CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The University of Nottingham

Discussion Paper No. 2009-15

Robin P. Cubitt,
Michalis Drouvelis,
Simon Gächter
and Ruslan Kabalin

August 2009

Moral Judgments in Social Dilemmas: How Bad is Free Riding?

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/economics/cedex/ for more information about the Centre or contact

Karina Terry
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0) 115 95 15620
Fax: +44 (0) 115 95 14159
karina.terry@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/economics/cedex/papers/index.html

Moral Judgments in Social Dilemmas: How Bad is Free Riding?

Robin P. Cubitt, University of Nottingham*

Michalis Drouvelis, University of York

Simon Gächter, University of Nottingham, CESifo and IZA

Ruslan Kabalin, University of Nottingham

14th August 2009

In the last thirty years economists and other social scientists investigated people's normative views on principles of distributive justice. Here we study people's normative views in social dilemmas, which underlie many situations of economic and social significance. Using insights from moral philosophy and psychology we provide an analysis of the morality of free riding. We use experimental survey methods to investigate people's moral judgments empirically. We vary others' contributions, the framing ("give-some" vs. "take-some") and whether contributions are simultaneous or sequential. We find that moral judgments depend strongly on others' behaviour; and that failing to give is condemned more strongly than withdrawing all support.

Keywords: moral judgments, framing effects, public goods experiments, free riding.

* Corresponding author. School of Economics, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, United Kingdom. Email: robin.cubitt@nottingham.ac.uk.

1. Introduction

A striking recent development in public economics has been its increasing use of data on people's normative attitudes, obtained with surveys and questionnaires. For example, views about distributive justice and the demand for redistribution have recently been examined by Fong (2001), Gaertner, et al. (2001), Corneo and Grüner (2002), Faravelli (2007), Gaertner and Schwettmann (2007), and Corneo and Fong (2008).[1]

In this paper, we extend the empirical investigation of people's normative views to a different economic context, namely social dilemma (public goods) games, and to a different type of normative view, namely moral judgment.[2] More specifically, we report an experiment that, using techniques adapted from moral psychology, explores how individuals judge the morality of a free rider in a social dilemma game.

A social dilemma arises when members of a group share the benefits of a common resource but each of them has to decide individually whether, or how much, to contribute to its provision. Contribution is costly to the contributor but helpful to all other group members. Thus, a social dilemma isolates a conflict between personal interest, which militates for free riding, and the collective interest, which requires contribution. The frequent occurrence of social dilemmas in economic and social life makes them important for economics and social science; and the conflict of interests that they embody also makes them potentially fruitful for empirical analysis of moral judgments. In fact, there are arguments to the effect that the conception of morality itself evolved in response to cooperation problems our ancestors had to deal with (e.g., Ridley (1996); Binmore (2005); Hauser (2006); Gintis, et al. (2008), Krebs (2008)).

Previous research has shown that people experience negative emotions towards free riders and that some are willing to incur costs to punish self-interested behaviour (e.g., Fehr and Gächter (2000); Fehr and Gächter (2002); Fehr and Fischbacher (2004); Cubitt, et al. (2008); Gächter and Herrmann (2009)). However, little is known

[1] For earlier contributions, see Schokkaert and Lagrou (1983), Yaari and Bar-Hillel (1984), and Gaertner (1994). See Konow (2003) and Gaertner (2009) for overviews.

[2] Moral judgments can be "defined as evaluations (good vs. bad) of the actions or character of a person that are made with respect to a set of virtues held to be obligatory by a culture or subculture" (Haidt (2001), p. 817).

about people's moral judgment (i.e., their purely normative views) of those who follow their own self-interest in social dilemmas. Although it seems that many people dislike free riders when directly affected by them, it does not follow that free riding is perceived to be *morally* reprehensible. Croson and Konow (2009) provide evidence of a difference between normative judgments reached from the standpoints of "stakeholder" and impartial observer in the context of dictator games; and the same could be true in social dilemmas. Thus, we ask: when judgment is not confounded with self-interest from being an affected party, is free riding still judged to be wrong? And, if so, what determines how severe a transgression it is seen as? And, how is the judgment formed?

In our study, subjects (n=538) responded to a questionnaire in which they were confronted with hypothetical scenarios involving a two-player public goods game. In various endings of these scenarios, one player always free rides; while the other contributes different amounts to the common resource in different scenarios For each scenario separately, subjects were asked to express their positive or negative moral rating towards the free rider without themselves being involved in the decision situation. Thus, as they were merely observers, their judgments should represent impartial moral evaluations of the free rider.

Our experimental design manipulates three aspects of the scenarios. First, as already noted, we manipulate the behaviour of the non-judged player, to see whether subjects' moral judgments of the free rider depend on this. Our second manipulation investigates how moral judgments depend on the order of moves of the players in the scenarios. In particular, we explore whether the sensitivity of judgments of the free rider to the behaviour of the non-judged player is affected by whether the scenario specifies that the free rider knew the other player's behaviour before choosing his own. Third, we explore whether moral judgments are sensitive to contextual cues provided by the framing of the decision problem. This is interesting because previous evidence shows that behaviour can be strongly influenced by such cues (see Tversky and Kahneman (1981); Levin, et al. (1998)). The framing manipulation we look at has a Give (contributing to the common resource) versus Take (withdrawing from the common resource) form. This manipulation is common in previous studies of social

dilemma games in psychology and economics,[3] but its impact on moral judgments reached in that context has not been studied, to our knowledge.

Our findings demonstrate that free riding is always perceived as a morally blameworthy action, except for one case in which it is seen as morally praiseworthy. The exceptional case is the one, which we will call "ratting on a rat", in which the judged free rider moves second, after observing that the other player has free ridden too. We provide evidence that, irrespective of whether moves are simultaneous or sequential, the higher is the other player's contribution, the more negative is the moral rating assigned to the free rider on average. Interestingly, this pattern of judgments is also observed at an individual level for a substantial minority of subjects in the simultaneous case, as well as for an overwhelming majority in the sequential case. Finally, we also find a strong framing effect in moral evaluations: other things equal, subjects condemn withdrawing support from the public good less than the corresponding equivalent action of failing to contribute to it.

We see these findings as a contribution not just to economics but also to the emerging literature in moral psychology and empirical moral philosophy (see Haidt (2001), Nichols (2004), Haidt (2007), Nado, et al. (2009) for recent syntheses). This literature has investigated how people arrive at moral judgments in a number of contexts. By extending this literature to cover free riding in social dilemmas, we make a contribution that is both conceptual and empirical. We give an analysis of the social dilemma problem from different ethical perspectives; and we argue that our findings shed useful light on two major accounts of how people form moral judgments that feature in the literature: the reason-based model and the emotion-based model.

Our paper is organised as follows. The next section describes the design, the motivation, and the procedures for our experiment. As part of this, it analyses different perspectives on the morality of the free rider in our scenarios; and introduces the two models of how moral judgments are formed. Section 3 describes and discusses the main results. Section 4 concludes.

[3] See, e.g. the one-shot experiments by Brewer and Kramer (1986); McDaniel and Sistrunk (1991); McCusker and Carnevale (1995); Sell and Son (1997); van Dijk and Wilke (2000); Rege and Telle (2004); Dufwenberg, et al. (2006), and the repeatedly-played public goods experiments by Andreoni (1995); Sonnemans, et al. (1998); Park (2000).

2. Design, Hypotheses and Procedures

2.1 Design

In our experiment, each subject responded to a questionnaire requiring her to report her moral judgment of a player in hypothetical scenarios. There were four treatments, each defined by a different questionnaire. Each subject responded to the questionnaire for one treatment only. Before explaining the differences between them, we first explain the points which the questionnaires had in common.

Each questionnaire described a decision problem for two fictitious players, named Person A and Person B; and then gave some possible endings, each of which specified players' choices and their consequences. A *scenario* comprises a description of a decision problem and an ending. Each questionnaire consisted of five scenarios with the same decision problem, but different endings.

In all scenarios, the players were the two members of a group playing a voluntary contributions game. Within each questionnaire, the behaviour of Person A varied across scenarios, but Person B was always a complete free-rider. After each ending, the subject was asked, as a detached observer, to rate the morality of Person B on a scale ranging from -50 (extremely bad) to +50 (extremely good).[4] Thus, in each treatment, we can test within-subjects for the impact of the behaviour of the non-judged player on the moral rating assigned to the free rider. All other tests are between-subjects and involve comparisons of subjects' responses across treatments.

There were two treatment variables: the framing used to describe the decision problem; and the order of moves in that problem. Each variable had two possible values: "Give" and "Take" for framing; and "Simultaneous" and "Sequential" for order of moves. Each was manipulated independently, yielding four treatments: Give-Simultaneous, Take-Simultaneous, Give-Sequential, and Take-Sequential.

