Journal of Economic Behavior and Organization 149 (2018) 74–87

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Experimental methods: Measuring effort in economics experiments

Gary Charness[a,*], Uri Gneezy[b,c], Austin Henderson[b]

[a] Department of Economics, University of California, Santa Barbara, United States
[b] Rady School of Management, University of California, San Diego, United States
[c] CREED, University of Amsterdam, United States

ARTICLE INFO

Article history:
Received 4 November 2017
Revised 24 February 2018
Accepted 26 February 2018

JEL codes:
B49
C90
C91
C92

Keywords:
Stated effort
Real effort
Timing of decisions
Goal-orientation
Experimental methodology

ABSTRACT

The study of effort provision in a controlled setting is a key research area in experimental economics. There are two major methodological paradigms in this literature: stated effort and real effort. In the stated-effort paradigm the researcher uses an "effort function" that maps choices to outcomes. In the real-effort paradigm, participants work on a task, and outcomes depend on their performance. The advantage of the stated-effort design is the control the researcher has over the cost of effort, which is particularly useful when testing theory. The advantage of the real-effort design is that it may be a better match to the field environment, particularly with respect to psychological aspects that affect behavior. An open question in the literature is the degree to which the results obtained by the two paradigms differ, and if they do, why. We present a review of methods used and discuss the results obtained from using these different approaches, and issues to consider when choosing and implementing a task.

1. Introduction

Understanding when and how individuals exert effort is critical to many questions in economics. While a large literature in experimental economics studies effort provision, different approaches have been used to operationalize it experimentally. Experimental economists have primarily utilized two methodological paradigms: stated effort and real effort. There is limited theoretical and/or experimental evidence to guide researchers in deciding which task to use. Furthermore, there are many real-effort tasks and ways to implement them.

With stated effort, the choice of "effort" involves clear numerical costs and benefits. In a typical implementation, participants are presented with a menu that displays a discrete selection of effort levels (e.g., from 1 to 10) and a corresponding list of costs. These costs often influence the profits of another subject, as in a gift-exchange situation (Fehr et al., 1993, 1997; Charness, 2004), or in a tournament involving effort (Müller and Schotter, 2010; Bull et al., 1987). The advantage of the stated-effort approach is that there is no uncertainty regarding an individual's cost of effort. A potential drawback of the

* Corresponding author.
E-mail addresses: charness@econ.ucsb.edu, gary.charness@ucsb.edu (G. Charness), ugneezy@ucsd.edu (U. Gneezy), austin.henderson@rady.ucsd.edu (A. Henderson).

https://doi.org/10.1016/j.jebo.2018.02.024
0167-2681/

method is that simply choosing a number may not capture the field environment and the psychological forces involved in putting forth actual effort.

Real-effort tasks measure the behavior of participants given specific observable tasks, such as solving mazes (Gneezy et al., 2003), solving anagrams (Charness and Villeval, 2009), adding series of two-digit numbers (Niederle and Vesterlund 2007), counting the number of zeros in a large grid (Abeler et al., 2011), transcribing meaningless "greek" letters (Augenblick et al., 2015), and cracking walnuts (Fahr and Irlenbusch, 2000). The effort could be physical, as in folding pieces of paper and stuffing envelopes, cognitive, as in solving a series of math equations, or creative, as in writing stories or packing quarters. The advantage of the real-effort method is that it is closer to the psychology of working. For example, the cost of effort might vary over time: solving mazes might be fun initially, but might gradually become less motivating. A potential drawback is that the researcher does not know the cost of effort (and perhaps not even the sign of the effort cost; Gross et al., 2015) for participants, so that testing theories is more challenging.

A key purpose for conducting a laboratory experiment is to use the advantages of a controlled environment to learn about an economically-interesting phenomenon. We identify several dimensions that are important when deciding about effort measurement, such as the timing of the effort decision, the existence of goal-oriented decision-making, and the particulars of decisions over effort and money. Our aim is to help organize the considerations involved in both picking the methodology best suited to the research question at hand and understanding the key limitations of that methodology.

2. Stated-effort experiments

Testing specific models is a central focus of many effort experiments, and this typically requires experimental control over the relevant components of the theory. One needs a clear mapping from the cost of effort to the resulting productivity. Models may rely upon specific characterizations of the properties of the cost of effort function. For example, the cost of function it may be linear such that each unit of effort has the same associated cost, or it could be convex, such that the cost of each additional unit of effort is increasing. Such properties may be important to the predictions of specific models.

Smith (1976) introduced and argued for induced value, which forms the logical basis for stated effort. Although many economic experiments make use of the induced-value paradigm, we focus here on papers that explicitly used it (at least in motivation) to study effort. The gift-exchange game using induced values and stated effort was first tested in Fehr et al. (1993) and has led to important insights and has had great impact on our understanding of labor relations. In a simplified version of this game, a firm chooses a wage between 0 and 100, and the firm's earnings are determined by (100-w)*e. The worker's earning is the wage less the cost of the effort level chosen. This is the cost-of-effort schedule:

e	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$c(e)$	0	1	2	4	6	8	10	12	15	18

This method is useful when considering social preferences, since the relationship between the firm's payoff and the worker's payoff is completely known to the worker and the "sacrifice" of freely-chosen higher effort provides clear benefits.

Stated effort is also useful for testing models in tournament settings. For example, Müller and Schotter (2010) consider the prize structure in contests, testing the Moldovanu and Sela (2001) model that shows the optimal structure depends on whether the cost-of-effort function is convex or not. The experimental results show that low-ability workers tend to "drop out" and provide little or no effort (this is not part of the equilibrium in the theoretical model), while high-ability workers provide excessive levels of effort, so that there is a bifurcation of effort. Nevertheless, the firm overall receives the expected amount of effort. The cost of effort was implemented as either a linear function or a quadratic function of the "decision number" (effort). The 2×2 experimental design also varied whether one prize or two prizes were awarded for the group of four participants. It seems clear that one would be unable to test this model with real effort, since the cost of effort would be unknown for each individual.

We list below a number of prominent papers that use a stated-effort methodology, by their research areas, main findings, and significance. Several previous and more extensive literature reviews examined experiments that used stated effort in specific fields such as labor (Charness and Kuhn, 2011) and coordination (Devetag and Ortmann, 2007). A number of experimental public-goods games (reviewed in Chaudhuri, 2011), trust games (reviewed in Johnson and Mislin, 2011), and principal-agent games (e.g., Charness and Dufwenberg, 2006; Brandts et al., 2016) also use the logic of stated effort, but are not explicitly about effort. Our list is neither meant to be exhaustive or an attempt to rank the most important papers, but rather to highlight how stated-effort has been used productively in a variety of research areas. For more detail, we refer people to the literature reviews mentioned above.

