inform the firm of the existence of defective freight trailers in need of maintenance. While drivers are formally required to transmit TMRs, no enforcement mechanism (either direct or indirect) was in place at the trucking company during the period of the study. For this reason, and because there are significant opportunity costs associated with sending TMRs, drivers may be tempted to refrain from doing so whenever the mechanical defect is not serious enough to prevent them from completing a dispatch, or when the defect is detected at the end of a dispatch when the driver no longer needs the trailer. On the other hand, reporting an observed defect is a service to other drivers who may be later assigned the defective trailer and who would then incur a costly delay. Thus, transmitting a TMR is akin to an act of costly voluntary cooperation towards fellow drivers.

Notice that this field decision situation shares important similarities with the laboratory setting we use to measure drivers' cooperativeness. Notably, drivers work alone and seldom encounter other drivers, and neither the firm nor other drivers can in practice identify those drivers who fail to report an observed mechanical defect, so there are no repeated interactions, and drivers cannot generally use TMRs to develop a reputation for cooperation.³ Moreover, as in the experiment, decisions about sending a TMR are made in isolation, using satellite uplinks installed in a driver's truck tractor. Also importantly, by reporting an observed mechanical defect drivers are cooperating with fellow drivers who are anonymous, just as in the laboratory experiment.

Of course, there are also some important differences between the lab and field settings. First, drivers in the experiment knew that their cooperation decisions were being monitored and recorded by the experimenter, while data on drivers' field cooperativeness were collected under much less obtrusive conditions. Moreover, while the decision situation in the lab was framed in abstract terms as a two-person money-sending task, the field decision situation is embedded in

³ We discuss the job conditions in detail in Section 2.3, below.

the natural environment in which drivers operate. Both factors have been conjectured to be potential sources of differences between other-regarding behavior observed in the lab and in the field (Harrison and List 2004; Levitt and List 2007).

In Section 3 we show that, despite these differences, our laboratory measure of drivers' cooperativeness successfully predicts variations in the extent to which drivers transmit TMRs to the firm in the months following the experiment. Drivers who are classified as Free-Riders in the experiment send significantly fewer TMRs to the firm than Conditional or Unconditional Cooperators. We do not find differences in TMR transmissions between Conditional and Unconditional Cooperators. These results are robust to controls for potential differences in exposure to the risk of a defective trailer (measured as the number of trailer assignments given to each driver), socio-demographic, and operational differences across drivers. Thus, our findings provide support to the view that parallelism exists between other-regarding preferences measured in the lab and other-regarding preferences manifested by the same subjects under field conditions which exhibit some structural similarities to the lab setting.

In Section 4 we exploit an additional piece of information in our data to examine whether the parallelism between lab and field cooperativeness extends to a related but distinct field decision setting. All drivers employed at the firm who had participated in the laboratory experiment were sent a short electronic survey on behalf of the experimenters in the months following the experiment. The survey questions were administered weekly via satellite, and drivers could respond using the same satellite uplink units that they also used for the TMRs. During the informed consent for the initial data collection during their training drivers were informed these surveys would be sent to them later, on the job, and they were asked to respond to them. They knew that the surveys were not financially incentivized and that there was no formal requirement to respond, and, as with the choice of sending a TMR, drivers faced significant opportunity costs in responding. Thus, as with the TMRs, choosing to respond to the satellite survey is akin to an act of costly voluntary cooperation. However, whereas sending a TMR is primarily a service to other fellow drivers, drivers know that responding to the survey is beneficial to the experimenters. The question is whether our lab measure of drivers' cooperativeness can also predict cooperativeness in this field situation, where the main beneficiaries of drivers' cooperation belong to a different social category relative to the lab setting and the previous field situation. We find that our result on drivers' cooperativeness in the

lab does *not* generalize to this second field setting. Although the survey response rate of Free-Riders is slightly lower than that of Conditional and Unconditional Cooperators, the effect is statistically insignificant.

The upshot of our study is that insights gained from lab experiments can be successfully extrapolated to some naturally-occurring decision settings in the field, but they may not generalize to *every* field decision setting. On the one hand, our findings suggest that factors and conditions that are often idiosyncratic to the laboratory environment (e.g., the high level of scrutiny of subjects' actions, or the abstractness of laboratory decision situations) are not insurmountable barriers to the generalizability of laboratory findings. On the other hand, our results raise questions about the cross-situational stability of other-regarding preferences. As we discuss in Section 5, one potential approach to these questions may be found in recent theoretical and empirical work proposing that individuals' propensity to engage in costly other-regarding behavior may be associated with social norms that vary across decision contexts and conditions, and in particular, with the social identity of the beneficiaries of one's pro-sociality (Akerlof and Kranton 2000; Akerlof and Kranton 2005; Chen and Li 2009; Barr, Lane et al. 2015; Chang, Chen et al. 2015; Gintis and Helbing 2015).

2. Data

2.1. The Sample & the Data Collection Process

The data used in this study were collected from a sample of trainee truck drivers employed by a large motor carrier headquartered in the Midwest of the United States as part of a larger project in behavioral personnel economics (Burks, Carpenter et al. 2008). The firm operates in the "long distance truckload" segment of for-hire motor freight, which means its primary business involves hauling full trailer loads of freight from shippers directly to receivers, across medium to long distances (Burks, Belzer et al. 2010).⁴

The experimental data was collected at a driver training school operated by the trucking company, where subjects were recruited to the study. Subjects participated while in the middle of a two-week residential training program designed to earn them a commercial driver's license and

⁴ In 2007 there were 36,600 firms in the two relevant industry categories, General Commodity and Special Commodity Long Distance Truckload (U.S. Census Bureau 2012; U.S. Census Bureau 2012). These firms employed about one-third of persons in the occupational category Heavy and Tractor-trailer Truck Drivers, which contained almost 1.8 million individuals overall in 2008 (Bureau of Labor Statistics 2012).

to qualify them to operate a truck-tractor and semi-trailer combination freight vehicle (the “18-wheeler” or “tractor-trailer” of US popular culture (see Burks, Belzer et al. (2010)) for the firm doing the training. The drivers in our sample were first approached at the company training school at the beginning of a class day by one of the authors (Burks), who asked them to participate in an experimental study comprised of several decision tasks and questionnaires, including follow-up questionnaires.⁵ In total, 1,065 drivers participated in the experimental study. All subjects participated in a social dilemma game, described in detail in the next subsection, which will be the basis for our *lab measurement* of drivers’ cooperativeness.

In addition to the experimental data, the trucking company provided rich follow-up data about drivers’ on-the-job performance in the months following the laboratory experiment. This included data on drivers’ satellite communications with the trucking company via a Qualcomm satellite uplink installed in each truck tractor. Work assignments and related information are sent to the driver, and the driver sends information back to the firm, through this unit, which is mounted in the cab and has looks like a small laptop. (Such a unit is displayed in Figure 1; see Llaneras, Singer et al. (2005, page 23 and following) for operational details.)

⁵ The experiment was set up as two two-hour-long blocks that subjects spent doing tasks with the researchers, either on computers or with paper and pencil, with a short break in between. At the beginning of each two-hour-long block subjects received a fixed payment of \$10 for their participation, and could earn additional money in the course of the experiment depending on their performance. Individual earnings from the whole process ranged from \$21 to \$168, with an average of \$53. Data was collected on 23 Saturdays over the course of nine months, between December 2005 and August 2006, two sessions per Saturday, with 20 to 40 subjects per session. The Informed Consent process offering study participation was run with the entire class of students who were at the midpoint of training on each Saturday. The participation rate was very high: 91% of subjects approached chose to join the study. See Burks, Carpenter et al. (2008) for more details on the other tasks used in the experiment and on the experimental protocol.



Figure 1: Qualcomm satellite uplink. (Source: Llaneras, Singer et al. (2005))

The flow of satellite communications messages is very large, and not part of the normal human resource or operational data provided by the firm. For the study of cooperation we requested that the trucking company record the incidence of transmission of “Trailer Maintenance Reports” messages, which are used by drivers to alert the company about freight trailers in need of maintenance or mechanical repair. As explained in detail in Section 2.3, we will use these satellite messages to derive our *first field measurement* of drivers’ cooperativeness. Moreover, the firm also agreed to administer a weekly follow-up survey to the drivers on behalf of the experimenters via the same satellite link used for the transmission of Trailer Maintenance Reports. As we discuss in Section 4, we will use responses to the survey questions as our *second field measurement* of drivers’ cooperativeness.

We requested the firm to record satellite messages over a follow-up period from December 2005 through late 2008. However, internal delays in implementing this request mean we actually have records running from June 18th, 2006 to October 26th, 2008. Because of employment separations before collection of the satellite message data began, only 655 of the 958 drivers who completed training⁶ are observed in the satellite uplink data, as the others exited before June 18, 2006. We have complete data on all the variables of interest (described in the following subsections) for all but 10 of these drivers: thus, our sample consists of 645 drivers in total.

⁶ 107 of the initial 1065 drivers did not complete training and so we have no data on their on-the-job behavior.

2.2. Cooperativeness in the Lab

Our lab measurement of drivers' cooperativeness is based on decisions made in a sequential version of the prisoner's dilemma (PD).⁷ At the outset of the game two players, labeled Person 1 and Person 2, are each allocated \$5. Person 1 moves first and chooses an amount $s_1 \in \{\$0, \$5\}$ to send to Person 2. Person 2 learns Person 1's decision and then chooses an amount $s_2 \in \{\$0, \$1, \$2, \$3, \$4, \$5\}$ to send back to Person 1. Any amount sent by either player is doubled by the experimenter. The game ends after Person 2's decision, and payoffs are computed as:

$$\pi_i = \$5 - s_i + 2 * s_j$$

for $i, j \in \{1, 2\}$, and $i \neq j$.

In the experiment subjects played this game once, and subjects made decisions in both roles knowing that the final assignment to roles would be determined randomly at the end of the experiment. On a first screen subjects were asked to make a decision in the role of Person 1 and on a second screen a decision in the role of Person 2. Person 2's decisions were elicited using the strategy method: subjects had to specify the amount they intended to transfer to Person 1 both for the case where Person 1 had sent \$0 and for the case where Person 1 had sent \$5. Thus, subjects were asked to make three decisions in total: one in the role of Person 1 and two in the role of Person 2.⁸ Once all decisions had been made, subjects were anonymously and randomly matched with another participant, were randomly assigned a role, and were shown their payoffs according to the decisions they had made in that role. On average, the 645 drivers in our final sample earned \$8.40 from the game, with a minimum of \$0 and a maximum of \$16.