To explain the Give versus Take manipulation, we fix for simplicity on the Simultaneous order of moves. In the Give frame, the decision facing each player was how much to contribute to a group project. The description of the decision problem and the first ending for the Give-Simultaneous treatment were as follows:

[4] Ratings were selected by using a mouse to move a slider on a computer screen. The slider was initially positioned at a rating of zero. Subjects had to click on the slider in order to activate it and were not allowed to proceed unless they had done so. This procedure was intended to prevent subjects from reporting a judgment of zero accidentally, while allowing them to do so after reflection.

Imagine a group that consists of two group members, Person A and Person B. Each group member receives an endowment of 20 tokens and has to decide how many tokens to keep for himself and how many to contribute to a group project. Each token he keeps for himself has a value of one pound for him. Each token contributed to the group project has a value of 1.50 pounds to the project. The total value of the project is divided equally between the two group members. So, each token contributed to the project earns both group members 0.75 pounds each. The total income of a group member is the sum earned from tokens kept for himself and his share of the earnings of the group project. Each group member decides simultaneously, that is, without knowing what the other one has done.

A) Assume that Person A contributes 0 tokens to the group project and Person B contributes 0 tokens to the group project. Therefore, the value of the group project is 0 pounds and, thus, as a result of their contributions, Person A's total income is 20 pounds and Person B's total income is 20 pounds. How do you rate ***Person B's*** *morality?*

As noted above, the five scenarios in the Give-Simultaneous questionnaire differed from each other only in respect of Person A's behaviour. Person A's contribution was 0 tokens (as shown) in the first scenario, rising to 20 tokens in increments of 5 over the other four scenarios.

In the Take frame, the decision facing each player was how much to withdraw from a group project. The description of the decision problem and first ending for the Take-Simultaneous treatment were as follows:

Imagine a group that consists of two group members, Person A and Person B. There are 40 tokens in a group project. Each group member has to decide how many, up to a maximum of 20, of these tokens to withdraw for himself and how many to leave in the group project. Each token he withdraws for himself has a value of one pound for him. Each token left in the group project has a value of 1.50 pounds to the project. The total value of the project is divided equally between the two group members. So, each token left in the project earns both group members 0.75 pounds each. The total income of a group member is the sum earned from tokens withdrawn by himself and his share of the earnings of the group project. Each group member decides simultaneously, that is, without knowing what the other one has done.

A) Assume that Person A withdraws 20 tokens from the group project and Person B withdraws 20 tokens from the group project. Therefore, the value of the group project is 0 pounds and, thus, as a result of their withdrawals, Person A's total income is 20 pounds and Person B's total income is 20 pounds. How do you rate ***Person B's*** *morality?*

As with the Give frame, the only difference between the five scenarios in a Take frame questionnaire was the behaviour of Person A. Person A's withdrawal was 20 tokens (as shown) in the first scenario, declining to 0 in decrements of 5 over the remaining four scenarios.

It is important to note that the Give and Take frames differ only in respect of the description of the decision problem. There is no difference between the two frames in terms of the feasible sets of monetary outcomes available to a player. In each frame, each player controlled the final destination of 20 tokens, each of which could be

allocated either to himself (earning £1 for him) or to the project (earning £0.75 for each player). To emphasise this similarity, we will use the term "effective contribution" below to refer to the tokens allocated by a player to the project, regardless of whether this was by means of contributing or not withdrawing.

In addition to the Simultaneous treatments, we ran two treatments (one with the Give frame, and one with Take) in which the non-judged player moved first. Each questionnaire for these Sequential treatments was obtained from the corresponding Simultaneous one by replacing the last sentence of the first paragraph with "Assume that Person A decides first and Person B observes Person A's choice before making his own decision." In all other respects, Sequential questionnaires were identical to the corresponding Simultaneous ones.

2.2 Motivation and hypotheses

The philosophical and psychological literatures suggest two broad accounts of how individuals might arrive at their moral judgments which, for convenience, we call the *reason-based model* and the *emotion-based model*, respectively. The former can be seen as a descendent of rationalist traditions in philosophy, whereas the latter has more affinity with naturalistic traditions.[5]

The reason-based model sees an individual's moral judgments as the result of deliberation in which the prior moral principles she endorses are applied to the case at hand. On this account, hypotheses about how subjects' judgments will vary across scenarios would be conditional on assumptions about their prior moral principles and, in particular, about whether those principles imply that the differences between our scenarios are morally relevant.

In contrast, the emotion-based model sees emotions and intuitions as the drivers of moral judgments. On this view, moral judgments express sentiments. Moral judgments are caused by quick moral intuitions which may be followed by *ex post* moral reasoning.[6] On the emotion-based model, whether and how far subjects report

[5] Descartes, Leibniz and Kant are major figures in the rationalist tradition (though what we call here the reason-based model does not require agents to endorse specifically Kantian moral principles); the naturalistic tradition derives from Hume and, especially in relation to ethics, from Smith. For more recent discussions of the philosophical and psychological background on moral judgments, see e.g., Haidt (2001); Nichols (2004); Doris and Stich (2005); Hauser (2006); Joyce (2006); Prinz (2006); Prinz (2007); Krebs (2008); Sinnott-Armstrong (2008); DeScioli and Kurzban (2009); Nado, et al. (2009).

[6] For experimental evidence, see Greene, et al. (2001) and Wheatley and Haidt (2005). For overviews, see Haidt (2001); Greene and Haidt (2002) and Haidt (2007).

different judgments across scenarios would depend on whether there are differences in the nature and intensity of the emotional responses cued by them.

As indicated above, our experimental design manipulates the framing of the decision problem (Give versus Take); the order of moves in that problem (Simultaneous versus Sequential); and the behaviour of the non-judged Person A. To facilitate our discussion of how these manipulations would be seen by the two models of judgment, we begin by giving names to certain hypotheses.

We refer to the view that moral judgments are insensitive to the Give versus Take manipulation as the *frame insensitivity hypothesis*.

Correspondingly, the *independence hypothesis* asserts that the moral rating of Person B is independent of Person A's effective contribution. Note that the independence hypothesis could hold in either Simultaneous or Sequential treatments, but (as we will see) the arguments that would motivate it in the two cases may be different. If the independence hypothesis holds under one order of moves, but not the other, this would induce a difference between Sequential and Simultaneous treatments for some otherwise identical scenarios.

If judgments of Person B are sensitive to Person A's effective contribution, it seems most likely that this will take the form that the higher is Person A's effective contribution, the less favourable is the moral rating assigned to Person B. We refer to this as the *increasing condemnation hypothesis*.[7] We focus on this (potential) direction of effect as, although Person B's effective contribution is always 0 tokens, the effective contribution of Person A rises across the successive endings of each questionnaire, leading to outcomes that are progressively less favourable to Person A and more favourable to Person B, both in relative and absolute terms. Each increment of 5 tokens in Person A's effective contribution reduces Person A's monetary payoff by £1.25, while increasing that of Person B by £3.75. Thus, although each player receives £20 in the first scenario of each questionnaire, by the last scenario, Person A receives £15 and Person B £35.

[7] This is not to say that only one direction of impact of Person A's effective contribution on the moral rating assigned to Person B is possible. A decreasing condemnation hypothesis is also conceivable if subjects think that a certain level of total contribution should be provided and only condemn Person B for his failure to provide that part of the required total contribution which Person A has not provided.

The reason-based model

The implications of the reason-based model depend on the prior ethical principles that subjects endorse and, especially, on whether these are consequentialist.[8]

With consequentialist principles, the reason-based model makes a clear prediction in respect of framing. For a *consequentialist* ethical theory, the moral value of an action derives from a comparison of its consequences with other feasible ones; so re-describing the decision problem, holding the actual and feasible consequences constant, should have no impact on the moral value of the action. Thus, if our subjects endorse any form of consequentialism that sees the morally relevant consequences in our scenarios as determined by the monetary outcomes, the reason-based model predicts that the frame insensitivity hypothesis will hold. For the remainder of the paper, by "consequentialism" we intend a form of the doctrine that has this implication.[9] Thus, if we observe a difference between judgments in the Give and Take frames, the reason-based model would have to interpret it as evidence of subjects endorsing prior ethical principles that are not consequentialist in the sense just described.

The consequentialist argument for the frame insensitivity hypothesis requires the morally relevant consequences of Person B's free-riding to be determined by the monetary outcomes, but it does depend on how broadly those monetary outcomes are construed.

If they take a *narrow* view, subjects could see the consequences of Person B's action as consisting only of the payments determined by the tokens in his own control. This would imply that the "consequence" of Person B making an effective contribution of zero tokens is the same across *all* scenarios. Then, in addition to frame insensitivity, the independence hypothesis would hold in Simultaneous

[8] Blackburn (2008), p. 74, defines consequentialism as the view that the "value of an action derives entirely from its consequences". For an extensive philosophical discussion, see Sinnott-Armstrong (2006); for a discussion from an economic point of view see Sen (1987).

