Journal of Economic Behavior and Organization 176 (2020) 321–336

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

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

Mental accounting of public funds – The flypaper effect in the lab☆

Johannes Becker*, Daniel Hopp, Michael Kriebel

Institute of Public Economics, University of Münster, Wilmergasse 6–8, Münster 48143, Germany

ARTICLE INFO

Article history:
Received 20 November 2019
Revised 13 March 2020
Accepted 30 April 2020

JEL classification:
C92
D72
H31

Keywords:
Mental accounting
Flypaper effect
Lab experiments

ABSTRACT

We report evidence from a series of laboratory experiments that focus on mental accounting of 'public funds'. Groups of three players decide upon how much to redistribute within the group. We measure the preference to redistribute when transfers are made either out of individual accounts (the players' own money) or out of a common account (the group's money). Since the common account is dissolved after each round and paid out to individuals, its size should not affect the decision to redistribute. The experiment is designed to control for anchoring as a potential confounding effect. We find that the size of the common account significantly affects redistribution behavior. Specifically, the transfer increases in the size of the common account. This effect is significantly more pronounced when the transfer is paid out of the common account (instead of the individual account). We interpret these findings as evidence for a flypaper effect due to mental accounting.

1. Introduction

Empirical evidence suggests that voters are more likely to agree to government expenditures out of excess tax revenue than out of new taxes. Precisely speaking, a positive shock in available public funds will result in higher public expenditures than if an equivalent shock occurred in the form of private income. In this sense, money sticks where it hits, which is why Arthur Okun coined the term 'flypaper effect' (Inman, 2009). The flypaper effect is hard to reconcile with standard models of policy-making. According to these models, optimal policy should be expected to equate the social marginal utility of public funds and that of private funds. It should not matter where these funds come from.

Most of the empirical work on the flypaper effect sets out to provide robust evidence for the effect in real world settings. Only a few papers aim at identifying distinct *causes* of the flypaper effect – be it (distorted) politics or behavioral biases on the voters' side. However, due to data constraints, it is often hard to disentangle the different theoretically identified driving forces from each other. A laboratory experiment provides the opportunity to cleanly isolate one single explanation and shut down any other confounding influences.

In this paper, we report evidence from a lab experiment that allows focussing on one specific explanation for the flypaper effect: mental accounting (Thaler, 1985). The concept of mental accounting supposes that households (or voters) assign

☆ We thank Bruno de Borger, Lydia Mechtenberg, Ben Meiselman, Stefanos Tsikas, Valentin Wagner and participants at conferences and seminars in Cologne, Dortmund, Mannheim, Muenster, Lake Tahoe, Paris, Freiburg and Zurich for valuable comments and suggestions. The usual disclaimer applies.

* Corresponding author.
E-mail address: johannes.becker@wiwi.uni-muenster.de (J. Becker).

https://doi.org/10.1016/j.jebo.2020.04.026
0167-2681/

different meanings to different kinds of funds, e.g. privately owned funds and public funds. Hines and Thaler (1995) use this approach to theoretically explain the flypaper effect.

We consider potential flypaper effects in a situation with collective redistribution. Redistribution (or: a fair distribution of income) is one of the most important public goods in modern societies and especially suitable for analysis in economic lab experiments. We use a standard redistribution experiment (as in Tyran and Sausgruber, 2006; Höchtl et al., 2012) in which groups of three players are formed. Two of them ('the rich') receive an amount of money, the third ('the poor') does not. All three players vote on how much is given to the poor player and the intermediate vote (the median voter's proposal) is realized. As a novel element, we split the players' total endowment in an *individual account* and a *common account* which is dissolved and distributed to the rich players' individual accounts at the end of each round after the transfer was given to the poor player. We measure the impact of a variation of the *common account* on the transfer proposals holding constant the participants' total wealth. Since the variation in the *common account* is independent of the variation in the overall economy size, the relative size of the *common account* should, in principle, be completely irrelevant for the transfer proposal. However, if the *common account* is mentally accounted for in a different way than the players' individual accounts (e.g. for purpose of redistribution), the size of the *common account* may have an impact on the transfer proposal even if the players' total endowment is not altered.

In this setting, the mental accounting effect may potentially be confounded with an anchoring effect (Tversky and Kahneman, 1974). Specifically, the size of the *common account* may just be a reference point used for 'anchoring' (and, thus, driving the size of) the proposals. To control for this, we consider two different treatments in order to precisely identify mental accounting and rule out anchoring as the (only) driving force of our results. In the first treatment, the transfer is paid from the rich players' *individual accounts*. In the second treatment, the transfer is paid out of the *common account*. Since the money in the *common account* belongs to the rich players anyway, it should not matter in which way the money is taken from them. Put differently, both treatments are outcome-equivalent, i.e. a given transfer leads to the same final account balances in both treatments.

We find that the transfer proposals and the actually realized transfers are significantly affected by the size of the *common account* (i.e. the public funds). This effect is particularly pronounced when transfers are paid out of the *common account*. To be precise, an increase in the common account by one standard deviation, i.e. nine units of the lab currency, increases the proposal by up to one unit, which corresponds to 17% of the median proposal. Especially the poor voters demand more redistribution when the *common account* is large. Our results prove to be robust in a number of checks. Among others, we change the voting mechanism to a random dictator regime. Whereas the number of extreme proposals (give nothing and full redistribution) increases, the average effect remains unaffected. Finally, to deal with potential concerns about the independence of observations, we restrict the sample to one observation per participant from the first treatment period. This between-subject analysis confirms our main findings.

Our paper adds to the literature by replicating the flypaper effect in the lab – in the absence of all confounding institutional influences – and, thus, informs about concepts like mental accounting, framing and fungibility (to be reviewed in the next section). Our results imply that, even with sound and undistorting political institutions, voters may prefer an increase in public spending when public funds are available but reject them if they imply tax rate hikes. It follows that, depending on where the revenue shock occurs, voter populations with identical preferences may end up with different levels of public spending.

The remainder of the paper is organized as follows. The next section reviews the literature on the flypaper effect and mental accounting. Section 3 describes the experimental design and setup and outlines the hypotheses. Section 4 reports the results of the experiment in the basic design and a range of extensions and robustness checks. Section 5 concludes and discusses some implications.

2. Related literature

In this section, we review the literature on the two main concepts of this paper: the flypaper effect and mental accounting.

The flypaper effect was discovered in the late 1960s when economists estimated the public spending response to different forms of public income like taxes, grants etc. (see Inman, 2009 for a brief historical account). Specifically, public spending barely increases in response to changes in private income whereas it strongly increases when the income change occurs in the public budget (e.g. due to federal grants).[1] Arthur Okun then coined the term 'flypaper effect' reflecting that money sticks where it hits. The early evidence of the flypaper effect, summarized by Gramlich (1977), Inman (1979), Fisher (1982), and Hines and Thaler (1995), has been subject to both, theoretical and econometric criticism. First, there might be straightforward institutional reasons why public funds (e.g. grants from the central government) may not be distributed via tax cuts to households, which is the case e.g. with matching grants (Baker et al., 1999; Brooks and Phillips, 2008). Second, the flypaper effect may well be explained within the neoclassical model of second-best taxation, taking into account the

[1] The flypaper effect is not restricted to the realm of public finance. For instance, Jacoby (2002) analyzes the flypaper effect of a school feeding program on caloric intake within families. Choi et al. (2009) consider the household's portfolio choice and detect a flypaper effect which may be due to mental accounting.

deadweight loss of taxation (Hamilton, 1986; Dahlby, 2011).[2] Third, higher private income is, in some cases, associated with a lower need for public expenditures (Hamilton, 1983); the empirical finding of a flypaper effect may thus be an artifact because the reference point is flawed. Fourth, Knight (2002) argues that grants may be endogenous; controlling for the endogeneity, he shows that the flypaper effect vanishes.

More recent studies take these theoretical and methodological points of criticism into account – and still yield contradictory results.[3] This raises the question under which circumstances a flypaper effect occurs. It is helpful to differentiate between supply side and demand side factors (Bailey and Connolly, 1998). Supply side factor explanations focus on the policy-makers' side, whereas demand side factor explanations argue with the voters' behavior.

On the supply side, there is a broad range of theoretical (political economy) approaches that all have in common that voters have standard preferences which make the flypaper effect undesirable, but the political system (or simply politics) prevents these preferences from being followed. The number of papers is too large to be reviewed here; useful overviews can be found in Bailey and Connolly (1998), Gamkhar and Shah (2007) and Inman (2009).[4]

On the demand side, the flypaper effect is assumed to be caused by the voters' behavior. Courant et al. (1979) and Oates (1979) assume that voters misunderstand their fiscal choice setting ('fiscal illusion').[5] Filimon et al. (1982) assume that voters cannot see through the veil of government budgets and, therefore, are not able to prevent governments from increasing spending. Hines and Thaler (1995) argue that voters may fall prey to political framing.[6] Apart from these arguments based on wrong, distorted or missing information, the flypaper effect may exist because the voters want the government to behave in a way that a flypaper effect occurs (or, at least, do not want to avoid it).[7] Accordingly, Hines and Thaler (1995) point to mental accounting as a cause for the flypaper effect. Mental accounting[8] is generally defined as the "process by which people formulate (...) problems for themselves" (Barberis and Thaler, 2003). The concept of mental accounting has been applied to consumer choices,[9] portfolio choice,[10] development aid[11] and tax avoidance[12] – and may also be applied to fiscal decision problems[13] in order to explain the flypaper effect. This effect occurs if voters mentally account for private funds in a different way than for public funds. Accounts may be associated with different purposes: Public funds are 'meant' to be spent on public goods, private funds should be spent on private consumption goods. This implies that the fungibility of public funds may be limited. If public funds are mentally accounted for by voters in a different way than private funds, the propensity to spend on certain goods may differ across types of funds.