3. Real-effort experiments

Researchers have used different real-effort tasks in laboratory and extra-laboratory (lab-in-the-field) settings. In Table 2, we present a partial list of real-effort tasks used in these types of settings; we then qualitatively evaluate these based on

Table 1
A range of stated-effort studies.

Abbreviated citation	Research area	Experimental design	Main experimental finding and significance
Van Huyck et al. (1990)	Coordination	The article examines a class of tacit pure coordination games with multiple equilibria, which are strictly Pareto ranked. It reports experiments that provide evidence on how human subjects make decisions under conditions of strategic uncertainty	Inefficient play is typically the result in these games. This is not the result of conflicting objectives or to asymmetric information. "Instead, coordination failure results from strategic uncertainty: some subjects conclude that it is too 'risky' to choose the payoff-dominant action and most subjects focus on outcomes in earlier period games.
Fehr et al. (1993)	Labor	A 2-stage design wherein first some participants ("employers") made wage offers which other participants ("workers") could choose to accept. The sellers then made a decision of how much effort to exert.	Workers responded to higher wage offers with higher effort, providing support for the fair-wage hypothesis.
Charness (2000)	Labor	Participants were either "employers" or "employees," and the wage of the employee was either presented to them as resulting from a random process or assigned by the experimenter.	Participants only responded with very high effort levels when the wage offers were seen as being made by random processes, indicating that the perceived responsibility of a wage rate is behaviorally important.
Brown et al. (2004)	Labor	Participants were either assigned as "firms" or "workers," and firms offered contracts (either to individual workers or publicly) with a wage and desired effort level. Treatments varied whether there was 3rd party contract enforcement, and there were 15 rounds, with stable identities throughout.	Stable long-term relationships between firms and employees emerged even in absence of 3rd party contract enforcement. Successful relationships, common in the no 3rd party enforcement condition, had both generous rent sharing and high effort from the beginning. With 3rd party enforcement, most interactions were one-shot.
Charness et al. (2006)	Principal-agent	The principal chooses to hire an agent or take an outside option. The agent decides whether to exert effort, with 5/6 chance of success if effort is chosen. In the communication treatments, the agent can send a free-form message to the principal.	Communication leads to more agents being hired and better performance by the agents. Statements of intent (promises) appear to have commitment power. The results are seen through the lens of guilt aversion.

(*continued on next page*)

Table 1 (*continued*)

Abbreviated citation	Research area	Experimental design	Main experimental finding and significance
Brandts et al. (2007)	Labor	Participants were either "managers" or "employees" and participated in a minimum effort game, in which all payoffs depended upon the lowest effort exerted by one of the employees. Managers were able to pick bonuses for employees corresponding to increases in the minimum effort. Treatments varied the level of communication between managers and employees—no communication, one-way managerial messages to employees, or two-way communication.	Allowing communication, and the particularities of the communications, were more important for overcoming coordination failures than good monetary incentives.
Chaudhuri et al. (2009)	Coordination	Participants play in a series of minimum effort games, in which groups of 8 participants play 10 rounds of the game, then give advice to their successors who play after them. Successors are also in some treatments able to see the messages and actions of their predecessors.	Efficient outcomes were more likely when the advice from each predecessor to a round was made common knowledge, rather than given to one successor each.
Müller et al. (2010)	Labor	Participants competed in an effort tournament, in which they were first randomly assigned their type and then made an effort allocation decision. The cost of effort function was varied across treatments between linear and convex, and the payoff structure was varied across treatments to either give a prize to those who exerted the highest or the two highest effort levels.	Low-type workers exerted less than theoretically-predicted effort, and high-type workers exerted greater than predicted effort. This bifurcation of effort contradicts the hypothesis that in such an effort tournament that effort should be a continuous and increasing function of ability.
Brandts et al.(2016)	Principal-agent	The principal can hire an agent, with either a rigid or flexible contract, to perform a task. If hired, the agent chooses a quality level (high, normal, or low), where non-normal quality is costly for the agent. The higher the quality, the better for the principal. There is a 50% chance of a cost shock.	Free-form communication is very effective at producing flexible contracts that achieve efficiency (high quality) and that take into account the cost shock. Everyone earns more with free-form communication, although restricted communication is ineffective.

Table 2
Some real-effort experiments.

Data entry	First use	Description	Prod.	Diff. imp	S/A	Learn	Comments
Library data entry	Gneezy and List (2006)	Participants entered information from a stack of books into a computerized database.	Yes	High	Low	Low	May be difficult to consistently find new productive data entry tasks. Task has been used many times. Counting the number of errors provides an additional measure of effort.
Classifying reviews	Bushong and Gagnon-Bartsch (2016)	Participants listen to Amazon book reviews, classifying them as either endorsing or criticizing the group. An annoying noise can be played to increase the cost of effort.	No*	High	Low	Low	Use of annoying noise gives researcher qualitative control over the cost of effort. Straightforward implementation through Amazon mTurk.
Visual search							
Counting zeros	Abeler et al. (2011)	Participants are given a table with 150 randomly ordered 0′s and 1′s, and asked to count the correct number of 0′s. A typical implementation is to count as many tables as possible within a time period.	No	Low	Low	Low	Can be implemented in several ways: by requiring the correct number of 0′s on a table to proceed, or by allowing errors and then not giving credit for an incorrect table.
Counting sevens	Mohnen et al. (2008)	Participants are given a block of random numbers and must count the number of 7′s in the block.	No	Low	Low	Low	Same as above, but potentially less difficult and thus a lower cost of effort.
Puzzles							
Packing quarters into boxes	Ariely et al. (2009)	Participants must pack 9 metal quarter-circles into a wooden box, a feat that can only accomplished with a particular arrangement of the metal pieces, within some time period. Performance is measured by amount of time to solve.	No	Med.	Med.	High	Simple to implement outside of a laboratory. Some participants may enjoy the task.
Labyrinth	Ariely et al. (2009)	Participants navigate a ball through a wooden maze on a plane by tilting the plane on 2 planes, while avoiding trap holes in the maze. Success is measured by number of trap holes passed within some time period.	No	Med.	Med.	High	Simple to implement outside of a laboratory. Some participants may enjoy the task. Luck may play a role in success in small sample sizes.
Solving mazes	Gneezy et al. (2003)	Participants solve computerized mazes by navigating a marker through a maze using the arrow keys.	No	Low	Med.	Low	Puzzles may have unequal difficulty, and some participants may enjoy the task. Task has been used many times.
Tetris-like game	Augenblick et al. (2015)	Participants must complete 4 rows of Tetris: blocks of various shapes descend slowly from top of screen and fall into place at the bottom. But descent rate does not increase, and there is no progression in the difficulty of the game.	No	Low	Low	Low	Participants are very likely familiar with the task, and some may enjoy it despite efforts to make it unenjoyable. By construction, participants cannot increase effort within a time period, and can only increase effort by increasing the amount of time they work.
Computerized tower of hanoi	Rutström and Williams (2000)	Participants on a computer play a game in which the goal is to move "disks" of various sizes onto "pegs" such that a larger disk is never placed on a smaller disk.	No	Low	Med.	High	Some participants may enjoy solving the puzzle. Can only be used once. Researcher has little control over the cost of effort.

