Exploring Encouragement, Treatment and Spillover Effects Using Principal Stratification, With Application to a Field Experiment on

Encouragement, experience and spillover effects in a field experiment on teens' museum attendance

Laura Forastiere[1]*, Patrizia Lattarulo*[2]*, Marco Mariani*[2]*, Fabrizia Mealli*[1]*, and Laura Razzolini*[3]

[1] *Department of Statistics, Informatics and Applications, University of Florence ,* [2] *IRPET - Regional Institute for Economic Planning of Tuscany ,* [3] *Department of Economics, Virginia Commonwealth University*

Abstract

This paper revisits results from a field experiment conducted in Florence, Italy to study the effects of incentives offered to high school teens to motivate them to visit art museums and to identify best practices to transform this behavior into a long run cultural consumption. Students belonging to a first group of classes receive a flier with basic information and opening hours of a main museum in Florence, Palazzo Vecchio. Students in a second group of classes receive the flyer and a short presentation conducted by an art expert. Students in a third group of classes, in addition to the flyer and the presentation, receive also a nonfinancial reward in the form of extra-credit points towards their school grade. Taking a Principal Stratification approach, we explore the causal pathways that may lead students to increase their future museum attendance. Within the strata defined by compliance to the three forms of encouragement, we estimate associative and dissociative principal causal effects, that is, effects of the encouragement on the primary outcome, long run cultural consumption, that are associative or dissociative with respect to the effects of the encouragements on the Palazzo Vecchio visit. This analysis allows to interpret these effects as ascribable either to the encouragements, or to the museum visit, or to classroom spillovers. To face identification issues, estimation is performed with Bayesian inferential methods using hierarchical models to account for clustering. The main findings of the analysis are as follows: what seems to matter the most is the motivational incentive (i.e., the presentation), rather than the induced experience, i.e., the Palazzo Vecchio visit.

1. INTRODUCTION

In this paper we report of a field experiment conducted to study the effects of incentives offered to high school teens to motivate them to visit art museums. This is a novel territory for field experiments. Apart from Lattarulo et al. (2016), whose results we revisit, and Kisida et al. (2014), there is no other similar study in this area of empirical research. Three types of incentives were used, as students received three forms of encour-

agement. Students belonging to a first group of classes received a flyer with basic information and opening hours of a main museum in Florence: Palazzo Vecchio. Students in a second group of classes received the flyer and a short presentation conducted by an art expert, aimed at enhancing the students' curiosity about museum visits in general and Palazzo Vecchio in particular, by portraying museum visits as intriguing and entertaining experiences. Students in a third group of classes, in addition to the flyer and the presentation, received also a non-financial reward in the form of extra-credit points towards their school grade. We analyze the effects of the three incentives on later museum visits. Future visits to museums may be induced directly by the form of encouragement received, which may directly stimulate personal motivation toward museum attendance (Lattarulo et al., 2016), or indirectly from the fact that the teenagers who carried out the visit to the proposed museum had a satisfactory experience that they try to repeat somewhere else (Kisida et al., 2014). They might also be induced by within-classroom spillovers, as the effects of encouragements and experience might propagate through interpersonal ties to the benefit, or to the detriment, of students that are insensitive to the encouragements and choose not to perform the proposed visit. We try to assess the effects generated by the encouragements alone (i.e., not through the museum visit) and those triggered by the experience of visiting the proposed museum. In other words, taking a principal stratification approach, we assess the effects on later museum visits directly caused by the encouragement - and in particular, the presentation by the art expert that elicits intrinsic motivation in students - and the effects indirectly caused by the experience itself of visiting Palazzo Vecchio. To deal with identification issues, estimation is performed with Bayesian inferential methods using hierarchical models to account for clustering.

Our study is an example of a *Cluster Randomized Encouragement Design* (CED), where encouragement is randomized at the level of a cluster of subjects (classes) but compliance is at the individual level (individual visit to Palazzo Vecchio). Interesting questions concerning mechanisms can be raised in this typical non-compliance setting. Encouragements, such as incentives, are used when a treatment cannot be enforced for ethical or practical reasons. Encouragement interventions can be conceived to foster a behavioral change of the target population, that is to increase the probability of adoption of a beneficial treatment or decrease the likelihood of a negative behavior. Hirano et al. (2000) were the first to apply the Principal Stratification ap-

proach (Frangakis & Rubin, 2002) to encouragement designs to estimate intention-to-treat effects within principal strata with and without exclusion restriction assumptions (Imbens & Angrist, 1994; Angrist et al., 1996; Imbens & Rubin, 1997; Mealli & Pacini , 2013). Oftentimes, in fact, the encouragement is itself source of alternative behaviors that may affect the outcome even without involving a change in the treatment received. Also when the major interest lies in the effect of the encouragement on the treatment uptake and, then, on the outcome, the investigation of the underlying mechanisms through which the encouragement program achieves its goal is important for both descriptive and prescriptive reasons. Indeed, such analysis primarily makes it possible to assess whether the intervention is working the way we expect it to, that is by changing the behavior in terms of treatment uptake, with the treatment having an effect on the outcome, and secondly it allows probing other potential mechanisms of the cluster intervention, such as those directly linking the encouragements to the outcome.

CEDs with individual non-compliance can be found relatively frequently in many field experiments (Sommer & Zeger, 1991; McDonald et al., 1992; Hirano et al., 2000; **?**; among others). Frangakis, Rubin & Zhou (2002) extended previous work with PS to account for clustering using Bayesian hierarchical models for inference. In this context, Forastiere et al. (2016) apply concepts of mediation analysis to general non-compliance settings, with the treatment being the intermediate variable.

CEDs are intriguing because they can give rise to several mechanisms that it can be worthwhile to investigate. In fact, not only the outcome may depend or not on a change in the treatment uptake as in classical principal stratification settings, but also, most of the times, encouragements lead to an overall behavioral change in the cluster that can substantially affect the outcome (Forastiere et al., 2016). In other words, since the encouragement is randomized at the cluster level, social interactions occurring among students in the same class may give rise to what in the literature is referred to as *interference* or *spillover effects* (Sobel, 2006; Hong & Raudenbush, 2006; Hudgens & Halloran, 2008; Tchetgen Tchetgen & VanderWeele, 2012). Specifically, friends' behaviors adopted as a result of the encouragement assignment, affect not only their own outcomes but also those of the entire class.

The paper is organized as follows. Section 2 describes the experimental design and presents the research

questions of interest. Section 3 sets the research questions, Section 4 presents the method used in the analysis, and finally Section 5 reports the results and conclusions.

2. EXPERIMENTAL DESIGN AND RESEARCH QUESTIONS

As described in Lattarulo et al. (2016), the Palazzo Vecchio field experiment was run in Florence, Italy, at three different points in time during 2014. It involved 266 students from 15 different classes and from three different high schools [1]. All students attended the fourth year of high school and were aged 17-18. All high schools offered a similar program of studies involving a mix of humanities, mathematics and scientific subjects (Liceo Scientifico) [2]. All students and teachers were told that they were participating in a study on teens' cultural consumption and were not informed that they were part of an experiment with different forms of encouragement. Students were offered the opportunity to visit Palazzo Vecchio (PV), one of the most popular museums in Florence. The entrance to the museum is free to individuals 17 and younger. For students over 17 years old, a free ticket was provided, to guarantee equal access to all participants. The timeline of the experiment was as follows:

- At $t = 1$, all students received one of three forms of encouragement to visit PV and information on students' personal characteristics was collected with a questionnaire;
- At $t = 2$, information about whether students visited or not PV was collected;
- At $t = 3$ (6 months later), the students reported the number of visits made to other museums starting from t=2.

The classes were randomly assigned to three encouragement levels:

1. $A = F$: students received a flyer containing basic information, as opening hours and a brief description of the museum and a short text written by the experimenters stating the importance of museum attendance;

[1] 294 students received the initial encouragement. However, only 266 completed the final step of the field experiment.

[2] Liceo Scientifico is, by far, the most popular choice among Italian students that wish to follow a general education program. This high school provides no vocational education, as its main purpose is to prepare students for college.

