American Economic Journal: Economic Policy 2015, 7(4): 103–131
http://dx.doi.org/10.1257/pol.20130290

Identity-Based Policies and Identity Manipulation: Evidence from Colonial Punjab[†]

By Guilhem Cassan*

I study identity-marker based policies and demonstrate the existence of identity manipulation. I analyze the impact of the Punjab Alienation of Land Act (1901), a caste-based legislation, on identity manipulation. Using data from the colonial census, I show that a movement of identity manipulation took place in response to the law. I estimate that in 20 years, 7.5 percent of the population that had an incentive to do so manipulated its identity. I then use an accounting exercise to estimate that the amount of mistargeting caused by this manipulation is between 3.9 percent and 8.2 percent of the land distributed. (*JEL* J15, N35, N45, N55, Q15, Q24, Z13)

Several countries use identity markers such as skin color or caste as a basis for public policies. Most notably, the United States has an "affirmative action" policy for ethnic minorities, Brazil has quotas in certain universities by skin color, while India has the largest "reservation" program in the world for low castes and tribes (the "Schedules Castes" and "Scheduled Tribes," as well as "Other Backward Classes"). Those policies take social identity markers as proxies for economic or social status: since the groups targeted are on average poorer/less educated/discriminated against, having policies based on those markers might be efficient in a context in which obtaining information on, say, income, is costly. However, this type of policy relies on the assumption that these identity markers are easily identifiable and difficult to manipulate. This paper questions this assumption, since substantial mistargeting of policies may well occur if it turns out to be wrong. In addition, this paper also provides an estimate of the extent of mistargeting linked to identity manipulation. The main contribution of this paper is thus to give the first estimates of such manipulation, convincingly linking a specific caste-based policy to

* University of Namur, Centre of Research in the Economics of Development (CRED), 8 rue du Rempart de la Vierge, Namur, 5000, Belgium (e-mail: guilhem.cassan@unamur.be). Previous versions of this paper have circulated under the title "British Law and caste-identity manipulation: the Punjab Alienation of Land Act." I am grateful to Véronique Bénéï, Luc Behaghel, Denis Cogneau, Lakshmi Iyer, Eliana La Ferrara, Sylvie Lambert, Ian Talbot, Marie Anne Valfort, Ekaterina Zhuravskaya, the anonymous referees as well as several seminar and conference participants for helpful comments. I thank Dave Donaldson for having shared the data on birth place and death rates. I am grateful to the Laboratoire d'Economie Appliquée-INRA and the Centre pour la Recherche Économique et ses Applications (CEPREMAP) for funding the data collection. This paper is produced as part of the project "Actors, Markets, and Institutions in Developing Countries: A Micro-empirical Approach," a Marie Curie Initial Training Network (ITN) funded by the European Commission. The maps of this article have been made with the Philcarto software: http://philgeo.club.fr.

† Go to http://dx.doi.org/10.1257/pol.20130290 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

caste-identity manipulation, as well as describing the extent of mistargeting linked to this manipulation. In order to do so, the case of caste-based legislation in colonial Punjab, a major Province of British India, is presented. The Indian context provides probably the best case study of such identity-based policies, because the Indian identity-based policies concern a wide segment of the population, are intensely debated, and are suspected to be diverted (notably) by caste-identity manipulation.

A second contribution of this paper is to build a dataset that allows the following of castes' population for a 40-year period (1881–1921) at the district level. This data enables me to evaluate the causal effect of a caste-based policy, the Punjab Alienation of Land Act, on caste-identity manipulation in colonial Punjab. By creating a category of "agricultural castes," the membership of which was almost compulsory to acquire land, this law created a very strong incentive to manipulate caste identity in order to claim membership in an agricultural caste.[1]

I estimate that 20 years after the law was passed up to 7.5 percent of the nonagricultural-caste population had manipulated its caste identity. In addition, an accounting exercise using a novel dataset of land ownership by caste and district from 1911 to 1931 enables me to estimate the extent of mistargeting that this caste-identity manipulation led to between 3.9 and 8.2 percent of the land distributed by the Punjabi government during that period.

In order to measure identity manipulation, I identify abnormal variations in the evolution of the population of each caste group. The identification strategy used compares the growth rates of the population of the "agricultural" and "nonagricultural" castes before and after the Alienation of Land Act went into effect. Thus, it relies not on a common trend assumption but on a weaker assumption of stability of trend differences. Additional robustness checks consist of placebo tests and a triple difference with areas in which the law was not in force (the Princely States of Punjab). Alternative demographic interpretations of the abnormal evolution of the population of the two caste groups following the Alienation of Land Act are carefully ruled out. In particular, the effects of migration and of demographic transition are demonstrated to be too weak to explain the results. Moreover, it is shown that this movement of caste-identity manipulation was heterogeneous, and was stronger in the districts in which the "agricultural" status granted additional benefits. The extent to which the land distributed by the government to "agricultural castes" members that was actually obtained by identity manipulators is determined by means of an accounting exercise.

A large body of literature has focused on mistargeting and manipulation in various contexts. Athreya and Somanathan (2008) for example, study the case of spatial misallocation of public goods. Mistargeting due to the use of proxy means tests has also been recently studied by Alatas et al. (2012), who show that they can lead to a mistargeting of up to 30 percent because of their inherent imprecision. In addition, Camacho and Conover (2011) show that manipulation of the proxies was at play in response to the implementation of a proxy-based social program in Colombia.

[1] Note that similar laws were also passed after the Independence to protect certain specific groups (for example, Maharashtra's "Restoration of Lands to Scheduled Tribes Act" of 1974), and were also suspected of creating widespead identity manipulation (Guha 2003).

Manipulation of the variables used to identify individuals by policymakers has also been studied in the taxation literature, with manipulation of taxable income in response to tax rates being one of the standard examples (Saez 2010). Identity-based policies, however, are often assumed to be less subjected to this type of issue, and the economics literature has so far failed to deal with this concern. Indeed, identity is often thought as being given at birth, easily observable, and almost not alterable.[2] As such, identity provides a good candidate for the identification of policy beneficiaries, particularly when it is correlated with economic or social conditions, as is often the case, given that many identity groups often suffer specific discrimination.

However, a good body of evidence points to identity manipulation in response to identity-based policies both in the Indian context[3] and in other countries.[4] However, the extent of identity manipulation, and of the public-policy mistargeting to which it can lead, has been neglected by the economics literature even though there is a strong suspicion that identity manipulation diverts the positive discrimination policy benefits from their beneficiaries on a large scale.[5]

By focusing on the link between public policies and identity manipulation, this paper is linked to the literature on "identity economics" (Akerlof and Kranton 2000), and on identity formation (Caselli and Coleman 2013; Botticini and Eckstein 2007; Bodenhorn and Ruebeck 2003; and Rao and Ban 2007), and adds to this literature by being the first to use panel data.

Section I of the paper presents the law and places it in its historical context; Section II describes the data; Section III describes the causal impact of the Punjab Alienation of Land Act on caste composition; Section IV rules out alternative interpretations of the results; Section V documents the heterogeneity of caste identity manipulation across districts, and Section VI proposes a theoretical framework in which the results may be interpreted. Finally, Section VII presents an accounting exercise in order to estimate the amount of mistargeting that this identity-manipulation might have created.

I. Historical Background

A. *The Punjab Alienation of Land Act*

By the end of the nineteenth century, the indebtedness of the landowners had become a concern for the British authorities: "One of the most significant domestic problem confronting the Indian government … was the growing indebtedness of

[2] Such "primordialist" assumptions are strongly contested in political science by the "constructivists." See, for example, Chandra (2012) for a collection of essays on the question and Posner (2005) for an in-depth case study of ethnic identification in Zambia.

[3] See, for example, the recent scandal of fake caste certificates for admission in higher education institutions in Delhi (*The Hindu* 2011), or, for the case of Scheduled Tribes, the work of Guha (2003).

[4] A large body of sociological literature has documented abnormal variations of ethnic groups in censuses: see, for example, Lieberson and Waters (1993) on American whites or Nagel (1995) on Native Americans. Francis and Tannuri-Pianto (2013) also study the case of self-identification as black in Brazilian universities following the implementation of quotas for blacks.

[5] Indeed, in India, this movement took such an extent that it was termed as "demand for disadvantage" by Somanathan (2008), while Dudley-Jenkins (2003) wrote of "the proliferation of people claiming to be in one of the various backward categories."

the cultivating classes and a concomitant transfer of landed property … to urban moneylenders" (Barrier 1966). This concern was particularly keen in the Province of Punjab, since the Indian army recruited largely from the province (Tai Yong 2005), and more specifically from the landowning castes.[6] Hence, avoiding rural agitation in that province was a prime concern and "… the driving force behind government attempts to find a solution to debt and land transfer was fear for its own position …" (Barrier 1966), as "widespread land alienations, many feared, would lead to rural revolt" (Gilmartin 1988). The Punjab Alienation of Land Act, which went into effect in June 1901, created an "agricultural caste" category: a member of an agricultural caste could transfer the ownership of his land (be it by sale or by mortgage) only to a member of an agricultural caste.[7] Since the members of the agricultural castes accounted for most of the land owners, the members of the nonagricultural castes wanting to acquire land were almost completely prevented from doing so, as only a very small amount of land was available for them to buy.[8] The Alienation of Land Act thus essentially used caste as a proxy for land ownership, ignoring the land owners among the nonagricultural castes and the landless among the agricultural castes.