(*continued on next page*)

Table 2 (*continued*)

Data entry	First use	Description	Prod.	Diff. imp	S/A	Learn	Comments
Memory							
Simon	Ariely et al. (2009)	An electronic device flashes a sequence of colored lights and corresponding sounds that the participant must duplicate.	No	Low	Med.	Low	Simple to implement outside of a laboratory. Differences in short-term memory may confound interpretation.
Recall last 3-digits	Ariely et al. (2009)	An experimenter reads a sequence of digits to the participant, then suddenly stops and asks the participant to recall the last 3 digits that were read.	No	Low	Med.	Low	Simple to implement outside of a laboratory. Differences in short-term memory may confound interpretation.
Physical challenge							
Dart ball	Ariely et al. (2009)	Participants throw a tennis ball at a target with attached Velcro patches to which the tennis ball will adhere.	No	Med.	High	Med.	Simple to implement outside of a laboratory. Some participants may enjoy the task.
Roll-up	Baumeister (1984)	Participants must maneuver a ball into a target hole by spreading apart then pushing together two metal rods	No	Med.	High	Med.	Simple to implement outside of a laboratory. Some participants may enjoy the task.
Running	Gneezy and Rustichini (2004)	Participants run twice along a 40 m track.	No	Low	High	Low	Easy to explain to children and implement. High variance in ability for adults makes it difficult to interpret results as due to variance in effort.
Hand dynamometer	Imas (2014)	Participants squeeze a specially calibrated dynamometer that requires them to exert a steady amount of pressure over a long period of time.	No	High	Low	Low	Requires some special equipment and calibration. Calibration means that variance in strength is accounted for and thus results are more easily interpreted as effort. Researchers have a high degree of control over the level of effort required from participants.
Clicking on a target	Houy et al. (2016).	Participants must click on the center of a target within 8 seconds while random perturbations move mouse pointer.	No	Low	Low	Low	Researcher has a high degree of control over amount of effort needed to succeed by changing the magnitude of the perturbations.
Repetitive task							
Sorting and counting coins	Bortolotti et al. (2009)	Participants must sort and count a number of coins worth 1, 2,5, and 10 Euro cents within a given time interval.	No*	Med.	Low	Med.	Possible some participants may have experience with task due to cashier experience.
Cracking walnuts	Fahr and Irlenbusch. (2000)	Participants are given a pile of walnuts and nutcracker and must produce some mass of cracked walnuts in a given time.	Yes	High	Low	Low	Not much researcher control over cost of effort, unless some participants get better tools than others?
Filling envelopes	Konow (2000)	Participants fold letters, stuff them into envelopes, and place them through a slot in a sealed box.	Yes	High	Low	Low	Can be difficult to find an appropriate reason to need envelopes stuffed. Task has been used many times with minor variations.
Sliders	Gill and Prowse (2011)	Participants are presented with "sliders" which they must click and drag to the center of a bar.	No	Low	Low	Low	Researcher has a high degree of control over the amount of effort.
Ball catching	Gächter et al. (2015)	Participants click a "left" or "right" to move a "tray" in order to catch balls on a screen that fall at fixed time intervals. The number of clicks and balls caught are recorded.	No	Low	Low	Low	Researcher has a high degree of control over the amount of effort and the cost of effort, as the cost per click can be easily manipulated.
Dragging a ball on a screen	Heyman and Ariely (2004)	Participants drag a ball across a screen, at which point it disappears and new one appears. Do as many as possible.	No	Low	Low	Low	Intuitively seems very frustrating, sine there is no discernable progress.

(*continued on next page*)

Table 2 (*continued*)

Data entry	First use	Description	Prod.	Diff. imp	S/A	Learn	Comments
Typing Alternative Keys	Swenson (1988)	Participants receive some amount of income per keystroke, typing alternatingly "!" and the return key.	No	Low	Low	Low	Task might require very little attention, so cost of effort could be low.
Repeatedly typing paragraph	Dickinson (1999)	Participants exactly type out the same paragraph over and over.	No	Low	Low	Low	Requires more attention than above. Multiple dimensions of effort, as errors can be measured.
Decoding							
Transcribing greek letters	Augenblick et al. (2015)	A row of random and blurry Greek letters appears on a screen; the participant replicates it by clicking on a list of "greek" letters.	No	Low	Low	Low	
Decoding character strings (Computer cards)	Chow (1983)	Participants are given a set of pre-punched computer cards and a decoding key that they use to translate the card punches to a character string.	No	Med.	Low	Low	
Encoding 3-letter words into numbers	Erkal et al. (2011)	Participants are given a table that codes unique numbers to each letter of alphabet, then is presented with a list of words and must convert the words into their numerical codes.	No	Low	Low	Low	
Decoding a number from a letter grid	Lévy-Garboua et al. (2009)	Participants are given a grid of letters and a decoding key, and they convert the letters into numbers.	No	Low	Low	Low	
Solving CAPTCHAs	McMahon (2015)	Participants solve as many CAPTCHAs (text distorted in a way so as to be unreadable to standard computerized text scanners) as possible within a given time period.	No	Low	Low	Low	
Cognitive							
Summing large matrices	Corgnet et al. (2011)	Participants are given 36 numbers in a matrix and must sum them. Notably, they did so for 100 minutes.in this experiment	No	Low	Med.	Low	Some participants may enjoy solving math problems, some might have math anxiety.
IQ test	Gneezy and Rustichini (2000)	Participants are presented with an IQ test and must provide correct answers.	No	Low	High	Low	Some participants may be intrinsically motivated to perform well.
Adding 2-digit numbers	Niederle and Vesterlund. (2007)	Participants add a series of 2-digit numbers in a given time period.	No	Low	Med.	Low	Some participants may enjoy solving math problems, some might have math anxiety.
Impossible math problem	Heyman and Ariely (2004)	Participants are given a grid of numbers and told they must select a group of numbers that add up to 100. However, task is impossible, as no combination of numbers does so. Effort is measured as the time spent on the task before giving up.	No	Low	Med.	High	
Miscellaneous							
Door-to-door fundraising	Gneezy and Rustichini (2000)	Participants go door-to-door collecting donations for charitable causes.	Yes	High	High	Low	Might be confounded by individual preferences for the task or skill at fundraising.
Numerical optimization (multi peaked)	Montmarquette et al. (2004)	Participants search for the highest value of a one or more peaked function displayed in a two-dimensional space by clicking a button repeatedly or continuously to uncover the space. Different buttons uncover the space at different rates.	No	Low	Low	Low	High control over the cost of effort by changing the costs attached to the buttons.