Identifying distinct explanations for the flypaper effect requires (at least) disentangling supply factors (politics) from demand factors (voters' behavioral bias). This is a challenging task and often seems hard to achieve with real world data. Therefore, it does not seem surprising that only a few papers provide (indirect) evidence in this regard.[14] For instance, Inman (2009) cites a number of studies that find that the flypaper effect is bigger in larger jurisdictions.[15] He concludes that "this evidence is sufficient to reject a strict version of the mental accounting explanation" (p. 7) and, instead, points to

[2] As long as there is an excess burden of taxation, the benevolent planner will set a level of public goods that is below the first-best level. If this efficiency cost per unit of tax revenue is increasing at a sufficiently steep slope, an increase in (non-tax) funds may lead to an over-proportional increase in spending.

[3] Knight (2002), Gordon (2004) and Lutz (2010) demonstrate that the flypaper effect is close to zero, at least in the long run. In contrast, a number of studies find robust evidence of the flypaper effect, among them Card and Payne (2002), Hoxby (2001), Baicker (2001), Singhal (2008), Dahlberg et al. (2008), Allers and Vermeulen (2016) and Leduc and Wilson (2017).

[4] Recent examples include Roemer and Silvestre (2002) who build a model in which the flypaper effect occurs as an outcome of multi-party political competition and Singhal (2008) who stresses the impact of political interest groups.

[5] To be specific, a grant is misinterpreted as reducing the marginal price of public goods (although it only lowers the average cost).

[6] "The choices to the public are not framed as between spending the money or cutting taxes, but rather how should the money be spent." (Hines and Thaler, 1995, p. 223)

[7] Vegh and Vuletin (2015) provide a novel argument for a 'rational' flypaper effect. With risky private and public income and incomplete markets, the response of precautionary savings to private income may be different from the response to public income – which may explain the flypaper effect finding.

[8] The basic idea of mental accounting as an explanation for non-standard behavior is established by Thaler (1985, 1990, 1999). The related theory of narrow bracketing is developed i.a. by Tversky and Kahneman (1981) and Rabin and Weizsäcker (2009).

[9] See e.g. Heath and Soll (1996), Milkman and Beshears (2009), Hastings and Shapiro (2013). Abeler and Marklein (2017) report that participants in a field experiment change their consumption behavior depending on the way income is provided (in-kind vs. cash). Clingingsmith (2015) compares the amounts given in a dictator game when the endowment is earned versus when it is received as a windfall and finds that income from these sources is only partially fungible (see also Cherry (2001)). Goerg et al. (2017) show that if income is provided by a party that is perceived as more similar to the recipient, senders give less in a dictator game. In contrast to these papers, we vary the type of account in which the endowment is provided.

[10] See e.g. Choi et al. (2009).

[11] See e.g. Van de Sijpe (2013) who finds that revenue from development aid is not fungible and therefore has the impact the donors intended. Schady and Rosero (2008), among others, evaluate the effectiveness of child benefit by analysing whether sources of household income matter.

[12] See e.g. Fochmann and Wolf (2016).

[13] "We know that changes in housing wealth, pension wealth and future income have very different effects on consumption than equivalent present value changes in current liquid assets or income. So, it should be no great surprise that households violate fungibility in evaluating their political leaders." Hines and Thaler, 1995, p. 223).

[14] Some papers attribute the flypaper effect to demand side factors like mental accounting without checking for confounding supply side factors, see e.g. Heyndels (2001), Heyndels and Van Driessche (1998, 2002).

[15] Romer et al. (1992) and Gordon (2004) find this for school districts and Ladd (1993) as well as Singhal (2008) for state governments.

politics as the most promising explanation. While the exclusive focus on politics receives some support by Singhal (2008)[16] and Leduc and Wilson (2017)[17], other studies show that flypaper effects need not be associated with government slack. For instance, Allers and Vermeulen (2016) show that grants are fully capitalized in housing prices (which suggests that grants do not increase government waste and makes rent-seeking an improbable explanation for the flypaper effect).[18]

To the best of our knowledge, there is no empirical study that cleanly rules out either supply factors or demand factors. In other words, it can usually not be excluded that both, voters' behavioral bias *and* distorted politics, play a role. Accordingly, recent studies often remain vague about the relative influence from both sides. For instance, Gennari and Messina (2014) write: "Politics are confirmed to play a crucial role in local budgeting processes (...). But the presence and the size of the flypaper effect does not seem to be fully ascribed to a misalignment between local policy makers and population. Demand side factors – like fiscal illusion – or behavioral phenomena are likely to play a substantial role." (p. 341). And Allers and Vermeulen (2016) add: "While empirical evidence of the flypaper effect is widespread, its explanation remains unclear." (p. 115).

3. Experimental design, hypotheses and setup

3.1. Design

The literature review above shows that existing studies have, so far, found it difficult to disentangle demand side factors from supply side factors. In this paper, we contribute to the literature by providing clean evidence that there is a demand side driven flypaper effect – using a lab experiment. A lab experiment has the well-known disadvantage of potentially lacking external validity, but it offers the opportunity to shut down all other channels (especially supply side factors) that prevent a clear cut identification of demand side factors in real world studies. In the lab economy, there are, obviously, neither institutional restrictions on how to spend the money nor an efficiency cost of taxation. Moreover, there is neither politics (supply side factors) nor active framing by policymakers in the sense that the funds are either used to cut taxes or to finance a public good. We thus are able to unambiguously identify behavioral biases on the voters' side as the driving force behind a flypaper effect in the lab economy. Due to the limits on external validity, our results can be interpreted as a 'proof of concept'. That is, we show that, even in the sterile environment of the lab where all confounding factors are controlled for or shut down, a flypaper effect may occur. If this is the case, we argue, it is plausible that this behavior plays a role outside of the lab.

We adopt the setting of a classical redistribution game for the experiment. Groups of three players are formed. Two of the players are 'rich' and have an individual endowment, one of them is 'poor' and has no endowment. Each player votes on the amount transferred to the poor player. In such a setting, it has been shown that, although any strategic incentive is absent, the average individual tends to vote for non-zero contributions to the poor player.[19] We introduce, as a novel element, a *common account* which is distributed to the rich players at the end of each round (which makes it effectively part of the rich players' individual fund). By construction, individuals should be indifferent with regard to the size of the *common account*.

We can now describe a flypaper effect in this environment. A flypaper effect occurs whenever the response of public spending (e.g. for redistribution, as in our setting) to an (unexpected) increase in the public budget (e.g. a grant from the federal government) is larger than to an increase in the taxpayers' private budgets. In the framework of our experiment, the public budget is the *common account*. If a change in the *common account* size affects the voting behavior, although it should be completely irrelevant for it, we call this a flypaper effect. In other words, a flypaper effect in our lab economy is identified if money provided in the *common account* is more likely to be spent on 'common purposes' (here: giving to the poor).[20]

We focus on mental accounting as the driving force behind the empirical phenomenon of the flypaper effect. While non-fungibility of funds is a necessary and sufficient condition for the flypaper effect, it is necessary but not sufficient for mental accounting. In other words, there may – in principle – be alternative explanations. In what follows, we discuss these alternatives and how we identify them in the experiment.

[16] Singhal (2008) considers the impact of political interest groups and shows strong flypaper effects for state income due to settlement agreements with the tobacco industries (presumably due to the effort of these interest groups). After these settlement agreements, states substantially increased spending on tobacco prevention and control programs.

[17] Leduc and Wilson (2017) provide some "suggestive evidence" (p. 258) that political contributions by the private public works sector to state policy makers partially drive the flypaper effect.

[18] Langer and Korzhenevych (2019) do not find evidence that transfers to German municipalities increase general administrative expenditures; rather they are mainly used for social expenditures which they interpret as reflecting voters' preferences.

[19] For instance, Engel (2011) derives a giving ratio of .283 in a meta-study on dictator games. In ultimatum games the mean offer is about 40% of the total endowment (Tisserand, 2014).

[20] A related approach is chosen by Fosgaard et al. (2014) who consider a public good game where participants can give to or take away from the public good. Although these two mechanisms are economically equivalent, the cooperation changes. The authors show that this is mainly due to changed beliefs about the other participants' behavior – an aspect which is of minor importance in our study.

To start with, players may misunderstand their choice environment ('fiscal illusion') and misinterpret public funds as exclusively designated for public purposes etc. We take a number of measures (to be outlined below) to increase the transparency of the experiment in order to make sure that it is not a lack of understanding that drives our results.