As in Anderson, Burks et al. (2013), our measure of cooperativeness uses the choices made in the role of Person 2.⁹ The use of the strategy method means we observe two decisions from each participant in the role of Person 2: one for the case where Person 1 behaves uncooperatively and sends \$0, and one for the case where Person 1 is cooperative and sends \$5. This allows us to classify subjects into three well-defined types depending on how cooperatively they respond to

⁷ The computerized tasks were programmed and implemented with the software z-Tree (Fischbacher 2007).

⁸ Before making their decisions subjects were also asked to predict others' behavior in each possible decision situation, and received an additional \$1 per correct answer (see Burks, Carpenter et al. (2008)). Thus, the maximum possible earnings from the PD game task are \$18.

⁹ Compared to decisions in the role of Person 2, it is more difficult to infer cooperativeness from Person 1's choices since these may also reflect considerations about the profitability of cooperating, consensus effects, etc. (see, e.g., Gaechter, Nosenzo et al. (2012)).

Person 1's actions.¹⁰ Subjects who behave uncooperatively and choose the own payoff-maximizing action (return \$0) irrespective of the amount sent by Person 1 are classified as "Free Riders". Subjects who choose the most cooperative action available (send back the maximum, or \$5) if Person 1 sends \$5, but behave uncooperatively and send back \$0 upon receiving \$0 from Person 1 are classified as "Conditional Cooperators". Finally, subjects who always choose the most cooperative action irrespective of what the first-mover sends to them are classified as "Unconditional Cooperators". This approach allows us to classify 65% of the subjects. To assign the remaining participants to a type category we calculate, for each subject, the Euclidean distance between his or her decisions and the decisions that each of the three types just defined would make, and then assign the subject to the least distant type category.¹¹ We can thus classify all but 12 subjects: these participants are classified separately as "Others". Figure 2 shows the result of this classification for the 645 drivers in the sample. About 24% of the drivers are classified as Free Riders, 47% as Conditional Cooperators, 27% as Unconditional Cooperators, and 2% as Others. In the rest of the paper, we ignore the latter category, creating a final sample of 633.

¹⁰ Note that in order to classify subjects according to their cooperativeness, it is important to observe Person 2's behavior in both subgames. Observing behavior in only one subgame is not sufficient. For example, observing that Person 2 sends \$0 when Person 1 sends \$0 does not reveal whether she is a "conditional cooperator" who defects when the first-mover defects, or whether she is exclusively motivated by own payoff maximization. The use of the strategy method solves this problem by allowing us to observe how Person 2 responds to both possible decisions of Person 1.

¹¹ Formally, the Euclidean distance between a subject's choices and the Free Rider type is computed as $D_{FR} = \sqrt{(x_{\$0} - 0)^2 + (x_{\$5} - 0)^2}$, where $x_{\$0}$ and $x_{\$5}$ are the amounts returned by the subject when the first-mover sends \$0 and \$5, respectively. Analogously, we compute the distance from the Conditional Cooperator and Unconditional Cooperator types as $D_{CC} = \sqrt{(x_{\$0} - 0)^2 + (x_{\$5} - 5)^2}$ and $D_{UC} = \sqrt{(x_{\$0} - 5)^2 + (x_{\$5} - 5)^2}$, respectively.

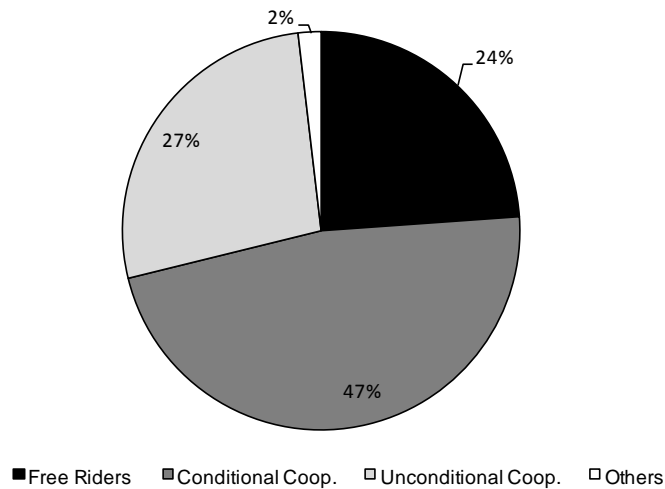


Figure 2: Cooperativeness in the Lab – Classification of Types

2.3. Cooperativeness in the Field

As our first measurement of drivers’ cooperative behavior in the field we use the on-the-job activity of transmitting Trailer Maintenance Reports (TMRs) to the trucking company. These reports are used by drivers to inform the firm that a freight trailer needs maintenance or requires repair. To understand why this is a relevant measure of cooperation we need to give some background information about the work situation. As is typical with firms operating in the truckload segment of the US motor freight industry (Section 2.1 and Burks, Belzer et al. (2010)), the cooperating firm operates about three times as many semi-trailers as it does truck tractors, so that most of the loading and unloading of trailers can be done by the personnel of customer firms that are shipping or receiving freight while the driver and a tractor are not present, but have moved on to another trailer and load. Thus, at any given time thousands of the firm’s trailers are parked at shipper or receiver locations, as well as at trucking-company-controlled parking lots or terminals around the continental US. Drivers typically take an empty trailer to a shipper’s location, drop it and hook up to a loaded trailer, then drive to the receiver’s location, drop the loaded trailer and hook to an empty, and repeat.

Trailers are quite durable, but periodically develop problems that require repair, such as flat tires, broken spring leaves, brake problems, or burned out running lights. The firm’s routine maintenance processes catch most such problems. For instance, when the driver fuels at a

company-operated terminal that provides truck services, the trailer is given a periodic maintenance inspection by a mechanic, who can also do spot repairs for non-major problems and can schedule major repairs for trailers that can be taken out of service.

Formally, a driver must file a TMR whenever a trailer is observed to be in a condition requiring repair, and all drivers learn this rule in their training. However, in practice TMRs are statistically rare: among our subjects the mean number of work assignments (our proxy for exposure to different trailers) is 198, over a mean number of weeks in which subjects are observed of 55.6, while the mean number of TMRs sent is only 2.5. This is for two reasons. The first is that the routine maintenance procedures are relatively effective, so drivers don't confront unexpected maintenance problems very often. The second is that most drivers typically only report a trailer maintenance problem when it directly affects their ability to complete a work assignment, because it is on average costly to a driver to send satellite message of any kind, including trailer maintenance reports. This is due to the interaction between the compensation system in place at the firm and the process and consequences of sending a satellite message to the firm.

All driver pay (except the weekly contribution to the health insurance premium by the firm) is in the form of piece rates, either cents per mile driven, or lump sums that are sometimes paid for specific extra activities that are required.¹² Thus, drivers face high-powered incentives to keep the truck in motion.¹³ Second, the cooperating firm serves a varied set of shipper and receiver locations spread all the way across the 48 states of the continental US. Thus, for each load assignment, the driver must select and travel over routes from a few hundred to several thousands of miles long while subject to a large array of potentially conflicting and partly stochastic constraints.¹⁴ The sequential nature of trips means that small errors or delays (especially early in a dispatch sequence) can cause larger later hold-ups, as delays with the current load assignment may cause drivers to miss their subsequent assignment, and requiring a new load assignment frequently means waiting. Thus, the driver's effective hourly wage depends on repeatedly solving this ongoing two-to-three day planning puzzle well, and any unexpected

¹² An example is a flat fee for weighing a trailer loaded by the shipper to make sure both the gross weight and the weight distribution across the rig's axles fall within regulatory limits.

¹³ Hence the common saying among drivers: "*you ain't earning if the wheels ain't turning*".

¹⁴ For example, these include issues such as shipper and consignee day and time requirements at the endpoints, restrictions on toll road use, congestion and weather conditions which can change over the course of a trip (the latter especially important during winter), and daily and weekly limitations on driving and on total work hours from the Federal Hours of Service regulations.

additional delay represents an opportunity cost that can reduce earnings.¹⁵

Sending any satellite message is potentially costly because the satellite uplink used to send messages has an interlock that prevents its use when the tractor is moving. Sometimes the driver is already delayed for some other reason and then sending a satellite message adds only the subjective cost of the effort involved (which is likely to be small), but on average this action is costly to the driver because of the necessity of spending extra time stopped on some occasions.

Every driver will send a TMR under one specific circumstance: that the current condition of the trailer to which the driver is assigned prevents its safe use to complete the current work task. This is bad news because reporting that a trailer is not currently serviceable may result in a potentially much larger delay and much larger opportunity cost, as it can disrupt the driver's work schedule. For example, if a driver observes a mechanical defect on an empty trailer he or she is assigned to move to a shipper location (where the next load is ready), and sends a TMR to report it, the driver will typically have a different trailer assigned to him or her. This may take time if dispatchers are busy and slow to respond with a new trailer assignment, and new empty trailers may not always be immediately available nearby.

But on other occasions, when the defect does not prevent the completion of an assignment, or when the defect is spotted at the end of an assignment and the driver could otherwise keep his or her tractor moving, a purely self-interested driver would be tempted to violate the rule requiring a TMR in the absence of some enforcement mechanism. And, signally, as a managerial decision, the firm did not attempt to track who followed this rule, or to enforce it, during the period of the study.

This was because, given the technology available at the time and the pattern of actual trailer use, the firm could not be sure who is responsible for not having reported a defect. Trailers are often hooked and moved by customer personnel when parked at customer locations, and the firm is often not entirely sure which of the firm's own drivers had last used any given trailer. Plus, many defects only show up over time, and could have been legitimately missed by a prior driver.¹⁶ In addition, the firm's operational organization separates driver supervision from trailer and load assignment. Driver supervision is carried out by a driver leader located at or near the

¹⁵ In fact, Burks et al. (2009) find that for this driver population the biggest single predictor of making it through the year of service after training that cancels the new driver's several thousand dollar training debt is basic cognitive skills, with a measure of planning ability giving the strongest prediction of three cognitive skill measures used.

