Unequal enforcement: How perceived ethnicity and written English proficiency affect welfare agencies’ responses

Preliminary Analysis Plan

Spencer Headworth∗
Viridiana Rios†

Objectives

Numerous studies have revealed systematic inequalities in government-individual relationships and access to state services and resources (Ayres and Siegelman 1995, Butler and Broockman 2011, Doleac and Stein 2013, Giulietti and Vlassopoulos 2015, Milkman et al. 2014, Turner et al. 2013, White et al. 2015). Other research has found less evidence of racial/ethnic discrimination in access to government programs for the poor (Einstein and Glick 2017).

The present study tests for discrimination in online systems for public reporting of suspected fraud in the Supplemental Nutrition Assistance Program (SNAP), examining the effects of race/ethnicity and written English proficiency. Our field experiment tests if public assistance fraud control entities respond to public fraud reports differently when those reports are raised by persons who are ostensibly Latina or white, and ostensibly native or non-native English speakers. The results shed new light on questions of between-group differences in access to state agencies. The study’s findings will also reveal how systematic differences in the ways fraud control cases are originated (or not originated) affect the outcomes of states’ fraud control activities.

This research extends the existing literature in three main ways. First, and most basically, this is the first study to directly examine individuals’ capacities to affect SNAP fraud control activity. Existing data offer some insights into how fraud control workers perceive reports of suspected fraud from members of the public, but do not offer systematic insights into how peoples’ claims on public agencies are adjudicated at this pivotal gatekeeping stage.

Second, existing research focuses primarily on how state agencies respond to people’s ostensible requests for information or services for themselves. Through experimentally varying the characteristics of the people who are ostensibly reporting SNAP fraud, our research addresses

∗ Purdue University, Department of Sociology
† Purdue University, Department of Political Science
a different type of state-individual interaction: how rule enforcement bodies respond to distinct categories of potential informants.

Third, in addition to testing for potential discrimination on the basis of race/ethnicity (Latina vs. white), our study adds written English proficiency as an independent variable. The ability to assess differences in agency responses across categories of both race/ethnicity and language skills illuminates new aspects of between-group differences in access to the state.

Research Design

In the United States, state-level public assistance agencies employ dedicated fraud investigation staff to police clients in SNAP, the largest federal means-tested nutritional assistance program. These investigators are civilian employees, usually located within public assistance agencies themselves. Their function, however, is closer to that of police officers than it is to conventional helping-oriented public assistance work. They are tasked with detecting, investigating, and substantiating clients’ alleged violations of program rules. Their ultimate objective is removing clients who have been determined to be in violation of rules from program participation, temporarily or permanently.

Fraud investigations usually begin with someone bringing a suspicion about a program participant or applicant to the attention of the fraud unit. These case referrals come from three main sources: 1) Referrals from public assistance agency workers; 2) referrals from other state agencies, including police departments; and 3) reports from members of the public. Fraud units maintain online submission systems and dedicated email addresses for members of the public to report suspected fraud; these public report channels provide a useful medium for testing fraud unit responses.

Our audit study (Baldassarri and Abascal 2017) tests real-world responses to systematically varied experimental conditions. Audit studies provide evidence of discrimination in areas such as hiring on the basis of race (Bertrand and Mullainathan 2004, Bertrand, Chugh, and Mullainathan 2005), sexual orientation (Tilcsik 2011), criminal record (Pager 2007, Pager, Bonikowski and Western 2009), and educational credentials (Gaddis 2015). Similarly, recent research on hiring at large law firms has found interactional effects of social class and gender (Rivera and Tilcsik 2016).

Rather than the resumes commonly used in labor market audit studies, our experimental instruments are pretextual written fraud reports submitted to online public fraud reporting
systems. Each pretextual report is brief (one paragraph) and makes reference to the same common type of alleged rule violation (an unreported income earner living in the home). In the reports, we vary reporters’ apparent race/ethnicity and written English proficiency.

Independent variables: Race/ethnicity and English proficiency

Building on documented success in previous studies using similar methods to examine different questions, we employ names highly associated with particular racial/ethnic groups as an indicator of reporter race/ethnicity. To date, many experimental designs have assumed that certain racialized names provide strong signals of race/ethnicity to those who read study instruments. There has been little effort, however, to confirm the validity of these operationalizations through measurement of different names’ strength as signals of race/ethnicity (Gaddis 2017a: 470). Thanks to Gaddis’ (2017a, 2017b) recent survey work, our study offers notable methodological strength in this area. Using Amazon’s Mechanical Turk, Gaddis asked respondents with U.S. addresses to indicate the race or ethnicity (if any) they associated with different names, thereby testing the strength of those names as race/ethnicity signals. From Gaddis’ results, we are able to design our experiment with a high degree of confidence in the validity of our operationalization of ethnicity via reporter names.