Moreover, the flypaper effect may be due to anchoring. Anchoring is defined as the process of evaluating a problem (or estimating a value) "by starting from an initial value that is adjusted to yield the final answer. The initial value, or starting point, may be suggested by the formulation of the problem" (Tversky and Kahneman (1974), p. 1128). In our context, this initial value may be the sizes of the common and the individual accounts. If the size of the *common account* affects the players' evaluation of what is the 'right' amount of public good, there may be a flypaper effect even without mental accounting. In the experiment, we will make sure that anchoring is not the driving force behind our findings. For this purpose, we introduce two treatments that differ in the mode of financing (i.e. which account is used for paying out the transfer) but not in the available information. Thus, if anchoring does play a role, it should affect both treatments equally which allows for cleanly identifying the role of mental accounting.

Confusion and anchoring are the two main alternative explanations apart from mental accounting. There are, however, concepts related to mental accounting, and it is important to clarify the relationship between these concepts.

The first of these related concepts is framing. "Framing refers to the way a problem is posed for the decision maker." (Barberis and Thaler (2003), p. 1071). So framing is done by the experimenter, mental accounting by the players. As is well-known, there is no environment free of framing; all we can do here is to make sure that the framing part does not preclude certain kind of behavior. To be specific, for there to be mental accounting there need to be at least two accounts which differ in some aspect. In the experiment, we vary the strength of framing to shed further light on this issue. In the baseline version of the experiment, we use the labels *common account* and *individual account* to frame the experiment (see Beatty et al. (2014) for recent evidence of the labeling effect). In a robustness check, we vary the labels in order to analyze the role of framing.

The second, closely related concept is salience. Accounts that differ in salience may be subject to mental accounting. In other words, salience may facilitate or mitigate mental accounting. Since we do not vary the salience of the two accounts across treatments (apart from a robustness check where the labeling effect is analyzed), we are confident that salience does not play a central role for the understanding of our results.

A third concept is loss aversion.[21] Loss aversion may affect our experiment only if the two accounts are treated differently and are, thus, subject to mental accounting. To be precise, the rich players could perceive their *individual account* as their endowment and the *common account* as a potential gain. A transfer out of the individual account is therefore considered as a loss, whereas a contribution out of the common account is a foregone gain. This does not contradict the mental accounting hypothesis, it rather is a specific version of it. Fortunately, our experiment allows for checking whether this version has explanatory power. The poor players should not be affected by loss aversion since none of the two accounts may be interpreted as their endowment. If the poor players do not respond to the *common account*, loss aversion is a plausible story to explain the flypaper effect in our context.

3.2. Hypotheses

Based on the considerations above, we can now state the following hypotheses which are tested in the lab. To start with, we test for unbiased behavior. In the experimental setting, the size of the *common account* (given the economy size) should be completely irrelevant for the transfer proposals.

Hypothesis 1 (unbiased behavior): *For a given economy size, the size of the common account does not affect the transfer proposal.*

In case that H1 is rejected, the question arises *why* the *common account* has an impact. Here, we consider two candidates: mental accounting and anchoring. Mental accounting refers to the source of the funding, i.e. which account is used to finance the transfer. With mental accounting, the proposals for a transfer taken out of the *individual account* are lower than those taken out of the *common account*. Anchoring refers to the size of the *common account* that may affect the level of transfers even in the absence of mental accounting. A larger *common account* makes participants "anchor" on a higher transfer level that is perceived to be fair.

Our approach is to vary the funding source for identical common account sizes, i.e. to hold the degree of anchoring stable while varying the environment for mental accounting. If this variation does not affect the transfer proposal, we will conclude that the flypaper effect is not due to mental accounting.

Hypothesis 2 (no mental accounting): *For a given size of the common account, a change in the funding source (the account) does not affect the transfer proposal.*

If, however, a variation in the funding source affects the transfer for a given *common account* size, we will interpret this as an indicator for mental accounting and will conclude that it is not anchoring (alone) that is responsible for the observed flypaper effect.

[21] Hines and Thaler (1995) themselves argue that loss aversion may explain the flypaper effect. If voters do not consider the public budget as "theirs", loss aversion implies that taxpayers are "much more sensitive to decreases in their welfare than to increases. This implies that the political cost of explicitly raising a tax is greater than the political benefit of an equivalent tax cut." (p. 223). This creates a policy discontinuity where additional need for public funds does not necessarily lead to tax hikes, and additional funds (grants) do not necessarily lead to tax cuts – again, a flypaper effect occurs.

3.3. Experimental setup

The experiment took place at the University of Muenster in November 2015, January and June 2016 and June 2018. 408 students (189 female and 219 male) from various disciplines (Business: 29.9%, Economics: 16.4, Law: 14.5, other: 39.2) drawn from a pool of students signed up in the online recruitment system ORSEE (Greiner, 2015) participated in 24 sessions. Each session was conducted with 15 or 18 participants who had no opportunity to communicate during the experiment. Written instructions were handed out on paper and read aloud by the experimenter. Thereafter, participants had to do two control tasks (see Online appendix for an English translation). Each session lasted approximately 80 minutes and finished with a short questionnaire. The questionnaire as well as the experiment were programmed in z-Tree (Fischbacher, 2007). On average, participants earned 17.17 Euros.

The experiment is a classic redistribution game with three participants in each group, augmented by a novel element: the *common account*. The game is played 24 times (periods) in each session with each period having five distinct stages.

Stage 1 (Group formation and role assignment) All participants are randomly and anonymously matched in groups of three players at the beginning of each period. With 24 periods and 18 (or 15) players, participants are expected to interact with the same person more than once (but do not know, of course, when this is the case). Each player is randomly assigned a 'type'. Within each group, two players are assigned the type *Green*, the third one the type *Blue*. Thus, each participant is expected to be *Blue* type player eight times and *Green* type player sixteen times during the experiment.[22] Types *Green* and *Blue* differ in endowments.[23]

Stage 2 (Endowments) The two type *Green* players receive a random but identical endowment of lab currency ("points", 30 points equal 1€) on their *individual account*, the type *Blue* player receives an endowment of zero. Next to their *individual account*, the two *Green* types are endowed with a random amount in a *common account*. Randomization is realized in two independent draws from a uniform distribution. First, we let the computer stochastically determine the size of the economy out of the interval [30, 60]. Then, the size of the *common account* is drawn from the interval [0, 30]. Thus, the *common account* can never be larger than the total size of the economy. The difference between the size of the economy and the *common account* is then split in half and assigned to the *Green* types' *individual accounts*.[24] All group members know each others' account balances as well as the *common account* balance.

Stage 3 (Transfer proposals, voting) At this stage, the group members vote on a transfer from the *Green* types to the *Blue* type player. All three players are asked to make a proposal on how large the transfer is supposed to be. From the three proposals, the intermediate one is realized (which is equivalent to median voter rule). Only the realized transfer will, at stage 5, be revealed to the group members. In theory, the median voter mechanism provides no incentive to overstate or understate the taste for redistribution and, thus, allows us to elicit the individuals' true preferences.[25]

Stage 4 (Financing of transfer) The treatments differ in how the transfer is financed. In the INDIVIDUAL treatment, each of the type *Green* players pays half of the transfer out of her *individual account*. After that, the *common account* is dissolved and paid out to the type *Green* players' *individual accounts*. If the initial *individual account* is lower than half of the transfer, it turns negative; the deficit is later on compensated by the payment out of the *common account*. The maximum transfer cannot exceed the overall endowments by the type *Green* players. In the COMMON treatment, the transfer is paid out of the *common account*. Afterwards, the remainder of the *common account* is paid out to the *Green* types' *individual accounts*. If the *common account* turns negative due to a high transfer, the deficit is compensated out of the *individual accounts*. One of the control tasks that have been solved by the participants before the experiment includes a case with a negative *common account* after the transfer has been paid (see Online appendix).

Stage 5 (Payoffs) At the end of each period, the realized transfer and all group members' final account balances are revealed to the whole group. Each player's period income (i.e. their final account balance) is stored. The aggregate income is paid out at the end of the experiment.

Since in both treatments, the two type *Green* players receive half of all money in the economy after deducting the transfer to the type *Blue* player, the size of the *common account* should be completely irrelevant for the transfer proposals. In other words, both type *Green* players should realize that they pay half of the transfer out of their pockets. Similarly, the type *Blue* player should not care about the source of the transfer (i.e. whether it is paid out of a *common account* or out of an *individual account*).

In the presence of mental accounting, a large *common account* may imply that the type *Green* players want to finance a large transfer and the type *Blue* player demands a large transfer. The *common account* (public funds) is then interpreted as being 'meant' to be used for group interests which is, here, redistribution.

[22] The fact that participants are of differing types over the course of the experiment weakens the motivation for redistribution – which may imply an underestimation of the effect. Since, however, we exploit the difference between treatments in which the motivation to redistribute is not altered, our results are, at least qualitatively, unaffected.

[23] The literature (Cherry et al. (2002) and Oxoby and Spraggon (2008)) suggests that redistribution out of windfall endowments is larger compared to redistribution out of earned income. While this may be the case, our results do not depend on the absolute inclination to give to the poor but on the difference between treatments, see below.

[24] This calibration ensures that for each of the *Green* type players the size of the *common account* equals the size of each of the *individual accounts* in expected values.