B. *Additional Benefits of "Agricultural Caste" Membership*

While the law provided substantial protection on the land market, the agricultural castes list was also used to identify the beneficiaries of various important land-related policies. Indeed, the Alienation Act was reinforced by the Punjab Pre-Emption Acts of 1905 and 1913 which gave preemption rights to members of agricultural castes. Even more importantly, "… this categorization … became the basis for eligibility for land grants in the canal colonies. For land distribution after 1900, the administration did not need to nominate specific groups, but could simply rule that in each selected district the agricultural castes, and those castes alone, were eligible" (Ali 1988).[9] Indeed, from the 1880s on, the colonial administration had dug canals, turning "6 million acres of desert into one of the richest agricultural regions in Asia" (Talbot 2007). As the government of Punjab was the owner of most of the land, it chose who was to become a "colonist," and, from 1902 on, allocated land grants on the basis of the agricultural caste status (Ali 1988). It can be seen in Figure 1 that the amount of land distributed by the Punjab government in the canal colonies was massive and accounted for the evolution of the land ownership of the agricultural castes in Punjab as a whole.

In a province in which the vast majority of the population lived in rural areas, being considered as a member of agricultural castes thus became critical after the Alienation of Land Act went into effect.

[6] Many of which were considered to be "martial races," see, for example, Vanden Eynde (2011) on the specificities of military recruitment in Punjab.

[7] See online Appendix 1 for the text of this Act.

[8] As underlined by Barrier (1966), the law was successfully enforced: "Sales to non agriculturists ceased after 1901." Other references emphasize the impact of the law on the nonagricultural castes, such as: "by means of this act moneylenders were practically wiped out of the land market" (Islam 1995).

[9] See also (Darling 1928): "For the most part, [the colonists] were taken from the Central Punjab, and only from the best agricultural tribes."

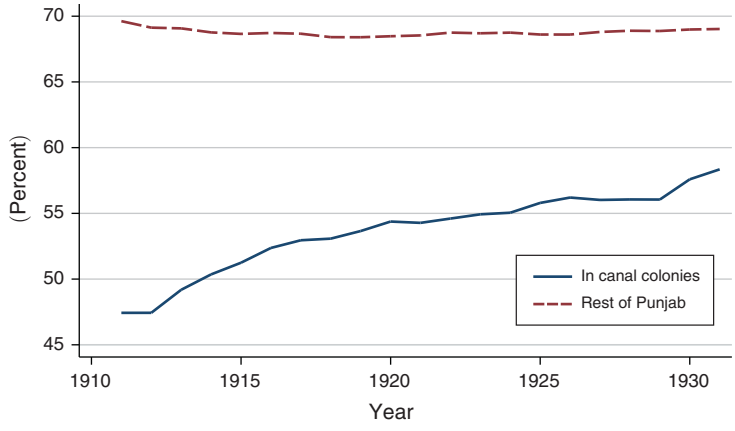

Figure 1. Evolution of the Share of Land Owned by Agricultural Castes, by Canal Colony Status

Source: Reports on the Land Revenue Administration of Punjab, 1911–1931

C. *Impact on Caste-Identity Manipulation*

Various reports of the administration mention the different ways in which attempts to avoid the Alienation of Land Act were made. The first and most obvious one was to lobby the authorities in order to include one's caste on the list of "agricultural castes."[10] Another way to evade the Alienation of Land Act was the use of "benami transactions:" using a member of an agricultural caste to buy or mortgage land for a member of a nonagricultural caste.[11]

The administrative reports also emphasize caste-identity manipulation: "… menials that have acquired money are attempting to get themselves recorded as agricultural tribes with a view to acquiring land …" (Punjab Government 1906), with "… cases of evasive attempt to change tribal designation from a nonagricultural to an agricultural tribe in order to defeat the provisions of the Act …" (Punjab Government 1908). This tendency was reported to be due to individual action, but also sometimes to the mobilization of a caste as a whole: "Frequent cases arise in which application is made by tribes not included in the group notified for the district to have the tribal designation altered to one so included" (Punjab Government 1909). The movement was so wide that specific recommendations were issued by the administration in order to prevent nonagricultural caste members from passing as agricultural.[12] Those attempts can also be found in various census reports, which emphasize a tendency of caste associations to make claims to the British

[10] Indeed, the number of castes considered as agricultural increased over time. See online Appendix 4. To rule out the possibility that the increase of agricultural caste population is due to increases in the number of castes considered as such, the 1921 list of agricultural castes is used throughout the paper. Any caste being on this list is taken to be an agricultural caste throughout the period.

[11] For example, in the Report on the Working of the Punjab Alienation of Land Act for 1908: "What are called benami transactions are reported from most districts. The money lender induces a member of an agricultural tribe … to take land on mortgage for the would-be borrower" (Punjab Government 1909).

[12] Those procedures are reproduced in online Appendix 2.

administration in order to be considered as agricultural: "the introduction of the Punjab Alienation of Land Act … has naturally stimulated … a tendency to claim an affinity with one or the other of the castes declared by Government as agricultural"[13] (Kaul 1912). This took place in a wider context of caste-identity manipulation all across India. Indeed, it has been widely documented[14] that far from being fixed, the caste system, under British rule, was evolving under the action of the caste associations (or caste "sabhas").[15]

II. Data

A. *Caste-Census Data*

To estimate the impact of the Punjab Alienation of Land Act on caste-identity manipulation, I have collected caste census data from 1881 to 1921. Indeed, every decennial census collected caste data, which was then tabulated at the district level. It has been documented that the census was part of the mobilization strategies of the caste associations, which very often claimed new caste names. This makes tracking single castes very difficult over time, as both classifications and names might change.[16] However, the Punjab Census data is of good quality from 1881 to 1921:[17] using the different census reports[18] and the Glossary of the Tribes and Castes of the Punjab and North-West Frontier Province (Rose 1911), I have been able to track the hundreds of changes in classification and names, and merge the newly created caste entries into "caste groups" that are comparable across censuses,[19] thus forming what I believe to be the first dataset of caste-group demography over time at such a disaggregate level.[20] However, the various modifications of district borders and the separation of the North West Frontier Province from Punjab in 1901 as well as the creation of the Delhi Province in 1911 led me to leave aside some districts while merging others in order to assure their comparability over time (see Figure 2).

Overall, I was able to follow 86 caste groups, 24 of which are agricultural[21] in at least one district. They represent from 97.7 percent to 99 percent of the population

[13] This claim persisted through time and can also be found in the Report on the Census of Punjab, 1931: "… on the present occasion more than ever before a tendency was noticeable in various localities,… to return a higher caste. One of the main reasons was a desire to be included in one of the agricultural tribes … to secure exemption from the provisions of the Punjab Alienation of Land Act" (Khan 1933).

[14] From Ghurye (1932), Srinivas (1966) to Dirks (2001), and Bayly (1999).

[15] Those associations were formed in order to "support social advancement" (Assayag 1995) and to gain access to the economic opportunities created by the British presence: "the associations began to press for places in the new administrative and educational institutions and for political representation" (Rudolph and Rudolph 1960).

[16] See Conlon (1981) on that matter.

[17] I do not use the 1871 and 1931 census because they do not report castes at as fine a level as the other years, and so do not allow all the castes to be tracked for those years.

[18] In particular, the census report of 1911 contains an "Ethnographic glossary of castes" listing many caste synonyms.

[19] See online Appendix 3 for the details of this grouping and its justification.

[20] Both geographically fine at the district level and fine at the caste level, since I follow caste groups, and not only "scheduled castes" and "scheduled tribes," as is usually the case in most datasets.

[21] More castes and tribes were actually considered agricultural, but, in order to be able to track them over time, I had to merge them either with other agricultural castes, or with nonagricultural ones (which bias the results downward, as I would then code the whole group as agricultural). The source used for this classification is Lal (1937). See online Appendix 4 for the list of agricultural castes.

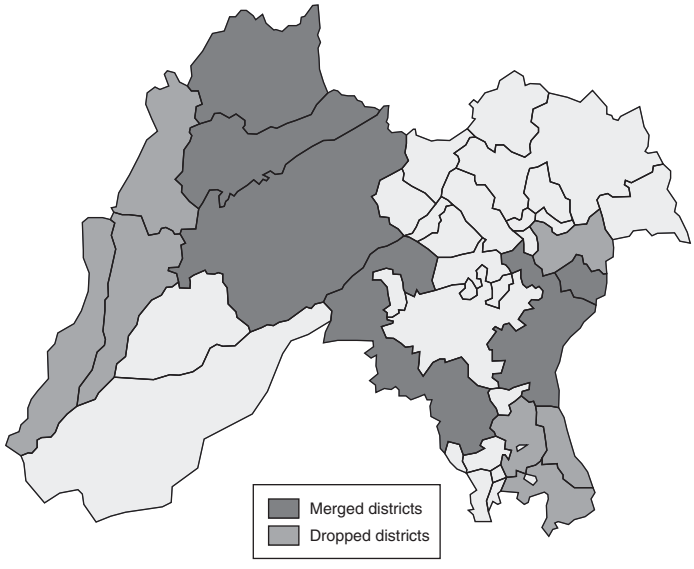

Figure 2. British Punjab: Dropped and Merged Districts

of the 33 districts and states I tracked over time, which themselves account for 88 percent of the population of the Province of Punjab. I have thus built a district-level panel of caste composition so the response of caste groups to the Punjab Alienation of Land Act can be studied at a very fine geographical level. The caste-identity manipulation estimated in this paper is a proxy for the caste-identity manipulation faced by the British bureaucracy, and is probably a lower bound, as there was no immediate incentive to pass as a member of an agricultural caste to the census administration. Indeed, since there was no link between the census and the administration in charge of the Alienation of Land Act, one would want to "lie" to the census only to maintain a coherence of caste declaration across British administrations.

B. *Descriptive Statistics*

The whole Province of Punjab had a population of 24.4 million in 1901, and covered an area of 354,634 square kilometers. It corresponds to the contemporary States of Punjab (Pakistan), Punjab (India), Himachal Pradesh (India), and Haryana (India). As for the rest of India, it was not entirely administered by the British, since some areas, the Princely States, were under the rule of local princes, and as such, were not subject to British law (see Iyer 2010 for more details and Figure 3 for their location). The population of the Princely States was 4.4 million, thus leaving 19.9 million under direct British rule. Hence, most of this paper deals only with the districts directly ruled by the British administration, but the Princely States are also used for triple-difference specifications.