¹⁶ For instance, the most common problem, a flat tire, could always have developed via a slow leak, and not been apparent at the time a trailer was dropped.

driver's home base, who is available 8 a.m. to 5 p.m. weekdays for coaching, disciplinary matters, resolving payroll issues, and the like. Driver load assignments are made by a team of hundreds of dispatchers operating from a central location with shifts around the clock, and to whom any given driver is essentially anonymous.

Thus, the firm's managers have historically chosen not to try to track whether TMRs are sent according to the formal rule by any specific drivers. This fact was fully known to all drivers, since there are many factors on which performance is tracked by driver supervisors, and this is specifically not included. Moreover, drivers who do not send a TMR to report an observed defect are also unlikely to be identified by other drivers who will subsequently hook up to the defective trailer, as drivers typically do not know who used a trailer before them.

However, since every driver understands that being assigned a defective trailer is bad news, every driver also understands that reporting an observed defect that could hold up someone else's later dispatch would be a service to others. Thus, each time a driver encountered a defective trailer on the job during the data collection period that did not require immediate repair for the driver's own purposes, the driver was effectively playing a one-shot game of costly cooperation with anonymous other drivers. As with the choice of sending back \$5 in the social dilemma game, reporting a defective trailer under these circumstances fundamentally represents a form of voluntary, non-strategic, and costly cooperation towards fellow drivers.

The trucking company recorded all TMRs sent by drivers between mid-June 2006 and October 2008. Thus, for each driver employed by the trucking company within this time frame, we have an exact count of the number of TMRs he or she transmitted to the firm. As mentioned, we also can observe the number of messages assigning the hooking or dropping of a trailer sent to each driver during this time frame. We use this measurement as a proxy for the level of exposure to potentially unserviceable trailers. So while we cannot tell, any more than the firm's managers can, which specific TMRs represent cooperative actions, what we can do is look for statistical differences in TMR sending behavior, adjusting for a measure of on-the-job exposure to potentially TMR-relevant situations.

2.4. Socio-demographic & Work-related Controls

We administered a questionnaire during the economic experiments at subject intake into the study to obtain information on the drivers' socio-demographic characteristics. In the analysis

of Sections 3.2 and 4 we will use this information to account for the existence of observable differences across the drivers in our sample. Table 1 presents a summary of the socio-demographic characteristics for the 633 drivers in our final sample.

Table 1: Summary of Socio-Demographic Controls (N = 633)

Age, <i>mean (min. – max.)</i>	37.2 (21 - 68)
Female (%)	9%
Non – White or Hispanic (%)	17%
Own Marital Status (%)	49%
Married or in marriage-type relation	19%
Separated/Divorced/Widowed	32%
Single	
Education Level (%)	
High School or lower	44%
Some College (no degree)	33%
Junior/Technical College	15%
College (4 yr.) or higher	08%
Household Income Level (%)	
\$0-\$10,000	39%
\$10,000-\$20,000	18%
\$20,000-\$30,000	14%
\$30,000-\$40,000	10%
\$40,000 or more	20%

Note: the variable “Household Income Level” was derived from the question “*Not counting your earnings, which range best fits the annual income you and your household have from other sources?*”

The data on drivers’ on-the-job activities provided by the trucking company also provides important work-related characteristics on a week-by-week basis. These data include operational measures such the location of the driver’s home terminal, the type of work in which the driver is primarily engaged in a given week, as well as the driver’s date of separation, if applicable. All of these may affect the probability that a driver encounters a trailer that either has or that develops a serious defect that was not reported already. Table 2 summarizes these measurements averaged across all drivers in the sample and across all weeks of employment.

Table 2: Summary of Operational Characteristics (N = 633)

Number of TMRs sent, <i>mean</i>	2.5
Number of Work Assignments, <i>mean</i>	198
Tenure (in weeks), <i>mean</i>	55.6
Terminal of Operation (%)	
Small City (site of training school)	0.37
Big City	0.28
Medium City #1	0.14
Medium City #2	0.22
Medium City #3	0.18
Large company located in the East U.S.	0.15
Large company located in the West U.S.	0.14
Other terminal	0.17
Type of Work (%)	
System	0.54
Dedicated	0.53
Team	0.11
Intermodal	0.08
Other	0.12

Note: drivers sometimes change the terminal out of which they operate as well as the type of work they are assigned to. The percentages shown for the “Terminal of Operation” and “Type of Work” variables represent the proportion of drivers who have *at least* one observation in a given category and, thus, do not necessarily add up to 100%. All cities are in the Midwestern United States.

The terminal of operation is the home base to which drivers return, as often as every day, but more commonly once every two or three weeks, and is associated with specific runs that originate or terminate there. Two specific terminals, the two “large company locations”, are important because the trailers there are owned and maintained primarily by a customer, so TMRs to the cooperating firm are very rare at those locations. The two most common types of work are “System,” which is the archetypal long distance truckload firm driving job described earlier in this section, randomly moving from one customer location to another over the continental US, and “Dedicated,” which is work in which the driver is assigned exclusively to the service of a particular large customer.¹⁷

¹⁷ “Intermodal” is shorter-haul work taking specialized trailers or containers between customer locations and rail yards, where the freight is placed on railcars for the long distance part of the freight movement. “Team” work is when two drivers operate the same truck, one driving while the other rests.

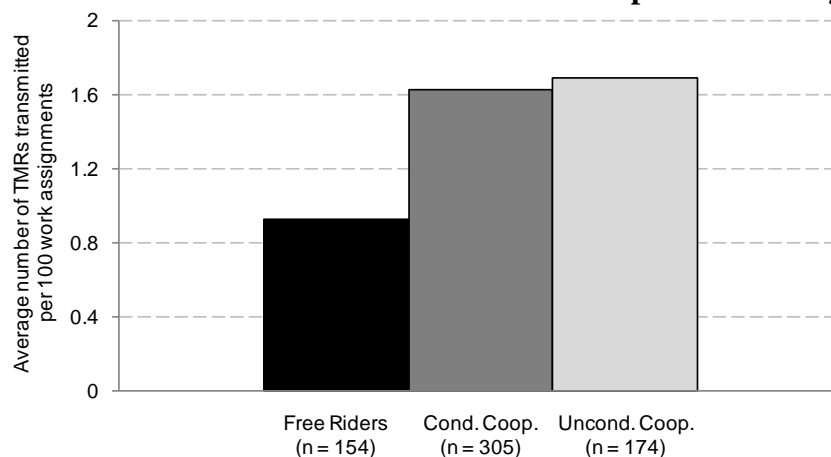
3. Results - Cooperativeness in the Lab Predicts Cooperativeness in the Field as Measured by TMR Sending Rates

In our sample of 633 drivers we observe a total of 1,608 TMRs sent to the firm between June 18th, 2006 and October 26th, 2008. About 50% of the drivers never transmitted a TMR to the firm within the period under study, about 25% transmitted between 1 and 2 TMRs, while the remaining 25% transmitted 3 or more TMRs. The median driver transmitted 1 maintenance report to the firm in the period under study, while the average driver transmitted approximately 2.5 reports. In the remainder of this section we examine the link between TMR transmissions and drivers' cooperativeness as measured in the lab experiment.

3.1. Descriptive Analysis

As a first step, we focus on the aggregate number of TMRs transmitted by drivers to the firm during the period under study. To control for potential differences in exposure to unserviceable trailers, we calculate for each driver the *rate of TMR transmissions per work assignment*. This is the ratio between the total number of TMRs that a driver transmitted to the company in the period under study and the total number of freight trailers that he or she was assigned to hook or drop during the same time frame. Figure 3 shows the average number of TMR transmissions per one hundred work assignments across the three main cooperativeness types classified in the lab experiment.

Figure 3: Rate of TMR Transmissions across Cooperativeness Types



The average rate of TMR transmissions is about 0.92 for drivers who are classified as Free Riders in the experiment. The rate is almost twice as high for drivers who are classified as

Conditional Cooperators (1.62) and Unconditional Cooperators (1.69). Treating each driver as an independent observation and using two-sided Wilcoxon rank sum tests, we reject the hypothesis that the rate of TMR transmissions of Free Riders is the same as that of Conditional Cooperators ($z = 2.85$; $p = 0.004$), or Unconditional Cooperators ($z = 2.24$; $p = 0.025$). We do not find statistically significant differences in the rate of TMR transmissions between Conditional and Unconditional Cooperators ($z = 0.43$; $p = 0.663$).

In summary, this descriptive statistical analysis shows that the drivers who are most cooperative in the laboratory experiment are also those who show the highest degree of cooperativeness on the job by transmitting TMRs to the trucking company at qualitatively higher rate per work assignment. Our next step is to check the robustness of this result by using regression analysis that allows us to control for observable differences across drivers.

3.2. Regression Analysis

Our regression analysis is based on a panel data set composed of drivers' transmissions of TMRs in all the weeks between June 18th, 2006 and October 26th, 2008. Our panel data set is unbalanced as drivers were not always on duty during the time frame under study, and not all drivers were employed by the trucking company for the whole period.¹⁸ Overall, we have data on 24,926 driver-week observations for the period under consideration.

In the regressions we use as dependent variable the total number of TMRs transmitted by a driver in a given week of duty. Because this variable is censored at zero by construction, we utilize a Tobit estimator. We present three regression models. To control for potential differences in exposure to unserviceable trailers, in all models we include as a regressor the variable "Num. of Work Assignments", which measures the total number of trailers a driver hooked or dropped in each week of duty. In Model I we use as additional regressors two dummy variables for the drivers classified as Conditional Cooperators and those classified as Unconditional Cooperators (note that the reference category are drivers classified as Free Riders). In Model II we add the set of controls for socio-demographic characteristics listed in Table 1, and in Model III we add the

¹⁸ As mentioned in Section 2.1, some drivers separate from employment before the firm began giving us satellite message data. Drivers occasionally disappear from the data for a week when they have time off at home. Since we know the level of exposure to the chance of a defective trailer in each week of observation we are confident that the differential presence or absence of specific drivers in specific weeks of the data is not driving our results. We also looked for potential selection effects through differential exits by testing for any effect of tenure on TMR sending rates using flexible specifications, and find none (results available from the authors upon request).

set of work-related controls listed in Table 2. To account for the panel structure of the data, we add driver-level random effects to all regression models. Table 3 reports the regression results.^{19,20}

[TABLE 3 HERE]

The regression analysis confirms that drivers' propensity to transmit TMRs is positively related to their lab cooperativeness. Starting with Model I, the coefficients on the Conditional Cooperator and Unconditional Cooperator variables are both positive and significantly different from zero at the 1% and 5% level, respectively. The estimates show that these two types of cooperators transmit to the firm about 0.5 TMRs more than Free Riders, the omitted category. The magnitude of the effect is similar across Conditional and Unconditional Cooperators ($\chi^2=0.71$, $p = 0.398$).²¹ While the magnitude of the Conditional and Unconditional Cooperators variables diminishes slightly when we add socio-demographic and operational controls in Models II and III, in both models the effects remain positive and statistically significant at the 1% or 5% level. If we consider that the mean number of TMRs sent by our subjects is 2.5, the regression results suggest that Conditional and Unconditional Cooperators are cooperative about 20% more often than Free Riders.