[9] When faced with counter-examples, a possible defensive move for advocates of consequentialism might be to reinterpret the consequences to include factors previously not seen as part of them. If "contributing no tokens" and "withdrawing 20 tokens" are interpreted as acts with different consequences, perhaps because one leads to the "consequence" that a withdrawal has been made and the other does not, then a framing effect in our design would be compatible with subjects making consequentialist judgments, in the reinterpreted sense. However, if taken to the limit, this reinterpretation strategy risks abolishing any distinction between an action and its consequences, so making it a truism that acts should be judged solely by their consequences. Our concern in this paper is the *empirical* question of whether free riding is judged by its economic consequences (which in our scenarios are monetary), not with the *philosophical* question of how far consequence re-description is a legitimate defence strategy for consequentialist ethical theory.

treatments, and in Sequential treatments, and with no difference between Simultaneous and Sequential.

However, if subjects take a *broad* view and see the monetary consequences of Person B's free riding as including all monetary payments that arise in a given scenario, the independence hypothesis might fail (at least in Sequential treatments). For example, it would be consistent with a broad view for the difference between (or ratio of) the payoffs to each player to be seen as a morally-relevant feature of the consequences of Person B's free riding. If subjects are consequentialist in the broad sense, and averse to unequal outcomes, this creates the potential for the increasing condemnation hypothesis to hold. This is because any increase in Person A's effective contribution would tilt relative payoffs (further) in Person B's favour, if Person B continues to free ride. If this is seen as an undesirable outcome, then the obligation on Person B to avoid it may strengthen; and, if so, one would expect Person B to be condemned more strongly for continuing to free ride. Thus, broad consequentialism can generate, out of an attitude towards unequal outcomes, a view that Person B ought to match Person A's effective contribution. We will call such an obligation one of *indirect reciprocation*.[10]

If subjects reason in this way, one might well expect the increasing condemnation hypothesis to hold in Sequential treatments, since Person B must be held responsible for (what a broad consequentialist sees as) the different, and known, consequences of his actions across the five scenarios in a Sequential treatment. When the consequences are worse, Person B can be blamed more heavily.

It is harder to formulate moral principles that rationalise conformity with the increasing condemnation hypothesis in Simultaneous treatments too, as they would have to license condemning Person B differently, given different effective contributions by Person A, even though Person B is neither responsible for nor knows Person A's choice. We will call the principle that an agent cannot be condemned on the basis of matters which he neither controls nor knows the *responsibility doctrine*. If subjects endorse this principle then, even if they are otherwise inclined to view consequences broadly, the independence hypothesis would hold in Simultaneous treatments.

[10] Note that we use this term differently from the usage in the literature on the evolution of cooperation (e.g., Nowak and Sigmund (1998)).

However, there are ethical views which violate the responsibility doctrine and might account for increasing condemnation even in Simultaneous treatments. At first sight, this may seem a strange property for moral principles. But, within a broad consequentialist framework, rationalisation for it can be found in the doctrine of *moral luck*, discussed by Nagel (1976) and Williams (1981).

According to this doctrine, an agent can be blamed for outcomes of their actions to which chance, or other matters outside their control, have contributed. As an example, Nagel argues that a driver who has negligently failed to check his brakes "would *have* to blame himself" (emphasis added) much more if a child runs into the road and is killed than if no situation arises which requires sharp braking, even though his negligence is the same in each case, and he neither predicted nor had any control over the child's action.[11] By analogy, in the current context, one might see it as bad moral luck for Person B if Person A makes a non-zero effective contribution, so turning his own free riding into unilateral free-riding, but still allow that this contributes to the culpability of his action – especially as it is a foreseeable result of Person B free-riding that he might prove to be the only free rider.[12]

Finally, note that there is nothing in the reason-based model that requires subjects to be consequentialist. If subjects endorse deontological moral principles instead, then the reason-based model predicts that they would apply those principles to form their judgments. *Deontological* views see the moral status of an action as flowing, not from its consequences, but from its intrinsic properties. For example, it might be seen as intrinsically wrong to commit murder even if, by some bizarre twist of fate, one could actually bring about net beneficial consequences by doing so.

In our context, if the intrinsic properties of Person B's free riding are to be distinguished from consequences, it seems inevitable that they can take no account of

[11] Interestingly, Nagel goes on to argue that, although the doctrine that one can only be blamed for matters in one's control is superficially plausible, it threatens to undermine much of normal moral judgment.

[12] Re-describing Person B's free riding as unilateral free riding may seem to elide Person B's action with Person A's. However, it is actually quite common to hold an individual responsible for a joint consequence of his own and another's action; indeed, arguably, this is sometimes almost unavoidable. One would normally hold President Kennedy's assassin responsible for his *death*, not just for the passage of some bullets down Dallas's Elm Street, even though the latter was not a sufficient condition for the former. Further, the gunman would probably still be held responsible for the death even if he had never intended to kill the President, but merely to take a few random shots along the street. The analogy is that, just as killing someone is a foreseeable, if contingent, result of taking pot shots down Elm Street, so being the only free rider is a foreseeable, if contingent, result of free riding. The moral luck doctrine suggests that, if being a unilateral free rider is particularly bad, then one can be blamed more strongly for free riding when it turns out ex post that one has done so unilaterally.

Person A's action when Person B is unaware of it. Thus, deontological forms of the reason-based model would lead us to expect the independence hypothesis to hold in Simultaneous treatments.

Whether it would also hold in Sequential treatments would depend on whether the intrinsic properties of Person B's free riding are sensitive to Person A's choice, when that is known to Person B. If not (for example, because free riding is seen as intrinsically and unconditionally wrong) then the independence hypothesis would be expected to hold in Sequential treatments, as well as in Simultaneous. But, a different view might rationalise increasing condemnation, for example if taking revenge on Person A when he has free ridden, or failing to reward him when he has contributed, are seen as intrinsic properties of Person B's free riding in different scenarios. If the former is morally acceptable but the latter is not then this view, which we will call *direct reciprocation*, would require the increasing condemnation hypothesis, but only in Sequential treatments.

Finally, if withdrawing and withholding support are seen as intrinsically different actions, application of deontological moral principles could also account for frame sensitivity of judgments of the free rider.

The emotion-based model

On the emotion-based model, it is not necessary to delve into such tricky terrain to account for violations of frame insensitivity and/or the independence hypothesis, because the model does not require moral judgments to flow from coherent principles. Instead, it sees them as cued by emotional responses to the scenario as a whole.

Evolutionary theorists argue that moral judgments may be situation-specific and frame-dependent (e.g., Krebs (2008), p. 116). Consistent with this argument, the emotion-based model suggests that subjects' judgments would be driven by gut-reactions to whole scenarios which, in turn, may be sensitive to seemingly incidental features of them. For example, the emotional response to a player whose effective contribution is zero might differ according to whether this free riding arises from complete failure to contribute to the project or from maximal withdrawal of support from it. This is possible even though the consequences are the same, and even in the absence of a prior ethical theory that licenses the distinction.

Similarly, it is easy to imagine how Person A's choice might affect a subject's emotional response to a scenario in which Person B free rides, even when their

choices are simultaneous. The subject might be more angered, or disgusted, or saddened, by a scenario in which Person B free rides the larger is Person A's effective contribution. For example, negative emotional responses to payoff inequality or to inequality of contributions could bring this about.

Although it is quite possible that emotional responses would vary between the Simultaneous and Sequential versions of otherwise identical scenarios, the emotion-based model need not confine conformity with the increasing condemnation hypothesis to the Sequential treatment, since emotional responses are responses to the whole scenario. Gino, et al. (2008a) provide evidence of an outcome bias in ethical judgment. Such a bias arises when the assessment of an action is influenced by ex post information about its outcome that was not available to the decision-maker. In our context, a similar bias might take the form of Person B being condemned most strongly when it turns out his free riding was unilateral, even though he did not know that at the moment of choice. This would lead one to expect conformity with the increasing condemnation hypothesis in Simultaneous treatments, the use of the word "bias" indicating that, on this view, there need be no principled justification for the phenomenon.[13]

Table 1 summarises the discussion of this section. For each of the two models of moral judgment, the Table indicates the factors that determine whether, and how, manipulation of the framing of the decision problem and of the non-judged player's behaviour should affect judgments of Person B's free riding. (For manipulation of Player A's choice, we distinguish between Sequential and Simultaneous treatments.) A bullet-point saying that the framing of the decision problem has "No effect, if …." indicates assumptions under which the frame insensitivity hypothesis should hold. Similarly, a bullet-point saying that Player A's choice has "No effect if …." indicates assumptions under which the independence hypothesis should hold.

[13] Gino, et al. (2008b) corroborate the outcome bias and also provide evidence of other "seemingly irrelevant" factors, such as the identifiability of victims, affecting moral judgments. As with chance outcomes, whilst such factors may seem irrelevant to a sober assessment, they could contribute powerfully to emotional responses.