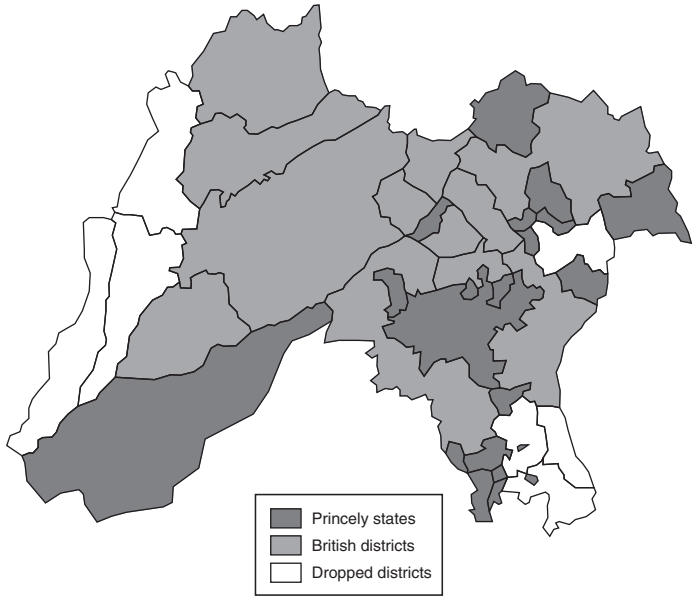

Figure 3. British Punjab: Princely States and British Districts

Table 1—Descriptive Statistics: Districts and States of Punjab, 1901

	British districts	Princely states
Mean population (SE)	1,408,241 (1,081,661)	207,298 (357,096)
Mean population/km2 (SE)	291 (175)	194 (127)
Mean urban population (SE)	10.8 % (0.05)	9.9 % (0.08)
Number of districts/states	12	21

Source: Report on the Census of Punjab, 1901. The figure refers to the districts made comparable over time.

The Province of Punjab was essentially rural, with 89 percent of the population living in a rural area.[22] Hence, most of its population was directly affected by the Alienation of Land Act. Urban residents were also affected if they wanted to own land.

Within the British districts, the population was roughly cut in half between agricultural and nonagricultural castes, as shown in Figure 4. However, the differential evolution of the populations of the two groups is very striking: while the agricultural castes grew slower before 1901, after the law was enacted, the growth rate differential with the nonagricultural castes reversed. Note too, that, from 1901 on, the overall population did not increase as rapidly as it did before. This was due to several demographic shocks affecting the province, which I shall discuss later on.

[22]The urban population is defined as "every municipality of whatever size; all civil lines not included within municipal limits; every cantonment; and every other continuous collection of houses, permanently inhabited by not less than 5,000 persons, which the Provincial Superintendent may decide to treat as a town for census purposes" (Risley 1903).

FIGURE 4. EVOLUTION OF THE POPULATIONS OF AGRICULTURAL VERSUS NONAGRICULTURAL TRIBES IN BRITISH DISTRICTS OF PUNJAB, 1881–1921

Source: Reports on the Census of Punjab, 1881 to 1921

Overall, the evolution of the population of the two caste groups is coherent with the Alienation Act leading to a movement of caste-identity manipulation. The trends in the evolution of the agricultural and nonagricultural caste populations seem to have inverted after 1901.

III. Empirical Approach

A. *Nonagricultural Castes as a "Control" Group*

That only certain castes were considered as "agricultural" by the Alienation of Land Act does not permit one to apply a simple double difference strategy, since the common trend assumption cannot be made here. Indeed, as agricultural castes were not randomly selected, and exhibit systematic differences from nonagricultural ones, they are on average larger (Table 2) and their growth rate before 1901 was on average lower than that of nonagricultural castes, as illustrated in Figure 4. Hence, running a difference-in-differences on levels of population would lead to an underestimation of the amount of identity manipulation.

To account for this, I will compare the growth rates of the populations of the two caste groups before and after the Alienation of Land Act. In this case, the identification relies on the much weaker hypothesis that the differences in the growth rates of the population of agricultural castes versus nonagricultural castes before and after 1901 would have remained stable in the absence of the law but not that their growth rates themselves were similar.

Hence, I will run regressions of the form:

$$(1) \qquad \ln(pop_{idt}) - \ln(pop_{idt-1}) = constant + \beta agr_{id} + \gamma post1901_t + \delta agr_{id} \times post1901_t + \eta X_{dt} + \epsilon_{it}.$$

The growth rate (approximated by the difference in log) of the population of caste i in district d (if the regression is at the district level) during each of the four decades

Table 2—Descriptive Statistics: Population of Castes by Agricultural Status, 1901

	Agricultural castes	Nonagricultural castes
Mean population (standard deviation)	506,789 (930,499)	86,496 (200,510)
Observations	26	91

Source: Report on the Census of Punjab, 1901

t is thus regressed on agr_i, a dummy indicating whether caste i will be considered as an agricultural caste after 1901 in district d; $post1901_t$, a dummy taking a value of 1 when the decade is in the 1901–1921 interval and 0 in the 1881–1901 interval; and X_{dt}, a set of district dummies and district dummies interacted with decade dummies, to control for any possible district-specific change in trend (if the regression is at the district level). As small castes might tend to have more extreme growth rates, all regressions are weighted by the population of the caste in 1881.[23]

I use two main specifications of this regression. In specification 1, I regress the growth rate of the caste population at the British Punjab level, while in specifications 2 and 3, I regress it at each British district level, which allows in specification 3 to control for any district-specific change in trend that might have been driving the results.

While this method is formally identical to a double-difference on the growth rates, this is not precisely the case, because nonagricultural castes are also treated: the growth rate of their population will be affected by the Alienation of Land Act, since some of their members would have passed to the other caste group. The interpretation of the coefficient on $post1901_t$ is thus different from a classic difference-in-differences. Indeed, it is composed not only of the counterfactual change in the growth rate of the agricultural castes after 1901 absent the Alienation of Land Act (as in a typical difference-in-differences), but also of the decrease in the growth rate of the nonagricultural caste population after 1901 in response to the Alienation of Land Act. However, this does not affect the interpretation of the coefficient of interest, $post1901 \times agricultural$, which is the evolution of the difference in the growth rate of the population of agricultural from that of nonagricultural castes. This also might raise concerns about the way in which the residuals are correlated. Indeed, as the same units are observed several times before and after treatment, serial correlation is an obvious issue, which indicates the need to cluster at the caste level (Bertrand, Duflo, and Mullainathan 2004). However, since there was passing from caste to caste, an across-caste correlation of the residuals is not to be excluded, so clustering at the $district \times decade$ level would be indicated. In order to take into account those two concerns simultaneously, all the regressions made at the district level are two-way clustered at the $district \times decade$ and caste levels (Cameron, Gelbach, and Miller 2011).

As can be seen in Table 3, the very precisely estimated coefficient on the interaction between agricultural and $post1901$ is positive and significant in all specifications.

[23] The results are robust to the choice of another year.

Table 3—Impact of the Alienation Act

	Province level	District level	
	(1)	(2)	(3)
*post*1901 × *agr*	0.0799***	0.0789***	0.0751***
	(0.0186)	(0.0122)	(0.0222)
agr	−0.0177	−0.0174	−0.00564
	(0.0126)	(0.00571)	(0.0146)
*post*1901	−0.109***	−0.116***	
	(0.0136)	(0.0269)	
Decade FE	No	No	Yes
District FE	No	No	Yes
Decade × District FE	No	No	Yes
Observations	429	2,640	2,640
Adjusted R^2	0.210	0.068	0.264

Notes: Weighted OLS regressions of castes' population growth rates by decade. Standard errors are clustered at the caste level in column 1 and are two-way clustered at the district-decade and caste levels in columns 2 and 3.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Thus, the average agricultural caste saw its difference in growth rate with the average nonagricultural caste increase by around 8 percentage points for every decade after 1901. This points to a very strong effect of the Punjab Alienation of Land Act on caste-identity manipulation: while the nonagricultural castes were growing slightly more rapidly before 1901, after that date the growth rate differential was completely changed. However, the negative coefficient on the *post*1901 dummy might underline the fact that, after 1901, the average caste tended to see its population increase on average less. This leads us to suspect the existence of demographic shocks that would affect Punjab after 1901.

B. *Demographic Shocks*

Hence, one might argue that the results obtained are not due to caste-identity manipulation, but only that those demographic shocks affected more the nonagricultural castes than it did the others. Indeed, as a matter of fact, between 1901 and 1921, various episodes of epidemic occurred with plague, malaria, and influenza killing millions.