(*continued on next page*)

Table 2 (*continued*)

Data entry	First use	Description	Prod.	Diff. imp	S/A	Learn	Comments
Numerical optimization (single peaked)	Van Dijk et al. (2001)	Participants search for the highest value of a single-peaked function displayed in a two-dimensional space by clicking a button repeatedly or continuously to uncover the space. Different buttons uncover the space at different rates.	No	Low	Low	Low	Less effort than above.
Creativity							
Creating new product ideas	Girotra et al. (2010)	Participants are tasked with identifying new product concepts for a given market, e.g., sporting goods that might be sold to students.	No*	Low	Med.	Low	Easy to implement and explain. Several degrees of researcher freedom with regards to quantity vs. quantity of creative effort. Some participants may have skills such as marketing or business experience that could be confounds. Fairly realistic.
Creating words from letter sets (short)	Charness and Villeval (2009).	Participants are given a set of 7 letters and must form as many words as possible within a given time period	No	Low	Med.	Low	Easy to implement and explain. Easy to judge output. Some participants may enjoy the task, so less control over cost of effort. Not so realistic.
Expressing words with materials	Laske and Schröder (2016)	Participants given set of materials consisting of a string, 2 O-rings, 4 wooden sticks, and 12-colored glass pebbles, and must construct representations of words with the materials.	No	High	Med.	Low	
Writing a story	Charness and Grieco (2018)	Participants wrote a story about a future city, something they would like to invent, or using specified words.	No	Low	Med.	Low	Easy to implement. Quality is subjective.
Using combinations of math operations	Charness and Grieco (2018)	Participants were given a number and designated a series of math operations to transform it to another designated number.	No	Low	Low	Low	Easy to implement. Quality is subjective.
Designing a Rebus Puzzle	Kachelmeier et al. (2008)	Participants are asked to create a rebus puzzle made with words and/or pictures with a hidden and non-obvious solution.	No	Low	Med.	Low	
Designing a Rebus Puzzle	Erat and Gneezy (2016)	Participants are asked to create a rebus puzzle made with words and/or pictures with a hidden and non-obvious solution.	No	Low	Med.	Low	

* Asterisks indicate that the task is not necessarily productive, but could be modified in a straightforward manner to make the effort useful.

several criteria relevant to either practical research considerations or to the suitability of the task for addressing the specific research question.

On the far end of realistic effort-provision experiments are studies that directly look at performance in a specific area of interest, such as the impact of incentives on exercise (Charness and Gneezy, 2009) or the impact of incentives on effort provision to support children's education (Glewwe et al., 2010). Research that directly deals with complex behavioral patterns such as exercise habits or support for education may well provide more convincing conclusions about those specific behaviors than analogous behavior in a laboratory, but is costly to conduct and difficult to generalize to other behavior.

The difficulty of implementation is a question of whether the method requires specific materials or preparation on the part of the researcher. In Table 2 we term a task "Low" to indicates that it can be run through a computer or with minimal materials, "Med." (Medium) indicates that there are some special materials or preparation required, but the overall burden is otherwise not high, and "High" indicates the need for a significant investment in preparing for or conducting the experiment.

The productivity column indicates whether the task requires participant to do work that has outside value, such as cracking walnuts that can later be sold (Fahr and Irlenbusch, 2000) or entering presumably-useful research data (Gneezy and List, 2006; Dutcher et al., 2015; Charness et al., 2016). If the task can be modified in a straightforward manner to make the output genuinely useful, an asterisk is placed next to "No." For example, the sorting and counting of coins in

Bortolotti et al. (2009) could be the sorting of loose change from a business so that it could be deposited at a bank. Although some tasks not marked with an asterisk could conceivably be made productive, we only marked cases where it would be easiest.

A skill or ability confound means that participants would be likely to have greater variance in performance at a task, meaning that larger sample sizes are needed to capture treatment effects. This may also be a theoretical confound, as in some cases we might expect higher skill at performing a task (e.g., throwing a ball at a target, as in Ariely et al. 2009) to be correlated with higher enjoyment of the task, and thus less net cost of effort. In addition, there is evidence that tasks produce different emotional responses and effort provision may be affected by different emotions (Lezzi et al., 2015). Learning is also an issue for some tasks, as participants may improve at translating effort into productivity over time, again making the link between observable actions and theory tenuous.

Control over the cost-of-effort function, seen as one of the major advantages of the chosen-effort paradigm, has been addressed primarily through qualitative means, for example by juxtaposing results from "easy" and "hard" real-effort tasks, although some such as the ball-catching task (Gächter et al., 2015) add quantitative control as well.

4. Practical differences between stated effort and real effort

In this section we mention some dimensions on which stated effort and real effort differ in practical terms. To the extent that realism is an important characteristic (perhaps for external validity in labor settings), these considerations tend to favor real-effort designs.[1] Of course, one must consider the trade-off between the value of knowing the effort cost and the heterogeneity involved with actual task performance. We discuss the timing of decisions, planned actions versus actual behavior, and differences between time and money.

4.1. Timing of decisions

While stated effort is a one-time decision (even in a repeated game, it is so in each period), real effort is a dynamic process in which the participant may change their effort while performing the task. In the stated-effort paradigm, each participant typically makes one immediate decision when choosing an effort level. In real-effort experiments, which naturally occur over time, the impact of the treatment on effort exertion is not always the same throughout. Effort levels over a period of hours may vary in a way that can drastically change the conclusions. Effects that appear consistently in a short-term setting may or may not ultimately produce changes in effort levels in a setting with more duration. In this section we discuss several studies that look at effort exertion over time and suggest reasons for why effort levels might change over time, including learning, shifting emotional states, and limitations in sustained effort expenditure.

A plausible explanation of variation in effort levels over time, particularly the existence of sometimes-temporary shifts in effort, is related to hot-versus-cold decision-making (Loewenstein and Schkade, 1999; Loewenstein, 2005). For example, participants who have just received a gift may feel a transient "rush" of gratitude that impels them to reciprocate. Once this rush fades, so does the increased effort. For example, participants in Gneezy and List (2006) were recruited to perform data entry with an advertised hourly wage rate for six hours of work, split into two 3-hour sessions separated by a lunch break. Before beginning work, participants in one treatment were surprised with a higher-than-expected wage. The immediate response was an increase in effort relative to a control group that received the advertised wage. However, the effort waned after the lunch break, eventually reaching the same level for both treatments. Gneezy and List (2006) found similar results in a door-to-door fund raising experiment.

Several questions about this interpretation remain: how long do specific emotional states continue to influence behavior? Is the emotional effect present only in the short-term, as in Gneezy and Imas (2014) who show how anger can affect strategic behavior in the short run, but that this effect vanishes after a ten-minute cooling-off period? Yet it is possible that some emotional states are strong enough to push effort levels for long periods of time. For example, does negative reciprocity, which seems to produce stronger psychological effects than positive reciprocity (Offerman, 2002; Charness, 2004), produce a permanent shift in effort expenditure? Kube et al. (2013) extend the Gneezy and List (2006) design to include a negative wage surprise, finding a persistent and significant negative reciprocity as measured by the decrease in effort. The absolute magnitude of the decrease in effort relative to the control treatment was twice as large as the highest gap between the positive and control treatments.