2. $A = P$: in addition to the previous flyer and text, students received a short presentation about the exhibit conducted by an art expert from the museum;
3. $A = R$: in addition to the previous flyer, text and presentation, students received a non-financial reward in the form of extra-credit points towards their final school grade.

A complete description and discussion of the three forms of encouragement can be found in Lattarulo et al. (2016). The flyer used was Palazzo Vecchio's official brochure and it was distributed to all students. The flyer provided the weakest possible incentive to perform the visit. The presentation encouragement was aimed at stimulating the students' intrinsic motivation to visit a museum, while the reward provided an additional extrinsic stimulus. Following Ryan and Deci (2000), the basic distinction between intrinsic and extrinsic motivation refers to doing something because it is inherently interesting and/or enjoyable (intrinsic), versus doing something to earn or avoid a specific separable outcome (extrinsic). The assignment of classes to the different types of encouragement was done by clusters at the class level following a completely randomized design, thus guaranteeing the independence between treatments' assignment and potential outcomes, which is required to identify causal effects. The 15 classes were randomly assigned to the three treatments in groups of five [3]. Table 1 reports some descriptive statistics about the students in each encouragement group. In spite of randomization, background characteristics are not always well balanced across the groups, which is not surprising given the relatively small number of classes and students involved in the experiment. For example, the proportion of males is higher in the classes assigned to the reward encouragement, while students in the classes assigned to the presentation seem to be more keen on museums and more interested in humanities than the others.

As shown in Table 1, following the encouragements, 53 students out of 266 (20%) did visit PV (M=1). Among the three encouragements, the reward accounted for the highest number of visits. Since the focus of the paper is on future cultural consumption, in Table 1, we provide information on the outcome variables of interest: the count of subsequent museum visits and the share of students performing at least one future visit. Figure 1 shows histograms for the count of subsequent visits under the three encouragments. From Table 1, the pre-

[3] The size of the experiment (15 classes) was primarily driven by budget and organizational constraints, which made it impossible to boost the number of classes involved. More complex randomization approaches to improve ex-ante covariate balance and precision with small samples – matched pair designs, blocking/stratifications or re-randomizations (e.g (Bruhn and McKenzie., 2009))– were not possible, because the experimenters had no a-priori knowledge of students' characteristics in each class.

	A = F	A = P	A = R	Overall
PRIOR TO THE ENCOURAGEMENT				
Male (1/0)	0.19	0.29	0.53	0.34
Already visited Palazzo Vecchio (1/0)	0.66	0.75	0.72	0.71
Count of museums visited last year	3.31	4.75	3.54	3.86
GPA (0-10)*	6.73	6.82	6.98	6.84
Uninterested in humanities (1/0)	0.36	0.15	0.41	0.31
Parental education (1-5)**	3.18	3.47	3.53	3.39
AFTER THE ENCOURAGEMENT				
No. of students performing the proposed visit (M=1)	3	10	40	53
Pr(M=1)	0.03	0.11	0.44	0.20
E(Y), Y is the count of later museum visits	1.49	4.39	3.00	2.95
Pr(Y>0)	0.36	0.86	0.91	0.71
No. of classes	5	5	5	15
No. of students	89	87	90	266

* GPA is continuous on a 1-10 points scale.
** Levels of parental education: 1 if both parents completed only compulsory education, 2 if only one of them completed high school, 3 if both completed high school, 4 if only one of them completed university, 5 if both parents completed university.

Table 1: Descriptive statistics on students prior to and after the encouragement. Proportions or means

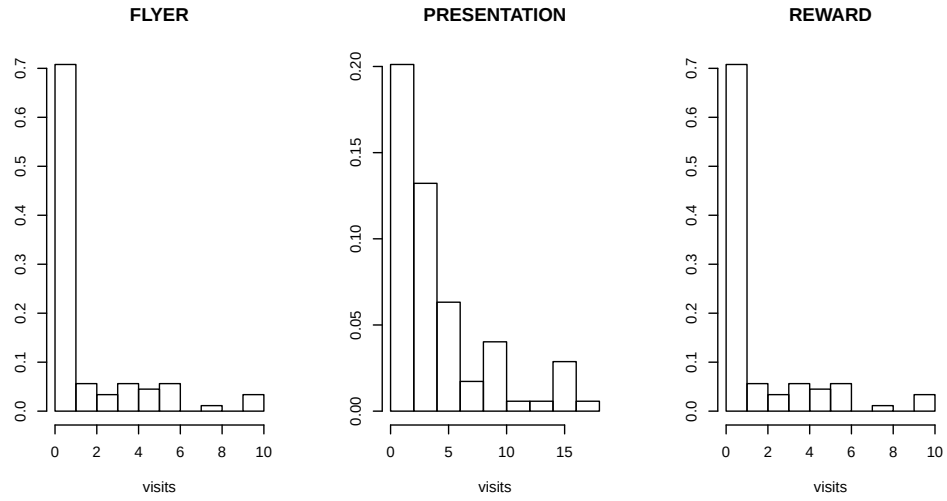


Figure 1: Outcome Histograms

sentation is the form of encouragement associated with the greater average number of future visits (4.39 per student). If contrasted with the lower average number of visits brought about by alternative encouragements,

the presentation seems to work better. These results are akin to those already found, taking the class as unit of analysis, in Lattarulo et al. (2016). If we look at the share of students performing at least one future visit, the reward turns out to be slightly more effective than the presentation, and both appear to be more effective than the flyer alone. It has to be stressed, however, that such gross effects of encouragements include not only the direct motivational stimulus towards later attendance provided by the encouragement itself, but also possible motivational reinforcements provided by the experience at PV and by classroom spillovers. The research question asked in this paper is what better motivates future visits. We will try to understand the effects of three motivational components: the incentives initially provided with encouragements that might raise future attendance in an unchanneled fashion, the possible motivational boost linked to having experienced the proposed museum visit, and spillovers. To tackle this problem, we use a principal stratification approach. This method accounts for different types of students and classifies them into different latent strata. In fact, strata are defined on the basis of the encouragement received and post-encouragement behavior, i.e., whether or not the student would have visit Palazzo Vecchio under the three encouragement conditions.

3. NOTATIONS AND DEFINITIONS

In this section we will give formal definitions of the aforementioned effects in the potential outcomes framework (Rubin, 1974, 2005). The setting consist of $j = 1,\dots,J$ classes and $i = 1,\dots,N_j$ students in each class with a total of N units uniquely denoted by the pair of indices ij. Let A_j denote a three-level cluster encouragement assignment, where clusters consist of classes, so that $A_j = R$ if class j is assigned to receiving the flier, $A_j = P$ if class j is assigned to receiving the flier and the presentation, and $A_j = R$ if class j is assigned to receiving the flier, the presentation and the reward. Let $M_{ij} \in \{0,1\}$ denote whether the student did not or did perform the proposed visit to Palazzo Vecchio and and $Y_{ij} \in \mathscr{Y}$ the final outcome variable for unit i in class j. Let also introduce a vector of covariates, $\mathbf{X}_{ij}$, that is, a vector of covariates of unit i in class j. Finally let $\mathbf{A}$, $\mathbf{M}$ and $\mathbf{Y}$ be the $(J \times 1)$-dimensional vector of encouragement assignments and the $(N \times 1)$-dimensional vectors of treatment received and outcomes, respectively.

Although classes are the units of assignment to the three encouragements, students are the ultimate units of analysis. The treatment concerns the visit at Palazzo Vecchio, $M_{ij} = 1$ if student i in class j visited Palazzo

Vecchio and $M_{ij} = 0$ if not. In terms of the outcome, let Y_{ij} be the number of museum visits occurred during the six months prior to the follow-up interview of student i belonging to class j. Note that throughout the paper we will use the term "individual" to refer to the lowest level of the analysis, which in this case refers to the students. We now introduce notation for the primitive potential outcomes. Let $M_{ij}(\mathbf{A})$ denote the potential visit at Palazzo Vecchio of student i under assignment vector $\mathbf{A}$. Similarly let $Y_{ij}(\mathbf{A},\mathbf{M})$ denote the potential outcome that student i in class j would have experienced if $\mathbf{A}$ and $\mathbf{M}$ were the vectors of assignments and treatments received in the whole population.