Table 1: Analysis of impact of manipulations on moral judgments

	Description	Impact of		
		Framing of decision problem	Player A's choice (Sequential treatments)	Player A's choice (Simultaneous treatments)
Reason-based model (following rationalist tradition in moral philosophy)	Judgments arise from application of prior moral principles to case in hand	• No effect, if principles are consequentialist. • Frame sensitivity possible if principles are deontological and intrinsic properties of not giving and taking are different.	• No effect, if principles are narrow consequentialist. • Increasing condemnation possible, if principles are broad consequentialist. • If principles are deontological, depends whether intrinsic properties of B's free riding are sensitive to A's choice.	• No effect if principles are narrow consequentialist or deontological. • If principles are broad consequentialist, depends whether subjects endorse responsibility doctrine or moral luck doctrine.
Emotions-based model, (following naturalistic tradition in moral philosophy and psychology)	Judgments arise from instinctive emotional reactions to case in hand, which may be rationalised by ex post reasoning	• No effect, if Give and Take frames cue same emotions. • Frame sensitivity possible if Give and Take frames cue different emotions.	• No effect, if emotions are unaffected by A's choice. • Increasing condemnation possible if effective contribution by A cues negative emotions towards free rider.	• No effect, if emotions are unaffected by A's choice. • Increasing condemnation possible if effective contribution by A cues negative emotions towards free rider.

2.3 Procedures

We recruited participants from among University of Nottingham students using the ORSEE software (Greiner (2004). In total, we sent 2,718 email invitations, resulting in 538 participants. Once a subject registered to take part, they were directed to the experiment's website. Subjects were allocated automatically to one of the four treatments, in a rotating sequence by time of registration for the experiment. After assigning their moral ratings, subjects were asked to give a brief verbal explanation of them.

Each subject saw only the questionnaire for the treatment they were assigned to. They could either respond immediately to the questionnaire, or exit and return to it any time before the closing date of the experiment (which was one week after invitations were sent out). Subjects returning later could still only see the questionnaire they had been assigned to initially. Subjects were omitted from the data analysis if they failed to complete a questionnaire by the closing date.[14] To counter the possibility of multiple submissions from the same subject, only one registration was permitted from a given invitation. Our use of ORSEE recruitment software, rather than an open internet experiment, enabled us to build in this safeguard, as well as giving us the demographic information on participants held in the ORSEE database.

It is inherent to our study that we could not incentivise task-responses, but we could incentivise participation. We comment on these features in turn.

Our objective was to study subjects' impartial moral attitudes. A questionnaire-based approach was appropriate for this purpose because any means of tying payments to subjects' responses would introduce a potential confound. [15] In particular, we wished to elicit the judgments that subjects would give in the role of a

[14] This resulted in the following number of participants: Give-Simultaneous – 135; Take-Simultaneous – 138; Give-Sequential – 128; Take-Sequential – 137.

[15] The use of survey-based methods and questionnaire-based experiments is standard in the study of mental states and social attitudes. The recent economics literature on self-reported happiness relies on non-incentivised responses (see Clark, et al. (2008), for a survey and the August 2008 *Journal of Public Economics* symposium for recent examples). Perhaps more directly relevant here is the fact that non-incentivised attitudinal social surveys are used to study preferences for redistribution (e.g. Corneo and Grüner (2002); Fong (2001); Corneo and Fong (2008)); and the World Values Survey (http://www.worldvaluessurvey.org) includes, among other attitudinal questions, questions on people's moral judgments of behaviour in real-world public goods contexts, like evading taxes and dodging fares. Attitudinal questions have also been deployed in research on social capital (e.g., Knack and Keefer (1997)). For use of questionnaire-based experimental methods to study perceptions of fairness and distributive justice, see Kahneman, et al. (1986); Anand (2001); Gächter and Riedl (2006); and the references in the introduction.

disinterested observer. In the words of Konow (2003), in a review of the empirical literature on judgments of fairness: "the primary goal is to track the values of the impartial spectator rather than the implicated stakeholder" (p. 1191). This precluded having subjects be participants in the voluntary contributions game: hence our use of hypothetical scenarios. Allowing subjects to assign financial penalties or rewards to the players in the scenarios, even hypothetically, would have confounded moral attitudes with attempts to bring about particular distributional consequences: hence our use of pure judgment tasks rather than – say – reward or punishment tasks. As our judgment tasks are moral judgment tasks, as opposed – say – to mathematical puzzles or judgments of distance, there are no objectively "right" or "wrong" answers to them. So, we could not reward subjects for judging correctly. Finally, rewarding subjects for making judgments that conform to particular ethical theories, or to our own ethical views, or to average opinion, would all have introduced obvious biases, relative to the motivation for our experiment. Our aim was to elicit subjects' *own* actual judgements, not their beliefs about which judgments would be rewarded or are held by others.[16]

Given the absence of task-related incentives, we felt that it might be difficult to generate a sufficient number of participants, without some participation incentive. On the other hand, having a substantial reward for participation might have attracted subjects unwilling to give considered responses and only willing to do the minimum necessary to obtain the reward. It might also disproportionally attract people for whom pecuniary concerns are particularly important. In the light of these considerations, we used two approaches in parallel. Prior to issuing invitations, we divided our potential subject pool randomly into two equal sub-groups: one for which there would be no payments at all ("No-Payment experiment") and one in which a random participation fee was provided ("Payment" experiment), in the form of entry to a lottery. The latter provided some protection against low participation, while conducting both experiments enabled us to check for any effect of the participation incentive on task responses. All subjects participating in either experiment were informed about the importance of answering the questionnaire as precisely and honestly as possible and that all responses would remain confidential. Subjects invited

[16] See Krupka and Weber (2008) for an experimental technique where people are rewarded for correctly guessing which norms *other* people hold. Their interest is in eliciting what people *think* the *social* norm is whereas we are interested in the *individual's* own moral judgments.

to the Payment experiment were told that those who completed the questionnaire would be entered into a prize draw, conducted publicly with two prizes of £50.[17]

3. Results

Before turning to our main questions, we consider the impact of the different participation incentive schemes on participation and task-responses. This is a useful preliminary for what follows, as well as of some independent interest.

We sent the same number of invitations to participate in each experiment; and, in response, 306 subjects completed a questionnaire in the Payment experiment, compared with 232 in the No-Payment experiment. This suggests that paying a random participation fee can be an effective (and cheap) way to increase the response rate. A Probit regression analysis reported in the Appendix (Table A1) supports this conclusion.

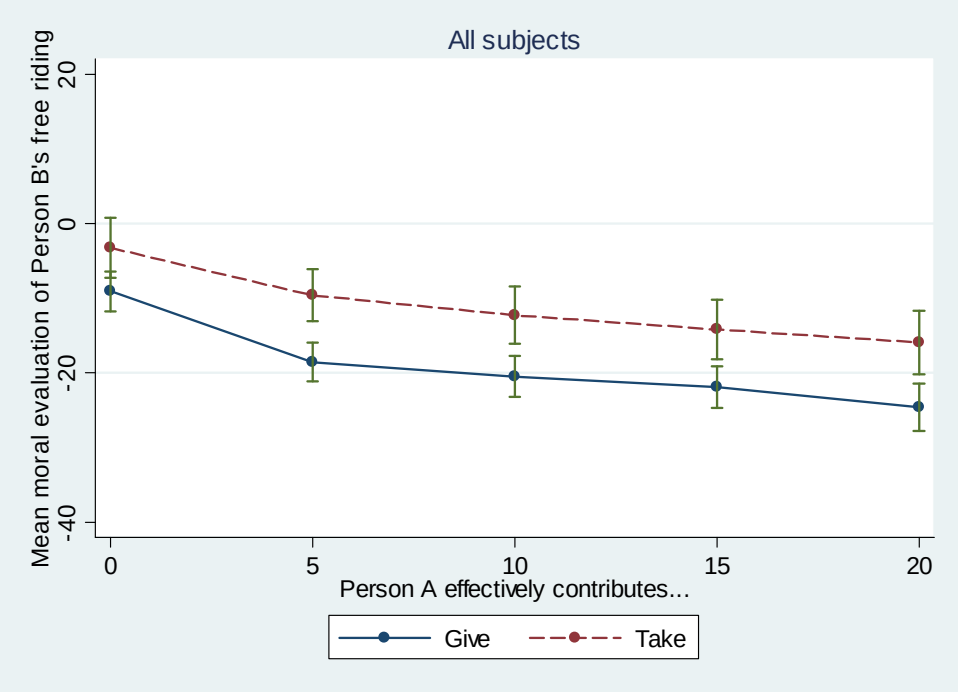
More importantly, the coefficient on the Payment variable in the regression analysis of moral evaluations reported in the Appendix (Table A2) is not statistically significant. Thus, it does not seem that the difference between the two experiments had any important impact on responses. In our view this indicates that there is no selection bias between those who participate in the two experiments. We therefore proceed below by pooling the data.