To account for this, I will resort to the neighboring Princely states, which were not subject to the British legislation (see Iyer 2010 for more details). Arguably, the states of Punjab faced the same epidemics as the British districts due to their close proximity but were not affected by the Punjab Alienation of Land Act. This provides a counterfactual that enables me to control for the demographic shocks of the period. The castes located in the Princely States of Punjab were, indeed, similar to the castes of the British districts; they were subject to the same epidemics, but were not affected by the Alienation of Land Act. Hence, if the variations in caste group populations observed in British Punjab are to be attributed to the Alienation of Land Act, I would expect the Princely States caste group not to exhibit any specific change

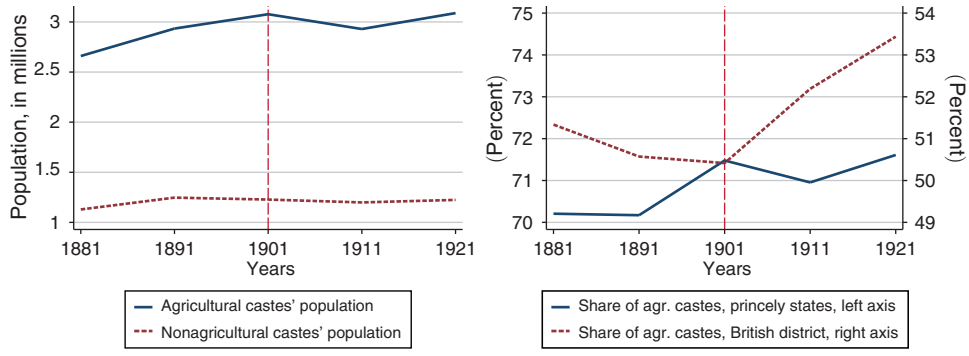

FIGURE 5. EVOLUTION OF THE POPULATIONS OF AGRICULTURAL VERSUS NONAGRICULTURAL TRIBES IN THE PRINCELY STATES OF PUNJAB, 1881–1921

Source: Reports on the Census of Punjab, 1881 to 1921

around 1901 as was, indeed, the case in British Punjab. We can see in Figure 5 that the populations of both agricultural and nonagricultural tribes exhibit relatively similar trends throughout the period in the Princely States.[24]

I will thus estimate regressions of the form:

$$\begin{aligned}(2)\quad \ln(pop_{idt}) - \ln(pop_{idt-1}) = {} & constant + \beta agr_i + \gamma post1901_t \\ & + \delta agr_{id} \times post1901_t + \rho british_d \times agr_{id} \\ & + \pi agr_{id} \times post1901_t \times british_d \\ & + \eta X_{dt} + \epsilon_{itd}.\end{aligned}$$

I use the same notation as in model 1, $british_d$, a dummy indicating that district d is a British district, X_{dt} either the interaction of $british_d$ and $post1901$ (and their main effects) or the interaction of district dummies and decade dummies (and main effects), in district-level specifications. We can see in Table 4 that the coefficient on $post1901 \times british \times agricultural$ is significant and positive in all the specifications, and is comparable to the within-British-districts specifications. Hence, it appears that the tendency for agricultural castes to grow relatively more rapidly than the nonagricultural ones after 1901 than before is specific to British districts. This confirms that the results obtained are not driven by asymmetric demographic shocks but by the impact of the Alienation of Land Act itself. Even more so, the estimated impact of the Alienation of Land Act with this identification strategy is roughly

[24] Note that the Princely States cannot be considered as perfect counterfactuals. As the work of Iyer (2010) has shown, the Princely States and the British districts differed in systematic ways. However, as the identification strategy used does not rely on a common trend assumption, for the Princely States not to be a valid counterfactual, one would need to argue that the difference in the growth rates of agricultural and nonagricultural caste populations would react differently in the British district and in the Princely States when exposed to similar shocks.

TABLE 4—BRITISH DISTRICTS VERSUS PRINCELY STATES

	Province level	District level	
	(1)	(2)	(3)
*post*1901 × *British* × *agr*	0.102***	0.127**	0.126*
	(0.0329)	(0.0606)	(0.0702)
*post*1901 × *agr*	−0.0222	−0.0488	−0.0517
	(0.0207)	(0.0542)	(0.0606)
agr	0.0338***	0.0434	0.0440
	(0.0123)	(0.0427)	(0.0472)
British × *agr*	−0.0514***	−0.0595	−0.0486
	(0.0157)	(0.0432)	(0.0492)
*post*1901 × *British*	−0.0612**	−0.0915	
	(0.0255)	(0.0654)	
*post*1901	−0.0477**	−0.0233	
	(0.0183)	(0.0560)	
British	0.0528***	0.0610*	
	(0.00980)	(0.0347)	
Decade FE	No	No	Yes
District FE	No	No	Yes
Decade × District FE	No	No	Yes
Observations	753	5,725	5,725
Adjusted R^2	0.168	0.051	0.222

Notes: Weighted OLS regressions of castes' population growth rates by decade. Standard errors are clustered at the caste level in column 1 and are two-way clustered at the district-decade and caste levels in columns 2 and 3.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

consistent with the results of the first one, with an implied impact ranging between 10 to 13 percentage points.

Appendix A proposes an additional test, allowing the coefficient on *post*1901 × *agricultural* to vary according to the level of exposure to the diseases of the district, while Appendix B also performs additional placebo tests, and shows that the only year during which the growth rates of agricultural and nonagricultural castes diverged was around 1901. Hence, it seems clear that the Alienation of Land Act had a causal impact on the caste composition of Punjab.

C. *Evaluating the Number of Caste Identity Switchers*

Evaluating the number of caste switchers from the regression tables, however, is not straightforward. Indeed, as both "treatment" and "control" are, in fact, treated, the coefficient on *post*1901 cannot be used to directly infer the counterfactual growth rate of agricultural castes, as discussed earlier. As a matter of fact, the growth rate of caste i during decade d can be written (absent the error term):

$$(3)\quad g_0\mathbf{1}(post_d) + g_1\mathbf{1}(post_d)1(agricultural_i) + g_2\mathbf{1}(post_d)\mathbf{1}(nonagricultural_i) + g_3\mathbf{1}(agricultural_i) + g_4\mathbf{1}(nonagricultural_i).$$

Where g_0 is the change in the growth rate of all castes after 1901, g_1 (respectively, g_2) the increase (or decrease) in the growth rate of agricultural castes (or nonagricultural) due to the entry (or exit) of caste members from nonagricultural castes (or to agricultural castes), g_3 (respectively, g_4) the growth rate of agricultural (or nonagricultural) castes throughout the periods. As $\mathbf{1}(agricultural_i) = 1 - \mathbf{1}(agricultural_i)$, the former expression simplifies to

$$(4) \qquad (g_0 + g_2)\mathbf{1}(post_d) + (g_1 - g_2)\mathbf{1}(post_d)\mathbf{1}(agricultural_i) + (g_3 - g_4)\mathbf{1}(agricultural_i) + g_4.$$

Hence, both the coefficients on *post* and *post* × *agricultural* contain g_2, i.e., the change in the growth rate of the nonagricultural caste due to identity manipulation. Thus, the coefficient on *post* × *agricultural* cannot be directly used to compute the number of persons manipulating their identity. However, we know that $g_1 \times pop\_agricultural_d = -g_2 \times pop\_nonagricultural_d$ where $pop\_agricultural_d$ is the population of the agricultural castes at the beginning of decade d (and respectively for $pop\_nonagricultural_d$); the number of persons leaving the nonagricultural castes must equal the number of persons entering the agricultural castes. Finding the values of g_0, g_1, and g_2 is then just a matter of resolving an equation system of three equations with three unknowns (see online Appendix section 6). Hence, in 1921, there was a total of 656,177 caste identity switchers, which represents 3.8 percent of the total British Punjab population or 7.5 percent of the nonagricultural caste population.

IV. Ruling Out Alternative Interpretations

However, it is unclear how the impact of the law should be interpreted: while the qualitative evidence taken from the census and administrative reports point to caste-identity manipulation, I cannot yet rule out other interpretations. In this section, I propose to rule out the two main alternative interpretations of the results: first, that the caste composition of migration might have changed in reaction to the Alienation of Land Act, and second, that the law created better living conditions for the castes it targeted, thus inducing those castes to enter the demographic transition earlier than the other ones.

A. *Change in the Caste Composition of Migration*

Indeed, a very plausible interpretation of the results would be migration. After the Alienation of Land Act was passed, members of the castes that would be considered as agricultural in the British districts of Punjab had an incentive to migrate from their place of origin to a British district of Punjab in order to benefit from the status that the law gave them. To rule out this interpretation, I use the birthplace statistics of the census. At around 5 percent, the share of the persons residing in a British district but not born there is relatively small. However, the data is not decomposed by caste and does not give the year of migration, while it is only the caste composition

changing by migration taking place after 1901 that could bias the results. In order to compute the migration taking place during that period, I would need to know how many of those who were not born in a British district and residing in such a district in 1901 were still present in 1911 or 1921 (and symmetrically for the emigrants from British districts). The Vital Statistics of India provide yearly district level data on the number of deaths in Punjab. Thus, I can compute the migration taking place between the Princely States and British districts of Punjab between 1901 and 1921 as

$$Immi_{jdt} = pop_{jd}^{x} - pop_{jd}^{x-10} \times \prod_{i=x-10}^{x} survivalrate_{id}.$$

With $Immi_{jdt}$ being the number of immigrants coming from district j to district d during decade t (with j a Princely State or a district located outside of Punjab and d a British district of Punjab, and t either 1901–1911 or 1911–1921), pop_{jd}^{x} and pop_{jd}^{x-10} the number of persons born in district j and enumerated in district d in year $x = 1911$ or 1921, and $survivalrate_{id}$ is 1 minus the death rate of district d in year i. The number of emigrants from British districts to non-British districts[25] can be similarly calculated.

To check if migration is indeed driving the results, I then recomputed the variations of population of each caste group, subtracting the population of immigrants from the population of agricultural castes and adding the population of emigrants, assuming that the migrants are distributed across the different castes proportionally to their respective sizes.[26] Hence, I make the extreme assumption that after 1901 all of the immigrants into British Punjab are agricultural caste members while all of the emigrants leaving British Punjab are nonagricultural caste members. I then reproduce the first identification strategy (described in model 1). Table 5 reports the results and shows that even under the extreme assumption used in this robustness check, the coefficient on *post*1901 × *agricultural* is still positive and significant.

B. *Better Economic Conditions for Agricultural Castes Due to the Alienation of Land Act*

Another possible interpretation could be demographic transition. To assess the validity of this interpretation, I will look at the age composition of each type of caste. If this interpretation is correct, the structure of the age pyramid would appear to be different for each type of caste. Two scenarios (and/or any combination of the two) can be conceived: one in which the fertility rate of the agricultural castes increases dramatically (or the child-death rate decreases dramatically), and one in which the death rates of the older population of agricultural castes decreases. The

[25] As the district of birth of Punjabis enumerated outside of Punjab is not known, the Punjabi emigrants are allocated to each district proportionally to the district's share in the total population. Also, as the data on death rate is not available outside of British Punjab, the death rate of each Punjab's Princely State is assumed to be equal to British Punjab's average death rate; while outside of Punjab, it is assumed that all persons born in Punjab and enumerated after 1901 outside Punjab migrated after 1901.