As expected, in all models we observe a positive and significant relation between TMR transmissions and number of work assignments, which we use to control potential different exposure to unserviceable trailers. Among the controls for socio-demographic characteristics included in the regressions, we identify two effects that are robust across Models II and III: the propensity to transmit TMRs is higher among older drivers, and among drivers who spent more

¹⁹ We also ran negative binomial random-effects regressions using the count of TMRs transmitted by a driver in a given week of duty as the dependent variable and "Num. of Work Assignments" as the exposure variable. These alternative regressions are reported in Table A3.1 in Appendix 3 and yield very similar results to the Tobit specifications.

²⁰ Note that in Model I the reference subject type is: Free Rider. In Model II the reference subject type is: Free Rider, Male, Married, White (Non-Hispanic), Education level High School or lower, Income category \$0-\$10k. In Model III the reference subject type is Free Rider, Male, Married, White (Non-Hispanic), Education level High School or lower, Income category \$0-\$10k, operating out of the Small City terminal (site of the training school operated by the trucking company), system driver.

²¹ It is not entirely surprising that we cannot identify systematic differences in the field behavior of Conditional and Unconditional cooperators: while in the lab we can distinguish between these two types since we control their beliefs about the cooperativeness of others, in the field we cannot observe subjects' beliefs about other divers' cooperativeness.

years in education. The result that cooperativeness is positively related to age is in line with several other studies using non-student samples (e.g., (List 2004)).

Model III introduces controls for operational differences across drivers. The regression shows that there are some geographical location and work type effects: in particular, as noted in Section 2.4, drivers operating out of terminals controlled by a large customer corporation located in the East and West of the United States send significantly fewer TMRs than drivers domiciled in the same city as the training school operated by the trucking company. These results show that, while socio-demographic characteristics and work variations affect TMR sending in intuitive ways, our main results are robust to the inclusion of these additional controls.

4. When Generalizability Fails – the Case of Cooperativeness towards the Experimenter

The results presented in the previous section show that drivers who behaved most cooperatively in the lab experiment were also those who behaved most cooperatively on the job. Thus, the qualitative results on individuals' other-regarding preferences that we obtained in a laboratory environment do generalize to the field, and do so for a large number of subjects who are observed after the lab measurement, over a significant length of time (mean of 55.6 weeks), and while performing work over a broad variety of geographic locations. In this section, we examine the robustness of the lab-field link established in the previous section, by examining whether our lab measurement of drivers' cooperativeness can also predict their cooperative behavior in a related, but distinct field situation.

To address this, we use an additional piece of information available in our data. As part of the original study design, the trucking company agreed to administer a follow-up survey to all drivers who participated in the laboratory experiment and who were currently employed at the firm. The survey consisted of two questions, and it was administered weekly to drivers via the same satellite uplink units that they used to transmit the maintenance reports.²² Because both the survey and the TMR use a preformatted message template, the time and effort involved in sending each type of message are similar.²³

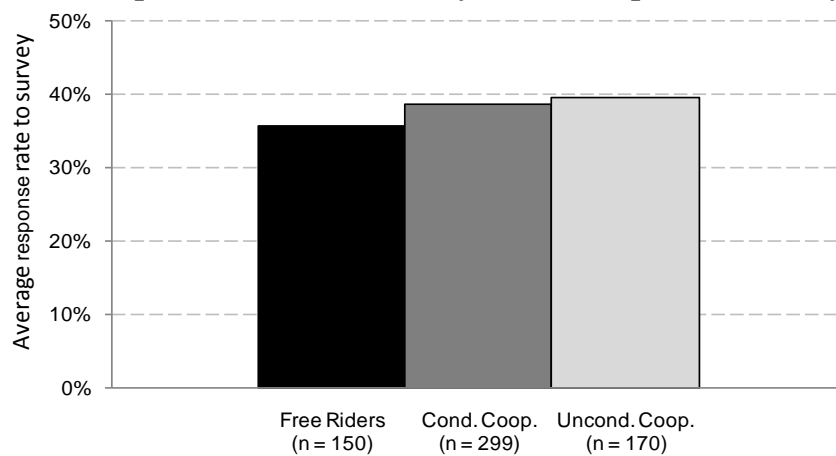
²² The first survey question was about job satisfaction and was elicited with a Likert scale, and the second about the pay miles the driver expected to make the following week. The latter is analyzed in Hoffman and Burks (2014).

²³ As explained earlier, these satellite units are installed on drivers' tractors. Thus, all drivers who were on duty (i.e. were assigned to a tractor) in a given week of employment received the two-question survey. If a driver was not on duty in a given week he or she did not receive the survey. In the following data analysis we only use weeks in which drivers were on duty.

During the Informed Consent process all the methods that would be used for data collection were explained to potential subjects, including the weekly satellite survey, so drivers were aware that the survey was sent on behalf of the experimenters that conducted the initial study at the training school, and they knew that there was no material consequence associated with responding (or not responding) to the questions. On the other hand, responding to the survey was on average costly in the same manner as sending TMRs (see details presented in Section 2.3, above). Thus, as in the case of TMRs, on average using the satellite units to respond to the survey is an act of costly and voluntary cooperation on the part of the drivers. However, in contradistinction to the TMR case, the main beneficiaries of the drivers' cooperation are not fellow drivers, but the experimenters. Thus, by analyzing drivers' response rates to the follow-up surveys we can explore whether our results on drivers' cooperativeness in the lab generalize to the field even when an important feature of the field decision context—the social identity of the main beneficiary of the acts of cooperation—differs from the lab decision context.

We have data on survey responses from 619 drivers over 85 weeks within the period running from June 18th, 2006 to May 18th, 2008.²⁴ The average response rate in the sample is 38%, with about 4% of the drivers responding to all the surveys that they were sent and about 12% of the drivers responding to none. Figure 4 shows the average response rate of the three main cooperativeness types classified in the laboratory experiment.

Figure 4: Response Rate to the Survey across Cooperativeness Types



²⁴ We do not have data on survey responses for 16 weeks during this period due to a lapse within the firm in transmitting them. Also, we do not have any data for 14 drivers who were included in our final sample used for the TMR analysis, as they were employed only during missing weeks, or were inadvertently left out of the transmission process by the firm.

The average response rate of drivers who were classified as Free Riders in the experiment is about 36%. The response rate of drivers classified as Conditional and Unconditional Cooperators is somewhat higher than this, averaging 39% and 40% respectively. However, these differences are small, especially if compared with the much larger variation across cooperativeness types in TMR transmissions (see Figure 3). In fact, treating each driver as an independent unit of observation and using two-sided Wilcoxon rank sum tests, we cannot find any statistically significant difference between Free Riders' response rate and the response rate of Conditional Cooperators ($z = 1.16$, $p = 0.246$) or Unconditional Cooperators ($z = 1.17$, $p = 0.244$). Conditional Cooperators' response rate is also not significantly different from that of Unconditional Cooperators ($z = 0.26$, $p = 0.792$).

As with the analysis of TMR message sending, we further examine the differences in satellite survey responses across cooperativeness types using regression analysis, to establish whether accounting for any observable differences might change this result. Our dependent variable assumes value 1 if a driver responded to the survey he or she was sent in a given week, and value 0 if the driver did not respond. Similarly to the regressions reported in Table 3, we start with a model that only includes as regressors two dummy variables for drivers classified as Conditional and Unconditional Cooperators. We then augment this model by adding socio-demographic controls (Model II) and operational controls (Model III). Table 4 reports the results of logit regressions with random effects at the driver level.²⁵ Results are displayed as factor changes in the odds of responding to the survey.²⁶

[TABLE 4 HERE]

In all models, the point estimates suggest that Conditional and Unconditional Cooperators have slightly higher odds of responding to the survey than Free Riders. However, in none of the models are the estimated values statistically different from zero (in the models containing the control variables the lowest p-value is 0.585). Overall then, we find no statistical evidence that

²⁵ In Model I the reference subject type is: Free Rider. In Model II the reference subject type is: Free Rider, Male, Married, White (Non-Hispanic), Education level High School or lower, Income category \$0-\$10k. In Model III the reference subject type is Free Rider, Male, Married, White (Non-Hispanic), Education level High School or lower, Income category \$0-\$10k, operating out of the Small City terminal (site of the training school operated by the trucking company), system drivers.

²⁶ These have the standard interpretation as multipliers on the odds of a response from a driver in the reference category, so values greater than 1 imply an increase, and values less than 1 imply a decrease, in the odds of responding.

those who behaved more cooperatively towards fellow drivers in the lab experiment also behave more cooperatively towards the experimenters, in a response setting that is similar to the response setting for TMR sending.

5. Discussion & Conclusions

In this paper we examine the generalizability of laboratory findings about other-regarding preferences to field settings by combining data from the laboratory and data on behavior in the field from a sample of truck drivers employed by a large U.S. motor carrier. We measure drivers' cooperativeness in a laboratory experiment using a sequential version of the prisoner's dilemma. We then use this measurement to predict drivers' cooperativeness in two naturally-occurring decision situations in the field.

We report two main findings. First, we show that measurement of the individual differences in other-regarding preferences generated in lab experiments can be successfully extrapolated outside the lab under some circumstances: drivers who acted more cooperatively towards fellow drivers in the experiment were also more cooperative towards fellow drivers in the field. This positive correlation between lab and field behavior emerges despite the existence of clear differences across lab and field settings in factors, such as the level of scrutiny and the context that embeds the decision situation, which have been conjectured to hinder the generalizability of laboratory findings to non-lab settings (e.g., Levitt and List (2007)).