3.1 How is free riding judged? The Simultaneous Case

We begin our main analysis with the Simultaneous treatments, in which, in each scenario, Person A and Person B decide without knowing the action of the other. Here, and below, the main tool for our analysis is the mean "moral evaluation function" (MEF). This is an aggregate measure of the moral ratings that subjects assigned to the free rider (Person B), expressed as a function of the effective contribution levels of his non-judged counterpart (Person A). Recall that effective contribution is the number of tokens contributed to the group project (in the Give frame) or, correspondingly, the number of tokens left in the group project (in the Take frame). Figure 1 shows the mean MEF, for each of the two Simultaneous treatments.

[17] Participants were given the date, time, and venue of the draw and invited to attend; they were also told that the winners would be contacted by email if they did not attend, so that payment was not conditional on attendance.

Figure 1. The moral evaluation function in the Simultaneous treatments



The horizontal axis indicates Person A's effective contribution, measured in number of tokens. The vertical axis indicates the average moral rating that subjects assigned to Person B, who is always a complete free rider. On this axis, the point 0 denotes that free riding is perceived to be of no moral significance. Ratings below 0 imply that subjects perceive free riding as morally blameworthy; whereas ratings above 0 imply that subjects perceive free riding as morally praiseworthy. The 95% confidence intervals for the mean moral evaluation in each of the five scenarios of each treatment are also shown.

Three features of Figure 1 are particularly striking (all of them are statistically significant according to econometric analysis documented in the Appendix, Table A3, first column). First, the average moral rating of Person B's free riding is negative in all cases shown, suggesting that subjects do regard the decision problem in the scenarios as having a moral dimension and free-riding as a blameworthy act. Second, the MEF for the Give treatment is always below that for the Take treatment indicating that, on average, subjects condemn total failure to contribute to the group project more strongly than complete withdrawal of support from it. Third, for each frame, the MEF is negatively sloped: the free rider is condemned more strongly the greater the other player's net contribution, even though moves are simultaneous.

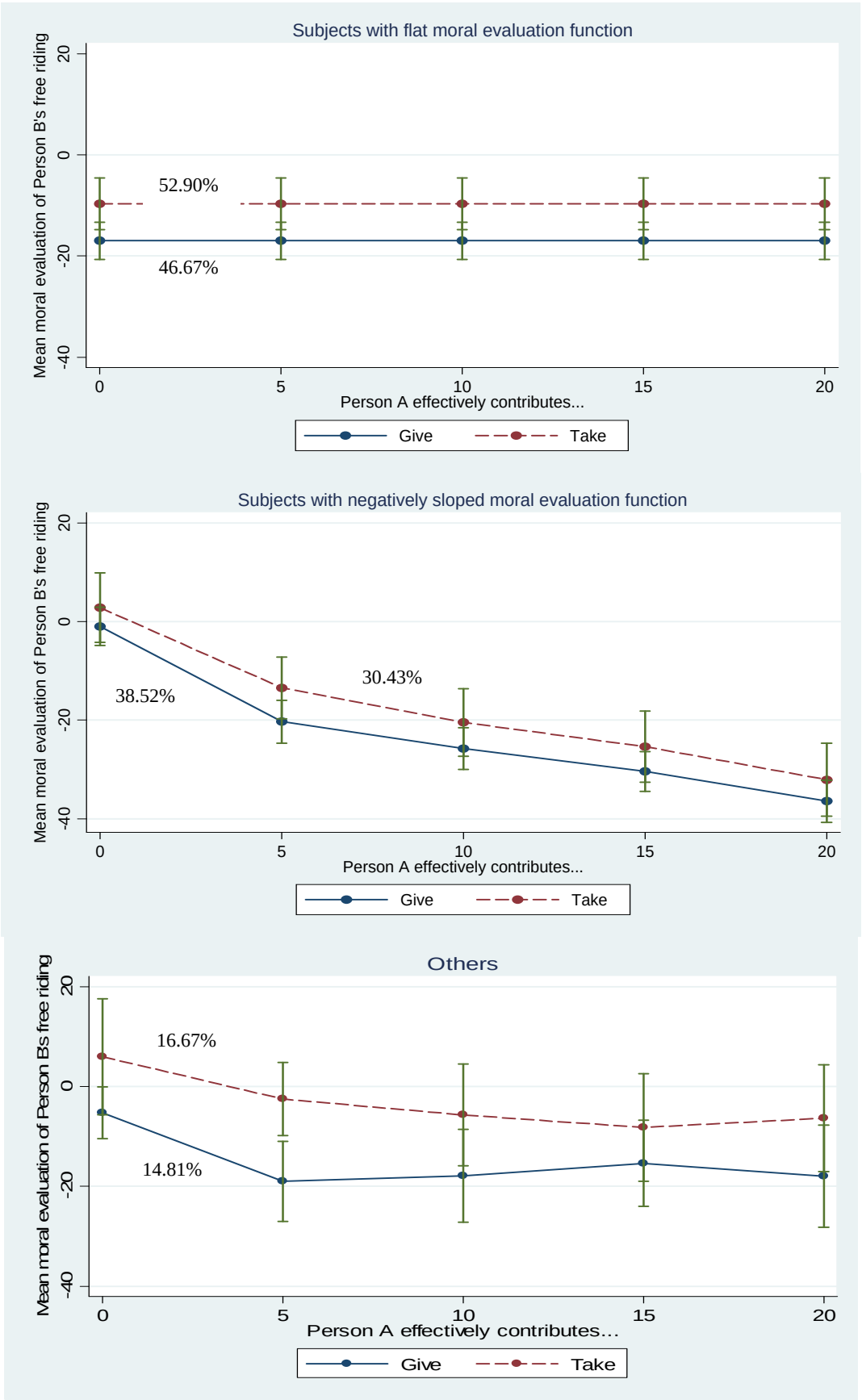
To understand the observed pattern of judgments better, we divided subjects into three categories (response patterns): (1) subjects with a *flat* MEF, (2) subjects with a *negatively sloped* MEF, and (3) "*others*", including non-monotonic subjects and subjects with a positively sloped MEF.[18] The mean MEFs for the Give and Take treatments, for each of these three response patterns, are shown in the three panels of Figure 2, respectively. The percentage of subjects in the relevant treatment falling in a given category is shown, as are the 95% confidence intervals for the mean moral evaluation in each of the five scenarios.

The largest category, accounting for 52.9 and 46.7 percent of people in Take and Give treatments, respectively, consists of those whose MEF is flat across the five effective contribution levels of Person A. The overwhelming majority of subjects with a flat MEF assigned a negative rating to the free rider, and the average is indeed highly significantly negative in both treatments. (Only 12.8 percent of subjects thought free riding is of no moral significance and therefore assigned a zero rating across all scenarios.) The second largest category is those subjects for whom free riding is more reprehensible the greater Person A's effective contribution (38.5% and 30.4% of subjects in the Give and Take treatments, respectively). The third category ("Others") comprises a minority of 14.8 and 16.7 percent of the subjects who have neither flat nor monotonically decreasing ratings.

The classification of subjects also structures the econometric evidence given in the second, third and fourth columns of Appendix, Table A3, which report separate regressions for each category. It turns out that subjects with a flat MEF and subjects classified as "others" are (at least weakly) significantly more condemning in the Give than in the Take treatment, whereas for subjects with negatively sloped MEFs frame insensitivity holds because we find no statistically significant difference across frames. In sum, the existence of a framing effect in our aggregate data for the Simultaneous treatments can be attributed largely to those subjects who condemn free riding equally across scenarios, or to "Others".

[18] Non-monotonic subjects refer to those whose MEF is strictly negatively sloped in one range and strictly positively sloped in another.

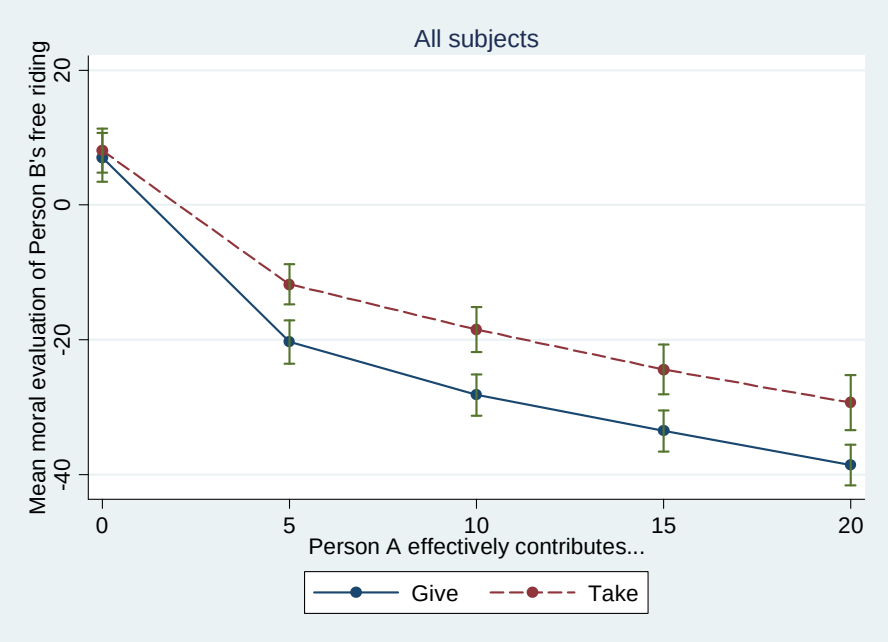
Figure 2. The moral evaluation function for each rating pattern in the Simultaneous treatments



3.2 Do sequential moves make a difference?

We now turn to the Sequential treatments, where Person B observes Person A's choice before making his own decision. Figure 3 shows the mean MEF for all subjects facing these treatments, taken together. As before, 95% confidence intervals of mean evaluations are shown for each scenario. Supporting regression analyses can be found in Table A4 in the Appendix.