[26] That is, I subtract x percent of the population of a district's immigrants from the population of an agricultural caste representing x percent of the district's agricultural castes' population, the opposite exercise being done for emigrants and nonagricultural castes.

Table 5—Migration Robustness Check

	Province level	District level	
	(1)	(2)	(3)
*post*1901 × *agr*	0.0545***	0.0516***	0.0474*
	(0.0186)	(0.0152)	(0.0264)
agr	−0.0177	−0.0162**	−0.00452
	(0.0126)	(0.00631)	(0.0156)
*post*1901	−0.0949***	−0.101***	
	(0.0136)	(0.0259)	
Decade FE	No	No	Yes
District FE	No	No	Yes
Decade × District FE	No	No	Yes
Observations	429	2,850	2,850
Adjusted R^2	0.169	0.052	0.243

Notes: Weighted OLS regressions of castes' population growth rates by decade. Standard errors are clustered at the caste level in column 1 and are two-way clustered at the district-decade and caste levels in columns 2 and 3.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

first scenario would result in the base of the age pyramid being relatively larger for agricultural castes, the second scenario would have the top of the pyramid being relatively larger for the agricultural castes.

The census reports give the composition by age of certain castes for the entire Province for the year 1911.[27] The 58 castes for which this information was reported represent 90 percent of the total population of the Province of Punjab in 1911. One can see in Figure 6 that the age structures of the two caste groups appear to be very similar, which would indicate that the Act did not have a large impact on the fertility and infant death rates of the agricultural castes. Indeed, the share of children under the age of 11 (born from 1900 to 1911) in the agricultural castes is slightly lower than the share of the same age group in the nonagricultural castes, thus invalidating a fertility rate/decrease in child death rate scenario to account for the increase in the share of agricultural castes in the population. However, the top of the pyramid is slightly larger for agricultural castes than it is for nonagricultural castes: with 22.93 percent of the agricultural castes' population against 22.27 percent of the nonagricultural caste population. A back of the envelope calculation is sufficient to rule out any major role of this difference in the evolution of the share of the agricultural castes. Indeed, if the share of the persons aged 39 years old or more in the agricultural castes' population was to be 22.27 percent, this age group's population would be 104,000 less than it is.[28] Overall, this would mean a decrease of the total agricultural castes' population for which the age data is available of 0.85 percent, i.e., a decrease of the share of the total population of agricultural castes in 1911

[27] The data also exists for the year 1921, but is not reported here, as the age categories do not allow to distinguish the age groups born after 1901 from those born before as cleanly as the 1911 data allows. Such information was unfortunately not present in previous census years.

[28] This amount is found using this calculation: $Population_{CF39+} = Population_{39-} \times 22.27\%/(1 - 22.27\%)$.

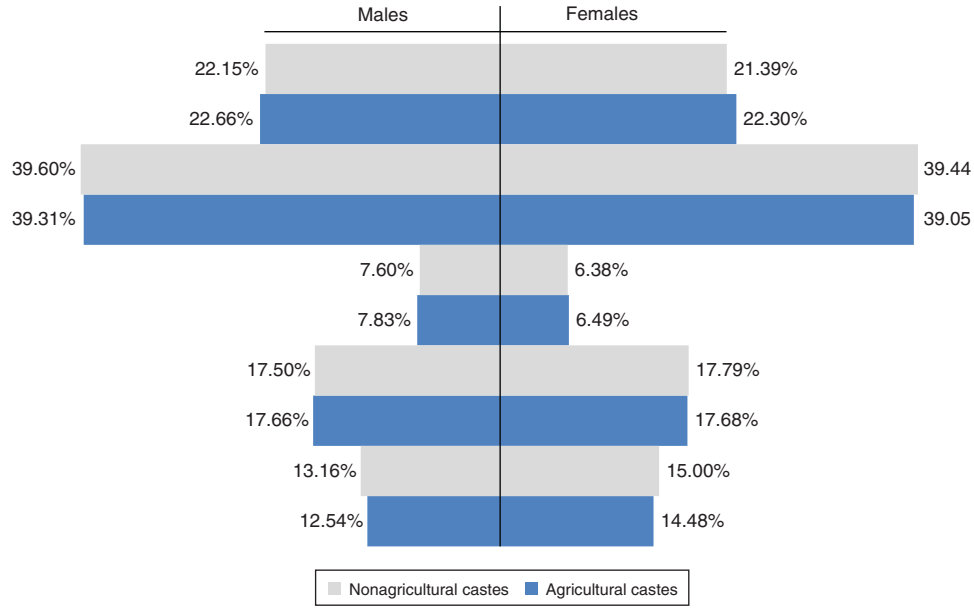

FIGURE 6. AGE PYRAMID BY AGRICULTURAL STATUS, 1911

Source: Report on the Census of Punjab, 1911

of 0.21 percentage points. However, the share of the agricultural castes' population increased by 1.75 percent between 1901 and 1911. The scenario of the Alienation of Land Act reducing the elderly death rate is thus not sufficient to explain the total evolution of the agricultural castes' population.

It thus appears that the interpretations of the results in terms of migration or demographic change can clearly not explain all the variation observed. Hence, and in line with the observations of the British Census administration, the only remaining explanation is caste-identity manipulation.

V. Heterogeneity of the Effect: Access to the Canal Colonies

We have seen that one of the main advantages given by the agricultural caste status was the access to the land of the canal colonies. One of the specificities of the canal colonies was that they were built in almost desert areas.[29] Hence, the grantees had to come from other regions, so migration played a major role in their development.[30] Indeed, the population of the canal colonies vastly increased between 1881 and 1921: the population of the districts in which they are located surged from 5 million to 7.9 million (+57 percent), while the rest of Punjab remained

[29] "These areas […] were practically desert waste supporting no settled population" (Paustian 1930).

[30] "According to the Chenab Colony's final colonization report, the population of the area grew from 112,000 in 1891 to over 1.1 million in 1911, of which the majority were migrants from other parts of the Punjab." (Gilmartin 2004).

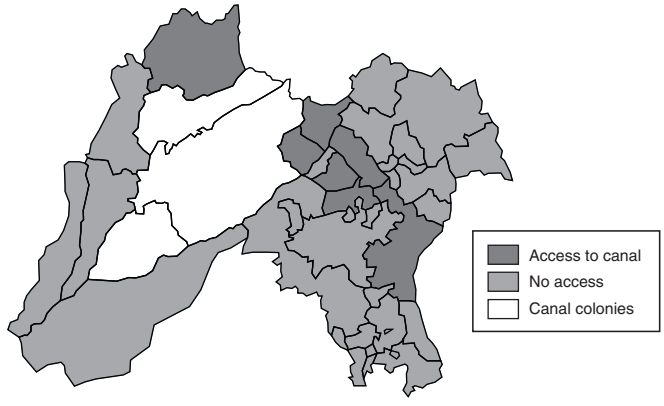

FIGURE 7. DISTRICTS WHOSE POPULATION IS ELIGIBLE TO LAND IN THE CANAL COLONIES

Source: Ali (1988)

relatively stable (+7 percent). This indicates a vast migratory movement within British districts towards the canal colonies.[31] An interesting feature of the process of the colonization of this area of Punjab is that the Punjab Government not only chose the recipients of the land grants with respect to their caste identity from 1901 on, but also the districts of origin of the "grantees" from the beginning of the colonization scheme. Indeed, among the objectives of the colonization was to "provide relief from population congestion..." (Ali 1988). Hence, only certain districts had access to the canal colonies. Figure 7 shows the districts eligible for the canal colonies land according to (Ali 1988). Being a member of an agricultural caste thus gave access to benefits depending on the district of residence, with agricultural caste status yielding a much greater economic advantage in the districts whose residents were eligible to the canal colonies' land.

This suggests that the returns to the agricultural status differed across districts. This pattern allows us to test for heterogeneous effects, using a specification separating the eligible districts from the others:

$$\begin{aligned}(5)\quad \ln(pop_{idt}) - \ln(pop_{idt-1}) = {} & constant + \beta agr_i + \gamma post1901_t + \lambda access_d \\ & + \delta agr_i \times post1901_t + \rho access_d \times agr_i \\ & + \pi agr_i \times post1901_t \times access_d \\ & + \eta post1901 \times access_{dt} + \epsilon_{itd}.\end{aligned}$$

With the same notation as in model 1 and $access_d$ a dummy indicating whether district d had access to the canal colonies, or was itself a canal colony. Table 6 gives

[31] "…the Punjab witnessed a major migration from Central Punjab into the newly opened canal colonies of Western Punjab" (Gilmartin 2004).

Table 6—Heterogeneity of the Effect: Access to the Canal Colonies

	Access level	District level	
	(1)	(2)	(3)
*post*1901 × *access* × *agr*	0.0466***	0.0460***	0.0528**
	(0.0177)	(0.00821)	(0.0249)
*post*1901 × *agr*	0.0389***	0.0370***	0.0289
	(0.0132)	(0.00857)	(0.0189)
agr	−0.0297***	−0.0279***	−0.00989
	(0.00911)	(0.00599)	(0.0163)
access × *agr*	0.0147	0.0136**	0.00625
	(0.0109)	(0.00540)	(0.0255)
*post*1901	−0.0694***	−0.0672*	
	(0.00735)	(0.0344)	
access	−0.00426	−0.00861	
	(0.00704)	(0.0358)	
*post*1901 × *access*	−0.0452***		
	(0.0154)	(0.0453)	
Year dummies	No	No	Yes
District FE	No	No	Yes
Year × District FE	No	No	Yes
Observations	720	2,850	2,850
Adjusted R^2	0.190	0.067	0.251

Notes: Weighted OLS regressions of castes' population growth rates by decade. Standard errors are clustered at the caste level in column 1 and are two-way clustered at the district-decade and caste levels in columns 2 and 3.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

the results of the regression. The first column at the "access level" aggregates the population of each type of British districts, those that had no access to the canal colonies, and those that had access to them/were a canal colony, which, in effect splits the British Punjab in two parts. It can be seen that, while the movement of caste-identity manipulation was widespread throughout the districts of Punjab, as indicates the coefficient on *post*1901 × *agricultural*, it is much greater for the districts that had access to the canal colonies, as is indicated by the positive and significant coefficient on *post*1901 × *access* × *agricultural*.[32]

It is thus shown that caste identity change was up to three times as large in the districts that had access to the land distributed in the canal colonies.