[25] However, recent studies show (e.g. Strulovici (2010)) that participants who are not exclusively motivated by money may exhibit strategic behavior. For that reason, we check the robustness of our results to a change in the voting mechanism implementing a random dictator regime (see Section 4.3).

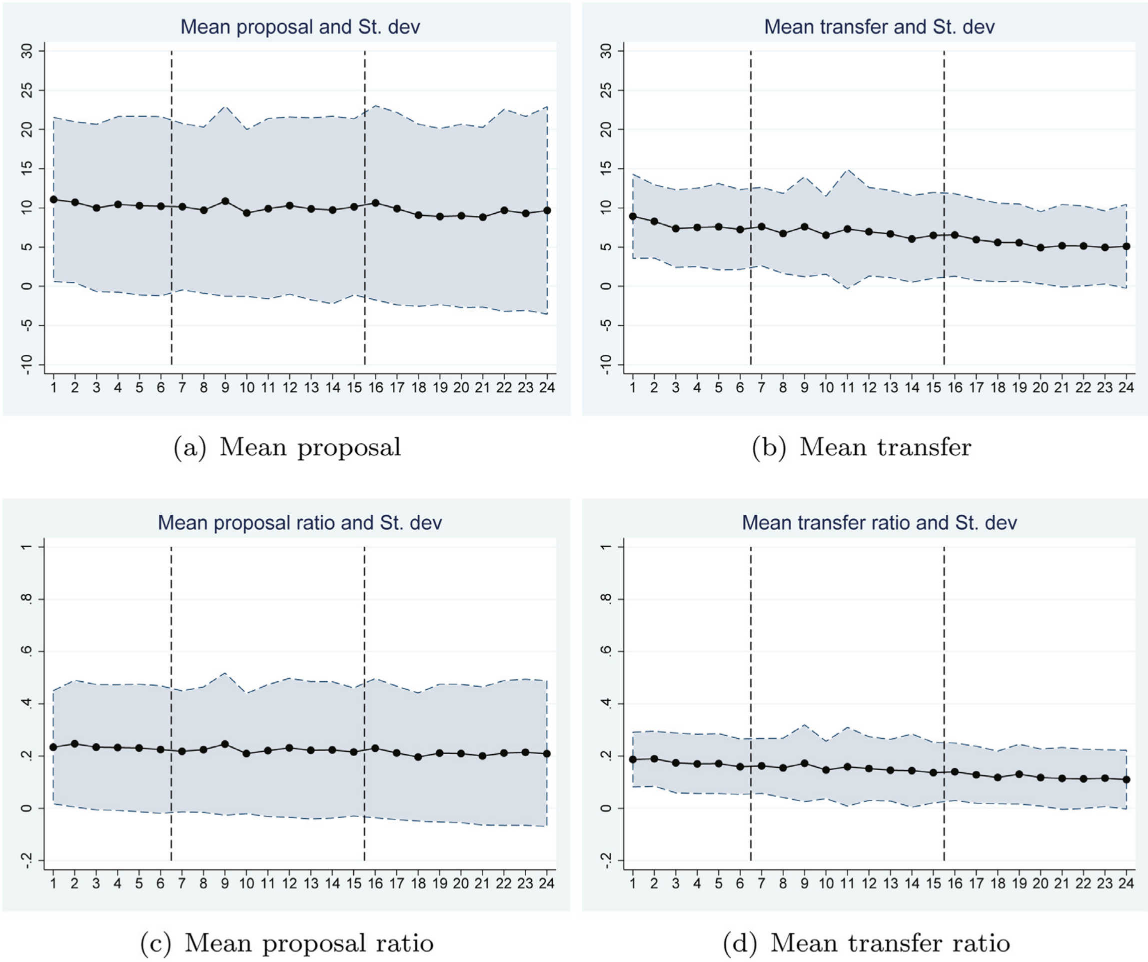


Fig. 1. Mean proposals and mean transfers as absolute values ((a) and (b)) and as a fraction of the size of the economy ((c) and (d)).

However, the literature suggests that individuals sometimes anchor their decisions on some salient, though irrelevant number. Such an irrelevant number may be the *common account*. In order to rule out that this number is the only driving force behind the transfer proposals, we use the variation within treatments. Note that, in both treatments, there is a *common account*. The only difference is that the transfer is either taken out of the *individual account* (INDIVIDUAL treatment) or the *common account* (COMMON treatment). In terms of payoffs, this differentiation is meaningless; in terms of mental accounting, it may affect the transfer proposal.

In the first six periods, the *common account* is endowed with zero points. We refer to this phase as the CONTROL treatment which allows us to collect information on the participants' propensity to give. In periods 7–15, participants are either in the INDIVIDUAL treatment or the COMMON treatment, in periods 16–24, the treatment is switched. We have thus between-subject and within-subject variation.

4. Results

4.1. Basic results

Fig. 1 shows (a) the mean proposal over the 24 periods and (b) the mean realized transfer. To increase comparability between averages, we scale the proposal and the transfer by the size of the economy (equal to sum of the initial endowments of type *Green* players plus the *common account*), see panels (c) and (d). In period 1, the average proposal (ratio) is 11.1 points (23.4%). There is a slight downward trend, but even in period 24, the players still propose on average 9.7 points (20.9%) to the poor player (the type *Blue* player).[26] The realized transfer is somewhat smaller (since it is usually a type *Green* player

[26] In Appendix A.3 we provide the mean proposal ratios over time separated by type.

Table 1
Descriptive statistics

	N	Participants	Proposal (SD)	Proposal ratio (SD)	Extreme proposal ratios	
					zero	max
Total	9792	408	9.91 (11.53)	0.22 (0.25)	0.23	0.04
Contr.	2448	408	10.45 (10.91)	0.23 (0.24)	0.18	0.04
COM	3672	408	9.83 (11.95)	0.22 (0.26)	0.25	0.04
IND	3672	408	9.62 (11.50)	0.22 (0.25)	0.25	0.03
Female*	4536	189	9.54 (8.99)	0.21 (0.20)	0.06	0
Male*	5256	219	10.22 (13.34)	0.23 (0.29)	0.17	0.03
COM-IND**	4752	198	9.28 (11.50)	0.21 (0.25)	0.12	0.02
IND-COM**	5040	210	10.50 (11.53)	0.23 (0.25)	0.11	0.02
Green						
Contr.	1632	408	5.61 (5.62)	0.13 (0.12)	0.28	0
COM	2448	408	4.16 (5.40)	0.09 (0.11)	0.37	0
IND	2448	408	4.36 (5.74)	0.10 (0.13)	0.37	0
Blue						
Contr.	816	408	20.13 (12.38)	0.45 (0.26)	0	0.12
COM	1224	408	21.17 (13.32)	0.47 (0.29)	0	0.13
IND	1224	408	20.13 (12.85)	0.45 (0.28)	0	0.10
	N	Participants	Transfer (SD)	Transfer ratio (SD)	Extreme transfer ratios	
					zero	max
Total	3264	408	6.58 (5.44)	0.15 (0.12)	0.15	0
Contr.	816	408	7.82 (5.13)	0.18 (0.11)	0	0
COM	1224	408	6.12 (5.41)	0.14 (0.12)	0.17	0
IND	1224	408	6.20 (5.55)	0.14 (0.12)	0.16	0
COM-IND	1584	198	5.64 (5.34)	0.13 (0.12)	0.19	0
IND-COM	1680	210	7.46 (5.39)	0.17 (0.12)	0.11	0

Notes: Proposal/transfer ratio as share of proposed/transferred points of economy size. Std. Dev. in parentheses.
Zero (max) proposal/transfer as share of proposals/transfers with zero (maximum) points of all proposals/transfers.
*: A Mann–Whitney *U* test fails to reject the null hypothesis that average proposal ratios are equal across gender.
**: A Mann–Whitney *U* test fails to reject the null hypothesis that average proposal ratios are equal between sequences.

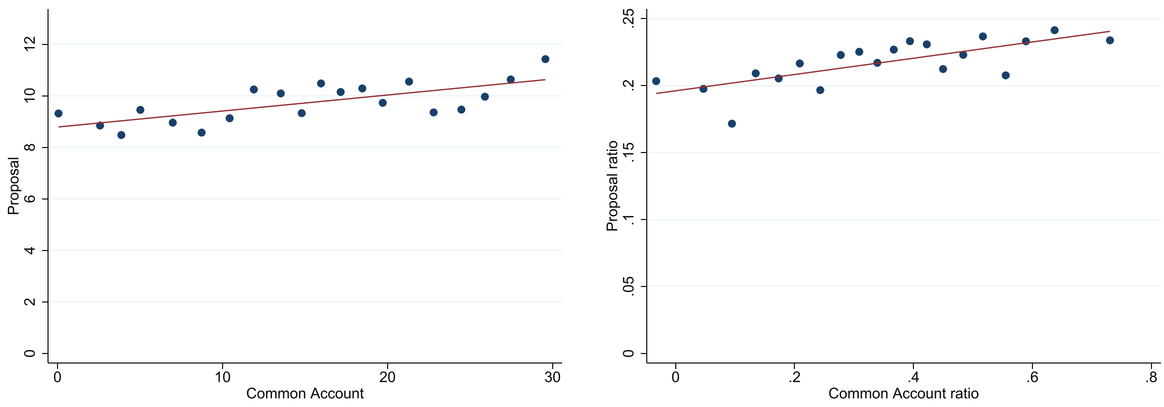
making the median proposal) and the downward trend is more pronounced. However, transfers are on average still more than 5 points (10%) in the last period, leaving the type *Green* player with a bit less than 45% of the economy.