However, and this is our second and most distinctive main finding, we also show that results obtained in a lab experiment measuring social preferences may not generalize to *every* decision situation: in a similar choice setting, responding on the satellite unit, the individual differences in the lab behavior of drivers did not predict individual differences in responses that benefitted counterparts from a different social identity or category—the experimenters. We are not the first to report either parallels or divergences between other-regarding behaviors in the lab versus in the field. But we believe we are the first to show both in a context in which a reasonable conjecture about the source of the variation in choices across essentially identically structured field decisions made by the same subjects is changing the social identity of the beneficiary of an act of costly cooperation - from an anonymous member of one's own social category to an anonymous member of an out-group.

Our results add to the findings of a rapidly growing literature that, using a combination of lab and field data from the same subjects, documents positive correlations between other-regarding behavior in the lab and in comparable, naturally-occurring social situations (Karlan 2005; Benz and Meier 2008; Baran, Sapienza et al. 2010; Barr and Zeitlin 2010; Carpenter and Myers 2010; Fehr and Leibbrandt 2011; Lamba and Mace 2011; Barr, Packard et al. 2014; Carlsson, Johansson-Stenman et al. 2014; Galizzi and Navarro-Martinez 2015).²⁷ Karlan (2005), for instance, uses second-mover behavior in a trust game experiment to measure trustworthiness among participants in a microcredit program in Peru. Using savings and loan outcome data from the same individuals, he finds lower defaults and higher voluntary savings among the individuals who were more trustworthy in the experiment.²⁸ Similarly, Fehr and Leibbrandt (2011) show that Brazilian shrimp fishermen who behave more cooperatively in a public-goods game experiment also use fishing instruments that are less harmful for the fishing grounds.²⁹ Benz and Myer (2008), Barr and Zeitlin (2010), and Carpenter and Myers (2010) find that dictator game giving correlates with other-regarding behaviors in various naturally-occurring social situations.³⁰ Barr, Packard et al. (2014) find that individual behavior in a public goods game correlates with participation in school accountability institutions and national elections in Albania. Carlsson, Johansson-Stenman et al. (2014) measure individual contribution behavior at four different points in time over several years, both in a lab experiment and in various field situations. They find strong correlations between contributions across the four situations.

However, there are some variations in the findings from this literature. For example, using a sample of subsistence farmers in Sierra Leone, Voors, Turley et al. (2011) and Voors, Turley et al. (2012) report no correlation between behavior in a standard public goods game experiment and contributions to a community project fund for their village. Galizzi and Navarro-Martinez

²⁷ See also Camerer (2015), Galizzi and Navarro-Martinez (2015) and Bowles and Polania-Reyes (2012) for exhaustive reviews.

²⁸ Baran, Sapienza et al. (2010) also use a trust game experiment with MBA students at the University of Chicago. They find that students' behavior in the game predicts their donations to the university at the end of the program.

²⁹ Lamba and Mace (2011) find a positive correlation between public goods game behavior and the sharing of a locally valued resource (salt) with people from one's village in India. Relatedly, Rustagi, Engel et al. (2010) and Carpenter and Seki (2011) find that groups whose members are more cooperative (as measured using public goods game experiments) are better in forest management and have a higher fishing productivity, respectively. Kosfeld and Rustagi (2015) find that real-world leaders' behavior in a third-party punishment game explains the relative success of groups in forest management.

³⁰ Benz and Meier (2008) find a positive correlation between dictator game giving and charitable giving by students at the University of Zurich. Barr and Zeitlin (2010) find that dictator giving by primary school teachers in Uganda negatively correlates with their absenteeism from work. Carpenter and Myers (2010) find that dictator giving is positively related to the decision to volunteer as a firefighter as well as firefighter training hours.

(2015) find no systematic relation between several lab measures of other-regarding preferences and various field behaviors related to donating and helping others. Moreover, some differences between the lab and the field are reported also in the studies that do find support for the external validity of lab experiments. For example, Karlan (2005) reports data from a step-level public goods game experiment conducted with the same participants used in the trust game, and finds no relation between saving and borrowing behavior and choices made in the public goods game, which he attributes to the lower similarity between this game and repayment choices, in contrast to the more direct similarity of the trust game.³¹

This variability is not unique to lab-field comparisons; laboratory studies have also shown similar variations in the pattern of individual differences in other-regarding behavior when attempting to generalize across distinct lab settings. For example, Kurzban and Houser (2005) find that individual behavioral types derived from four-person public goods games that are similar to those defined in the present study are stable across different experimental settings. Relatedly, in two online experiments, Capraro, Jordan et al. (2014) and Peysakhovich, Nowak et al. (2014) show that behavior in social dilemma games (prisoner's dilemma and public goods games) correlates with dictator giving and behavior in the trust game. By contrast, Herrmann and Orzen (2008) find that subjects' cooperativeness in a prisoner's dilemma laboratory experiment does not predict their investment choices in a rent-seeking laboratory game. Müller, Sefton, et al. (2008) report a lab experiment where subjects play a sequence of five two-stage voluntary contribution games. These authors classify subjects based on their cooperativeness in the game, and find that the classification of about two-thirds of participants varies from game to game.³²

Let us then step back and attempt to summarize the present state of this literature. First, the fact that choices made by an individual about other-regarding behavior in one setting may or

³¹ Along with the positive correlation between dictator giving and charitable donations, Benz and Meier (2008) also report that students who never made any donation in the field donated positive amounts in the experiment. This, however, may be due to differences across settings in the action space (binary decision vs. incremental donations) and the endowment used for the donations (earned vs. windfall money). Laury and Taylor (2008) report mixed evidence of a correlation between public good game behavior and contributions to a naturally-occurring public good, as the existence of a link between lab and field decisions depends on how they measure subjects' altruism. List (2006) and Stoop, Noussair et al. (2012) also report evidence lab results that do not generalize to the field, albeit not in the context of within-subject studies – see Camerer (2015) for a discussion of these studies.

³² Using a within-subject design Blanco, Engelmann et al. (2011) observe subjects' behavior in an ultimatum game, a dictator game, a sequential-move prisoners' dilemma game and a public goods game. While they find significant correlations across some decision games (e.g. second-mover behavior in the prisoners' dilemma game is positively correlated with dictator game and public goods game behavior), they also report lack of correlations across other decision settings (e.g. dictator game behavior is not correlated with public goods game behavior).

may not predict his or her other-regarding behavior in a different setting is not idiosyncratic to lab-field comparisons. As shown by the papers cited above it can also occur in generalizing from one lab setting to another, and, as Camerer (2015) argues, from one field setting to another.

Second, in the existing studies observed within-subject variability across contexts in other-regarding behavior does not so far seem to be systematically related to factors that have been cited by skeptics of the external relevance of lab measures of social preferences as notably different between lab and field, such as the higher scrutiny or abstractness of task framing in the typical lab measure.³³ For example, Fehr and Leibbrandt (2011) observe a correlation between lab and field cooperativeness despite the fact that their subjects face an abstractly-framed public goods game in the lab and a naturally-framed cooperation problem in the field (the overexploitation of fish resources). Similarly, in our study we find a significant link between lab and field cooperativeness in the case of TMR transmissions despite the fact that in the laboratory experiment drivers made choices under relatively high-scrutiny conditions and in an abstract, neutrally-framed decision setting, while the field setting involved a naturally-framed decision situation in which the level of scrutiny was much lower. Moreover, there are no substantial differences in scrutiny and context across the TMR transmission and satellite survey settings to explain why we find a link between lab and field behavior in the former setting but not in the latter.³⁴

So, how should we understand this variability in whether or not other-regarding behavior generalizes from one context to another? Those who are skeptical that social preferences should play a causal role in microeconomic explanations similar to that of other standard economic or decision-theoretic preferences (such as those for consumer goods, for risk-taking, or for the time pattern of monetary payments) may suggest that such variability is evidence that stable other-regarding preferences do not exist. We here suggest an alternative view: that it is at least as sensible to presume social preferences are stable, but to seek other factors that vary by context that might affect their expression in each context.

³³ And, as Camerer (2015) has pointed out, though the lab may have “high scrutiny,” there are lots of field situations which embody quite significant scrutiny, and this may occur in association with higher stakes than in labs; see Camerer (2015), Section 3.B.

³⁴ Differences in scrutiny and context seem also unable to explain the lack of correlation between game and field behavior in Voors, Turley et al. (2011) and (2012) given that game and field decisions were collected using the same elicitation procedure. The authors note that important differences across the two settings are the size of the stakes and the fact that, relative to the standard public goods game, contributing to the community project benefits a larger number of people (the entire village).

For instance, a reasonable conjecture is that observed individual differences in other-regarding behavior in the lab (or indeed, in any specific setting) originate from stable differences in the willingness of subjects to forgo individual material benefits when doing so is required in order to comply with social norms, i.e. shared understandings about the social appropriateness of different choices in a given situation (Elster 1989; Ostrom 2000). But, in this framework, the relevant social norms, or what does or does not constitute socially appropriate behavior, may *vary* across decision contexts. Factors affecting this may include the conditions under which a choice is made, how the decision situation is framed, and most crucially, the social identity of the decision-maker and that of the individuals with whom the decision-maker interacts (Akerlof and Kranton 2000; Bicchieri 2006; Huck and Müller 2007; Krupka, Leider et al. 2011; Barr, Lane et al. 2015; Chang, Chen et al. 2015).³⁵ For example, different norms of cooperation may apply when individuals are confronted with in-group and out-group members (Hedinger and Goette 2006; Goette, Huffman et al. 2012).

On this interpretation we would expect the lab measure of individual differences in costly cooperation towards other drivers to predict individual differences in the willingness to engage in similar behavior towards drivers in the field, especially since we measured these differences just as new-to-the-occupation trainees were being socialized into their new social identity as tractor-trailer drivers all employed at the same firm. However, the lab measure would only predict such differences in behavior towards the experimenters if the norms for driver behavior towards social scientists are the same as those towards fellow drivers. Since we are not in the drivers' in-group, this seems unlikely, and makes the failure of the lab measure to predict differences in satellite survey responses unsurprising.