Hennig-Schmidt et al. (2010) investigate the fair wage-effort hypothesis over time, as well as in a more typical short-term laboratory setting using a real-effort task. In the longer setting, participants performed a data-entry task in two discrete one-hour sessions separated by a month, with an expected show-up fee and hourly wage. When participants arrived for the second session, they were either paid the expected wage or given a pay increase of 10% or 40%. When participants believed they were providing a surplus to the employer, a wage increase significantly increased effort, and when they did not, a wage increase had no effect.

Kosfeld and Neckermann (2011) examine the impact of a non-monetary reward on effort in a data-entry setting wherein participants received a fixed wage for two hours of work. The possibility of a non-monetary reward given to the best per-

[1] There are some topics, such as creativity, in which it is precisely the behaviorally-interesting particularities around the topic that a stated-effort design would miss, and so it seems to us that using a real-effort design is necessary.

former, in this case a signed card from the director of the organization benefitting from the data-entry task, led to sustained levels of higher effort, a result driven by a small number of highly-productive workers who stood a reasonable shot of winning the prize.

The conclusion from this section is clear: researchers should consider whether time is relevant to the research question. Longer-term experiments can help in capturing aspects of the decision making process. If the duration of the behavioral response to a stimulus is relevant to the theoretical importance of a phenomenon, then the experimental methodology should reflect this. Emotional states are particularly likely to change over specific events, such as eating a meal or sleeping. Such events can "reset" emotional states, rapidly accelerating the pace at which the impact of an emotional state on actions decreases. Further research in this domain might look at effort levels over increasingly long durations, potentially through field experiments in real workplaces where participant behavior can be measured over time.

Similarly, further research is also needed on the persistence of stated-effort decisions over time. An experiment could perhaps test whether there is any difference in stated-effort decisions over a time span in which there has been found to be significant changes in real-effort provision.

4.2. Planned actions versus actual behavior

Even when individuals have a strong intention to meet goals they have decided to pursue, they may fail to do so because they do not effectively deal with self-regulatory problems – goal striving may not be enough by itself (see Gollwitzer and Sheeran, 2006 for a review). Individuals allocate their effort by planning future effort (e.g., setting a goal) and by exerting current effort. Effort planning involves scheduling future behavior, whether in informal circumstances such as planning a gym routine or a study session, or in more formal circumstances such as creating a work schedule or negotiating a contract. A plan to allocate effort in the future is purposeful; for example, a gym routine is planned because the expected effort cost from going to the gym is outweighed by the expected benefits of improved health.

Buser and Peter (2012) find that people have problems with scheduling when required to perform multiple tasks (Sudoku and Word Search). In three treatments, participants either are required to perform these tasks sequentially, required to multi-task, or they can organize and schedule the work as desired. People who were required to multi-task perform significantly worse than those who were required to work sequentially. It is interesting that participants who were allowed to create their own schedule also perform significantly worse, suggesting that scheduling is an important aspect of productivity. A final result goes against the stereotype that females are better at multi-tasking than males, since their performance is reduced by just as much as men when required to multi-task and are even *less* likely to multi-task when free to choose.[2]

When individuals consider exerting current effort, pressures outside of goal-seeking may also be in force. For example, the unpleasantness of actually exercising in a gym may discourage one from following through with his or her plan. The goal-oriented valuation that drove initial goal-setting competes with more myopic valuation systems for influence over behavior at the time of action.

The economics literature in domains such as savings and healthy behavior describes the difficulties people experience with following through on plans. Individuals state a preference for saving or exercising, but often fail to follow through. An outgrowth of these observations is the development of commitment devices and other behavioral tools or strategies, including social incentives, which align an individual's future incentives with their current incentives (Thaler and Benartzi, 2004; Ashraf, et al., 2006; Kast et al., 2012).

When designing an experiment, it is appropriate to consider the degree of difficulty individuals might have in following up on their planned behavior. Stated effort may measure the desire to attain a goal, such as winning a tournament (e.g. Müller and Schotter, 2010). However, stated effort may fail to be predictive of actual effort. When using existing research to make predictions about external phenomena, the interpretation of results from either stated-effort or real-effort should be carefully considered in the light of whether desire is likely to translate into action.

4.3. Differences between effort and money

Individuals may not always behave similarly when making decisions over money and effort. We consider three empirical patterns here for their relevance to selecting a methodology or interpreting results from effort experiments: individuals can exhibit a preference for donating effort rather than money in charitable giving; exhibit differently-shaped time preferences over money and effort; and money can crowd out motivation from other sources and change the nature of a social interaction. As the stated-effort task is fundamentally a decision over money, a concern is that some divergences might exist between results obtained from stated-effort and real-effort tasks in these domains. Developing an encompassing theory to explain why decisions over money and effort may not always be equivalent is beyond the scope of this paper. We limit ourselves to simply presenting these patterns along with examples from oft-studied domains. This should help researchers become aware of these behavioral differences and hopefully lead to useful formal models.

[2] On the (emerging) topic of multi-tasking, Offerman and van der Veen (2015) create a dual-task environment in which one task involved making public-good contributions and the other involved keeping a randomly-moving red dot inside a box on the screen. They consider how people react to either a slow or quick increase of a subsidy for contributions to the public good. With the dual task, people seem to fail to react to a series of small changes in the decision problem.

That individuals can exhibit a preference for effort exertion over monetary donation is sometimes referred to as the "volunteering puzzle" (Handy and Katz, 2008).[3] Consider a lawyer who volunteers in a soup kitchen. If the lawyer's goal is to maximize the amount of food served at the soup kitchen, then spending an hour working at her occupation and then donating the wages to the soup kitchen is much more effective than working at the kitchen. The donated money could be used to employ several lower-skilled workers in her place.

A potential driver of donations of effort or time rather than money may be differences in the warm-glow (Andreoni 1990) attained from donating. Individuals may derive utility from effort for reasons of social signaling, self-image, or the pleasure of performing the task itself. Andreoni et al. (1996) present a model wherein individuals derive utility from donations that is separable over money and effort, which has been supported by subsequent experimental evidence. Brown et al. (2013) find that participants in a laboratory experiment are more likely to donate and donate more when they can work directly for a charity rather than work for themselves and later donate to the charity. Even when participants in this study could freely toggle between working for themselves and working for charity and wages for self were 33% higher, they still give substantially more time to charity. It seems that donations of effort are more motivated by private warm-glow than monetary donations.