Assumption 1. *Cluster-level SUTVA for the encouragement assignment*
Cluster-level Stable Unit Treatment Value Assumption (SUTVA) for the encouragement assignment consists of two parts:

(i) An student's potential outcomes and potential values of the intermediate variable do not vary with encouragements assigned to classes other than the studentÂŠs own class, i.e. $M_{ij}(\mathbf{A}) \equiv M_{ij}(A_j)$ and $Y_{ij}(\mathbf{A},\mathbf{M}) \equiv Y_{ij}(A_j,\mathbf{M}_j)$, where $\mathbf{M}_j$ is the vector of dimensions $N_J \times 1$ of treatment received by individuals of class j.

(ii) For each class there are no different versions of each encouragement level. Formally:

$$\text{if } A_j = A_j' \text{ then } M_{ij}(A_j) = M_{ij}(A_j') \text{ and if } A_j = A_j' \text{ and } \mathbf{M}_j = \mathbf{M}_j' \text{ then } Y_{ij}(A_j,\mathbf{M}_j) = Y_{ij}(A_j',\mathbf{M}_j')$$

Cluster-level SUTVA is an extension of the individual-level SUTVA introduced by Rubin (1978, 1980, 1990) to settings with cluster-level assignments and individual-level intermediate variable. Yet it is worth noting that part (i) of the assumption requires that the outcome Y_{ij} of student i in class j does not vary with the encouragement conditions or treatments received in other classes. This assumption is made plausible by the fact that the classes involved in the experiment were located in different schools, in different buildings of a same school or, when this was not possible, on different floors of a same building. However the previous assumption does not rule out the possibility of spillover effects of the intermediate variable within classes, that is Y_{ij} can be affected by the treatment received by other units of the same cluster j. Under cluster-level SUTVA we can use the notation $M_{ij}(A_j)$ and $Y_{ij}(A_j,\mathbf{M}_j)$.

Note that the only observable potential outcome is the one where, if A_j were set to a, the treatment received by all the students in class j were left to the value it would take under encouragement condition a, that is $Y_{ij}\big(a,\mathbf{M}_j(a)\big)$. In other words, we can only observe for each student the potential outcome (count of subsequent visits) corresponding to the encouragement to visit Palazzo Vecchio actually received and to the response to this

encouragement by the student him/herself and his/her classmates. Throughout we will use the notation $Y_{ij}(a)$ for potential outcomes of this type, and we will only use these potential outcomes, thus avoiding the use of so-called a priori counterfactual outcomes.

Based on these potential outcomes, the overall average effect of cluster encouragement intervention a_k versus cluster encouragement intervention a_h on the individual outcome $k, h = F, P, R$, referred to as *Intent-to-Treat Effect* (ITT) is defined as the following contrast:

$$ITT_{kh} := \mathrm{E}\big[Y_{ij}(a_k)\big] - \mathrm{E}\big[Y_{ij}(a_h)\big] \tag{3.1}$$

4. PRINCIPAL STRATIFICATION APPROACH

Principal stratification has been first introduced by Frangakis and Rubin (2002), in order to address post-treatment complications in an experimental setting. Its use in mediation analysis has been proposed as a way to relax the sequential ignorability assumption but still being able to yield valid causal inference of what VanderWeele (2008) called the principal strata direct effects (Gallop et al., 2009; Elliott et al., 2010; Page, 2012; Mattei & Mealli, 2011).

The units under study can be stratified in subpopulations, the so-called *Principal Strata*, defined according to the potential values of the actual treatment received:

$$S^{m_F m_P m_R} := \big\{i : M_{ij}(F) = m_F, M_{ij}(P) = m_P, M_{ij}(R) = m_R.\big\} \tag{4.1}$$

In our study, students can be divided in principal strata based on their behavior in terms of visiting Palazzo Vecchio under the three encouragement conditions. Under each encouragement level, each student may perform or not perform the proposed visit to Palazzo Vecchio. However, since only one of the students' potential responses to the three encouragements is observed, i.e., the one associated with the encouragement actually received, these subpopulations are latent, in the sense that, in general, it is not possible to identify the specific subpopulation a unit i belongs to. Let S_{ij} be the indicator of the latent group to which subject i belongs. When A_j has three levels and M_{ij} are binary there are eight strata $S_{ij} \in \{S^{000}, S^{001}, S^{011}, S^{111}, S^{100}, S^{110}, S^{010}, S^{101}\}$. Strata membership can also be referred to as compliance status. Table 2 outlines the eight possible principal strata and their denomination.

Assumption 2. *Monotonicity of Compliance*

M(A=F)	M(A=P)	M(A=R)	Stratum ID	Stratum Label
1	1	1	A	Always Takers (S^{111})
0	1	1	B	Presentation Compliers (S^{011})
0	0	1	C	Reward Compliers (S^{001})
0	0	0	D	Never Takers (S^{000})
1	0	0	E	Defiers (S^{100})
0	1	0	F	Defiers (S^{010})
1	1	0	G	Defiers (S^{110})
1	0	1	H	Defiers (S^{101})

Table 2: Theoretical types of students defined according to their compliance behavior to the encouragement received

Monotonicity of encouragement assignment on treatment receipt requires

$$M_{ij}(F) \leq M_{ij}(P) \leq M_{ij}(R) \quad \forall i, j$$

This assumption rules out the presence of the last four strata. It conveys the idea that every student that would perform the proposed visit to Palazzo Vecchio in response to a certain level of encouragement would do the same if he/she had received a higher level of encouragement. It conveys that there is no student who would visit Palazzo Vecchio with the flier only would not visit Palazzo Vecchio when receiving the other two enhanced incentives, and no student who would visit Palazzo Vecchio with the flier and the presentation would not visit Palazzo Vecchio when receiving also the reward.

In what follows we will maintain Assumption 2. As mentioned previously in our study this assumption is plausible not only because the incentives are incremental, but also because it is not contradicted by the data. In fact the proportion of students going to Palazzo Vecchio is increasing with the incremental incentives. The four principal strata can be named as follows. *Never takers* are those who do not perform the proposed visit to Palazzo Vecchio whatever encouragement they receive. On the contrary, *always takers* are those who perform this visit whatever encouragement they receive. The other two strata refer to students that undertake the proposed museum visit provided the encouragement they receive is above a certain threshold. *Presentation compliers* are those who perform the proposed visit if they receive at least the presentation, otherwise they do not. In this stratum, there is a mixture of students that are sensitive to the intrinsic motivational stimulus

provided by the presentation and students that require the reward to go. Finally, *reward compliers* are those who undertake the visit to Palazzo Vecchio only if pushed to do so by an extrinsic motivational stimulus such as the reward promise.

Each observed stratum defined by the assigned encouragement and the treatment received corresponds, in general, to a mixture of principal strata. Table 3 shows these mixtures under the monotonicity assumption.

	M=0	M=1
A=F	Never Takers, P-compliers, R-Compliers	Always Takers
A=P	Never Takers, R-Compliers	Always Takers, P-compliers
A=R	Never Takers	Always Takers, P-compliers, R-Compliers

Table 3: Strata observed in the data and the theoretical types of students that might belong to these strata

Let $\pi_{m_F m_P m_R} := P(S_{ij} = S^{m_F m_P m_R})$ denote the probability of belonging to stratum $S^{m_F m_P m_R}$. Randomization of the encouragement assignment and the monotonicity assumption imply the following results:

$$\begin{gathered}
\pi_{100} = \pi_{010} = \pi_{110} = \pi_{101} = 0; \\
\pi_{111} = P(M_{ij} = 1 \mid A_j = F); \quad \pi_{000} = P(M_{ij} = 0 \mid A_j = R); \\
\pi_{001} = P(M_{ij} = 0 \mid A_j = P) - \pi_{000} = P(M_{ij} = 1 \mid A_j = R) - P(M_{ij} = 1 \mid A_j = P); \\
\pi_{011} = P(M_{ij} = 1 \mid A_j = P) - \pi_{111} = P(M_{ij} = 1 \mid A_j = R) - \pi_{111} - \pi_{001}
\end{gathered} \tag{4.2}$$

and

$$\begin{gathered}
E[Y_{ij}(F) \mid S_{ij} = S^{111}] = E[Y_{ij} \mid M_{ij} = 1, A_j = F] \\
E[Y_{ij}(R) \mid S_{ij} = S^{000}] = E[Y_{ij} \mid M_{ij} = 0, A_j = R]
\end{gathered} \tag{4.3}$$

We report in Table 4 the method of moments estimates of the quantities in (4.2) and (4.3).