Our study was not designed as a test of this approach, and we cannot, of course, rule out alternative explanations of our findings. But we propose that the overall pattern of results in the literature, along with the specific results reported here, suggest that future work should be focused on bringing norms and their context-dependence more fully within the scope of economic theory and economic experiments, while maintaining the current insights from experiments about other-regarding behavior, such as the presumptions that individuals may vary

³⁵ For recent studies providing empirical evidence on the existence and variability of social norms across decision situations see Krupka, Leider et al. (2011), Burks and Krupka (2012), Gaechter, Nosenzo et al. (2013), and Krupka and Weber (2013). More specifically, Barr, Lane et al. (2015) and Chang, Chen et al. (2015) relate individual behavior in redistribution games to social norms that are social-identity dependent.

in the extent to which they are motivated by social preferences, and that—other things held equal—the personal material cost (i.e. price) of other-regarding behavior will affect its incidence.

Table 3: Determinants of TMR Transmissions

	Model I	Model II	Model III
1 if Conditional Cooperator	0.607*** (0.001)	0.561*** (0.002)	0.551*** (0.000)
1 if Unconditional Cooperator	0.464** (0.024)	0.426** (0.038)	0.420*** (0.009)
Num. of Work Assignments	0.164*** (0.000)	0.163*** (0.000)	0.190*** (0.000)
1 if Female	-	0.367 (0.150)	0.284 (0.147)
1 if Non-White or Hispanic	-	- 0.141 (0.481)	- 0.282* (0.070)
1 if Single	-	0.444** (0.017)	0.232 (0.109)
1 if Separated or Divorced or Widowed	-	- 0.015 (0.940)	0.138 (0.374)
Age (in years)	-	0.017** (0.028)	0.015** (0.014)
1 if Education = Some College (no degree)	-	0.546*** (0.001)	0.433*** (0.001)
1 if Education = Junior or Technical College	-	0.468** (0.027)	0.315* (0.054)
1 if Education = College or higher	-	0.564** (0.027)	0.446** (0.022)
1 if Income Category = \$10k - \$20k	-	- 0.017 (0.932)	- 0.071 (0.645)
1 if Income Category = \$20k - \$30k	-	0.176 (0.441)	0.165 (0.356)
1 if Income Category = \$30k - \$40k	-	- 0.098 (0.704)	- 0.095 (0.640)
1 if Income Category = \$40k or higher	-	0.036 (0.855)	- 0.070 (0.645)
Tenure (in weeks)	-	-	- 0.001 (0.441)
1 if Terminal of Operation = Big City	-	-	- 0.032 (0.780)
1 if Terminal of Operation = Medium City #1	-	-	0.184 (0.253)
1 if Terminal of Operation = Medium City #2	-	-	- 0.220 (0.104)
1 if Terminal of Operation = Medium City #3	-	-	- 0.080 (0.662)
1 if Terminal of Operation = company in the East U.S.	-	-	- 4.320*** (0.000)
1 if Terminal of Operation = company in the West U.S.	-	-	- 3.316*** (0.000)
1 if Terminal of Operation = other terminal	-	-	- 0.504*** (0.001)
1 if Type of Work = Dedicated	-	-	- 0.332*** (0.003)
1 if Type of Work = Team	-	-	0.084 (0.684)
1 if Type of Work = Intermodal	-	-	-0.277 (0.122)
1 if Type of Work = Other	-	-	-0.039 (0.841)
Constant	-5.277*** (0.000)	- 6.352*** (0.000)	- 5.485*** (0.000)
<i>N. of observations</i>	24926	24926	24926
<i>N. of groups (drivers)</i>	633	633	633
<i>Log-likelihood</i>	-6266.43	-6252.96	-6099.49
<i>LR test: Cond. Coop. = Uncond. Coop.</i>	0.71, <i>p</i> = 0.398	0.65, <i>p</i> = 0.421	1.01, <i>p</i> = 0.315

Tobit random-effects models. In all models the dependent variable is the sum of TMRs transmitted in a given week of employment. P-values are reported in parentheses. Significance levels: * 10%; ** 5%; *** 1%.

Table 4: Determinants of Satellite Survey Responses

	Model I	Model II	Model III
1 if Conditional Cooperator	1.186 (0.271)	1.070 (0.655)	1.085 (0.585)
1 if Unconditional Cooperator	1.213 (0.267)	1.095 (0.594)	1.062 (0.723)
1 if Female	-	1.261 (0.292)	1.244 (0.318)
1 if Non-White or Hispanic	-	0.616*** (0.004)	0.626*** (0.005)
1 if Single	-	0.689** (0.017)	0.778 (0.106)
1 if Separated or Divorced or Widowed	-	0.659** (0.012)	0.622*** (0.004)
Age (in years)	-	1.017*** (0.009)	1.033*** (0.000)
1 if Education = Some College (no degree)	-	1.444*** (0.009)	1.442*** (0.008)
1 if Education = Junior or Technical College	-	1.183 (0.354)	1.118 (0.536)
1 if Education = College or higher	-	1.096 (0.685)	1.146 (0.543)
1 if Income Category = \$10k - \$20k	-	0.867 (0.404)	0.857 (0.364)
1 if Income Category = \$20k - \$30k	-	0.978 (0.906)	0.954 (0.804)
1 if Income Category = \$30k - \$40k	-	0.939 (0.768)	0.882 (0.555)
1 if Income Category = \$40k or higher	-	0.598*** (0.002)	0.555*** (0.000)
Tenure (in weeks)	-	-	0.988*** (0.000)
1 if Terminal of Operation = Big City	-	-	0.852* (0.068)
1 if Terminal of Operation = Medium City #1	-	-	0.917 (0.506)
1 if Terminal of Operation = Medium City #2	-	-	0.943 (0.568)
1 if Terminal of Operation = Medium City #3	-	-	0.999 (0.994)
1 if Terminal of Operation = company in the East U.S.	-	-	0.633*** (0.004)
1 if Terminal of Operation = company in the West U.S.	-	-	0.696** (0.021)
1 if Terminal of Operation = other terminal	-	-	0.901 (0.367)
1 if Type of Work = Dedicated	-	-	1.062 (0.516)
1 if Type of Work = Team	-	-	0.777 (0.108)
1 if Type of Work = Intermodal	-	-	0.686** (0.024)
1 if Type of Work = Other	-	-	1.035 (0.807)
<i>N. of observations</i>	17749	17749	17749
<i>N. of groups (drivers)</i>	619	619	619
<i>Log-likelihood</i>	-9730.26	-9705.70	-9583.89
<i>LR test: Cond. Coop. = Uncond. Coop.</i>	0.02, <i>p</i> = 0.881	0.03, <i>p</i> = 0.874	0.02, <i>p</i> = 0.878

Logit random-effects models. In all models the dependent variable assumes value 1 if the driver responded to the survey he or she received in a given week and value 0 otherwise. Results are displayed as factor changes in the odds of responding. P-values are reported in parentheses. Significance levels: * 10%; ** 5%; *** 1%.

Appendix 1: Instruction Script for the Social Dilemma Experiment (NOT FOR PUBLICATION)

What follows is the text of the script spoken by Burks to each group of subjects. Subjects also saw an abbreviated version of these instructions on their computer screens, while they were listening to the instructions. The table mentioned in this text is in Appendix 2.

ACTIVITY ONE: TWO-PERSON SENDING DECISION.

This activity does not take very long to do, but it is the most complicated thing to explain that we will do all day. So please bear with me as I give you the details.

You are going to make the decisions in this task ONCE. Some things you do today are going to be repeated, but the decisions in this task are not among the things we will repeat.

The BASICS OF THIS ACTIVITY ARE VERY SIMPLE, and so let me start there.

In this activity there are two different roles, Person 1 and Person 2. When we figure out your payoff you will be either a Person 1, or a Person 2, but not both. Each Person 1 will be matched with a Person 2 here in this room, but neither of you will ever know which specific other driver trainee you have been matched with.

Whether you are a Person 1 or a Person 2, the basics are the same. You will have a new amount of five dollars put in your account at the beginning of this activity by us. You have to decide whether to keep this five dollars, or to send it to the other person you are paired with. If you KEEP the money, it is yours at the end of the activity. If you SEND the money, we will double it, so that the person you send it to gets twice what you sent. Likewise, the person you are matched with will be making a SIMILAR decision about their five dollars. If they keep it, then it is theirs at the end of the activity, but if they send it to you, you will get twice what they sent.

So, that is THE BASIC OUTLINE. NEXT WE WILL LOOK AT THE DETAILS.

If you are a Person 1 your decision is simple. You have to decide whether to send your five dollars to Person 2, or to keep it. If you keep it, it is yours at the end of the activity, but if you send it to Person 2, we will *double* it, so that Person 2 actually gets \$10.

Now, Person 2 also gets to decide about sending money to Person 1. But there are a couple of special features to Person 2's decision.

The first special feature is that Person 2 doesn't just have a yes-no choice about sending the five dollars. Instead, Person 2 can send any exact dollar amount to Person 1. So Person 2 can send: \$0, \$1, \$2, \$3, \$4, or \$5. Just like before, however, whatever Person 2 keeps is theirs at the end of the activity, and whatever Person 2 sends will be *doubled* by the researchers.

Example: if Person 2 keeps \$4 and sends \$1 to Person 1, Person 1 will actually receive \$2,

Example: if Person 2 keeps \$2 and sends \$3 to Person 1, Person 1 will actually receive \$6.

The second special feature is that Person 2 gets to decide what to do under two different cases. The first case is how much they want to respond if Person 1 has not sent them money. This choice will be on the LEFT side of the Person 2 choice screen. The second case is how much they want to respond if Person 1 has sent them money. This choice will be on the RIGHT side of the Person 2 choice screen. In both cases the rules are the same: Person 2 can choose how many dollars to keep and how many to send, and whatever is sent is doubled.

TABLE OF PAYOFFS IS HANDED OUT. (table is provided following end of script text)

PLEASE DO NOT WRITE ON THIS SHEET, AS WE WILL RE-USE IT.

Look at payoff handout sheet.

Let's look at the top table. As you can see from looking at the first column (from Line 1 to Line 6), the top table is for the case in which Person 1 decides to send Person 2 \$0. If you look at the second column of the top table (from Line 1 to Line 6), you can see all of the possible choices Person 2 has about how to respond. Finally, in each line, if you follow the arrow to the right, you see two more columns that show what the two people, Person 1 and Person 2, make in earnings. So, Lines 1 through 6 show each response Person 2 can make to the decision of Person 1 to send \$0, and the payoffs each of them receive.