Figure 3. The moral evaluation function in the Sequential treatments



Comparison of Figure 3 with Figure 1 reveals some notable similarities and differences. Once again, for all scenarios where Player A's effective contribution is non-zero, the mean moral evaluation of the free rider, Player B, is negative – indicating condemnation. Also, the mean MEF for the Give frame again lies below that for the Take frame whenever Person A's effective contribution is non-zero, indicating that the framing effect observed in Simultaneous treatments is largely robust to a sequential order of moves.[19] The main qualitative differences between Figure 3 and Figure 1 relate to the slope of the MEF, which is steeper in the Sequential case. Especially when Player A's effective contribution is 10 or more, the average condemnation of Player B is notably stronger in Sequential treatments than in the corresponding Simultaneous treatments. Also, interestingly, there is a directional

[19] In Appendix, Table A4, the impact of framing in Sequential treatments is shown by the significance of the interaction terms.

difference in the judgment of Person B's free riding, when Person A free rides too. In contrast to the Simultaneous treatments, people in the Sequential treatments regarded it as morally *praiseworthy* for Player B to "rat on a rat".

A striking difference between the Sequential and Simultaneous treatments is in the relative size of the categories of subject. In contrast to the Simultaneous treatments, people with a negatively sloped MEF are by far the largest category in the Sequential treatments (88.3% and 77.4% in Give-Sequential and Take-Sequential, respectively). In stark contrast to the Simultaneous treatments, only very few subjects in the Sequential treatments (3.9% and 5.1% in Give-Sequential and Take-Sequential, respectively) have flat MEFs.

3.3. Discussion

Our main findings can be summarized as follows:

> *Finding 1*: On average, free riding is judged morally reprehensible in all cases, except that it is judged morally commendable to "rat on a rat" (i.e. to free ride knowing that the co-player has already free ridden).
>
> *Finding 2*: *Ceteris paribus*, failure to contribute to the public good is condemned more strongly, on average, than total withdrawal of support from the public good. This holds both in the Simultaneous and the Sequential treatment.
>
> *Finding 3*: On average, moral judgments conform to the increasing condemnation hypothesis (that a free rider is condemned more strongly the larger is the effective contribution of his co-player). In Simultaneous treatments, about half of subjects pass judgments on the free rider that are independent of the contribution of the other player. Yet, the overwhelming majority of subjects in Sequential treatments conform to the increasing condemnation hypothesis, as do a substantial minority in Simultaneous treatments.

Finding 1 suggests that public goods problems are perceived as having a moral dimension in the sense that, in general, subjects do not give neutral moral judgments of free riding. Then, the question is what drives the judgments that they do give.

Interpreted from the perspective of the reason-based model, our findings are indicators of the prior moral principles that subjects apply. For this view to account for Finding 2 would require that subjects endorse moral principles that distinguish between non-contribution and withdrawal, even when their monetary consequences are the same. Such principles would have to be non-consequentialist.

The reason-based model fits well with people who judge Person B equally across scenarios in the Simultaneous treatments, since doing so is consistent with the responsibility doctrine or with deontological moral principles. However, the reason-based model can explain the commendation of ratting on a rat and, more generally, the greater prevalence of increasing condemnation in Sequential treatments than Simultaneous treatments, indicated by Finding 3, *only* if a substantial number of subjects endorse principles that call for (direct or indirect) reciprocation. Even if they do, it would be difficult to account for the presence of a substantial minority who display increasing condemnation in Simultaneous treatments on the reason-based account unless some subjects are prepared, as a matter of principle, and in line with the moral luck doctrine, to condemn the free rider on the basis of actions taken by his co-player for which he is not directly responsible and of which he was unaware at the moment of choice.

The emotion-based model suggests a different explanation of our findings. According to this model, judgments are *ex post* rationalizations of emotional or affective reactions to the scenarios. Even for impartial observers, free riding might cue emotions such as anger, disgust, irritation, or milder forms of distaste. On this view, increasing condemnation could arise from stronger affective reactions to scenarios that seem particularly unequal or unfair on the non-judged player. A further possibility, particularly applicable to the Sequential treatment, is that Person A is seen as trusting Person B to reciprocate when he makes a non-zero effective contribution, and that subjects experience a negative emotional response to the betrayal of this trust.

If, in some judges, such reactions of distaste are cued more by relative effective contribution than by consideration of the facts known to the free rider, that would explain the presence of increasing condemnation in Simultaneous treatments (as well as in Sequential ones). Thus accounted for, the finding is in line with the outcome bias in ethical judgments identified by Gino, et al. (2008a). Indeed, from this perspective, the surprising feature of our findings is not so much the presence of subjects who conform to the increasing condemnation hypothesis in Simultaneous treatments as the fact that the modal group does not do so.

Emotional responses could also explain the positive evaluation of “ratting on a rat”, for example if there is a positive affective response to the first free rider getting what he deserved when the judged player free rides back.

A priori, and given the findings of research in moral psychology (Haidt (2001)) the emotion-based model seems a promising way to explain framing effects in judgments, as it is quite possible that details of the description of different scenarios might cue different emotional responses. However, to explain the direction of the framing effect that we have observed requires more than just this remark and is not straightforward.

The differences between corresponding Give and Take scenarios can be encapsulated by the following which, as an illustration, uses scenarios in which both effective contributions are zero:

***Each group member receives an endowment of 20 tokens and** has to decide **how many tokens to keep for himself** and how many to **contribute to a** group project.*	***There are 40 tokens in a group project**. Each group member has to decide **how many, up to a maximum of 20, of these tokens to withdraw** for himself and how many to **leave in** the group project.*
*Each token he **keeps** for himself has a value of one pound for him. Each token **contributed to** the group project has a value of 1.50 pounds to the project.*	*Each token he **withdraws** for himself has a value of one pound for him. Each token **left in** the group project has a value of 1.50 pounds to the project.*
*So, each token **contributed to** the project earns both group members 0.75 pounds each. The total income of a group member is the sum earned from tokens **kept for** himself and his share of the earnings of the group project.*	*So, each token **left in** the project earns both group members 0.75 pounds each. The total income of a group member is the sum earned from tokens **withdrawn by** himself and his share of the earnings of the group project.*
*Assume that Person A **contributes 0** tokens **to** the group project and Person B **contributes 0** tokens **to** the group project. Therefore, the value of the group project is 0 pounds and, thus, as a result of their **contributions**, Person A's total income is 20 pounds and Person B's total income is 20 pounds.*	*Assume that Person A **withdraws 20** tokens **from** the group project and Person B **withdraws 20** tokens **from** the group project. Therefore, the value of the group project is 0 pounds and, thus, as a result of their **withdrawals**, Person A's total income is 20 pounds and Person B's total income is 20 pounds.*

Passages in the Give frame are on the left, with the corresponding passages in the Take frame on the right. We have shown here only those sentences where there is some difference between the frames, suppressing all sentences which are common. The exact differences between corresponding sentences are shown by bold text. All of the main differences between the Give and Take frames are introduced in the first pair of passages shown; the subsequent ones are knock-on effects.

Our prior expectation was that, if framing made a difference, Player B's free riding would be condemned more strongly when the Take frame is used because, in this case, a zero effective contribution involves abrogating for himself some part of

the group project; whereas, in the Give frame, players are merely allocating their own endowment. This conjecture can also be supported by the theory of loss aversion (Tversky and Kahneman (1991)), if the initial status quo is taken as the reference-point. On this view, Player A suffers a loss as a result of Player B's action in the Take frame, but not in the Give frame. If subjects condemn the imposition of losses more strongly than the corresponding failure to grant a gain, Player B would be condemned more strongly in the Take frame, contrary to our Finding 2.

One possible explanation of Finding 2 is that subjects take Person B to have been given a gift (i.e. the endowment) in the Give frame and condemn him for not sharing it; whereas they see the players as having to fend for themselves in the Take frame and are disinclined to judge them harshly for doing so. A related possibility is that subjects see responsibility for the group project as more ambiguous in the Take frame than the Give frame, so cuing stronger moral responses in the latter case.