Comparing time preferences over money and effort, Augenblick et al. (2015) measure the shape of time preferences over money and consumption, operationalized as a period of time that must be spent working on a boring task. They find no evidence for present bias in money, but do find evidence for present bias in consumption in their two experiments. Additionally, participants exhibited a demand for commitment devices for effort, but not money. Money appears to be fungible between time periods, while effort does not. Bisin and Hyndman (2014) find present-bias over real effort in a field experiment in which students must complete tasks by a fixed deadline, and further find that demand for a self-imposed commitment device is stronger in students who describe themselves as less conscientious, indicating that they are sophisticated about their time-preferences.

Money and effort are not always interchangeable in social interactions. Introducing monetary exchanges into a social interaction can change the character of the social interaction, potentially crowding out other incentives. Consider asking your friend to come over and help you move your sofa to your new home. Paying your friend $20 at the end would seem odd, but telling him that you will be happy to help him whenever he will need help or buying him dinner would not. Gneezy and Rustichini (2000) and Heyman and Ariely (2004) find that low levels of monetary compensation can produce less effort than no monetary compensation. If one wishes to study a social interaction that is often denominated in terms of effort, using money as the currency of exchange may crowd out key factors relevant to decision making.

When designing an experiment, researchers should consider whether in the domain they study effort and money are interchangeable. As there does not yet appear to be data that describes the degree of interchangeability in many domains, further research that facilitates such comparisons would helpful for making inferences based on laboratory experiments and deciding which methodology is most appropriate.

5. Comparative studies

We found only few empirical investigations that directly compare results with parallel methodologies—both stated-effort and real-effort that are applied to the same treatment effect or decision-making environment. The treatment effect or environment itself is not chosen specifically as a test of comparability. Three studies Brüggen and Strobel (2007), Charness et al. (2016) and Dutcher et al. (2015) find general equivalence between the methodologies in the environment tested, and one study Lezzi et al. (2015) finds significant differences between the results obtained from the stated-effort task and several real-effort tasks.

Brüggen and Strobel (2007) used a gift-exchange game, with participants responding to a monetary transfer by either solving as many math problems as possible in five minutes or by selecting an effort level. There was evidence of positive reciprocation in both treatments, with higher average earnings and greater variance in the real-effort treatment. Charness et al. (2016) investigated the role of social comparisons (both for wages and wage-decision rights) on workers' performance. The main treatments involved stated effort, but an additional treatment featured an adding-numbers task. They find qualitatively similar results from both paradigms, with quantitatively similar earnings.

Dutcher et al. (2015) use a repeated public-goods setup with three treatments ("useful effort," "trivial effort," and "stated effort"). Participants were matched into groups of four for multiple periods; in each period they could either contribute money to the group fund (earning $0.40 for the group per unit) or keep the currency in an individual fund (earning $0.20 per unit). In the "useful effort" treatment it was made clear that the data entry contributed to a research project, whereas in the "trivial effort" treatment subjects were not given any context for the task. The maximum amount of data entry per period was capped at 10 lines, and in each period of the stated-effort treatment participants were given 10 tokens, allowing participants in all treatments to have access to a comparable number of tokens. There was no difference across treatments for either average contributions or trends in contributions.

[3] In 2015 there were over 7.9 billion volunteer hours provided by 62.6 million volunteers in the United States for an estimated value of $184 billion. (Corporation for National and Community Service, 2015), and the estimated value of monetary charitable donations was over $358 billion in 2014. (Giving USA, 2015).

Gächter et al. (2015) used a computerized ball-catching task, in which participants move a slider at some cost to catch balls dropping randomly from the top of a screen. The argument is that performing an activity, even one that requires almost no physical or mental effort, captures the relevant aspects of a real-effort task. They use this task to study team production, gift exchange, and effort tournament, and obtain results in line with stylized findings from previous studies which use stated effort.

Lezzi et al. (2015) directly compare the relationship between effort exertion and anxiety, risk preferences, and gender across the slider task (Gill and Prowse, 2011), adding numbers (Niederle and Vesterlund, 2007), counting zeros (Abeler et al., 2011), and the stated-effort task. Participants in each task competed in a two-person 10-round all-pay format tournament wherein the participant who exerted the highest effort in each round won the round. Men exerted higher effort than women in the slider task but not the other tasks, anxiety decreased performance in the counting-zeros task but not others, and risk aversion was positively associated with performance on the counting zeros task but not others. The authors conclude that the task specificity of their results indicates that researchers should be careful when generalizing their experimental findings. Conducting such experiments would provide valuable knowledge on the situations in which specific laboratory effort experiments make useful predictions according to the methods employed.

6. Conclusion

Designing experiments that test real-effort and stated-effort on some of the dimensions identified in this paper stand to help provide a stronger empirical basis for differentiating the situations where each methodology may be appropriate. An extension of this is to study whether different treatment effects observed in a laboratory context using stated- and real-effort map to equivalent differences in more realistic settings. That is, if there is a scenario in which it is found that stated-effort and real-effort methodologies produce different results, then testing which results more closely align with the field phenomena of interest would provide careful consideration of the settings to which we can expect stated effort or real effort to generalize.

Stated effort is quite useful in a variety of situations, particularly when one is interested in a task that can be done quickly and immediately. Knowing the cost of effort is critical in many cases, particularly when one is making social comparisons (usually regarding payoffs) or testing theory. But when the field setting involves sustained effort, as in most labor environments in the field, an experimental task involving real effort seems advisable for external validity. Since initial behavior may be driven by an early rush of emotion that fades quickly or since emotional or physical fatigue may well manifest over time, one must be careful when drawing conclusions from quick monetary choices. While to some degree having multiple periods can simulate periods of work, this would not seem to have the same psychological feel.

While it is desirable to have tasks with duration and real effort in an environment where effort must be supplied for a period of time, it may not be clear *ex ante* how long the duration must be with real effort in the lab. If Gneezy and List (2006) had only tested behavior in the 3-hour morning session, they would have concluded that a high-wage surprise leads to higher real effort. Having a second 3-hour session after a lunch break was crucial for the interpretation that positive feelings about this higher wage fade over time, perhaps as the sense of surprise fades and one's sense of entitlement grows.

A big question for experimenters is whether one gets different results with real effort and stated effort. Direct comparisons between results with stated and real effort are still scarce to date, but in several cases the effects are qualitatively similar. The relationship between effort or production and the other party's earnings is clear in these cases and this seems necessary for this equivalence. In fact, Hennig-Schmidt et al. (2010) find that "explicit cost and surplus information that enables an exact calculation of an employer's surplus from the work contract is a crucial prerequisite for a positive wage-effort relation." Ideally, one would like a real-effort task where there is not much variance in production across ability, so that there is a reasonably strong connection between effort and observed production.

The methodology used to measure effort in the laboratory should be appropriate to the specific research question under consideration. We have identified several considerations researchers should make to pick a methodology that best suits their needs, including the timing of the effort decision, the duration of the phenomenon, and goal orientation. Further, we provide a review of real-effort tasks along with qualitative assessments of methodological and logistical attributes.