[32] This result could be an artifact due to a within-British-districts migration towards the canal colonies in two steps: first, a member of an agricultural caste would move from a district not having access to the canal colonies to one such district, and second, such a person would migrate to the canal colonies. Online Appendix 5 rules out this interpretation.

VI. Model

A. *Identity Manipulation as a Rational Response to Identity-Based Policies*

This section proposes a simple model to help structure our thinking about the impact of this Act on identity manipulation and of its impact on efficiency and welfare. In order to do so, I will use a framework inspired by models of workfare à la Besley and Coate (1992), modeling the cost of identity manipulation as an entry cost into a government program. The model presented here will thus focus on one aspect of the Alienation of Land Act, its role in the distribution of the land of the canal colonies, leaving aside its impact of land sales and land mortgaging. I consider a risk neutral population composed of two groups A and $\bar{A}$, with population A and $\bar{A} = kA$. The government allocates a package of benefits to individuals on the basis of their group identity: only members of group A are allowed to apply. The government cannot discriminate between applicants apart from their group identity, and thus draws the beneficiaries randomly among applicants. In the context of the Punjab Alienation of Land Act, this package consisted of the various benefits granted on the land market, as well as access to land in the canal colonies. The benefits from obtaining the package depend on the skills of the individuals at exploiting it. Each group g has a distribution of skills such as $S_g^i \in \{S_g^h, S_g^l\}$ with $S_g^h > S_g^l > 0$. There is a proportion θ_g of type h in group g. In our context, those skills can be considered to be land ownership skills. The number T of packages to be allocated is fixed and $T < W$ with W the size of the highest skill group in the population. For simplification, I normalize T to 1. Individuals can choose to manipulate their identity to pass as members of group A in order to be able to apply to the policy. However, identity manipulation comes at a cost c_g. This cost can be thought of as the disutility derived from departing from the group's behavioral norm as in models à la Akerlof and Kranton (2000) in order to pass successfully as a member of caste A. Thus, $c_{\bar{a}} \geq 0$ while, by definition, $c_a = 0$. Hence, for an individual of ability i of group g, the utility U derived from applying to the policy is

$$U_g^i = \begin{cases} S_g^i - c_g & \text{if the package is obtained} \\ -c_g & \text{if the package is not obtained.} \end{cases}$$

It follows that all members of group A apply, and members of ability i of group $\bar{A}$ apply if $E(U_{\bar{a}}^i) \geq 0$.

As a result, there are three possible sizes of the applicant pool:

- If $c_{\bar{a}} > \frac{S_{\bar{a}}^h}{A(1 + k\theta_{\bar{a}})} = \bar{c}$, only group A applies for the good.
- If $c_{\bar{a}} \leq c = \frac{S_{\bar{a}}^l}{A(1 + k)}$, the whole population applies for the good.
- If $\bar{c} \geq c_{\bar{a}} > c$, only group A and type H of group $\bar{A}$ apply.

The thresholds $\bar{c}$ and $\underline{c}$ can be interpreted as the utility value attached to identity: for sufficiently high $c_{\bar{a}}$, individuals of type H (respectively, L) of group $\bar{A}$ are ready to renounce to an amount of utility $\bar{c}$ (respectively, $\underline{c}$) to avoid having to manipulate their identity. It is straightforward to see that the population P of manipulators is a function of $c_{\bar{a}}$:

$$P(c_{\bar{a}}) = \begin{cases} 0 & \text{if } c_{\bar{a}} > \bar{c} \\ k\theta_{\bar{a}}A & \text{if } \bar{c} \geq c_{\bar{a}} > \underline{c} \\ kA & \text{if } c_{\bar{a}} \geq c \end{cases}.$$

As a result, the model gives the following testable comparative statics:

PROPOSITION 1: *For a given level of $S^i_{\bar{a}}$, the higher $c_{\bar{a}}$ is, the lower identity manipulation. Identity-based policies and "primordialist" views of identity in general typically assume that $c_{\bar{a}}$ is so high that manipulation will be extremely marginal. Section III can be considered a test of this primordialist assumption.*

PROPOSITION 2: *For a given level of $c_{\bar{a}}$, the higher the returns from the policy for nontargeted groups $S^i_{\bar{a}}$, the higher the numbers of identity manipulators. As we've seen in Section V, the magnitude of identity manipulation was indeed increasing in proportion to the magnitude of the benefits granted by the agricultural status.*

B. *Efficiency and Welfare Discussions*

Apart from those predictions, this simple model also enables us to think formally about both the efficiency and welfare consequences of identity manipulation.

Efficiency.—Indeed, the efficiency-maximizing policy would be to distribute the package to the highest ability members of the population: members of type H of either caste A or $\bar{A}$. Let $\overline{S_g}$ denote the average ability of members of group g. The policy of allocating goods along caste lines instead of ability is efficiency decreasing, as $\overline{S_a} < S^h_a$ by assumption.

However, in the model, the cost of identity manipulation is similar to a work requirement in a workfare model: only individuals benefiting enough from the policy will be willing to incur this cost. Hence, this cost potentially pushes individuals with sufficiently high skills among $\bar{A}$ to self-select into applying,[33] with potentially efficiency-increasing consequences compared to the case of a caste-based policy without manipulation. In other words, as the policy chosen is only a second best allocation, certain ways of bypassing it might be efficiency-enhancing. Assuming

[33] In a world of perfect financial markets and unobservable skills, the easiest way to get to the first best would be to allocate the policy package by auction. However, in the rural Punjab at the beginning of the 20th century, it is likely that financial markets were highly imperfect.

that the efficiency-maximizing policies discussed above were not possible, we will use two benchmarks: a caste-based policy without manipulation and a nondiscriminatory policy. The average productivity of beneficiaries is a function of $c_{\bar{a}}$:

$$\text{Productivity of beneficiaries} = \begin{cases} \overline{S_a} & \text{if } c_{\bar{a}} > \bar{c} \\ \dfrac{\overline{S_a} + k\theta_{\bar{a}} S_{\bar{a}}^h}{1 + k\theta_{\bar{a}}} & \text{if } \bar{c} \geq c_{\bar{a}} > \underline{c} \\ \dfrac{\overline{S_a} + k\overline{S_{\bar{a}}}}{1 + k} & \text{if } \underline{c} \geq c_{\bar{a}} \end{cases}.$$

It is trivial to see that the efficiency consequences of identity manipulation depend on the relative skills of the manipulators and the nonmanipulators:

- If $\underline{c} \geq c_{\bar{a}}$, manipulation is efficiency enhancing if: $\overline{S_a} < \overline{S_{\bar{a}}}$.
- If $\bar{c} \geq c_{\bar{a}} > \underline{c}$, manipulation is efficiency enhancing if: $\overline{S_a} < S_{\bar{a}}^h$.

Hence, if $S_a^i = S_{\bar{a}}^i$ for all $i = h, l$, identity manipulation has an ambiguous effect on efficiency. Indeed, for intermediary levels of $c_{\bar{a}}$, manipulation is unambiguously efficiency enhancing. In that case, a caste-based policy with costly identity manipulation is preferable to both a non-caste-based policy and a caste-based policy with no manipulation, as it allows us to screen out the low-skill population among the $\bar{A}$ and increase the average ability of applicants.

On the contrary, for low $c_{\bar{a}}$, manipulation is efficiency enhancing only if $\theta_{\bar{a}} > \theta_a$. Given the design of the policy, this would be extremely surprising. In addition, for low $c_{\bar{a}}$, the caste-based policy is similar to a non-caste-based policy in terms of efficiency, as it does not manage to screen out anyone, but is more costly in terms of welfare as a portion of the population would incur a cost that could be avoided.

Welfare.—This leads us to discuss the welfare implications of a policy based on identity. We can easily compute the levels of welfare (computed as the sum of utilities) associated with the different levels of $c_{\bar{a}}$, and compare them to the welfare attained with a nondiscriminatory policy ($c_{\bar{a}} = 0$):

$$\text{Welfare}(c_{\bar{a}}) = \begin{cases} \overline{S_a} & \text{if } c_{\bar{a}} > \bar{c} \\ \dfrac{\overline{S_a} + k\theta_{\bar{a}} S_{\bar{a}}^h}{1 + k\theta_{\bar{a}}} - kA\theta_{\bar{a}} c_{\bar{a}} & \text{if } \bar{c} \geq c_{\bar{a}} > \underline{c} \\ \dfrac{\overline{S_a} + k\overline{S_{\bar{a}}}}{1 + k} - kAc_{\bar{a}} & \text{if } \underline{c} \geq c_{\bar{a}} \end{cases}.$$

It is straightforward to see that, for low levels of $c_{\bar{a}}$, a caste-based policy is welfare decreasing: the population of applicants is the same as without caste discrimination, but a share is incurring the cost of identity manipulation, for a total welfare loss of $kAc_{\bar{a}}$. Similarly, for high levels of $c_{\bar{a}}$, a caste-based policy is welfare

enhancing if $\overline{S_a} > \overline{S_{\bar{a}}}$, i.e., if the policy enables the selection of relatively highly skilled individuals without anyone having to incur the cost of identity manipulation. Finally, for intermediary levels of $c_{\bar{a}}$, the policy is welfare enhancing if $\frac{\overline{S_a} + k\theta_{\bar{a}} S^h_{\bar{a}}}{1 + k\theta_{\bar{a}}} > S^l_{\bar{a}} + \frac{1 + k \quad \theta_{\bar{a}}}{1 + k\theta_{\bar{a}}\, 1 - \theta_{\bar{a}}} A c_{\bar{a}}$, that is, if the policy permits the selection of relatively highly skilled individuals and the total cost of identity manipulation of high-skilled members of $\bar{A}$ is not too high.[34] Thus, the model allows us to formally think about the potential welfare and efficiency consequences of identity manipulation. In particular, it shows that identity manipulation is potentially welfare- and efficiency-increasing if it has an intermediate cost: high enough to prevent low-skilled individuals from applying while low enough to allow high-skilled individuals to pass as agricultural castes.