	$P(S_{ij})$	$E[Y_{ij}(F) \mid S_{ij}]$	$E[Y_{ij}(P) \mid S_{ij}]$	$E[Y_{ij}(R) \mid S_{ij}]$
Always Takers	0.034	5.667	·	·
Presentation Compliers	0.081	·	·	·
Reward Compliers	0.329	·	·	·
Never Takers	0.556	·	·	3.180

Table 4: Method of moments estimates of probabilities of principal strata membership and mean potential outcomes within principal strata

Even if these moment-based estimates are not adjusted by the residual observed unbalance of covariates

between the three groups, they reveal that students are mostly Reward Compliers or Never Takers. Always takers show a relatively high average number of post-encouragement visits even with the weakest form of incentive, which will motivate later an exclusion restriction type assumption for this subgroup.

4.1 PRINCIPAL CAUSAL EFFECTS

The overall effect of one cluster encouragement (say a_k) versus another cluster encouragement (say a_h) within each principal stratum (and eventually within levels of baseline covariates) is named *principal causal effect* (PCE) and is defined as:

$$PCE_{kh}(m_F, m_P, m_R) := \mathrm{E}\big[Y_{ij}(a_k) \mid S_{ij} = S^{m_F m_P m_R}\big] - \mathrm{E}\big[Y_{ij}(a_h) \mid S_{ij} = S^{m_F m_P m_R}\big] \tag{4.4}$$

ITT is then a weighted average of PCEs, with weights given by the conditional probability of belonging to each principal stratum:

$$ITT_{kh} = \sum_{m_F m_P m_R} PCE_{kh}(m_F, m_P, m_R) \cdot \pi_{m_F m_P m_R} \tag{4.5}$$

In principal strata where the visit to Palazzo Vecchio is unaffected by the encouragement, i.e. never-takers and always-takers, principal causal effect, $PCE(m, m, m)$ with $m \in \{0, 1\}$, are called *dissociative causal effects* (DCEs). As they cannot originate from having performed the suggested visit, DCEs may be a combination of two different types of effects: *pure encouragement effects*, that is, effects of the cluster encouragement through the direct modification of individual behaviors (Frangakis, Rubin & Zhou, 2002), and *spillover mediated effects*, due to mechanisms of *interference* by behavioral change in other students of the same cluster, both in terms of treatment receipt (classmates might decide to comply or not with the encouragement received) or in terms of actions occurring in the follow up period. It is impossible to disjoin pure encouragement and spillover mediated effects without resorting to additional assumptions. What counts here is that both types of DCEs originate from the encouragement and, therefore, their separation may turn out to be of moderate substantive interest. However, because in our experiment the reward promise that is added to the presentation applies only to the proposed visit of Palazzo Vecchio and not later in the follow up period, it follows that the DCEs of the reward encouragement versus the presentation alone can only be interpreted as spillover mediated effects (see Table 5).

On the contrary, PCEs for presentation (S^{011}) and reward compliers (S^{001}) can be of two different types:

associative causal effects (ACEs), possibly combining pure encouragement, spillover mediated and experience effects, only when we contrast encouragements a_k and a_h that change their decision to perform the suggested visit, that is, $M_{ij}(a_k) \neq M_{ij}(a_h)$; *dissociative causal effects* (DCEs), possibly combining pure encouragement and spillover mediated effects, when we contrast encouragements a_k and a_h that do not change their decision to visit PV, that is, $M_{ij}(a_k) = M_{ij}(a_h)$. Again, because in our experiment the reward promise that is added to the presentation does not apply in the follow up period, it follows that the PCE of the reward encouragement versus the presentation alone (Table 5): i) should be interpreted as a spillover mediated effect, for presentation compliers ; ii) should be the simpler combination of experience effects and spillover mediated effects that are triggered by the experience itself in the follow up period, for reward compliers. Again, separating experience and spillovers due to experience may turn out to be of moderate substantive interest. What counts here is that the latter PCE cannot originate from pure encouragements, which makes it possible to interpret it as a treatment-mediated effect.

	E[Y(P)]-E[Y(F)]	E[Y(R)]-E[Y(P)]	E[Y(R)]-E[Y(F)]
Always Takers	[D] Enc, Spill	[D] Spill	[D] Enc, Spill
Presentation Compliers	[A] Enc, Exp, Spill	[D] Spill	[A] Enc, Exp, Spill
Reward Compliers	[D] Enc, Spill	[A] Exp, Spill	[A] Enc, Exp, Spill
Never Takers	[D] Enc, Spill	[D] Spill	[D] Enc, Spill

Table 5: Interpretation of the effects in the strata corresponding to theoretical types of students. Each principal causal effect is disentangled into pure encouragement effects (Enc), spillover mediated effect (Spill) or the effect through personal experience of Palazzo Vecchio (Exp). [A] stands for associative principal causal effect; [D] stands for dissociative principal causal effect.

4.2 ESTIMATING PRINCIPAL CAUSAL EFFECTS

4.2.1 HIERARCHICAL MODELS

In this section we describe the models used for our analysis: a model for the outcome and a model for the principal strata membership. Because of the cluster-level randomization the use of the hierarchical framework is needed as correlation among individuals arising from common environmental factors and even reciprocal influence cannot be ignored. In our setting, individuals living in the same community are likely to show resemblance not only in terms of outcomes, but also in terms of individual treatment uptake. Furthermore, the level of resemblance in outcomes may vary across different individual strata. Correlation in cluster random-

ized trials with individual non-compliance has been intensively studied by Jo (2008), after Frangakis, Rubin & Zhou (2002), who were the first authors to accommodate in their analysis correlation in both outcome and non-compliance status.

POTENTIAL OUTCOME MODEL

In our study, the outcome of interest, Y_{ij}, is the number of museum visits that student i in class j has experienced during the six month prior to the follow-up interview. Given the histograms in Figure 1, we assume for the potential outcomes $Y_{ij}(a)$ a zero-inflated poisson (ZIP) model, which is a model for count data with excess zeros. The ZIP model can be viewed as a mixtures of two types of individuals. The first type gives Poisson distributed counts, which might contain zeros. The second type always gives a zero count. Let $\lambda_{ij} = \lambda_{ij}(a, S_{ij}, \mathbf{X}_{ij})$ be the Poisson mean and $p_{ij} = p_{ij}(a, S_{ij}, \mathbf{C}_{ij})$ be the probability of an individual being of the second type, both expressed as a function of the encouragement $A_j = a$, the principal stratum S_{ij} and the vector of covariates $\mathbf{X}_{ij}$. The ZIP model can, then, be formulated as follows:

$$P(Y_{ij}(a) = y | S_{ij}, \mathbf{X}_{ij}) = \begin{cases} p_{ij}(a, S_{ij}, \mathbf{X}_{ij}) + \big(1 - p_{ij}(a, S_{ij}, \mathbf{X}_{ij})\big) Pois\big(0; \lambda_{ij}(a, S_{ij}, \mathbf{X}_{ij})\big) & \text{if } y = 0 \\ \big(1 - p_{ij}(a, S_{ij}, \mathbf{X}_{ij})\big) Pois\big(y; \lambda_{ij}(a, S_{ij}, \mathbf{X}_{ij})\big) & \text{if } y > 0 \end{cases} \tag{4.6}$$

where $Pois\big(0; \lambda_{ij}\big) = exp(-\lambda_{ij})$ and $Pois\big(y; \lambda_{ij}\big) = \dfrac{e^{-\lambda_{ij}} \lambda_{ij}^{y}}{y!}$. In addition, we model the parameter $\lambda_{ij}(a, S_{ij}, \mathbf{X}_{ij})$ of the Poisson part using a hierarchical generalized linear model with a log link:

$$\begin{gathered} log\Big(\lambda_{ij}\big(a, S_{ij}, \mathbf{X}_{ij}\big)\Big) = \alpha_a^{S_{ij}T} + \boldsymbol{\beta}_a^{S_{ij}T} \mathbf{X}_{ij} + b_j \\ b_j \sim N(0, \sigma_b) \end{gathered} \tag{4.7}$$