4. Concluding remarks

This paper contributes to economic moral psychology by presenting an experimental investigation of the moral judgments that people make of an important form of economic behaviour: free riding in public goods games. Rather than free rider problems from the natural economy, we have used as our vehicle different scenarios involving a two-player voluntary contributions game, very similar to those typically used in experimental investigations. Such investigations have played a major role in generating stylized facts about the determinants of contributions to public goods (Ledyard (1995); Zelmer (2003); Gächter and Herrmann (2009)) and in inspiring theory development (Gintis (2003); Fehr and Schmidt (2006); Fehr and Gintis (2007); Gintis, et al. (2008)). Yet, whether people perceive a moral dimension to the interaction between them in public goods experiments and, if so, what their moral judgments are, has hitherto been unexplored.

Three main aspects of our findings stand out. First, and most fundamentally, the voluntary contributions game is seen as having a moral dimension: non-involved outside observers tend not to give neutral moral evaluations of a free rider. If this is a general feature of social dilemmas, it suggests that moral judgments might play a role in explaining cooperative behaviour and sanctioning of non-cooperators.

Second, the moral judgments passed on a free rider appear sensitive to the behaviour of the other agent – especially when the free rider was aware of the other

agent's choice but even, for a substantial minority of judges, when he was not. This is in line with the finding of an outcome bias in moral judgments, identified by Gino, et al. (2008a) and Gino, et al. (2008b) and with the moral luck doctrine. It suggests that ethical principles that call for some form of reciprocation and/or emotional responses cued by comparisons of contributions may be at work in driving moral judgments.

Third, we observe a framing effect: subjects, on average, condemn a total failure to contribute to the public good more strongly than the complete withdrawal of support from it that implies the same net consequences.

Putting these points together, one conclusion seems clear: our findings cannot be explained by subjects forming their moral judgments by applying simple consequentialist moral principles. The picture of moral judgments which emerges from our study is one in which they respond to features of the whole situation, not just to the consequences of the judged action, narrowly conceived. Whilst it is not impossible to reconcile this feature of our findings with the reason-based model, the totality of the findings seems to fit more easily with the emotions-based model.

Acknowledgements: We thank participants at the GATE Conference of the French Economic Association in Behavioural and Experimental Economics in Lyon, the XII Spring Meeting of Young Economists in Hamburg, the ESA Meeting in Rome, the 5th CREED-CeDEx-UEA meeting, the AHRC workshop on Culture and Mind in Sheffield, the Conference on New Directions in Welfare in Oxford, and Yuan Ju for useful comments. Financial support from the ESRC (PTA-030-2005-00608) and the University of Nottingham is gratefully acknowledged. This paper is part of the MacArthur Foundation Network on Economic Environments and the Evolution of Individual Preferences and Social Norms.

Appendix – Regression Results

Table A1. Does paying a random participation fee affect response rates?

Independent Variables	**Dependent Variable: Participation = 1; No-participation = 0**
Payment	0.054***
	(0.015)
Male	-0.013
	(0.015)
Obs.	2,718

*Notes: Probit estimation. Marginal effects listed. Robust standard errors are presented in parentheses. The variable "Payment" is a dummy variable equal to 1 for those subjects who participated in the "Payment" condition and 0 otherwise. The variable "Male" is a dummy variable equal to 1 for male subjects and 0 otherwise. ** denotes significance at the 5-percent level, and *** at the 1-percent level.*

Table A2. Does paying a random participation fee affect moral evaluations?

Independent Variables	**Dependent Variable: Moral evaluations of the free rider**
Person A contributes 5 tokens	-15.571***
	(0.923)
Person A contributes 10 tokens	-20.314***
	(1.046)
Person A contributes 15 tokens	-23.942***
	(1.133)
Person A contributes 20 tokens	-27.507***
	(1.224)
Payment	1.529
	(1.534)
Male	-0.857
	(1.531)
Constant	0.108
	(1.526)
Obs.	2,690

*Notes: OLS estimates. Robust standard errors are presented in parentheses and clustered on individuals. ** denotes significance at the 5-percent level, and *** at the 1-percent level.*

Table A3. Simultaneous treatments – Regression results

	Dependent variable: Moral evaluation of the free rider			
Person A contributes	All subjects	Subjects with negatively sloped function	Subjects with flat function	"Others"
5 tokens	-9.481***	-19.327***	–	-13.75***
	(1.374)	(2.483)		(4.273)
10 tokens	-11.422***	-24.788***	–	-12.65**
	(1.533)	(2.478)		(4.852)
15 tokens	-12.837***	-29.423***	–	-10.15**
	(1.601)	(2.445)		(3.914)
20 tokens	-15.541***	-35.442***	–	-12.75**
	(1.863)	(2.611)		(4.855)
Take	5.834**	4.367	7.066**	10.883*
	(2.437)	(4.055)	(3.187)	(6.162)
Male	-2.342	-3.025	-2.183	-4.961
	(2.229)	(3.644)	(3.245)	(5.161)
5 tokens × Take	3.126	3.089	–	5.272
	(1.923)	(3.856)		(6.538)
10 tokens × Take	2.400	1.527	–	0.998
	(2.200)	(4.124)		(7.171)
15 tokens × Take	1.902	1.256	–	-4.024
	(2.319)	(4.151)		(6.356)
20 tokens × Take	2.874	0.561	–	0.446
	(2.606)	(4.357)		(6.510)
Constant	-8.006***	-0.069	-15.721***	-2.769
	(1.557)	(2.121)	(2.361)	(3.478)
Obs.	1,365	470	680	215

*Notes: (1) OLS estimates. Robust standard errors are presented in parentheses and clustered on individuals. * denotes significance at the 10-percent level; ** at the 5-percent level, and *** at the 1-percent level. (2) Scenario dummies were excluded for the subjects whose MEF was flat.*

Table A3 documents OLS models, with the moral evaluation of Person B as the dependent variable. The explanatory variables include the effective contribution level of Person A (contributing 0 is the baseline), the dummy variable "Take", which equals 1 for the Take treatment, and 0 otherwise, and the dummy variable "Male", which equals 1 if subjects were male and 0 otherwise. We also control for slope differences by including as independent variables interaction terms between the dummy variable "Take" and each scenario separately. The regression coefficients in the "All subjects" column reveal that the dummy variable "Take" is statistically significant at the 5 percent level, implying that subjects are significantly more condemning in the Give than in the Take treatment, as the coefficient is positive and mean evaluations are in the negative range. In aggregate, subjects evaluate free riding as being an immoral act, as the coefficient of the constant is negative and statistically significant. The mean MEF is negatively sloped, *ceteris paribus*, since the coefficients of the four scenario dummy variables are all negative, and statistically different from zero and from each other (from F-test, p-value = 0.000), corroborating the increasing condemnation hypothesis.

Table A4 documents the same regression model as Table A3 for the Sequential treatments.

Table A4. Sequential treatments – Regression results

	Dependent variable: Moral evaluation of the free rider
Person A contributes	All subjects
5 tokens	-27.398***
	(2.102)
10 tokens	-35.25***
	(2.155)
15 tokens	-40.586***
	(2.148)
20 tokens	-45.602***
	(2.159)
Take	1.054
	(2.515)
Male	-0.731
	(1.967)
5 tokens × Take	7.596***
	(2.848)
10 tokens × Take	8.754***
	(3.072)
15 tokens × Take	8.148**
	(3.253)
20 tokens × Take	8.258**
	(3.387)
Constant	7.309***
	(1.968)
Obs.	1,325

*Notes: (1) OLS estimates. Robust standard errors are presented in parentheses and clustered on individuals. ** denotes significance at the 5-percent level, and *** at the 1-percent level. The second model takes out those whose evaluation function is non-decreasing ("flat" or "others"). Since only very few people correspond to these categories we only estimatee the MEF for the pooled sample.*

References

Anand, P., 2001. Procedural fairness in economic and social choice: Evidence from a survey of voters. Journal of Economic Psychology 22, 247-270.

Andreoni, J., 1995. Warm-Glow Versus Cold-Prickle - the Effects of Positive and Negative Framing on Cooperation in Experiments. Quarterly Journal of Economics 110, 1-21.

Binmore, K., 2005. Natural Justice. Oxford University Press, Oxford.

Blackburn, S., 2008. Oxford Dictionary of Philosophy, Second Edition Revised. Oxford University Pres, Oxford.

Brewer, M. B., Kramer, R. M., 1986. Choice behavior in social dilemmas: effects of social identity, group size, and decision framing. Journal of Personality and Social Psychology 50, 543-549.

Clark, A. E., Frijters, P., Shields, M. A., 2008. Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. Journal of Economic Literature 46, 95-144.

Corneo, G., Fong, C. M., 2008. What's the monetary value of distributive justice? Journal of Public Economics 92, 289-308.

Corneo, G., Grüner, H. P., 2002. Individual preferences for political redistribution. Journal of Public Economics 83, 83-107.