Perhaps the main added value of the stated-effort approach is allowing the researcher to connect the experimental results to theory. A researcher who wishes to learn whether behavior in an experiment is consistent with comparative-statics predictions of an equilibrium theory must know the function that maps the costs of effort to production. On the other hand, the main added value of the real-effort approach is the better connection to the psychology of effort, since one must be cautious in interpreting levels of behavior with stated effort.

It would be nice to have more papers with realistic real-effort tasks. At the same time, more papers that test interesting theories with the stated-effort approach would also be quite welcome. Furthermore, even unrealistic real-effort experiments would be useful to the extent that they help us identify interesting psychological mechanisms. Our goal is not to suggest that one methodology is superior, since both approaches clearly have their merits. To be clear, we have used both forms of effort elicitation techniques in our own research (and we have different views on the relative merits). Rather, our goal was to highlight some of the relevant parameters that researcher should consider when designing their method for measuring effort.

References

Abeler, J., Falk, A., Goette, L., Huffman, D., 2011. Reference points and effort provision. Am. Econ. Rev. 101 (2), 470–492.
Andreoni, J., 1990. Impure altruism and donations to public goods: a theory of warm-glow giving. Econ. J. 100 (401), 464–477.
Andreoni, J., et al., 1996. Charitable Contributions of Time and Money. University of Wisconsin-Madison, Working Paper.
Ariely, D., et al., 2009. Large stakes and big mistakes. Rev. Econ. Stud. 76 (2), 451–469.
Ashraf, N., Karlan, D., Yin, W., 2006. Tying Odysseus to the mast: evidence from a commitment savings product in the Philippines. Q. J. Econ. 121 (2), 635–672.
Augenblick, N., Niederle, M., Sprenger, C., 2015. Working over time: dynamic inconsistency in real effort tasks. Q. J. Econ. 130 (3), 1067–1115.
Baumeister, R.F., 1984. Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. J. Personality Social Psychol. 46 (3), 610.
Bisin, A., Hyndman, K., 2014. Present-Bias, Procrastination and Deadlines in a Field Experiment. National Bureau of Economic Research, Working paper No.w19874.
Bortolotti, S., Devetag, G., Ortmann, A., 2009. Exploring the Effects of Real Effort in a Weak-Link Experiment. University of Trento, Working Paper No. 0901.
Brandts, J., Charness, G., Ellman, M., 2016. Let's talk: how communication affects contract design. J. Eur. Econ. Assoc. 14 (4), 943–974.
Brandts, J., Cooper, D.J., 2007. It's what you say, not what you pay: an experimental study of manager-employee relationships in overcoming coordination failure. J. Eur. Econ. Assoc. 5 (6), 1223–1268.
Brown, M., Falk, A., Fehr, E., 2004. Relational contracts and the nature of market interactions. Econometrica 72 (3), 747–780.
Brown, A.L., Meer, J., Forrest Williams, J., 2013. Why Do People Volunteer? An Experimental Analysis of Preferences for Time Donations. National Bureau of Economic Research, No. w19066.
Brüggen, A., Strobel, M., 2007. Real effort versus chosen effort in experiments. Econ. Lett. 96 (2), 232–236.
Bull, C., Schotter, A., Weigelt, K., 1987. Tournaments and piece rates: an experimental study. J. Political Economy 95 (1), 1–33.
Buser, T., Peter, N., 2012. Multitasking. Exp. Econ. 15 (4), 641–655.
Bushong, B., Gagnon-Bartsch, T., 2016. Misattribution of Reference Dependence: Evidence from Real-Effort Experiments. Harvard University, Working Paper.
Charness, G., 2000. Responsibility and effort in an experimental labor market. J. Econ. Behav. Organiz. 42 (3), 375–384.
Charness, G., 2004. Attribution and reciprocity in an experimental labor market. J. Lab. Econ. 22 (3), 665–688.
Charness, G., Dufwenberg, M., 2006. Promises and partnership. Econometrica 74 (6), 1579–1601.
Charness, G., Grieco, D., 2018. Creativity and financial incentives. J. Eur. Econ. Aff. Forthcoming.
Charness, G., Cobo-Reyes, R., Lacomba, J., Lagos, F., Perez, J.M., 2016a. Social comparisons in wage delegation: experimental evidence. Exp. Econ. 19 (2), 433–459.
Charness, G., Gneezy, U., 2009. Incentives to exercise. Econometrica 77 (3), 909–931.
Charness, G., Kuhn, P., 2011. Lab labor: what can labor economists learn from the lab?. In: Handbook of Labor Economics, 4, pp. 229–330.
Charness, G., Cobo-Reyes, R., Sánchez, Á., 2016b. The effect of charitable giving on workers' performance: experimental evidence. J. Econ. Behav. Organiz. 131 (PA), 61–74.
Charness, G., Villeval, M.C., 2009. Cooperation and competition in intergenerational experiments in the field and the laboratory. Am. Econ. Rev. 99 (3), 956–978.
Chaudhuri, A., 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Exp. Econ. 14 (1), 47–83.
Chaudhuri, A., Schotter, A., Sopher, B., 2009. Talking ourselves to efficiency: coordination in inter-generational minimum effort games with private, almost common and common knowledge of advice. Econ. J. 119 (534), 91–122.
Chow, C.W., 1983. The effects of job standard tightness and compensation scheme on performance: an exploration. Account. Rev. 58 (4), 667–685.
Corgnet, B., Hernan-Gonzalez, R., Rassenti, S., 2011. Real Effort, Real Leisure and Real-Time Supervision: Incentives and Peer Pressure in Virtual Organizations, Working Paper, No.11-05.
Corporation for National & Community Service. Trends and highlights overview. https://www.nationalservice.gov/vcla/national.
Devetag, G., Ortmann, A., 2007. When and why? A critical survey on coordination failure in the laboratory. Exp. Econ. 10 (3), 331–344.
Dickinson, D.L., 1999. An experimental examination of labor supply and work intensities. J. Labor Econ. 17 (4), 638–670.
Dutcher, G., Salmon, T., Saral, K.J., 2015. Is "Real" Effort More Real? Working Paper.
Erkal, N., Gangadharan, L., Nikiforakis, N., 2011. Relative earnings and giving in a real-effort experiment. Am. Econ. Rev. 101 (7), 3330–3348.
Erat, S., Gneezy, U., 2016. Incentives for creativity. Exp. Econ. 19 (2), 269–280.
Fahr, R., Irlenbusch, B., 2000. Fairness as a constraint on trust in reciprocity: earned property rights in a reciprocal exchange experiment. Econ. Lett. 66 (3), 275–282.
Fehr, E., Kirchsteiger, G., Riedl, A., 1993. Does fairness prevent market clearing? An experimental investigation. Q. J. Econ. 108 (2), 437–459.
Fehr, E., Simon, G., Kirchsteiger, G., 1997. Reciprocity as a contract enforcement device: experimental evidence. Econometrica 65 (4), 833–860.
Gächter, S., Huang, L., Sefton, M., 2015. Combining "real effort" with induced effort costs: the ball-catching task. Exp. Econ. 19 (4), 687–712.
Gill, D., Prowse, V.L., 2011. A Novel Computerized Real Effort Task Based on Sliders. IZA Discussion Paper, No.5801.
Girotra, K., Terwiesch, C., Ulrich, K.T., 2010. Idea generation and the quality of the best idea. Manage. Sci. 56 (4), 591–605.
Giving USA (2015). Giving USA 2015: the annual report on philanthropy for the year 2014.
Glewwe, P., Ilias, N., Kremer, M., 2010. Teacher incentives. Am. Econ. J. 2 (3), 205–227.
Gneezy, U., Rustichini, A., 2004. Gender and competition at a young age. Am. Econ. Rev. 94 (2), 377–381.
Gneezy, U., Rustichini, A., 2000. Pay enough or don't pay at all. Q. J. Econ. 115 (3), 791–810.
Gneezy, U., List, J.A., 2006. Putting behavioral economics to work: testing for gift exchange in labor markets using field experiments. Econometrica 74 (5), 1365–1384.
Gneezy, U., Imas, A., 2014. Materazzi effect and the strategic use of anger in competitive interactions. Proc. Natl. Acad. Sci. 111 (4), 1334–1337.
Gneezy, U., Niederle, M., Rustichini, A., 2003. Performance in competitive environments: gender differences. Q. J. Econ. 118 (3), 1049–1074.
Gollwitzer, P.M., Sheeran, P., 2006. Implementation intentions and goal achievement: a meta-analysis of effects and processes. Adv. Exp. Social Psychol. 38, 69–119.
Gross, T., Guo, C., Charness, G., 2015. Merit pay and wage compression with productivity differences and uncertainty. J. Econ. Behav. Organiz. 117 (C), 233–247.
Handy, F., Katz, E., 2008. Donating behavior: if time is money, which to give? A preliminary analysis. J. Econ. Stud. 35 (4), 323–332.
Hennig-Schmidt, H., Sadrieh, A., Rockenbach, B., 2010. In search of workers' real effort reciprocity—a field and a laboratory experiment. J. Eur. Econ. Assoc. 8 (4), 817–837.
Heyman, J., Ariely, D., 2004. Effort for payment a tale of two markets. Psychol. Sci. 15 (11), 787–793.
Houy, N., Nicolai, J.P., Villeval, M.C., 2016. Doing Your Best when Stakes are High? Theory and Experimental Evidence. IZA Discussion Paper, No.9766.
Imas, A., 2014. Working for the "warm glow": on the benefits and limits of prosocial incentives. J. Public Econ. 114 (C), 14–18.
Johnson, N.D., Mislin, A.A., 2011. Trust games: a meta-analysis. J. Econ. Psychol. 32 (5), 865–889.
Kachelmeier, S.J., Reichert, B.E., Williamson, M.G., 2008. Measuring and Motivating quantity, creativity, or both. J. Account. Res. 46 (2), 341–373.
Kast, F., Meier, S., Pomeranz, D., 2012. Under-Savers Anonymous: Evidence on Self-Help Groups and Peer Pressure as a Savings Commitment Device. National Bureau of Economic Research, Working paper, No.w18417.
Konow, J., 2000. Fair shares: accountability and cognitive dissonance in allocation decisions. Am. Econ. Rev. 90 (4), 1072–1091.
Kosfeld, M., Neckermann, S., 2011. Getting more work for nothing? Symbolic awards and worker performance. Am. Econ. J. 3 (3), 86–99.