Example: look at Line 1. From the first two columns, this is the situation in which Person 1 sends \$0, and Person 2 responds by also sending back \$0. Following the arrow to the right, you can see that Person 1 makes \$5, because he keeps \$5, and gets \$0 from Person 2, and Person 2 also makes \$5, because he keeps all of his initial \$5, and also receives nothing from Person 1.

Example: look at Line 4. From the first two columns, this is the situation in which Person 1 sends \$0, and Person 2 responds by sending back \$3. Following the arrow to the right, you can see that Person 1 makes \$11, because he keeps \$5, and also gets \$3 doubled to \$6 from Person 2. But Person 2 makes \$2, because he sent \$3 of his initial \$5, but received nothing back.

Example: look at Line 6. From the first two columns, this is the situation in which Person 1 sends \$0, and Person 2 responds by sending back \$5. Following the arrow to the right, you can see that Person 1 makes \$15, because he keeps \$5, and also gets \$5 doubled to \$10 from Person 2. But Person 2 makes \$0, because he sent all of his initial \$5, but received nothing back.

Now let's look at the bottom table. This repeats the same pattern as the top table, except that, as you can see from looking at the first column (from Line 7 to Line 12), it is for the case in which Person 1 sends \$5 to Person 2. If you look at the second column of the bottom table (from Line 7 to Line 12), you can see all of the possible choices Person 2 has about how to respond. Finally, in each line, if you follow the arrow to the right, you see two more columns that show what the two people, Person 1 and Person 2, make in earnings. So, Lines 7 through 12 show each response Person 2 can make to the decision of Person 1 to send \$5, and the payoffs each of them receive.

Example: look at Line 7. From the first two columns, this is the situation in which Person 1 sends \$5, and Person 2 responds by sending back \$0. Following the arrow to the right, you can see that Person 1 makes \$0, because he sent all his initial \$5, and gets \$0 back from Person 2, but

Person 2 makes \$15, because he keeps all of his initial \$5, and also gets \$5 doubled to \$10 from Person 1.

Example: look at Line 9. From the first two columns, this is the situation in which Person 1 sends \$5, and Person 2 responds by sending back \$2. Following the arrow to the right, you can see that Person 1 makes \$4, because he sent all of his initial \$5, and gets \$2 doubled to \$4 back from Person 2. And Person 2 makes \$13, because he kept \$3 of his initial \$5, also gets \$5 doubled to \$10 from Person 1

Example: look at Line 12. From the first two columns, this is the situation in which Person 1 sends \$5, and Person 2 responds by sending back \$5. Following the arrow to the right, you can see that Person 1 makes \$10, because he sent all his \$5, and gets back \$5 doubled to \$10 from Person 2. And Person 2 also makes \$10, because he sent back all of his initial \$5, also got \$5 doubled to \$10 from Person 1.

Any questions now?

OK, now for the next to last special feature. This is very important. We are going to randomly assign the roles of Person 1 and Person 2 at the end of the activity, not at the beginning. So, we are going to ask everyone to make a decision first as a Person 1, and then second, a decision as a Person 2.

Let me repeat that: you will first make a decision IN CASE YOU ARE A Person 1, whether you will send \$5 or not. Then, on a new screen, you will also make a two-decision IN CASE YOU ARE A Person 2: how much to send back if you got \$0 (LEFT SIDE of the screen), and how much to send back if you got \$5 (doubled to \$10) (RIGHT SIDE of the screen).

So the way the payoffs will work is that first you will be matched by the computer with someone else here in the room. For example, #11 over here might be matched with #23 over there. Of course, let me remind you that we will never tell you with whom you were matched. Since everyone made both a Person 1 choice and then a Person 2 choice, once you have been matched, the computer will in effect flip a coin—it will randomly make one of you Person 1 and the other Person 2. Then it will look at your choices and those of the person you were matched with, and calculate your payoffs.

Finally, here is the last special feature. Before each decision screen there is another question. We are not only going to ask you what you want to do, I'd like to know what you think everyone else here today will do. So, we will also ask you to guess how others will handle the decisions you are about to make.

So for instance, right before the Person 1 screen asks you whether you will send \$5 to Person 2 or not, we will also ask you to guess what percent of the people here in the room will send \$5 as Person 1. We will pay you \$1 extra if your guess is close (+/- 5%) to what people actually do.

And, when you make your decision as Person 2, you will have to tell us how much you want to send to Person 1 both when Person 1 sent \$0 to you, and also when Person 1 sent \$5 to you.

Right before you make this decision, we will also ask you the average amount in dollars you think people in this room will send in each of these cases. We will pay you \$1 extra for each guess that is close (+/- \$.25) to what people actually choose.

So, to recap, you will have FOUR different screens of choices. FIRST, your best guess about the % of those here today who will send \$5 as Person 1. SECOND, your own decision in case the computer makes you a Person 1 for the payoffs. THIRD, your best guess about how much people here will send as Person 2, for the case when they got nothing (LEFT SIDE of the screen), and again for the case in which they received \$5 (RIGHT SIDE of the screen). FOURTH, your own choice in case the computer makes you a Person 2 for the payoffs, of how much to send back to Person 1 when Person 1 sent you \$0 (LEFT SIDE of the screen), and when Person 1 sent you \$5 (RIGHT SIDE of the screen).

Any questions now?

At the end of the activity, the computer will show you what your earnings are in total, including both from your guesses about others and from your choices as Person 1 or Person 2.

OK, let's look at the instructions on the first computer screen. When you are happy you understand them, please click "continue" or "OK" in the lower right-hand corner of the screen. That will take you to a waiting screen, and when everyone is there, we will move to the first of two practice question screens, to make sure you understand how the payoff table works.

Practice Screen 1: Person 1 sends \$0, and Person 2 responds by sending back \$1.
Line 2 (Person 1 gets \$7, Person 2 gets \$4)

Practice Screen 2: Person 1 sends \$5, and Person 2 responds by sending back \$3. Line 10
(Person 1 gets \$6, and Person 2 gets \$12)

Appendix 2: Table of PD Game Payoffs Given to Subjects (NOT FOR PUBLICATION)

	Person 1 Sends	Person 2 Sends		Person 1 Makes	Person 2 Makes
Line 1	\$0	\$0	→	\$5	\$5
Line 2	\$0	\$1	→	\$7	\$4
Line 3	\$0	\$2	→	\$9	\$3
Line 4	\$0	\$3	→	\$11	\$2
Line 5	\$0	\$4	→	\$13	\$1
Line 6	\$0	\$5	→	\$15	\$0
	Person 1 Sends	Person 2 Sends		Person 1 Makes	Person 2 Makes
Line 7	\$5	\$0	→	\$0	\$15
Line 8	\$5	\$1	→	\$2	\$14
Line 9	\$5	\$2	→	\$4	\$13
Line 10	\$5	\$3	→	\$6	\$12
Line 11	\$5	\$4	→	\$8	\$11
Line 12	\$5	\$5	→	\$10	\$10

Appendix 3 – Additional results (NOT FOR PUBLICATION)

Table A3.1 contains the results of negative binomial random-effects regressions where the dependent variable is the number of TMRs sent by a driver in a given week of employment and the independent variables are those included in Models I to III in Table 3 in the main text. The number of job assignments received by a driver in a given employment week is used as exposure variable in the regressions. Results are expressed as factor changes in the odds ratios.

Table A3.1 – Negative Binomial Regressions of TMR Transmissions

	Model I	Model II	Model III
1 if Conditional Cooperator	1.518 ^{***} (0.007)	1.497 ^{***} (0.009)	1.546 ^{***} (0.000)
1 if Unconditional Cooperator	1.404 ^{**} (0.049)	1.392 [*] (0.062)	1.402 ^{**} (0.018)
1 if Female	-	1.370 (0.147)	1.213 (0.260)
1 if Non-White or Hispanic	-	0.845 (0.321)	0.792 [*] (0.088)
1 if Single	-	1.262 (0.137)	1.217 (0.121)
1 if Separated or Divorced or Widowed	-	1.000 (0.998)	1.155 (0.284)
Age (in years)	-	1.012 [*] (0.089)	1.013 ^{**} (0.021)
1 if Education = Some College (no degree)	-	1.623 ^{***} (0.000)	1.520 ^{***} (0.000)
1 if Education = Junior or Technical College	-	1.429 ^{**} (0.046)	1.269 [*] (0.097)
1 if Education = College or higher	-	1.722 ^{***} (0.010)	1.574 ^{***} (0.007)
1 if Income Category = \$10k - \$20k	-	0.982 (0.911)	0.951 (0.702)
1 if Income Category = \$20k - \$30k	-	1.073 (0.722)	1.058 (0.719)
1 if Income Category = \$30k - \$40k	-	0.911 (0.676)	0.917 (0.634)
1 if Income Category = \$40k or higher	-	1.019 (0.912)	0.928 (0.578)
Tenure (in weeks)	-	-	0.999 (0.476)
1 if Terminal of Operation = Big City	-	-	1.014 (0.888)
1 if Terminal of Operation = Medium City #1	-	-	1.233 (0.143)
1 if Terminal of Operation = Medium City #2	-	-	0.849 (0.184)
1 if Terminal of Operation = Medium City #3	-	-	1.006 (0.970)
1 if Terminal of Operation = company in the East U.S.	-	-	0.011 ^{***} (0.000)
1 if Terminal of Operation = company in the West U.S.	-	-	0.040 ^{***} (0.000)
1 if Terminal of Operation = other terminal	-	-	0.633 ^{***} (0.001)
1 if Type of Work = Dedicated	-	-	0.740 ^{***} (0.003)
1 if Type of Work = Team	-	-	1.035 (0.850)
1 if Type of Work = Intermodal	-	-	0.779 [*] (0.076)
1 if Type of Work = Other	-	-	0.966 (0.843)
Constant	0.119 ^{***} (0.000)	0.048 ^{***} (0.000)	0.056 ^{***} (0.000)
<i>N. of observations</i>	24842	24842	24842
<i>N. of groups (drivers)</i>	633	633	633

Negative binomial random-effects models. In all models the dependent variable is the number of TMRs transmitted in a given week of employment (the total number of assignments completed in that week is used as exposure variable). P-values are reported in parentheses. Significance levels: * 10%; ** 5%; *** 1%.