Table 1 shows the proposal (ratio) for the total sample and different subsamples. The total number of observations is 9792. Note, though, due to the repetition and the repeated interaction within a session, the number of independent observations is smaller. We deal with the independence of observations further in Section 4.3. The average proposal ratio is 22.1% of the economy size. The proposal ratio is very similar across all treatments regarding both the average and the extreme proposals, the exception being that zero proposals are less frequent in the control treatment (periods 1–6). A quarter of participants proposes a zero transfer to the type *Blue* player; these are exclusively type *Green* and mostly male players. 4% in total (12% of the type *Blue* players) propose the full economy size.[27] Again, this kind of selfish proposing behavior is mostly found with male players.[28] Around half of all observations (4752) are from a sequence of treatments where the COMMON treatment comes first (periods 7–15) and the INDIVIDUAL treatment second (periods 16–24). The other half (5040) has the opposite sequence. Voting behavior between sequences does not fundamentally differ. The average ratio of realized transfers to the economy size is 15% and fairly similar across treatments.

So, while there is no treatment effect on the total average proposal, there is a sizeable impact of relative *common account* size variations on the proposal *within* treatments. Fig. 2 (a) shows the proposals (y-axis) for different *common account* sizes (x-axis) in a binned scatter plot controlling for the economy size and demographic characteristics. In the whole sample, the size of the *common account* is associated with larger amounts proposed to be given to the type *Blue* player. A similar correlation is found for proposals and *common accounts* as a fraction of the economy size (see Fig. 2 (b)).

[27] Those players also drive the large mean proposal ratio of 46% for the type *Blue* players. If we exclude these 12% the average proposal ratio decreases to 39%.

[28] However, on average women propose a slightly lower level of redistribution than men (21–23%).



(a) Absolute values (b) Relative values

Fig. 2. Proposal (ratio) and common account (ratio).

4.2. Analysis

To gain further insights, we now turn to regression analysis. Let i be the index of individuals, j the index of a group and t the time index. The proposal for a transfer to the *Blue* type is denoted by $Prop_{it}$ and the size of the economy by $Size_{jt}$.

A test for Hypothesis 1 is that the transfer proposal is unaffected by the *common account* size, holding the economy size constant. Accordingly, the variable of interest is the proposal, $Prop_{it}$.

In the first six periods, the *common account* is fixed at zero. We use this phase to collect information on the participants' general willingness to give. Controlling for all observed and unobserved heterogeneity (by using fixed effects estimation), the average proposal is 10.45 points, which corresponds to 23% of the median economy size. The mean realized transfer is 7.81 points which corresponds to 17% of the median economy size.

After period 6, the two type *Green* players have a *common account*, denoted by $Common_{jt}$. As indicated above, the size of the *common account* is determined independently from the size of the economy. We start by regressing the proposal on the *common account*, allowing the size of the economy to have an impact:

$$Prop_{it} = a_0 + a_1 Common_{jt} + a_2 Size_{jt} + \mathbf{bX}_i + \tau_i + \varepsilon \tag{1}$$

where $\mathbf{X}_i$ denotes the vector of player-specific socio-demographic variables and τ_i the unobserved time-invariant individual effect.

Table 2 reports the regression results. It shows that the size of the *common account* has a significant impact on the size of the proposal. Starting with a simple OLS estimation in specification (1), an increase in the *common account* by one standard deviation (9 points) increases the proposal by 0.56 points, which corresponds to 9% of the median proposal. Note that we control for the size of the economy, i.e. a variation of the *common account* does not, in itself, add to the potential consumption of the two type *Green* players.

The fact that the size of the *common account* has a significant impact rejects Hypothesis 1 (stated in Section 3.2). It is generally in line with the idea that the mental accounting of public funds differs from the one of private funds. So far, however, we cannot rule out that this effect is due to simple anchoring. Players may take the *common account* as some (arbitrary) reference at which their proposal is measured. To rule out anchoring as the only driving force behind this, we now make use of the two treatments, COMMON and INDIVIDUAL. In both treatments, the *common account* is a potential anchor, but only in the COMMON treatment, the transfer will be directly financed out of public funds, i.e. the *common account*.

Table 3 measures the impact of the *common account* on the proposal separately for each treatment. It illustrates the substantial difference between the treatments. The coefficient estimate on the *common account* is almost three times as large in the COMMON treatment compared to the INDIVIDUAL treatment and it is significant on the 1% level. To illustrate the magnitude of the effect, we calculate the elasticity of the transfer proposal with respect to the *common account*, $\frac{\partial Prop}{\partial Common} \frac{Common_{med}}{Prop_{med}}$. Evaluated at the median values, the elasticity is 0.28, i.e. a 10% increase in the *common account* increases the proposal by almost 3%. Column (2) of Table 3 shows that the effect of an increase in the *common account* holding constant the economy size is very similar to the effect of an increase in the economy size holding constant the *common account* in the COMMON treatment.

In what follows, we distinguish between the treatments by using a dummy variable in the full sample. Let COMMON denote a dummy that equals 1 if the participant is in the COMMON treatment and 0 if not. The regression equation is then given by

$$Prop_{it} = a_0 + a_1 Common_{jt} + a_2 Common_{jt} \times COMMON + a_3 COMMON + a_4 Size_{jt} + a_5 Size_{jt} \times COMMON + \mathbf{bX} + \tau_i + \varepsilon \quad (2)$$

The results are reported in Table 4. The simple OLS estimation in column (1) shows that the sum of the coefficients $a_1 + a_2$ is larger than a_1 alone, but imprecisely measured. In the subsequent estimation, we add some obvious controls (2), period dummies (3), and all available controls (4), before eliminating all individual-specific time-invariant heterogeneity by the use of fixed effects (5) and period dummys (6). Including period dummies, the impact of the *common account* in the COMMON treatment is more than twice as large as in the INDIVIDUAL treatment, and the difference is statistically significant on the 1% level. We may thus reject Hypothesis 2 (Section 3.2) and conclude that the observed flypaper effect is not due to anchoring alone. Adding the two coefficients, this implies that an increase in the *common account* by one standard deviation (9 points) increases the proposal by 1 point (i.e. 17% of the median proposal).[29]

Table 5 shows results for different types of players. Columns (1) to (3) reveal that the impact of the *common account* in the COMMON treatment is clearly significant for both type *Green* and type *Blue* players as well as for the median voter. The impact is more than twice the size for the type *Blue* players compared to the type *Green* players and the treatment effects are significantly different (see column 4). Thus, mental accounting seems to be especially important for the *demand* for redistribution. Type *Blue* players seem to shy away from demanding points out of the type *Green* players' *individual account* as those points are perceived as belonging to the type *Green* players. In contrast, the *common account* is not (yet) clearly assigned to one specific player and, thus, the propensity to demand higher redistribution is larger.

As Table 1 shows, the proposals are subject to heterogeneity. For instance, one third of the proposals are either extremely low or extremely high, i.e. these participants propose zero points as a *Green* player and demand the total amount of points as a *Blue* player. To distinguish different types of participants with respect to their proposal behavior, we split the sample into four parts. We assign types according to the mean proposal ratios, i.e. mean proposal as a fraction of economy size, as *Green* players in the first six rounds, in which the *common account* balance is zero. Analyzing these groups separately yields that the most selfish and the most generous group do not respond to the COMMON treatment. The treatment effect is mainly driven by the other half of the participants, who exhibit moderate redistribution preferences.[30]

4.3. Robustness checks

We do not find differences in behavior between men and women with respect to the impact of the treatment. However, mental accounting only seems to occur with participants with a below median Abitur grade in math. This finding is in line with other studies that show that some behavioral anomalies are related to skills (see e.g. Abeler and Marklein (2017)).[31] This difference in the exposition to the treatment effect along cognitive abilities may suggest that our results are driven by confusion about the experimental design rather than by mental accounting. We address this concern in several ways to rule out that the results are due to misunderstandings. First, participants might think that they cannot give more than the endowment in the *common account*. Before the experiment starts, we explicitly deal with this source of misunderstanding by providing a control task that highlights contributions which are larger than the size of the *common account* (see control tasks in the Online appendix). Reassuringly, in the experiment, a quarter of all proposals actually exceeds the *common account* of the respective period. Moreover, for this misunderstanding to have explanatory power, we would expect strong bunching of proposals at the *common account* size. The Appendix shows that there is only mild bunching, so this should be of minor importance. The distribution of the proposal to *common account* ratio is similar for both ability types and does not indicate that the aforementioned problem is relevant. Moreover, the bunching occurs in both treatments and, thus, leaves our main identification strategy unaffected (see A.2). Second, another reason for confusion may be that the source of the transfer alters after fifteen periods. However, if we restrict the sample to observations from periods 16 to 24 (where all changes in the financing mechanism have taken place), the results are still the same: an increase in the *common account* only affects the proposal in the COMMON treatment. Furthermore, we find stable shares of transfers of the economy size across treatments (see Table 1), i. e. potential misunderstandings concerning the difference between the transfer mechanisms does not affect the degree of redistribution. Accordingly, we are confident that our results are not driven by any misunderstanding of the instructions.