VII. Mistargeting of Land in the Canal Colonies

A. *Data Sources on Land Ownership in Colonial Punjab*

This section focuses on the land distribution made by the Punjab government in the canal colonies. The legislation on the land distribution by the Punjab government has its own history, and is detailed in online Appendix 7. Interested readers can also refer to Ali (1988), which is the main historical reference of this section. In order to document the mistargeting of government land due to caste-identity manipulation, I compiled land ownership data in colonial Punjab from two main sources: the annual reports on the Land Revenue Administration of the Punjab (published annually from 1862 to 1940) and the reports on the working of the Punjab Alienation of Land Act (published annually from 1902 to 1909). They contain information on annual sales at the district level (area sold). With the creation of the agricultural caste category, they distinguish between members of the agricultural castes and members of other castes from 1902 on. In particular, they distinguish the sales between and within the two types of castes for the entire period under scrutiny. From 1910 on, they also contain data on land ownership (area owned and number of owners), and distinguish three types of land owners: member of agricultural castes, member of other castes, and "shamilat, village abadi and Government property," which are various forms of communal ownership, the bulk of which is constituted by government property.[35] In order to acquire the ownership of their land, "colonists" had to have occupied the land allocated to them by the government for at least ten years.[36] Hence, land ownership data for the period 1911–1931 is relevant for the population that manipulated its caste identity during the 1901–1921 period.

These data thus allow one to specify very precisely the evolution of the land ownership of agricultural castes (in terms of the number of owners and the total area owned), and the origin of this evolution, (sales between caste groups or the

[34] One can easily verify that the cost dimension is an increasing function of $\theta_{\bar{a}}$.

[35] Online Appendix 8 details the construction of the data.

[36] Please refer to online Appendix 7 for the evolutions of the rules of land ownership acquisition for "colonists." For more details, Ali (1988) proposes a thorough analysis of the history of the canal colonies.

distribution of government's land in the "canal colonies"). As described in online Appendix 12, the total amount of sales between members of agricultural castes and nonagricultural castes accounted for only a tiny fraction of the land owned,[37] and cannot account for the evolution of the land ownership of the agricultural castes. The remaining part of this section will thus focus only on the transfers of land from the government to the agricultural castes, the data having been cleaned of the effect of land sales between caste groups.[38]

B. *An Estimate of Mistargeting of Land Distribution*

As the data are aggregated at the district level, it is not possible to link an individual who manipulated her caste identity to her access to government land. I will thus resort to an accounting exercise aimed at bounding the total mistargeting.[39] Hence, this section should be taken as illustrative evidence, and the estimates proposed interpreted with much more caution than the rest of the paper.

The method proposed in Section IIIC can be applied to the coefficients of Table 6 and used to estimate the number of caste identity switchers in the canal colonies' districts and in the districts having access to them. Table 6 gives descriptive evidence of two types of identity manipulation: the coefficient on $post1901 \times agricultural$ would give estimates of individuals who manipulated their identity to obtain the "standard" policy advantages offered to agricultural castes, while the coefficient on $post1901 \times access \times agricultural$ would give the number of manipulators wanting to obtain access to government land. Only the latter are of interest here.[40] For the purpose of this exercise, the coefficient on $post1901 \times access \times agricultural$ presented in Table 6, a simple correlation, is given here a causal interpretation. With the information on the evolution of the number of land owners, I can compute the share of manipulators among the agricultural caste members obtaining land grants from the government and from that, the share of land attributed to "manipulator." Online Appendix 15 gives the computation of the share of manipulators among the beneficiaries of the grant, which is estimated to be between 14.7 percent and 32.5 percent for 1911–1921 and 15 percent and 31 percent for 1921–1931. On Figure 8, panel A depicts the amount of land obtained by caste-identity manipulators every year between 1911 and 1931, while panel B shows the accumulation of land obtained by manipulators with respect to the total amount of land distributed. Hence, in 1931, the total land obtained by manipulators amounted to between 3.8 and 8.2 percent of the total area allocated by the government during that period.

[37] The total area sold from nonagricultural to agricultural castes between 1902 and 1931 represented only 1.3 percent of the land privately owned in 1931.

[38] For a detailed account of the data construction, see online Appendices 8, 9, 11, 13, and 14.

[39] By "mistargeting" I mean the allocation of land to beneficiaries not intended by the government.

[40] See online Appendix 6.2 for the detail of the calculation. It is to be noted that since it is only the total amount of caste identity manipulators that need to be equal between nonagricultural caste and agricultural castes, this method proposes two estimates for the number of caste identity manipulators that manipulated *in response to the access to government land*.

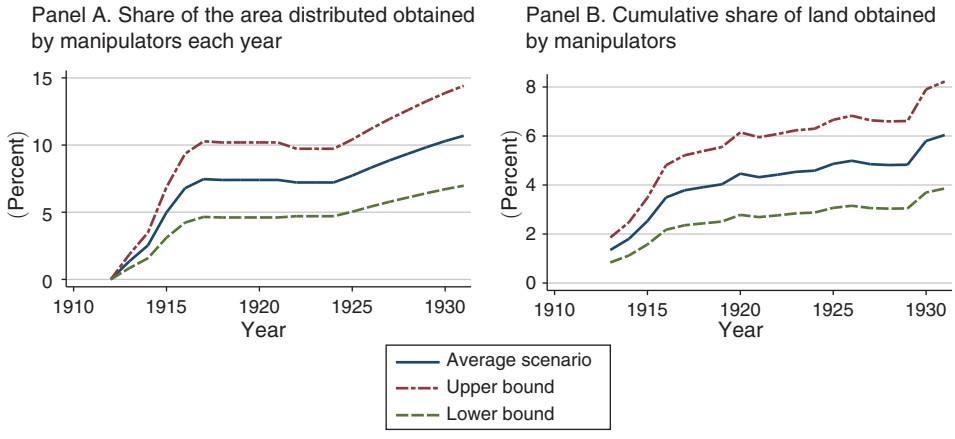

FIGURE 8. LAND OBTAINED BY CASTE IDENTITY MANIPULATORS

VIII. Conclusion

Using various identification strategies, this paper shows that the enactment of the Punjab Alienation of Land Act in 1901 created an agricultural caste category with almost exclusive access to land and in turn, created a large movement of caste-identity manipulation. Indeed, caste groups were given a very strong incentive to manipulate their caste identity in order to benefit from the Alienation of Land Act, and from 1901 on, the trend of the population of agricultural castes exhibited a relative increase of 8 to 12 percentage points per decade. As this effect only occured in the British districts of Punjab and did not vary with the exposure to the various epidemics of the period, I can rule out that the various demographic shocks of the period drive the results. Moreover, I show that neither migration nor demography can explain this evolution, so the results are driven by the ability of caste groups to manipulate their identity in response to administrative incentives. Up to 3.8 percent of the total population (7.5 percent of the population of the agricultural castes) manipulated its caste identity in order to benefit from the protection of the Punjab Alienation of Land Act. It is estimated that this identity manipulation led to a misallocation of government land of up to 8.2 percent. As shown in the model, it is possible that this identity manipulation, by allowing highly skilled individuals to apply, was both efficiency- and welfare-enhancing.

APPENDIX

A. *Disease Environment Robustness Check*

This Appendix complements Section IIIB by proposing an additional robustness check on the impact of the disease environment on the relative evolution of the two caste groups. The Reports on the Sanitary Administration of Punjab contain the yearly death rates of each British district. Thus, I can check the extent to which each district was affected by diseases and see if it is the districts that were the most

Table 7—Controlling for Diseases

	Controlling for death rate	
	(1)	(2)
*post*1901 × *agr*	0.0629***	0.0573***
	(0.00637)	(0.0215)
*post*1901 × *agr* × *deathrate*	0.0204	0.0153
	(0.0268)	(0.0318)
agr × *deathrate*	−0.0207	−0.0251
	(0.0282)	(0.0316)
agr	−0.00441	0.0126
	(0.00966)	(0.0176)
*post*1901	−0.0898***	
	(0.0346)	
deathrate	−0.00409	
	(0.0244)	
*post*1901 × *deathrate*	−0.0351	
	(0.0324)	
Decade FE	No	Yes
District FE	No	Yes
Decade × District FE	No	Yes
Observations	2,079	2,079
Adjusted R^2	0.075	0.274

Notes: Weighted OLS regressions of castes' population growth rates by decade. Standard errors are two-way clustered at the district-decade and caste levels.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

affected that saw their agricultural castes have their share in the population increase the most. I create the variable "deathrate" which centers and reduces the average death rates of each district over each decade, allowing for an interpretation of the coefficient on "deathrate" as the effect of an increase by a standard deviation of the death rates. Table 7 shows the results of the regression of the growth rates of caste groups on their agricultural status interacted with the difference with the average death rate. It can be seen that the coefficient on *post*1901 × *agricultural* is not affected, while the coefficient on *post*1901 × *agricultural* × *deathrate* is not significant, which indicates that the districts most affected by the epidemics do not exhibit a significantly different pattern than does the average district.