The parameters $\alpha_a^{S_{ij}}$ and $\boldsymbol{\beta}_a^{S_{ij}}$ depend on the principal stratum and on the encouragement level. b_j is a random intercept to account for cluster correlation. To reduce the complexity of the model, we remove the dependence on the covariates of the parameter $p_{ij}(a, S_{ij}, \mathbf{X}_{ij})$ of the binary part. Therefore, $p_{ij}(a, S_{ij}, \mathbf{X}_{ij}) = \phi_a^{S_{ij}}$.

We also assume that the two potential outcomes $Y_{ij}(F)$, $Y_{ij}(P)$ and $Y_{ij}(R)$ are independent, given the covariates and strata membership.

As far as the vector of covariates $\mathbf{X}_{ij}$ is concerned, we included the characteristics that are assumed to be predictive of the compliance behaviors and those where we have reported some unbalance across the encour-

agement groups (see Table 1): male (1/0), already visited Palazzo Vecchio (1/0),count of museums visited last year, GPA (continuous on a 1-10 points scale), personal interest in humanities (three levels: 1 if the student is interested only in humanities, 2 if she/he is interested both in humanities and science, 3 if she/he is uninterested in humanities), parental education (five levels: 1 if both parents completed only compulsory education, 2 if only one of them completed high school, 3 if both completed high school, 4 if only one of them completed university, 5 if both parents completed university).

PRINCIPAL STRATA MODEL

We modeled the principal strata membership using a *Multinomial Logit random effect model*:

$$ln\frac{P\left(S_{ij}=k|\mathbf{C}_{ij}\right)}{P\left(S_{ij}=S^{111}|\mathbf{C}_{ij}\right)}=\gamma_k+\boldsymbol{\delta}_k^T\mathbf{C}_{ij}+\mathrm{a}_{kj}$$
$$\mathbf{a}_j\sim N(\mathbf{0},\Sigma_a) \tag{4.8}$$

where $k=S^{000},S^{001},S^{011}$, using the stratum of always-takers as the reference category. a_{kj} are the random intercepts, to take into account cluster correlation in compliance behavior.

4.2.2 BAYESIAN INFERENCE

Given the latent nature of principal strata, principal causal effects are only partially identified, in that the observed data provide a possible set of values strictly included in the parameter space (Gustafson , 2010). Thus the posterior distribution of principal causal effects will in general converge to a non degenerate distribution with support equal to the identified set. However, in finite sample, if the prior distribution is proper, so will be the posterior distribution; the credible intervals for the principal causal effects will incorporate the uncertainty due to partial identifiability as well as that due to sampling variability (Tamer, 2010). For the estimation of the parameter vector $\boldsymbol{\theta}=\{\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_b,\boldsymbol{\gamma},\boldsymbol{\delta},\Sigma_a\}$, we used a Bayesian approach, which overcomes the problem of latent principal strata by using a two-stage Gibbs-sampling strategy that first samples the missing strata memberships S_{ij}, thereby allowing assessment of the distributions of Y_{ij} conditional on the complete data consisting of subpopulations without mixture components. This procedure is well know as *Data Augmentation* scheme (Tanner & Wong, 1987). See the appendix for the detailed Gibbs-Sampling procedure.

Given the posterior distribution of the parameters, we computed the finite sample posterior distribution of the principal causa effects by first imputing the principal strata membership from the estimated principal

strata model and then imputing the missing potential outcomes from the estimated outcome models.

To ease estimation we maintained the following assumption.

Assumption 3. Exclusion Restriction for Always-Takers

$$E[Y_{ij}(F)|S_{ij} = S^{111}] = E[Y_{ij}(P)|S_{ij} = S^{111}] = E[Y_{ij}(R)|S_{ij} = S^{111}]$$

This assumption implies that $PCE_{kh}(111) = 0$, $\forall k, h$.

There are two reasons that justify relying on this assumption. The first is that the low number of always-takers, as shown in Table 4, does not provide information for the estimation on the set of parameters for this stratum. Therefore, this assumption improves the stability of the estimation procedure. For the same reason, a possible violation of this assumption has little effect on the estimation of the rest of the parameters. The second reason is that we have estimated a high number of museum visits for the always-takers when assigned to the flyer only (See Table 4). Therefore the assumption of a small effect of other types of encouragements for this type of students seems plausible.

5. RESULTS

A GALLERY OF FOUR PORTRAITS

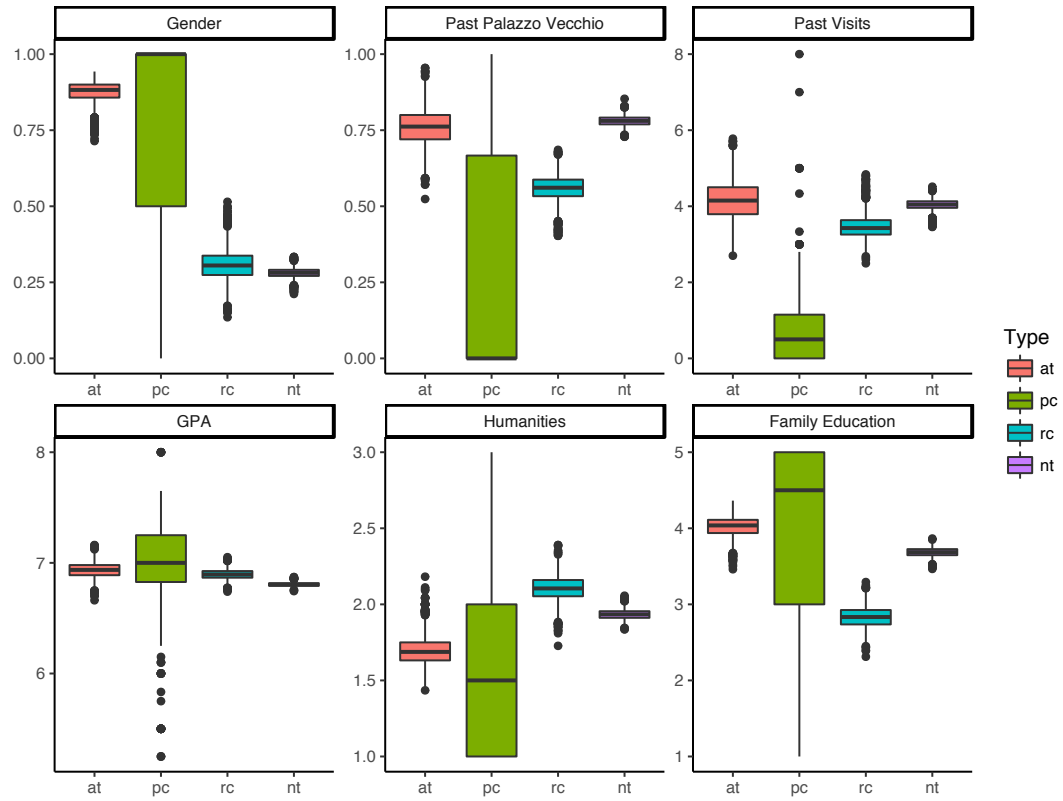
We first focus on the estimated probabilities of membership in the four principal strata, reported in Table 6. These probabilities were estimated by imputing for each student the principal strata membership according to his covariates. It is worth noting that these estimates differ from the method of moment estimates in Table 4 because of the postulated model, which also adjusts for the residual unbalance in the observed covariates within the three groups. Two strata have considerable size. The estimated probability that a student involved in the experiment is a never taker is 59.1% (the 95% credibility interval is 51.9-65.4%), the probability that a student is a reward complier is 31.7% (24.4-39.5%). The other two principal strata are much smaller. The probability of being an always taker is 9.1% (6.8-12%), that of being a presentation complier is almost for sure less than 1%.