Is there a 'labeling effect' (see e.g. Beatty et al. (2014))? That is, does the label '*common account*' drive the results by making participants propose higher transfers for public purposes? To address this concern, we change the labels from '*in-*

[29] We explicitly checked for whether the sequence of treatments is important. In both sequences, the *common account* significantly affects the proposal in the COMMON treatment on the 1% level. In the INDIVIDUAL treatment, there is an impact of the *common account* in rounds 7–15. A possible explanation is that the player's attention is drawn to the size of the *common account* after it is introduced but not used for funding. In any case, the effect becomes larger when the transfer is actually paid out of the *common account* in periods 16–24. The regression results are reported in Table 1 in the Online appendix.

[30] More specifically, we take the quartiles of the distribution of these means to form four groups: very selfish (average proposal ratio of .007), mildly selfish (.05), mildly generous (.12) and very generous players (.20). While the *very generous* group does not respond to the COMMON treatment, it does so to the size of the *common account* independent of treatment. The regression estimates are reported in Table 2 in the Online Appendix.

[31] See column (1) and (2) in Table 3 in the Online appendix for the comparison of these subgroups.

dividual account' into '*account 1*' and from '*common account*' into '*account 2*'. The results show that mental accounting does not seem to be driven by the labels of the accounts.[32]

Moreover, the median voter mechanism may be questioned. The literature provides some arguments that the voting mechanism may influence the results due to strategic behavior (see e.g. (Strulovici, 2010)).[33] We therefore replace the median voter procedure by the random dictator procedure. Note that this design is identical to the baseline design in all regards except for the decision rules. The results reported in Table 6 in the Online Appendix show that our main findings are not affected by the voting mechanism, which is in line with Höchtl et al. (2012).[34]

Finally, there may be concerns about the independence of observations. In each session, the 15 or 18 participants interacted with each other. This is a compromise between two goals: on the one hand, preserving as much as possible the one-shot game character of the game played in each period and, on the other hand, generating a sufficient number of observations. An obvious downside is that the observations in a given session are not perfectly independent and may therefore give rise to biased estimations. To cope with this concern, we restrict the sample to observations from period 7 (the first period with a non-zero *common account*). Reassuringly, the estimation results fully support our findings based on the full sample: The *common account* significantly increases the proposal.[35]

5. Conclusion

This paper proposes a lab experiment in order to identify a specific cause of the flypaper effect. With real world data, a precise identification of distinct causes is often difficult – if not impossible – since supply side factors (i.e. politics) and demand side factors (voters' behavioral bias) are hard to disentangle.

In the experiment, individuals decide on 'public spending' (here: redistribution to the poor) either out of their individual accounts or out of a common account (the 'public budget'). Since the common account belongs to the 'rich' individuals anyway, the size of the common account should not have any impact on the decision on how much to spend, i.e. give to the poor. However, the experiment shows that the transfer to the poor player increases in the size of the public budget. Thus, a positive shock in the common account has a stronger effect on spending than an equivalent shock in the private account – a flypaper effect occurs. By contrasting two settings in which the transfer is financed either out of the common or the individual account (a pure accounting difference), we rule out that this finding is due to anchoring.

The admittedly artificial situation in the lab somewhat restricts the scope for interpretation of our findings with regard to real world spending decisions. Therefore, the contribution to the understanding of real world flypaper effects depends on how strong one thinks the external validity of the above experiment is.

At a modest (or precautious) level, our finding may be interpreted as a 'proof of concept'. In a completely controlled and transparent environment, individuals opt for higher spending if the common account is larger, and vice versa. If mental accounting can be observed in this transparent, often repeated situation, it may at least establish the argument that this kind of non-standard behavior *may* affect real-world decisions. To be sure, in the (noisier) real world, institutions and politics may amplify or mitigate this behavior. Our findings do, however, lend conceptual support to empirical work (with real world data) that contradict Inman's (2009) conclusion that the existing "evidence is sufficient to reject a strict version of the mental accounting explanation" (p. 7). It should however be explicitly stated that the experimental findings do not allow for quantitative conclusions.

If one accepts, for a moment, our findings as relevant for real world decision-making – what are the policy implications of a flypaper effect due to mental accounting? First of all, it implies that the observed flypaper effects are not exclusively an indicator for a dysfunctional political system where politicians and bureaucrats engage in rent-seeking activities. Second, mental accounting makes welfare analysis more complicated. That is, depending on where the income occurs, an increase in public spending may be welfare-enhancing or not. Third, from the viewpoint of an upper-level government that decides on grants for lower-level governments (or an international organization that distributes aid), mental accounting can be used to affect the way voters *want to* spend the grant money (or aid).

[32] See Table 3 in the Online appendix for the regression results.

[33] For example, participants might hide behind the median voter and therefore take extreme positions, i.e. they propose either zero or the whole economy depending on their type. Shayo and Harel (2012) and Kamenica and Brad (2014) find factors influencing the voters' decision that go beyond someone's material motivation especially when the probability of being pivotal is very small.

[34] While mental accounting can be identified in all designs, the random dictator rule seems to lead to more extreme proposals (basic design: .301; random dictator design: .533). One possible explanation is based on the fact that, if the decisive individual is chosen randomly from the three group members, it is more likely that an extreme proposal is realized as a transfer. This is reflected by the share of extreme transfers of the total number of transfers (basic design: .180; random dictator design: .556). Therefore, in contrast to the median voter rule, the revelation of very selfish preferences is as likely as the revelation of more moderate preferences. As a consequence, participants with a moderate preference for redistribution may respond to the new information and adjust their behavior.

[35] See pages 14 and 15 in the Online Appendix for more details.

Appendix. A.1 Tables

Table 2
Effect of *Common Account* on Proposal (total sample)

	Dependent variable: *Proposal*					
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) FE	(6) FE
Common Account	0.0628***	0.0623***	0.0620***	0.0622***	0.0670***	0.0670***
	(0.0162)	(0.0163)	(0.0167)	(0.0167)	(0.0133)	(0.0148)
Economy Size	0.1437***	0.1430***	0.1415***	0.1418***	0.1451***	0.1438***
	(0.0158)	(0.0157)	(0.0162)	(0.0162)	(0.0126)	(0.0134)
Abitur Grade		−0.0007	−0.0007	0.0064		
		(0.0040)	(0.0040)	(0.0051)		
Female		−0.9019**	−0.9021**	−0.7229*		
		(0.4346)	(0.4352)	(0.4321)		
Economics		−0.5320	−0.5324	−0.5382		
		(0.4345)	(0.4351)	(0.4376)		
Constant	2.262***	3.100***	2.964**	4.286***	2.132***	1.9868**
	(0.7047)	(1.1051)	(1.2260)	(1.5846)	(0.6743)	(0.8210)
Period Dummys	NO	NO	YES	YES	NO	YES
Controls	NO	NO	NO	YES	NO	NO
Observations	7344	7344	7344	7344	7344	7344
Cluster/Groups	408/-	408/-	408/-	408/-	24/408	24/408

Notes: Standard errors in parentheses. In OLS estimates, standard errors are clustered by individual and in FE estimates by sessions. FE regressions include individual fixed effects.
* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Table 3
Comparison of treatments (FE)

	Dependent variable: *Proposal*	
	(1) INDIVIDUAL	(2) COMMON
Common Account	0.0401***	0.1102***
	(0.0141)	(0.0192)
Economy Size	0.1471***	0.1365***
	(0.0165)	(0.0149)
Constant	2.863***	1.784**
	(0.8336)	(0.7921)
Observations	3672	3672
Cluster/Groups	24/408	24/408

Notes: Standard errors in parentheses. Standard errors are clustered by session. Regressions include individual fixed effects. Estimates include period dummies.
* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Table 4
Effect of *Common Account* across treatments (total sample)

	Dependent variable: *Proposal*					
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) FE	(6) FE
COMMON Dummy	1.511	1.541	1.324	1.354	0.135	-0.168
	(1.473)	(1.462)	(1.478)	(1.475)	(0.688)	(0.725)
Common Account	0.0411**	0.0386**	0.0400**	0.0409**	0.0312**	0.0320**
	(0.0216)	(0.0215)	(0.0218)	(0.0217)	(0.0149)	(0.0151)
Common Account × COMMON Dummy	0.0434	0.0448	0.0412	0.0420	0.0732***	0.0709***
	(0.0330)	(0.0330)	(0.0328)	(0.0325)	(0.0187)	(0.0186)
Economy Size	0.165***	0.165***	0.160***	0.161***	0.158***	0.152***
	(0.0205)	(0.0204)	(0.0207)	(0.0207)	(0.0149)	(0.0162)
Economy Size × COMMON Dummy	−0.0443	−0.0454	−0.0391	−0.0400	−0.0238	−0.0160
	(0.0318)	(0.0316)	(0.0323)	(0.0322)	(0.0183)	(0.0190)
Constant	1.539	3.706***	3.716***	3.692**	2.026***	2.050**
	(0.938)	(1.331)	(1.424)	(1.635)	(0.713)	(0.863)
Period dummys	NO	NO	YES	YES	NO	YES
All controls	NO	NO	NO	YES	NO	NO
Observations	7344	7344	7344	7344	7344	7344
Cluster/Groups	408/-	408/-	408/-	408/-	24/408	24/408