B. *Placebo Tests*

Another test for the causal impact of the Punjab Alienation of Land Act on the caste composition of Punjab is a placebo test: it might be that the difference in the growth rates of the population of the two caste groups often switches sign and that such a change happened around 1901. To test for this eventuality, I resort to simple placebo tests, and show that the only time at which a significant change in the difference of the growth rates between agricultural castes and nonagricultural castes happened was around 1901. I reproduce the same specification as described in model 1, but using three different time windows: 1881–1901, with the turning

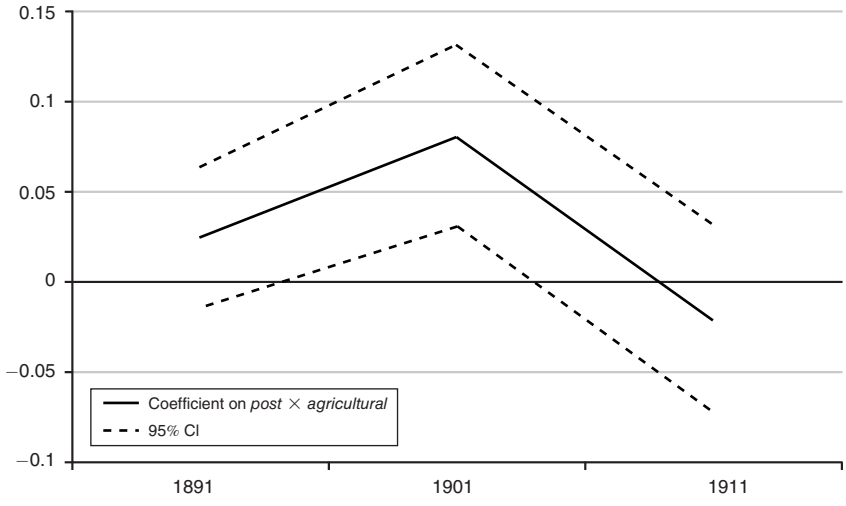

FIGURE 9. PLACEBO TESTS REGRESSIONS

point at 1891, 1891–1911, with the turning point at 1901 and finally 1901–1921 with the turning point at 1911. If it was really the Alienation Act that caused the change in the difference in the growth of the two caste groups, then only the coefficient on *post* × *agricultural* associated to the 1901 turning point should be positive and significant, while the two other turning points should have a small and nonsignificant coefficient. Figure 9 shows the three coefficients on *post* × *agricultural*, with a varying turning point. It can be seen that the only positive and significant coefficient is the one associated with 1901, which is in line with the Alienation of Land Act affecting the caste composition of Punjab.

REFERENCES

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115 (3): 715–53.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Julia Tobias. 2012. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." *American Economic Review* 102 (4): 1206–40.

Ali, Imran. 1988. *The Punjab Under Imperialism, 1885–1947*. Princeton: Princeton University Press.

Assayag, Jackie. 1995. "The Making of Democratic Inequality. Caste, Class, Lobbies and Politics in Contemporary India (1880–1995)." *Pondy Papers in Social Sciences* 18: 5–16.

Athreya, Siva, and Rohini Somanathan. 2008. "Quantifying Spatial Misallocation in Centrally Provided Public Goods." *Economics Letters* 98 (2): 201–06.

Barrier, Norman G. 1966. *The Punjab Alienation of Land Bill of 1900*. Durham, NC: Duke University.

Bayly, Susan. 1999. *Caste, Society and Politics in India: From the Eighteenth Century to the Modern Age*. Cambridge: Cambridge University Press.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119 (1): 249–75.

Besley, Tim, and Stephen Coate. 1992. "Workfare versus Welfare: Incentive Arguments for Work Requirements in Poverty-Alleviation Programs." *American Economic Review* 82 (1): 249–61.

Bodenhorn, Howard, and Christopher S. Ruebeck. 2003. "The Economics of Identity and the Endogeneity of Race." National Bureau of Economic Research (NBER) Working Paper 9962.

Botticini, Maristella, and Zvi Eckstein. 2007. "From Farmers to Merchants, Conversions, and Diaspora: Human Capital and Jewish History." *Journal of the European Economic Association* 5 (5): 885–926.

Camacho, Adriana, and Emily Conover. 2011. "Manipulation of Social Program Eligibility." *American Economic Journal: Economic Policy* 3 (2): 41–65.

Cameron, Colin, Jonah Gelbach, and Douglas Miller. 2011. "Robust Inference with Multiway Clustering." *Journal of Business and Economic Statistics* 29 (2): 238–49.

Caselli, Francesco, and Wilbur John Coleman. 2013. "On the Theory of Ethnic Conflict." *Journal of the European Economic Association* 11 (S1): 161–92.

Cassan, Guilhem. 2015. "Identity-Based Policies and Identity Manipulation: Evidence from Colonial Punjab: Dataset." *American Economic Journal: Economic Policy*. http://dx.doi.org/10.1257/pol.20130290.

Chandra, Kanchan, ed. 2012. *Constructivist Theories of Ethnic Politics*. Oxford: Oxford University Press.

Conlon, Frank F. 1981. "The Census of India as a Source for Historical Study of Religion and Caste." In *The Census in British India: New Perspectives*, edited by N. Gerald Barrier, 103–18. New Delhi: Manohar Publications.

Darling, Malcolm Lyall. 1928. *The Punjab Peasant in Prosperity and Debt*. Oxford: Oxford University Press.

Dirks, Nicholas B. 2001. *Castes of Mind: Colonialism and the Making of Modern India*. Princeton: Princeton University Press.

Dudley-Jenkins, Laura. 2003. *Identity and Identification in India: Defining the Disadvantaged*. London: Routledge.

Francis, Andrew M., and Maria Tannuri-Pianto. 2013. "Endogenous Race in Brazil: Affirmative Action and the Construction of Racial Identity among Young Adults." *Economic Development and Cultural Change* 61 (4): 731–53.

Ghurye, G. S. 1932. *Caste and Race in India*. London: Routledge and Kegan Paul.

Gilmartin, David. 1988. *Empire and Islam: Punjab and the Making of Pakistan.* Oakland: University of California Press.

Gilmartin, David. 2004. "Migration and Modernity: The State, the Punjabi Village and the Settling of the Canal Colonies." In *People on the Move: Punjabi Colonial and Post-Colonial Migration*, edited by Ian Talbot and Shinder Thandi, 3–17. Oxford: Oxford University Press.

Guha, Sumit. 2003. "The Politics of Identity and Enumeration in India c. 1600–1990." *Comparative Studies in Society and History* 45 (1): 148–67.

The Hindu. 2011. "Fake caste certificates racket for getting admission to DU detected." *The Hindu.* June 27.

Islam, M. Mufakharul. 1995. "The Punjab Land Alienation Act and the Professional Moneylenders." *Modern Asian Studies* 29 (2): 271–91.

Iyer, Lakshmi. 2010. "Direct Versus Indirect Colonial Rule In India: Long-Term Consequences." *Review of Economics and Statistics* 92 (4): 693–713.

Kaul, Pandit Harikishan. 1912. *Report on the Census of Punjab 1911, Pt. 2*. Government of India.

Khan, Ahmad Hasan. 1933. *Report on the Census of Punjab 1931, Pt. 1*. Lahore: Punjab Government.

Lal, Shadi. 1937. *Commentaries on the Punjab Alienation of Land Act*. Lahore: Lahore University Book Agency.

Lieberson, Stanley, and Mary C. Waters. 1993. "The Ethnic Responses of Whites: What Causes Their Instability, Simplification, and Inconsistency?" *Social Forces* 72 (2): 421–50.

Nagel, Joane. 1995. "American Indian Ethnic Renewal: Politics and the Resurgence of Identity." *American Sociological Review* 60 (6): 947–65.

Paustian, Paul W. 1930. *Canal Irrigation in the Punjab: An Economic Inquiry Relating to Certain Aspects of the Development of Canal Irrigation by the British in the Punjab*. New York: Columbia University Press.

Posner, Daniel N. 2005. *Institutions and Ethnic Politics in Africa*. Cambridge: Cambridge University Press.

Punjab Government. 1906. *Annual Report on the Punjab Alienation of Land Act, XIII of 1900, as amended by Punjab Act I of 1907: For the year ending the 30th September 1905*. Lahore: Civil and Military Gazette Press.

Punjab Government. 1908. *Annual Report on the Punjab Alienation of Land Act, XIII of 1900, as amended by Punjab Act I of 1907: For the year ending the 30th September 1907*. Lahore: Civil and Military Gazette Press.

Punjab Government. 1909. *Annual Report on the Punjab Alienation of Land Act, XIII of 1900, as amended by Punjab Act I of 1907: For the year ending the 30th September 1908*. Lahore: Civil and Military Gazette Press.

Rao, Vijayendra, and Radu Ban. 2007. "The Political Construction of Caste in South India." http://eml.berkeley.edu/~webfac/bardhan/e271_f07/rao.pdf.

Risley, Herbert Hope. 1903. *Report on the Census of India, 1901*. Kolkata: Office of the Superintendent of Government Printing.

Rose, H. A. 1911. *Glossary of the Tribes and Castes of the Punjab and North-West Frontier*. Lahore: Government Printing Press.

Rudolph, Lloyd I., and Susanne Hoeber Rudolph. 1960. "The Political Role of India's Caste Associations." *Pacific Affairs* 33 (1): 5–22.

Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (3): 180–212.

Somanathan, Rohini. 2008. "La Demande pour le désavantage." *Afrique Contemporaine* 2 (226): 111–31.

Srinivas, M. N. 1966. *Social Change in Modern India*. Bombay: Allied Publishers.

Tai Yong, Tan. 2005. *The Garrison State: Military, Government and Society in Colonial Punjab, 1849–1947*. London: Sage Publications.

Talbot, Ian. 2007. "The Punjab under Colonialism: Order and Transformation in British India." *Journal of Punjab Studies* 14 (1): 3–10.

Vanden Eynde, Oliver. 2011. "Military Service and Human Capital Accumulation: Evidence from Colonial Punjab." http://personal.lse.ac.uk/vandeney/Milservice_March11.pdf.