References

- Abeler, J. and D. Nosenzo (2015). "Self-Selection into Laboratory Experiments: Pro-social Motives versus Monetary Incentives." Experimental Economics 18(2): 195-214.
- Akerlof, G. A. and R. E. Kranton (2000). "Economics and Identity." The Quarterly Journal of Economics 115(3): 715-753.
- Akerlof, G. A. and R. E. Kranton (2005). "Identity and the Economics of Organizations." Journal of Economic Perspectives 19(1): 9-32.
- Akerlof, G. A. and R. E. Kranton (2010). Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being, Princeton and Oxford: Princeton University Press.
- Anderson, J., S. Burks, et al. (2013). "Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples." Experimental Economics 16(2): 170-189.
- Baran, N. M., P. Sapienza, et al. (2010). Can we infer social preferences from the lab? Evidence from the trust game.
- Bardsley, N., R. Cubitt, et al. (2010). Experimental Economics: Rethinking the Rules. Princeton and Oxford, Princeton University Press.
- Barr, A., T. Lane, et al. (2015). "On the Social Sp appropriateness of Discrimination." Manuscript, University of Nottingham.
- Barr, A., T. Packard, et al. (2014). "Participatory Accountability and Collective Action: Experimental Evidence from Albania." European Economic Review 68: 250-269.
- Barr, A. and A. Zeitlin (2010). Dictator games in the lab and in nature: External validity tested and investigated in Ugandan primary schools. University of Oxford Department of Economics. Oxford, UK, Economics Series Working Papers. CSAE WPS/2010-11.
- Benz, M. and S. Meier (2008). "Do people behave in experiments as in the field?—evidence from donations." Experimental Economics 11(3): 268-281.
- Bicchieri, C. (2006). The Grammar of Society: The Nature and Dynamics of Social Norms. Cambridge, Cambridge University Press.
- Blanco, M., D. Engelmann, et al. (2011). "A Within-Subject Analysis of Other-Regarding Preferences." Games and Economic Behavior 72(2): 321-338.
- Bowles, S. and S. Polania-Reyes (2012). "Economic Incentives and Social Preferences: Substitutes or Complements?" Journal of Economic Literature 50(2): 368-425.
- Bureau of Labor Statistics (2012) "SOC 53-3032: Heavy and Tractor-trailer Truck Drivers." Occupational Outlook Handbook 2010-11 Edition.
- Burks, S., J. Carpenter, et al. (2008). Using Behavioral Economic Field Experiments at a Firm: the Context and Design of the Truckers and Turnover Project. The Analysis of Firms and Employees: Quantitative and Qualitative Approaches. S. Bender, J. Lane, K. Shaw, F. Andersson and T. von Wachter. Chicago and London, NBER and University of Chicago 45-106.
- Burks, S. V., M. Belzer, et al. (2010) "Trucking 101: An Industry Primer." Transportation Research Circular Number E-C146 of the Transportation Research Board.
- Burks, S. V., J. Carpenter, et al. (2009). "Cognitive skills affect economic preferences, strategic behavior, and job attachment." Proceedings of the National Academy of Sciences 106: 7745-7750.

- Burks, S. V. and E. Krupka (2012). "A Multi-Method Approach to Identifying Norms and Normative Expectations within a Corporate Hierarchy: Evidence from the Financial Services Industry." Management Science 1(58): 203-217.
- Camerer, C. (2015). The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. Handbook of Experimental Economic Methodology. G. Frechette and A. Schotter. Oxford, UK, Oxford University Press.
- Capraro, V., J. J. Jordan, et al. (2014). "Heuristics guide the implementation of social preferences in one-shot Prisoner's Dilemma experiments." Scientific Reports 4: 6790.
- Carlsson, F., O. Johansson-Stenman, et al. (2014). "Social Preferences Are Stable over Long Periods of Time." Journal of Public Economics 117: 104-114.
- Carpenter, J. and C. K. Myers (2010). "Why volunteer? Evidence on the role of altruism, image, and incentives." Journal of Public Economics 94(11-12): 911-920.
- Carpenter, J. and E. Seki (2011). "Do Social Preferences Increase Productivity? Field Experimental Evidence from Fishermen in Toyama Bay." Economic Inquiry 49(2): 612-630.
- Chang, D., R. Chen, et al. (2015). "Social norms and identity dependent preferences." Working paper, School of Information, University of Michigan.
- Chen, Y. and S. X. Li (2009). "Group Identity and Social Preferences." American Economic Review 99(1): 431-457.
- Cleave, B. L., N. Nikiforakis, et al. (2013). "Is There Selection Bias in Laboratory Experiments? The Case of Social and Risk Preferences." Experimental Economics 16(3): 372-382.
- Elster, J. (1989). "Social Norms and Economic Theory." The Journal of Economic Perspectives 3(4): 99-117.
- Falk, A., S. Meier, et al. (2013). "Do Lab Experiments Misrepresent Social Preferences? The Case of Self-selected Student Samples." Journal of the European Economic Association 11(4): 839-852.
- Fehr, E. and A. Leibbrandt (2011). "A field study on cooperativeness and impatience in the Tragedy of the Commons." Journal of Public Economics 95(9-10): 1144-1155.
- Fischbacher, U. (2007). "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." Experimental Economics 10(2): 171-178.
- Gaechter, S., D. Nosenzo, et al. (2012). "Who Makes a Good Leader? Cooperativeness, Optimism, and Leading-by-Example." Economic Inquiry 50(4): 953-967.
- Gaechter, S., D. Nosenzo, et al. (2013). "Peer Effects in Pro-social Behavior: Social Norms or Social Preferences?" Journal of the European Economic Association 11(3): 548-573.
- Galizzi, M. M. and D. Navarro-Martinez (2015). On the external validity of social-preference games: A systematic lab-field study, Department of Economics and Business, Universitat Pompeu Fabra, Economics Working Papers.
- Gintis, H. and D. Helbing (2015). "Homo Socialis: An Analytical Core for Sociological Theory." Review of Behavioral Economics 2(1-2): 1-59.
- Goette, L., D. Huffman, et al. (2012). "The Impact of Social Ties on Group Interactions: Evidence from Minimal Groups and Randomly Assigned Real Groups." American Economic Journal: Microeconomics 4(1): 101-115.
- Harrison, G. W. and J. A. List (2004). "Field experiments." Journal of Economic Literature 42(4): 1009-1055.
- Hedinger, B. and L. Goette (2006). Cooperation in the cockpit: evidence of reciprocity and trust among Swiss air force pilots, University of Zurich Working Paper.

- Herrmann, B. and H. Orzen (2008). The Appearance of Homo Rivalis: Social Preferences and the Nature of Rent-seeking. CeDEx Discussion Paper Series at the University of Nottingham.
- Hoffman, M. and S. V. Burks (2014). "Training Contracts and Worker Overconfidence." University of Toronto Working Paper.
- Huck, S. and W. Müller (2007). Allais for all: revisiting the paradox.
- Karlan, D. (2005). "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions." American Economic Review 95(5): 1688-1699.
- Kosfeld, M. and D. Rustagi (2015). "Leader Punishment and Cooperation in Groups: Experimental Field Evidence from Commons Management in Ethiopia." American Economic Review 105(2): 747-783.
- Krupka, E., S. Leider, et al. (2011). A Meeting of the Minds: Contracts and Social Norms.
- Krupka, E. L. and R. A. Weber (2013). "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" Journal of the European Economic Association 11(3): 495-524.
- Kurzban, R. and D. Houser (2005). "Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations." Proceedings of the National Academy of Sciences of the United States of America 102(5): 1803-1807.
- Lamba, S. and R. Mace (2011). "Demography and ecology drive variation in cooperation across human populations." Proceedings of the National Academy of Sciences 108(35): 14426-14430.
- Laury, S. K. and L. O. Taylor (2008). "Altruism spillovers: Are behaviors in context-free experiments predictive of altruism toward a naturally occurring public good?" Journal of Economic Behavior & Organization 65(1): 9-29.
- Levitt, S. D. and J. A. List (2007). "What do laboratory experiments measuring social preferences reveal about the real world?" Journal of Economic Perspectives 21(2): 153-174.
- List, J. A. (2004). "Young, selfish and male: Field evidence of social preferences." The Economic Journal 114(492): 121-149.
- List, J. A. (2006). "The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions." Journal of Political Economy 114(1): 1-37.
- Llaneras, R. E., J. P. Singer, et al. (2005). Assessment of Truck Driver Distraction Problem and Research Needs. Washington, DC, National Highway Safety Administration: pp. 1-84.
- Müller, L., M. Sefton, et al. (2008). "Strategic behavior and learning in repeated voluntary-contribution experiments." Journal of Economic Behavior & Organization 67: 782-793.
- Ostrom, E. (2000). "Collective action and the evolution of social norms." The Journal of Economic Perspectives 14(3): 137-158.
- Peysakhovich, A., M. A. Nowak, et al. (2014). "Humans display a 'cooperative phenotype' that is domain general and temporally stable." Nature Communications 5.
- Rustagi, D., S. Engel, et al. (2010). "Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management." Science 330(6006): 961-965.
- Smith, V. L. (1982). "Markets as Economizers of Information - Experimental Examination of the Hayek Hypothesis." Economic Inquiry 20(2): 165-179.
- Smith, V. L. (1982). "Microeconomic Systems as an Experimental Science." American Economic Review 72(5): 923-955.

- Stoop, J., C. N. Noussair, et al. (2012). "From the Lab to the Field: Cooperation among Fishermen." Journal of Political Economy 120(6): 1027-1056.
- U.S. Census Bureau (2012) "NAICS 4841210: General Commodity Long Distance Truckload." 2007 Economic Census.
- U.S. Census Bureau (2012) "NAICS 4842300: Special Commodity Long Distance Truckload." 2007 Economic Census.
- Voors, M., T. Turley, et al. (2011). "Using Artefactual Field Experiments to Learn about the Incentives for Sustainable Forest Use in Developing Economies." American Economic Review (Papers and Proceedings) 101: 329-333.
- Voors, M., T. Turley, et al. (2012). "Exploring whether behavior in context-free experiments is predictive of behavior in the field: Evidence from lab and field experiments in rural Sierra Leone." Economics Letters 114(3): 308-311.