Notes: Standard errors in parentheses. COMMON dummy takes the value of 1 in the COMMON treatment and 0 otherwise. Basic controls (e.g. Abitur grade, gender) are included in columns 2 to 4. In OLS estimates, standard errors are clustered by individual and in FE estimates by sessions. FE regressions include individual fixed effects.
* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Table 5
Comparison of types (FE)

	Dependent variable: *Proposal*			
	(1) GREEN	(2) BLUE	(3) MEDIAN VOTER	(4) TYPES
Common Account	0.0252***	0.0483	0.0211	0.0861**
	(0.0097)	(0.0314)	(0.0127)	(0.0445)
Economy Size	0.0769***	0.3289***	0.1107***	0.3184***
	(0.0133)	(0.0355)	(0.0159)	(0.0371)
COMMON Dummy	−0.7084	4.2153**	−0.7288	3.7495*
	(0.8507)	(1.4649)	(1.1871)	(1.9676)
Common Account × COMMON Dummy	0.0506***	0.1299***	0.0843***	0.0829
	(0.0145)	(0.0339)	(0.0147)	(0.0539)
Economy Size × COMMON Dummy	−0.0055	−0.1215***	−0.0152	−0.0892*
	(0.0189)	(0.0342)	(0.0291)	(0.0481)
GREEN Dummy				−3.4758*
				(1.7199)
GREEN Dummy × COMMON Dummy				−5.8763**
				(2.6413)
Common Account × GREEN Dummy				−0.0811
				(0.0527)
Common Account × GREEN × COMMON				−0.0181
				(0.0655)
Economy Size × GREEN Dummy				−0.2490***
				(0.0395)
Economy Size × GREEN × COMMON				0.1098*
				(0.0592)
Constant	−0.632	7.015***	−1.115*	4.368**
	(0.66739)	(1.7249)	(0.6617)	(1.8171)
Observations	4896	2448	3047	7344
Cluster/Groups	24/408	24/408	24/408	24/408

Notes: Standard errors in parentheses. Standard errors are clustered by session. GREEN dummy takes the value of 1 for all choices of green players and 0 otherwise. Regressions include individual fixed effects. Estimates include period dummies.
* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

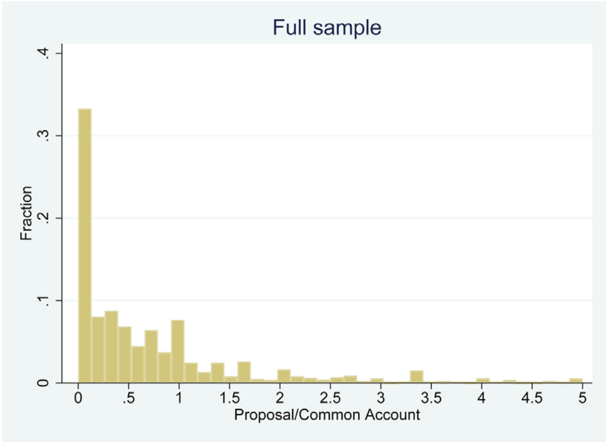


Fig. 3. Distribution of Proposal/Common account.

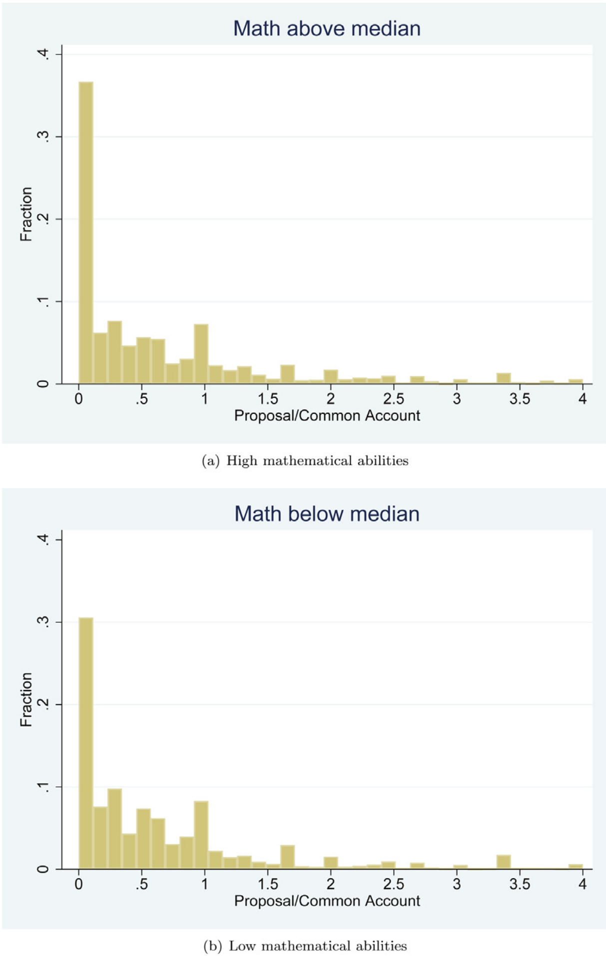


Fig. 4. Distribution of Proposal/Common account split by math abilities.

A.2 Bunching

Fig. 3 plots the frequency of proposals measured as a fraction of the *common account*. 26% of all proposals are larger than the size of the *common account* (i.e. the ratio is equal to or above unity). This fraction does not vary substantially across treatments. In the COMMON treatment the share of proposals exceeding the *common account* is 0.24, in the INDIVIDUAL treatment it is 0.27. Therefore, we are confident that our results are not driven by the misunderstanding that proposals must not exceed the *common account*.

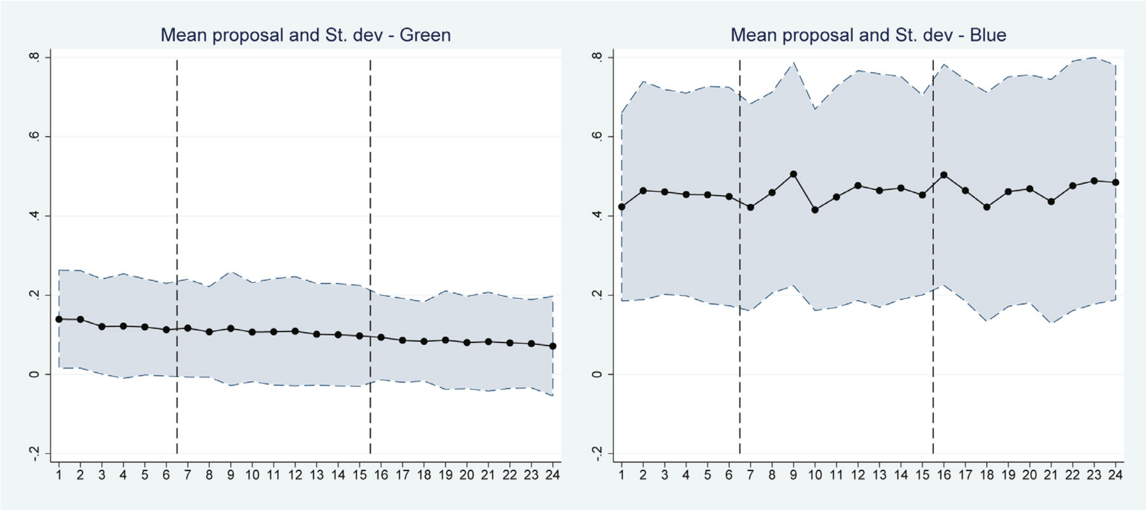


Fig. 5. Mean proposal ratio and SD by type.

Fig. 4 shows the distribution of the proposal to *common account* ratio for participants with high mathematical abilities and those with low mathematical abilities. The share of proposals exceeding the *common account* is 0.24 for the latter group and 0.27 for the former group. As similiar patterns occur we conclude that a misunderstanding of the financing mechanism does not explain our results.

A.3 Mean proposal ratio by type

Figure 5 shows the mean proposal ratio over the 24 periods for type *Green* players (left hand side) and type *Blue* players (right hand side). There is a slight downward trend for the type *Green* players and no clear trend for the type *Blue* players.