	Mean	SD	2.5%	5%	95%	97.5%
Always Takers	0.091	0.014	0.068	0.071	0.117	0.120
Presentation Compliers	0.001	0.003	0.000	0.000	0.004	0.008
Reward Compliers	0.317	0.039	0.244	0.256	0.383	0.395
Never Takers	0.591	0.034	0.519	0.530	0.643	0.654

Table 6: Estimated posterior probabilities of principal strata membership

In words, the overwhelming majority of the students taking part in the experiment are not really attracted by the proposed visit of Palazzo Vecchio. At most, they are available to perform the visit if they receive the reward promise, but the presentation alone hardly elicits sufficient motivation for the suggested visit experience. This does not mean, of course, that these students' later behavior is unaffected by the encouragements received or by the visit experience they did against a reward. What types of students are most likely to be never takers, always takers, presentation or reward compliers? We need to look at the characterization of principal strata to find it out. To this end, we derive from our model the posterior distributions of the proportions or means of covariates in the principal strata themselves. Figure 2 reports means and 95% credibility intervals of these distributions. Three of the portraits we obtain are like high-definition photographs, while one is more blurry and impressionistic. The portraits of never takers and reward compliers are clear-cut. The stratum of never takers

is composed of fairly successful female students from educated families, somewhat interested in humanities at school and museum-goers in their free time. Usually, they have already visited PV, and are not willing to revisit it. The stratum of opportunistic individuals, that of reward compliers, is also likely to host good female students whose interests include humanities, but these students have a less educated family background, and more seldom go to museums in their free time. They might not have visited PV earlier.



Note: Gender = 1 if male; Past Palazzo Vecchio = 1 if the student has already visited PV in the past; Past visits is the count of museum visits performed last year; GPA is continuous on a 1-10 points scale; Humanities has three levels: 1 if the student is interested only in humanities, 2 if she/he is interested both in humanities and science, 3 if she/he is uninterested in humanities; Parental education has five levels: 1 if both parents completed only compulsory education, 2 if only one of them completed high school, 3 if both completed high school, 4 if only one of them completed university, 5 if both parents completed university.

Figure 2: Distribution of Covariates by Principal Strata

Little uncertainty surrounds the characterization of always takers, who are very likely to be fairly successful male students from an educated family, moderately keen on humanistic disciplines at school, yet rather assiduous museum-goers in their free time. Despite they usually have already visited PV, these students are ready to comply even to the weakest encouragement, possibly motivated by a desire for cultural experiences that goes beyond one-spot museum consumption. The blurriest portrait is that of presentation compliers, also due to

the tiny size of the stratum. However, the idea we get from results is that they might be good male students, grown up in educated families, with modest or no interest at all in humanities and museums. They might have not visited PV earlier, but are ready to do it if their intrinsic motivation is appropriately stimulated.

PRINCIPAL CAUSAL EFFECTS

Of the four portraits above, we focus on those of never takers and reward compliers. The group of presentation compliers is so small that any conclusion can be drawn for this subgroup, whereas principal causal effects for always takers are ruled out by the exclusion restriction assumption. Tables 7 and 8 report the estimated causal effects for these principal strata, respectively, on two outcomes: the count of subsequent museum visits and the probability that this count is greater than zero. The posterior distributions of these principal causal effects are illustrated in Figures 3 and 4. A definition of causal estimands has been provided in Section 4.1, together with hints about their possible interpretation depending on the stratum they refer to. Estimates are based on imputations from the predictive posterior distributions of potential outcomes (see supplemental materials), based on 5000 Markov chain Monte Carlo (MCMC) iterations.

Let us consider never takers first. As explained earlier in this paper, any difference in subsequent museum attendance cannot be due to the experience, but only to the different contents of the encouragements to which never takers may be randomly assigned with their class, or to classroom spillovers that have arisen afterwards. The differential effect of the presentation relative to the flyer, at the mean, is 2.04 museum visits, and a 32% higher probability of performing at least one museum visit in the follow-up period. This effect is extremely likely to have a positive sign (96% and 100%, respectively), and suggests that the presentation worked in raising the students' motivation to attend museums later on, although it hardly worked in attracting the students to PV. Without resorting to additional assumptions, we cannot disentangle, here, the motivational reinforcement directly provided by the presentation and that due to the classroom spillovers induced by the presentation itself. However, given the very low probability of having always takers and presentation compliers as classmates, the spillovers stemming from the proposed museum visit of others might appear rather unlikely in this context, where the reward compliers that are there do not receive sufficient stimulus to perform this visit. The differential effect of the reward relative to the presentation is also positive on the count of future visits, 1.72,

	Never Takers						Reward Compliers					
	Mean	2.5%	5%	95%	97.5%	$Pr(\cdot > 0)$	Mean	2.5%	5%	95%	97.5%	$Pr(\cdot > 0)$
Presentation vs Flyer	2.04	-0.26	0.15	4.12	4.78	0.96	5.62	2.05	2.51	10.22	11.52	1.00
Reward vs Presentation	1.72	-1.80	-1.10	5.21	6.11	0.85	-2.27	-8.15	-6.95	1.79	2.62	0.17
Reward vs Flyer	3.77	1.02	1.41	7.12	8.21	1.00	3.35	0.72	1.08	6.34	7.42	0.99

Table 7: Estimation of Principal Causal Effects on Count Outcome

and on the probability of performing at least one museum visit in the follow-up period, 9%. Yet, these causal effects are slightly more uncertain, as witnessed by credibility intervals, although the associated probability to have a positive sign is remarkably high. At any rate, the idea is that classroom spillovers induced by the reward encouragement might be at work. As the probability of having reward compliers as classmates is considerable, this spillover might derive both from their visit to PV, and from the subsequent visits that the presentation has induced in classmates during the follow up. Finally, the differential effect of the reward relative to the flyer for never takers amounts to the sum of the two causal effects previously presented and confirms previous results and interpretation.

Let us focus on reward compliers now. Depending on the encouragement these students receive, any difference in their subsequent museum attendance can be due to encouragements, classroom spillovers, and also to the experience done at PV. The differential effect of the presentation relative to the flyer is positive, both in terms of the count of later visits and the probability that this count is greater than zero. This effect cannot originate from experience, as none of the two encouragements contrasted here is a sufficient nudge towards PV for a reward complier. Again, we are not able to say, without additional assumptions, how much of this causal effect is due to the pure encouragement directly provided by the presentation, and how much is due to the spillovers induced by the presentation itself. As the probability of having always takers and presentation compliers as classmates is relatively low, we may argue that spillovers, if any, should originate from visits performed by classmates in the follow-up period, rather than from the visit to PV that neither reward compliers nor never takers are available to perform when receiving the presentation alone. Whatever the pathway, the reinforcement of intrinsic motivation provided by the presentation is confirmed. The differential effect of the