Kube, S., Maréchal, M.A., Puppe, C., 2013. Do wage cuts damage work morale? Evidence from a natural field experiment. J. Eur. Econ. Assoc. 11 (4), 853–870.
Laske, K., Schröder, M., 2016. Quantity, Quality, and Originality: The Effects of Incentives on Creativity. CGS 7-1,Working Paper.
Lévy-Garboua, L., Masclet, D., Montmarquette, C., 2009. A behavioral Laffer curve: emergence of a social norm of fairness in a real effort experiment. J. Econ. Psychol. 30 (2), 147–161.
Lezzi, E., Fleming, P., Zizzo, D.J., 2015. Does it Matter Which Effort Task You Use? A Comparison of Four Effort Tasks When Agents Compete for a Prize, Working Paper.
Loewenstein, G., 2005. Hot-cold empathy gaps and medical decision making. Health Psychol. 24 (4S), S49.
Loewenstein, G., Schkade, D., 1999. Wouldn't it be nice? Predicting future feelings. Well-being Found. Hedonic Psychol. 85–105.
McMahon, M., 2015. Better Lucky Than Good: The Role of Information in Other-Regarding Preferences, Working Paper.
Mohnen, A., Pokorny, K., Sliwka, D., 2008. Transparency, inequity aversion, and the dynamics of peer pressure in teams: theory and evidence. J. Labor Econ. 26 (4), 693–720.
Moldovanu, B., Sela, A., 2001. The optimal allocation of prizes in contests. Am. Econ. Rev. 91 (3), 542–558.
Montmarquette, C., et al., 2004. Redesigning teams and incentives in a merger: an experiment with managers and students. Manage. Sci. 50 (10), 1379–1389.
Müller, W., Schotter, A., 2010. Workaholics and dropouts in organizations. J. Eur. Econ. Assoc. 8 (4), 717–743.
Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much? Q. J. Econ. 122 (3), 1067–1101.
Offerman, T., 2002. Hurting hurts more than helping helps. Eur. Econ. Rev. 46 (8), 1423–1437.
Offerman, T., van der Veen, A., 2015. How to subsidize contributions to public goods: does the frog jump out of the boiling water. Eur. Econ. Rev. 74 (C), 96–108.
Rutström, E.E., Williams, M.B., 2000. Entitlements and fairness: an experimental study of distributive preferences. J. Econ. Behav. Organiz. 43 (1), 75–89.
Smith, V.L., 1976. Experimental economics: Induced value theory. Am. Econ. Rev. 66 (2), 274–279.
Swenson, C.W., 1988. Taxpayer behavior in response to taxation: an experimental analysis. J. Account. Public Policy 7 (1), 1–28.
Thaler, R.H., Benartzi, S., 2004. Save more tomorrow™: using behavioral economics to increase employee saving. J. Political Economy 112 (S1), S164–S187.
Van Dijk, F., Sonnemans, J., Van Winden, F., 2001. Incentive systems in a real effort experiment. Eur. Econ. Rev. 45 (2), 187–214.
Van Huyck, J., Battalio, R., Beil, R., 1990. Tacit coordination games, strategic uncertainty, and coordination failure. Am. Econ. Rev. 80 (1), 234–248.