References

Abeler, J., Marklein, F., 2017. Fungiblity, labels, and consumption. J. Eur. Econ. Assoc. 15 (1), 99–127.
Allers, M.A., Vermeulen, W., 2016. Capitalization of equalizing grants and the flypaper effect. Reg. Sci. Urban Econ. 58, 115–129.
Baicker, K., 2001. Government decision-making and the incidence of mandates. J. Public Econ. 82 (2), 147–194.
Bailey, S.J., Connolly, S., 1998. The flypaper effect: identifying areas for further research. Public Choice 95, 335–361.
Baker, M., Payne, A., Smart, M., 1999. An empirical study of matching grants: the cap on CAP. J. Public Econ. 72, 269–288.
Barberis, N., Thaler, R., 2003. A Survey of Behavioral Finance. Working Paper.
Beatty, T.K.M., Blow, L., Crossley, T.F., O'Dea, C., 2014. Cash by any other name? Evidence on labeling from the UK winter fuel payment. J. Public Econ. 118, 86–96.
Brooks, L., Phillips, J.H., 2008. An institutional explanation for the stickiness of federal grants. J. Law Econ. Organ. 26 (2), 243–264.
Card, D., Payne, A.A., 2002. School finance reform, the distribution of school spending, and the distribution of student test scores. J. Public Econ. 83, 49–82.
Cherry, T.L., 2001. Mental accounting and other-regarding behavior: evidence from the lab. J. Econ. Psychol. 22 (5), 605–615.
Cherry, T.L., Frykblom, P., Shogren, J.F., 2002. Hardnose the dictator. Am. Econ. Rev. 92 (4), 1218–1221.
Choi, J.J., Laibson, D., Madrian, B.C., 2009. Mental accounting in portfolio choice: evidence from a flypaper effect. Am. Econ. Rev. 99 (5), 2085–2095.
Clingingsmith, D., 2015. Mental accounts and the mutability of altruism: An experiment with online workers. Working Paper.
Courant, P., Gramlich, E., Rubinfeld, D., 1979. The Stimulative Effects of Intergovernmental Grants: Or Why Money Sticks Where it hits. Fiscal federalism and grants-in-aid, Urban Institute Press, Washington, DC.
Dahlberg, M., Mörk, E., Rattsø, J., Ågren, H., 2008. Using a discontinuous grant rule to identify the effect of grants on local taxes and spending. J. Public Econ. 92 (12), 2320–2335.
Dahlby, B., 2011. The marginal cost of public funds and the flypaper effect. Int. Tax Public Financ. 18 (3), 304–321.
Engel, C., 2011. Dictator games: a meta study. Exp. Econ. 14 (4), 583–610.
Filimon, R., Romer, T., Howard Rosenthal, H., 1982. Asymmetric information and agenda control. J. Public Econ. 17, 51–70.
Fischbacher, U., 2007. Z-tree: Zurich toolbox for readymade economic experiments. Exp. Econ. 10 (2), 171–178.
Fisher, R., 1982. Income and grant effects on local expenditures: the flypaper effect and other difficulties. J. Urban Econ. 12, 324–345.
Fochmann, M., Wolf, N., 2016. Mental Accounting in Tax Evasion Decisions – an Experiment on Underreporting and Overdeducting. Working Paper available at SSRN: https://ssrn.com/abstract=2595070.
Fosgaard, T.R., Hansen, L.G., Wengström, E., 2014. Understanding the nature of cooperation variability. J. Public Econ. 120, 134–143.
Gamkhar, S., Shah, A., 2007. The impact of intergovernmental fiscal transfers: a synthesis of the conceptual and empirical literature. In: Boadway, R., Shah, A. (Eds.), Intergovernmental Fiscal Transfers. The World Bank, Washington, D.C.
Gennari, E., Messina, G., 2014. How sticky are local expenditures in Italy? Assessing the relevance of the flypaper effect through municipal data. Int. Tax Public Financ. 21 (2), 324–344.
Goerg, S.J., Johanson, D.B., Roger, J.D., 2017. Endowments, perceived similarity, and dictator giving. Econ. Inq. 55 (2), 1130–1144.
Gordon, N., 2004. Do federal grants boost school spending? Evidence from title i. J. Public Econ. 88, 1771–1792.
Gramlich, E., 1977. Intergovernmental grants: a review of the empirical literature. In: Oates, W.E. (Ed.), The Political Economy of Federalism, Lexington.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1 (1), 114–125.
Hamilton, B., 1983. The flypaper effect and other anomalies. J. Public Econ. 22, 347–362.
Hamilton, J., 1986. The flypaper effect and the deadweight loss from taxation. J. Urban Econ. 19, 148–155.

Hastings, J.S., Shapiro, J.M., 2013. Fungibility and consumer choice: evidence from commodity price shocks. Q. J. Econ. 128 (4), 1449–1498.
Heath, C., Soll, J.B., 1996. Mental budgeting and consumer decisions. J. Consum. Res. 23 (1), 40–52.
Heyndels, B., 2001. Asymmetries in the flypaper effect: empirical evidence for the flemish municipalities. Appl. Econ. 33, 1329–1334.
Heyndels, B., Van Driessche, F., 1998. Mental accounting in local public sector budgeting: an empirical analysis for the flemish municipalities. East. Econ. J. 24 (4), 381–394.
Heyndels, B., Van Driessche, F., 2002. How municipalities react to budgetary windfalls. Econ. Govern. 3 (3), 211–226.
Hines, J.R., Thaler, R.H., 1995. Anomalies: the flypaper effect. J. Econ. Perspect. 9 (4), 217–226.
Höchtl, W., Sausgruber, R., Tyran, J.R., 2012. Inequality aversion and voting on redistribution. Eur. Econ. Rev. 56 (7), 1406–1421.
Hoxby, C., 2001. All school finance equalizations are not created equal. Q. J. Econ. 116 (4), 1189–1231.
Inman, R., 1979. The fiscal performance of local governments: an interpretative review. In: Mieszkowski, P., Straszheim, M. (Eds.), Current Issues in Urban Economics, Baltimore.
Inman, R.P., 2009. The flypaper effect. In: Macmillan, P. (Ed.), The New Palgrave Dictionary of Economics. Palgrave Macmillan, London.
Jacoby, H.G., 2002. Is there an intrahousehold 'flypaper effect'? Evidence from a school feeding programme. Econ. J. 112 (476), 196–221.
Kamenica, E., Brad, L.E., 2014. Voters, dictators, and peons: expressive voting and pivotality. Public Choice 159 (1–2), 159–176.
Knight, B.G., 2002. Endogenous federal grants and crowd-out of state government spending: theory and evidence from the federal highway aid program. Am. Econ. Rev. 92 (1), 71–92.
Ladd, H., 1993. State responses to the TRA86 revenue windfalls: a new test of the flypaper effect. J. Policy Anal. Manag. 12, 82–104.
Langer, S., Korzhenevych, A., 2019. Equalization transfers and the pattern of municipal spending: an investigation of the flypaper effect in Germany. Ann. Econ. Financ. 20 (2), 737–765.
Leduc, S., Wilson, D., 2017. Are state governments roadblocks to federal stimulus? Evidence on the flypaper effect of highway grants in the 2009 recovery act. Am. Econ. J. 9 (2), 253–292.
Lutz, B., 2010. Taxation with representation: intergovernmental grants in a plebiscite democracy. Rev. Econ. Stat. 92 (2), 316–332.
Milkman, K.L., Beshears, J., 2009. Mental accounting and small windfalls: evidence from an online grocer. J. Econ. Behav. Organ. 71 (2), 384–394.
Oates, W., 1979. Lump-sum intergovernmental grants have price effects. In: Mieszkowski, P., Oakland, W. (Eds.), Fiscal Federalism and Grants-in-aid, Washington, DC.
Oxoby, R.J., Spraggon, J., 2008. Mine and yours: property rights in dictator games. J. Econ. Behav. Organ. 65 (3–4), 703–713.
Rabin, M., Weizsäcker, G., 2009. Narrow bracketing and dominated choices. Am. Econ. Rev. 99 (4), 1508–1543.
Roemer, J.E., Silvestre, J., 2002. The 'flypaper effect' is not an anomaly. J. Public Econ. Theory 4 (1), 1–17.
Romer, T., Rosenthal, H., Munley, V.G., 1992. Economic incentives and political institutions: spending and voting in school budget referenda. J. Public Econ. 49 (1), 1–33.
Schady, N., Rosero, J., 2008. Are cash transfers made to women spent like other sources of income? Econ. Lett. 101 (3), 246–248.
Shayo, M., Harel, A., 2012. Non-consequentialist voting. J. Econ. Behav. Organ. 81 (1), 299–313.
Van de Sijpe, N., 2013. The fungibility of health aid reconsidered. J. Dev. Stud. 49 (12), 1746–1754.
Singhal, M., 2008. Special interest groups and the allocation of public funds. J. Public Econ. 92, 548–564.
Strulovici, B., 2010. Learning while voting: determinants of collective experimentation. Econometrica 78 (3), 933–971.
Thaler, R., 1985. Mental accounting and consumer choice. Market. Sci. 4 (3), 199–214.
Thaler, R., 1990. Anomalies: savings, fungibility, and mental accounts. J. Econ. Perspect. 4 (1), 193–205.
Thaler, R., 1999. Mental accounting matters. J. Behav. Decis. Mak. 12 (3), 183–206.
Tisserand, J.C., 2014. The Ultimatum Game, a Meta-analysis of 30 Years of Experimental Research. Working Paper.
Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185 (4157), 1124–1131.
Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. Science 211 (4481), 453–458.
Tyran, J.-R., Sausgruber, R., 2006. A little fairness may induce a lot of redistribution in democracy. Eur. Econ. Rev. 50 (2), 469–485.
Vegh, C.A., Vuletin, G., 2015. Unsticking the flypaper effect in an uncertain world. J. Public Econ. 131, 142–155.