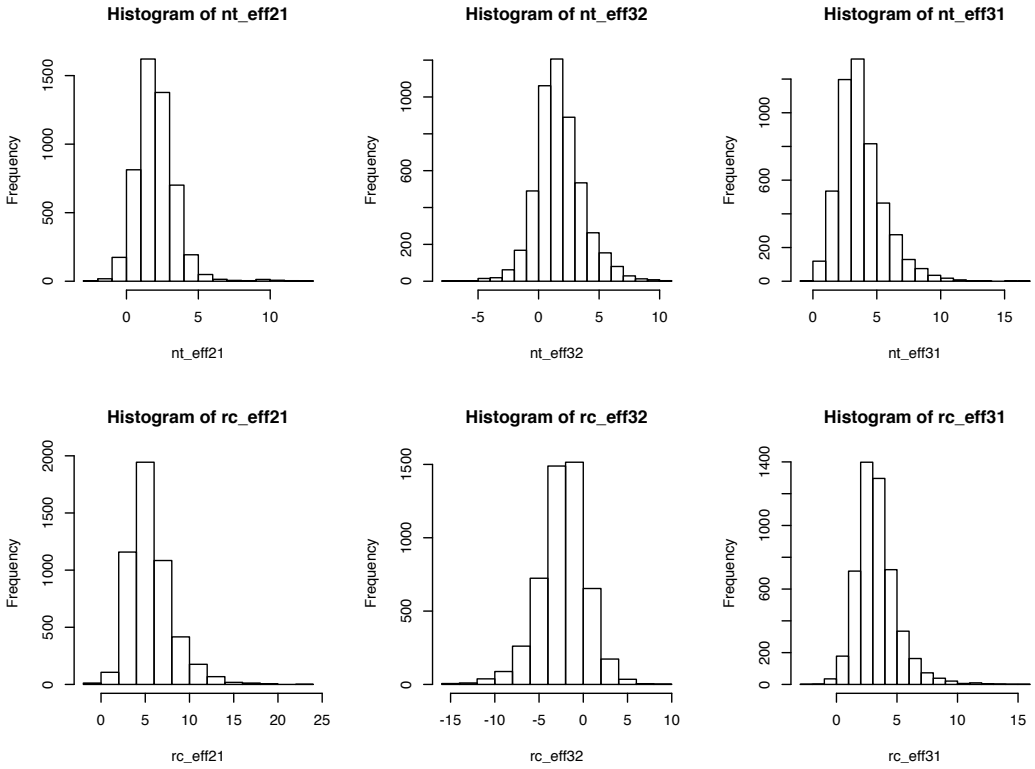


Figure 3: Histograms of Principal Causal Effects on Count Outcome

reward relative to the presentation represents the causal effect of the experience at PV and of the spillovers generated by this experience in the aftermath. This effect, although affected by some uncertainty, tends to be positive when performing at least one museum visit in the follow up is the outcome of interest (the mean of the posterior distribution is 8%), as witnessed by the probability of having positive effects in the posterior distribution (76%). However, after looking at the count of subsequent visits as an outcome, the picture becomes increasingly ambiguous. Here, the mean of the posterior distribution of the causal effect is negative, and the probability of having positive effects is low (17%). Both results are surrounded by some uncertainty, which suggests that reward compliers may react to experience in opposite ways. They basically mirror a shrinkage of the distribution of potential outcomes after the PV experience. This shrinkage is associated, on the one hand, to the reduction of no-attendance behavior and, on the other hand, to an even sharper reduction of stronger museum-goers. In words, the PV experience may induce museum attendance of some reward compliers. At the same time, the visit performed against a reward might also crowd out the later attendance of some others, by

anticipating it over time. Again, the differential effect of the reward relative to the flyer for never takers amounts to the sum of the two causal effects previously presented, and confirms that the presentation is likely to be a good motivator of subsequent museum attendance.

	Never Takers						Reward Compliers					
	Mean	2.5%	5%	95%	97.5%	$Pr(\cdot > 0)$	Mean	2.5%	5%	95%	97.5%	$Pr(\cdot > 0)$
Presentation vs Flyer	0.32	0.06	0.10	0.54	0.58	0.99	0.55	0.18	0.24	0.80	0.82	1.00
Reward vs Presentation	0.09	-0.07	-0.05	0.22	0.24	0.84	0.08	-0.11	-0.09	0.26	0.29	0.76
Reward vs Flyer	0.40	0.19	0.22	0.58	0.63	1.00	0.63	0.25	0.34	0.86	0.88	1.00

Table 8: Estimation of Principal Causal Effects on Binary Outcome

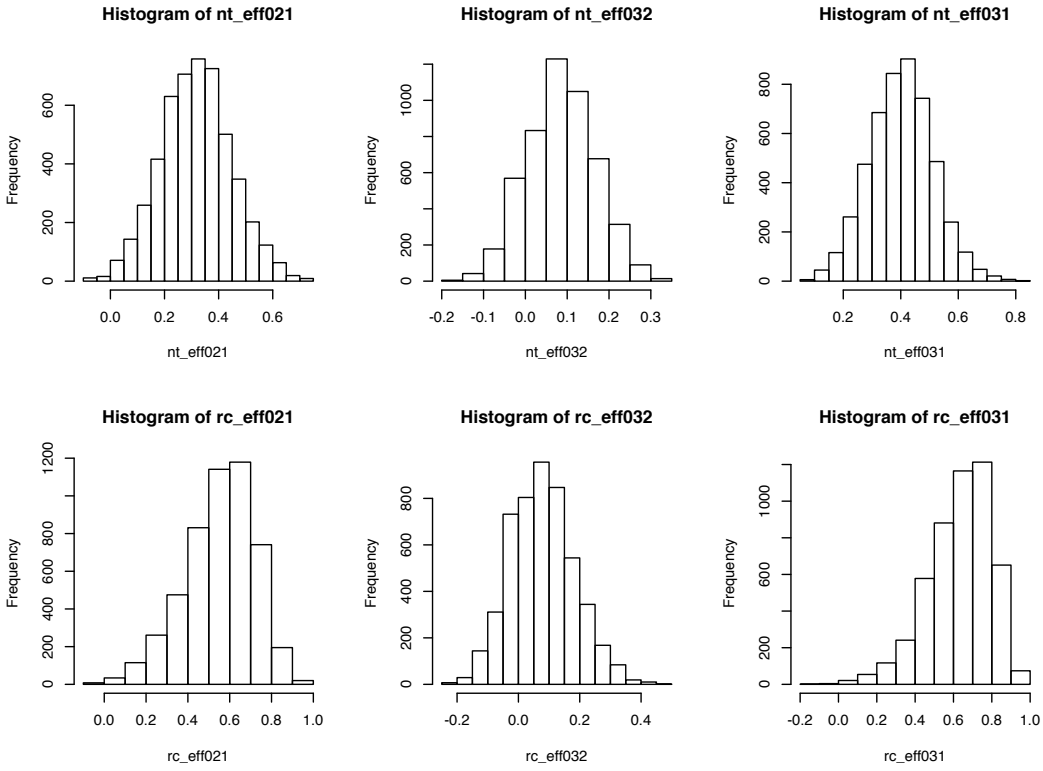


Figure 4: Histograms of Principal Causal Effects on Binary Outcome

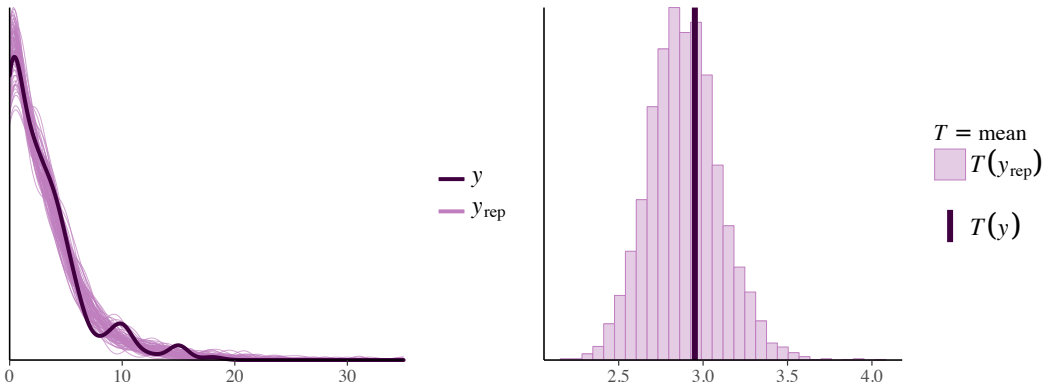


Figure 5: Posterior Predictive Distribution against Observed Distribution of Post Visits

POSTERIOR PREDICTIVE MODEL CHECKS

To check the model fit we performed posterior predictive model checks. We used the estimated model to draw potential outcomes under the observed encouragement level and compare them with the observed outcomes. Figure 5 (right) shows the comparison between draws of the posterior predictive distribution and the observed outcome distribution. We have also computed the posterior distribution of the mean and compared it with the observed mean (Figure 5, left). Using this testing procedure, we cannot reject the null hypothesis of the observed outcome coming from the estimated model.