Croson, R., Konow, J., 2009. Social preferences and moral biases. Journal of Economic Behavior & Organization 69, 201-212.

Cubitt, R., Drouvelis, M., Gächter, S., 2008. Framing and free riding: emotional responses and punishment in social dilemma games. CeDEx Discussion Paper No. 2008-02, University of Nottingham.

DeScioli, P., Kurzban, R., 2009. Mysteries of morality. Cognition 112, 281-299.

Doris, J., Stich, S., 2005. As a Matter of Fact: Empirical Perspectives on Ethics. In Jackson, F., Smith, M., (Eds.), The Oxford Handbook of Contemporary Philosophy. Oxford University Press, Oxford.

Dufwenberg, M., Gächter, S., Hennig-Schmidt, H., 2006. The framing of games and the psychology of strategic choice. CeDEx Discussion Paper 2006-20, University of Nottingham.

Faravelli, M., 2007. How context matters: A survey based experiment on distributive justice. Journal of Public Economics 91, 1399-1422.

Fehr, E., Fischbacher, U., 2004. Third-party punishment and social norms. Evolution and Human Behavior 25, 63-87.

Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. American Economic Review 90, 980-994.

Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. Nature 415, 137-140.

Fehr, E., Gintis, H., 2007. Human motivation and social cooperation: Experimental and analytical foundations. Annual Review of Sociology 33, 43-64.

Fehr, E., Schmidt, K. M., 2006. The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories. In Kolm, S.-C., Ythier, J. M., (Eds.), Handbook of the Economics of Giving, Altruism and Reciprocity. Elsevier B.V., Amsterdam, pp. 615-691.

Fong, C., 2001. Social preferences, self-interest, and the demand for redistribution. Journal of Public Economics 82, 225-246.

Gächter, S., Herrmann, B., 2009. Reciprocity, culture, and human cooperation: Previous insights and a new cross-cultural experiment. Philosophical Transactions of the Royal Society B – Biological Sciences 364, 791-806.

Gächter, S., Riedl, A., 2006. Dividing justly in bargaining problems with claims: Normative judgments and actual negotiations. Social Choice and Welfare 27, 571-594.

Gaertner, W., 1994. Distributive justice: Theoretical foundations and empirical findings. European Economic Review 38, 711-720.

Gaertner, W., 2009. Distributive Justice: An Overview of Experimental Evidence. In Anand, P., et al., (Eds.), The Handbook of Rational and Social Choice. An Overview of New Foundations and Applications. Oxford University Press, Oxford, pp. 501-523.

Gaertner, W., Jungeilges, J., Neck, R., 2001. Cross-cultural equity evaluations: A questionnaire-experimental approach. European Economic Review 45, 953-963.

Gaertner, W., Schwettmann, L., 2007. Equity, Responsibility and the Cultural Dimension. Economica 74, 627-649.

Gino, F., Moore, D. A., Bazerman, M. H., 2008a. No Harm, no foul: The outcome bias in ethical judgments. Harvard Business School Working Paper 08-080, available at SSRN: http://ssrn.com/abstract=1099464.

Gino, F., Shu, L. L., Bazerman, M. H., 2008b. Nameless + Harmless = Blameless. When Seemingly Irrelevant Factors Influence Judgment of (Un)ethical Behavior. Harvard Business School Working Paper 09-020, available at SSRN: http://ssrn.com/abstract=1238661.

Gintis, H., 2003. Solving the puzzle of prosociality. Rationality and Society 15, 155-187.

Gintis, H., Henrich, J., Bowles, S., Boyd, R., Fehr, E., 2008. Strong reciprocity and the roots of human morality. Social Justice Research 21, 241-253.

Greene, J., Haidt, J., 2002. How (and where) does moral judgment work? Trends in Cognitive Sciences 6, 517-523.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., Cohen, J. D., 2001. An fMRI investigation of emotional engagement in moral judgment. Science 293, 2105-2108.

Greiner, B., 2004. An Online Recruitment System for Economic Experiments. In Kremer, K., Macho, V., (Eds.), Forschung und wissenschaftliches Rechnen GWDG Bericht 63. Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen, pp. 79-93.

Haidt, J., 2001. The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment'. Psychological Review 108, 814-834.

Haidt, J., 2007. The new synthesis in moral psychology. Science 316, 998-1002.

Hauser, M., 2006. Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong. Ecco (HarperCollins), New York.

Joyce, R., 2006. The Evolution of Morality. The MIT Press, Cambridge.

Kahneman, D., Knetsch, J. L., Thaler, R. H., 1986. Fairness as a Constraint on Profit Seeking - Entitlements in the Market. American Economic Review 76, 728-741.

Knack, S., Keefer, P., 1997. Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. Quarterly Journal of Economics 112, 1251-1288.

Konow, J., 2003. Which is the fairest one of all? A positive analysis of justice theories. Journal of Economic Literature XLI, 1188–1239.

Krebs, D. L., 2008. Morality. An Evolutionary Account. Perspectives on Psychological Science 3, 149-172.

Krupka, E., Weber, R., 2008. Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary? IZA Discussion Paper No. 3860, Institute for the Study of Labor, Bonn.
Ledyard, J. O., 1995. Public goods: A survey of experimental research. In Roth, A. E., Kagel, J. H., (Eds.), The Handbook of Experimental Economics. Princeton University Press, Princeton, pp. 111-181.
Levin, I. P., Schneider, S. L., Gaeth, G. J., 1998. All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects. Organizational Behavior and Human Decision Processes 76, 149-188.
McCusker, C., Carnevale, P. J., 1995. Framing in Resource Dilemmas: Loss Aversion and the Moderating Effects on Sanctions. Organizational Behavior and Human Decision Processes 61, 190-201.
McDaniel, W. C., Sistrunk, F., 1991. Management dilemmas and decisions: Impact of framing and anticipated responses. Journal of Conflict Resolution 35, 21-42.
Nado, J., Kelly , D., Stich , S., 2009. Moral judgment. In Symons, J., Calvo, P., (Eds.), The Routledge Companion to Philosophy of Psychology. Routledge, Milton Park.
Nagel, T., 1976. Moral Luck. Proceedings of the Aristotelian Society: Supplementary Volumes 50, 137-151.
Nichols, S., 2004. Sentimental Rules: On the Natural Foundations of Moral Judgment. Oxford University Press, New York.
Nowak, M. A., Sigmund, K., 1998. Evolution of indirect reciprocity by image scoring. Nature 393, 573-577.
Park, E.-S., 2000. Warm-glow versus cold-prickle: a further experimental study of framing effects on free-riding. Journal of Economic Behavior and Organization 43, 405-421.
Prinz, J. J., 2006. The emotional basis of moral judgments. Philosophical Explorations 9, 29-43.
Prinz, J. J., 2007. The Emotional Construction of Morals. Oxford University Press, Oxford.
Rege, M., Telle, K., 2004. The impact of social approval and framing on cooperation in public good situations. Journal of Public Economics 88, 1625-1644.
Ridley, M., 1996. The Origins of Virtue: Human Instincts and the Evolution of Cooperation. Penguin Books, London.
Schokkaert, E., Lagrou, L., 1983. An empirical approach to distributive justice. Journal of Public Economics 21, 33-52.
Sell, J., Son, Y., 1997. Comparing public goods and common pool resources: three experiments. Social Psychology Quarterly 60, 118-137.
Sen, A., 1987. On Ethics and Economics. Blackwell, Oxford.
Sinnott-Armstrong, W., 2006. Consequentialism. The Stanford Encyclopedia of Philosophy (Fall 2008 Edition), Edward N. Zalta (ed.), <http://plato.stanford.edu/archives/fall2008/entries/consequentialism/>.
Sinnott-Armstrong, W. (ed), 2008. Moral Psychology. Vol. 1-3. The MIT Press, Cambridge.
Sonnemans, J., Schram, A., Offerman, T., 1998. Public good provision and public bad prevention: The effect of framing. Journal of Economic Behavior & Organization 34, 143-161.
Tversky, A., Kahneman, D., 1981. The Framing of Decisions and the Psychology of Choice. Science 211, 453-458.

Tversky, A., Kahneman, D., 1991. Loss Aversion in Riskless Choice - a Reference-Dependent Model. Quarterly Journal of Economics 106, 1039-1061.
van Dijk, E., Wilke, H., 2000. Decision-induced focusing in social dilemmas: give-some, keep-some, take-some, and leave-some dilemmas. Journal of Personality and Social Psychology 78, 92-104.
Wheatley, T., Haidt, J., 2005. Hypnotic disgust makes moral judgements more severe. Psychological Science 16, 780-784.
Williams, B., 1981. Moral Luck. Philosophical Papers 1973-1980. Cambridge University Press, Cambridge.
Yaari, M., Bar-Hillel, M., 1984. On dividing justly. Social Choice and Welfare 1, 1-24.
Zelmer, J., 2003. Linear public goods experiments: A meta-analysis. Experimental Economics 6, 299-310.