6. CONCLUDING REMARKS

This paper revisits results from a field experiment conducted to study what interventions may boost museum attendance of high school teens. This is a novel area of application of field experiments and contributes to the economic literature that studies the effects of the use of incentives to nudge behavior and, in particular, to increase cultural consumption. Within a randomized trial where randomization occurred at the class level, students received three encouragements of increasing strength. The idea that underlies this scheme of incentives is that not only later museum attendance could be directly raised by the verbal stimulation of intrinsic motivation, but also by the motivational reinforcement associated with museum experience. We explore causal pathways in this situation taking a principal stratification approach. With respect to relevant latent subgroups, this approach enables us to make causal claims regarding the direct effects of the presentation encouragement on the students' later museum attendance and the indirect effect due to the proposed museum experience. It has

been already highlighted by the literature that principal stratification is appropriate to focus on direct effects of encouragements, as these unambiguously coincide with so-called dissociative effects. Instead, associative effects do not coincide with the indirect effects conceived in mediation analysis (e.g. VanderWeele, 2011). To this regard, this paper shows how an appropriate experimental design may help focus on indirect effects also using the principal stratification framework, without paying the price of adding assumptions (Forastiere et al., 2016). The analysis yields the following results. We find that most students are not really attracted by the proposed visit of Palazzo Vecchio, but there is also a considerable share of students that undertake the experience when offered an extrinsic stimulus, such as the reward promise. The presence of a large subgroup of never takers, which accounts for the majority of students involved in the experiment, allows to voice the direct effect of the intrinsic motivational encouragement towards later museum attendance provided by the presentation. This causal effect is positive; it is also highly probable that it is reinforced by classroom spillovers. The presence of a considerable subgroup of students that comply with the reward promise, which applies only to the proposed visit of Palazzo Vecchio and not later in the follow up period, allows to voice the motivational reinforcement that the experience itself provides. This effect is rather ambiguous, in the sense that, on the one hand, the suggested visit may either induce or crowd out later attendance, with reward compliers possibly reacting to experience in opposite ways. This evidence suggests that it is the motivational encouragement, rather than experience, to play an important role in boosting teens' museum attendance.

Access and consumption of cultural goods are important for young adults and teenagers. They may exert effects that extend over a lifetime and benefits that are relevant both from a personal and a social point of view. Unfortunately, young adults and teens often show little interest and awareness in museum attendance. This study confirms that policies based on nudging individual behavior with appropriate incentives that stimulate the intrinsic component of personal motivation may be promising in terms of changing individual attitudes towards cultural consumption.

REFERENCES

ANGRIST, J. D., IMBENS,G. W., & RUBIN, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association*, 91, 444–472.

BRUHN, M., & MCKENZIE, D., (2009). In pursuit of balance: randomization in practice in development field experiments *American Economic Journal: Applied Economics*, 1(4), 200–232.

ELLIOTT, M. R., RAGHUNATHAN, T. E., & LI, Y. (2010). Bayesian inference for causal mediation effects using principal stratification with dichotomous mediators and outcomes. *Biostatistics*, 11, 353–372.

FORASTIERE, L., MEALLI, F., & VANDERWEELE, T. J. (2016). Identification and estimation of causal mechanisms in clustered encouragement designs: Disentangling bed nets using Bayesian principal stratification. *Journal of the American Statistical Association*, 111(514), 510–525.

FRANGAKIS, C. E. & RUBIN, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58, 21–29.

FRANGAKIS, C. E., RUBIN, D. B., & ZHOU, X. H. (2002). Clustered encouragement design with individual noncompliance: Bayesian inference and application to Advance Directive Forms (with discussion). *Biostatistics*, 3, 147–164.

GALLOP, R., SMALL, D. S., LIN, J. Y., ELLIOTT, M. R., JOFFEE, M., & TEN HAVE, T. R. (2009). Mediation analysis with principal stratification.. *Statistics In Medicine*, 28, 1108–1130.

GUSTAFSON, P. (2010). Bayesian inference for partially identified models.. *The International Journal of Biostatistics*, 6(2), Art. 17.

HIRANO, K., IMBENS, G. W., RUBIN, D. B., & ZHOU, X. H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics*, 1, 69–88.

HONG, G. & RAUDENBUSH, S. W. (2006). Evaluating Kindergarten Retention Policy: A Case Study of Causal Inference for Multilevel Observational Data. *Journal of the American Statistical Association*, 101, 901–910.

HUDGENS, M. G. & HALLORAN, M. E. (2008). Towards causal inference with interference. *Journal of the American Statistical Association*, 103, 832–842.

IMBENS, G. W. & ANGRIST, J. D. (1994). Identification and estimation of local average treatment effects. *Econo-*

metrica, 62, 467–476.

IMBENS, G. W. & RUBIN, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics*, 25, 305–327.

JO, B. (2008). Causal inference in randomized experiments with mediational processes. *Psychological Methods*, 13, 314–336.

KISIDA, B., GREENE, J. P., & BOWEN, D. H. (2014). Creating cultural consumers. The dynamics of cultural capital acquisition. *Sociology of Education*, 87(4), 281–295.

LATTARULO, P., MARIANI, M., & RAZZOLINI, L. (2016). Nudging museums attendance: a field experiment with high school teens. *Journal of Cultural Economics*, doi:10.1007/s10824-016-9285-6.

MATTEI, A. & MEALLI, F. (2011). Augmented designs to assess principal strata direct effects. *Journal of the Royal Statistical Society*, 73(5), 729–752.

MCDONALD, C., HIU, S., & TIERNEY, W. (1992). Effects of computer reminders for influenza vaccination on morbidity during influenza epidemics. *MD Computing*, 9, 304–312.

MEALLI, F. & PACINI, B. (2013). Using secondary outcomes to sharpen inference in randomized experiments with noncompliance.*Journal of the American Statistical Association*, 108(503), 1120–1131.

PAGE, L. C. (2012). Principal Stratification as a Framework for Investigating Mediational Processes in Experimental Settings. *Journal of Research on Educational Effectiveness*, 5, 215–244.

RUBIN, B. D. (1974). Estimating causal effects of treatments in randomized and non randomized studies. *Journal of Educational Psychology* 66, 688–701.

RUBIN, B. D. (1978a). Bayesian inference for causal effects. *Annals of Statistics*, 6, 34–58.

RUBIN, B. D. (1980). Comment on "Randomization Analysis of Experimental Data in the Fisher Randomization Tes" by D. Basu. *Journal of the American Statistical Association*, 75, 591–593.

RUBIN, B. D. (1986). Which Ifs have Causal Answers? Comment on "Statistics and Causal Inference" by P. Holland. *Journal of the American Statistical Association*, 81, 961–962.

RYAN, R. M. DECI, E. L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54–67.

SOBEL, M. E. (2006). What Do Randomized Studies of Housing Mobility Demonstrate?: Causal Inference in the Face of Interference. *Journal of the American Statistical Association,*, 101, 1398–1407.

SOMMER, A. & ZEGER, S.(1991). On estimating efficacy from clinical trials. *Statistics in Medicine*, 10, 45–52.

TANNER, M. A. & WONG, W. H.(1987). The Calculation of Posterior Distributions by Data Augmentation (with discussions). *Journal of the American Statistical Association*, 82, 528–550.

TAMER E. (2010). Partial Identification in Econometrics. *Annula Review of Econometrics*, 2, 167?195.

TCHETGEN TCHETGEN E. J. & VANDERWEELE T. J. (2012). On causal inference in the presence of interference.. *Statistical Methods in Medical Research*, 21, 55–75.

VANDERWEELE T. J. (2008). Simple relations between principal stratification and direct and indirect effects. *Statistics and Probability Letters*, 78, 2957–2962.

VANDERWEELE,T. J. (2011). Principal stratification - Uses and limitations. *The International Journal of Biostatistics*, 7(1), 1–14.