In his examination of white-typed female given names, Gaddis (2017a) finds numerous given names as strong signals of whiteness, with twenty given names out of forty-one read as “white” by more than 90% of respondents. His examination of Latina-typed female given names, though, reveals these names to be relatively weaker signals of race/ethnicity to a U.S. respondent base, with only two names out of eighteen read as “Latina” by more than 90% of respondents (Gaddis 2017b). However, pairing “Latina” given names with “Latina” surnames increases the signal’s strength dramatically: “correct” racial/ethnic identification exceeded 94% for all Latina given names when paired with Latina surnames (Gaddis 2017b). Adding “white” surnames also generally increased the strength of given names as signals of whiteness (Gaddis 2017a).

Based on these results, we use strongly racially/ethnically typed given name and surname pairs for both our ostensibly Latina and our ostensibly white reporters. We do not provide the specific names used for the experiment, to prevent the possibility of archived copies of our pretextual reports being identified.

Because our interest is in effects of reporter characteristics on report disposition, we do not use any specific names for the reported party that could imply a race/ethnicity. Instead, we use generic descriptors such as “coworker” and “neighbor.” As these descriptors suggest, we hold
the relationship between the reporter and the reported party constant at the approximate “acquaintance” level.

We also hold constant other major characteristics of the report that could potentially influence its disposition. All reported parties are described as female, and as having an unspecified number of children (but more than one). All reports allege suspected eligibility fraud in the SNAP program via the presence of an unreported income earner in the home (one of the most common allegations fraud control units investigate). We also held constant other basic components of the reports, such as approximate length. We do not provide the specific text of the experimental reports, to prevent the possibility of archived copies of our pretextual reports being identified.

We vary written English proficiency through modifying the text of the reports themselves. We began with “non-error” versions of report text, without errors of written English. Working from these versions, we introduced into each report written English errors. Because of our interest in potential interational effects of Latina ethnicity and lower English proficiency, we introduced errors of written English characteristic of native Spanish speakers to create “error” versions. We originally derived these errors from the list of such errors contained in Norman Coe’s (1987) chapter on speakers of Spanish and Catalan in Swan and Smith’s Learner English, and then refined them further with the assistance of an external consultant with expertise in the acquisition of English as a new language for native Spanish speakers.

Thus, each report exists as a pair: one “non-error” version and a matched “error” version. Combining the English proficiency variable with the ethnicity variable yields a two-by-two table with four different experimental conditions (see Table 1).

Table 1: Experimental conditions.

<table>
<thead>
<tr>
<th>White &amp; English proficiency</th>
<th>White &amp; English errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latina &amp; English proficiency</td>
<td>Latina &amp; English errors</td>
</tr>
</tbody>
</table>

Dependent variable: agency action on report
This project aims to determine the effects of fraud reporter characteristics on the likelihood that the report is taken seriously and acted upon by the fraud unit in a meaningful way. For each simulated report we submit, we will measure this response primarily as a simple 0/1 dummy variable: yes, the unit takes action, or no, it does not.

Fraud unit responses to reports are not immediately apparent. Public fraud reporters are not usually provided information about the action the fraud unit takes (or does not take) in following up on their reports. Therefore, we required a research design element to provide a signal of agency interest in the report.

To meet this requirement, we failed to include in each report key information that the fraud unit would need to move ahead with even the earliest stages of an investigation: no pretextual report included any name for the fabricated reported party, or any other potentially identifying information. A fraud unit response to the report with a request for the missing information indicates interest and constitutes a “1” on the “take action” outcome variable. Nonresponse, on the other hand, indicates non-interest, and constitutes a “0” on the variable. We will also measure the “quality of the answer” using two proxies: (a) the length of the response (longer is “higher quality” answer), and (b) time to respond (shorter is “higher quality” answer). We did not respond to any requests for additional information; when we received such a request, we logged and deleted it, and the course of that specific pretextual report was concluded.

Drawing on Author A’s prior interview-based study of fraud investigators and informal conversations with fraud unit representatives during the planning stage of this study, we expected variation in report screening approaches across fraud units and between individual report screening agents. Interviews with fraud investigators reveal that perceived unreliability among public reporters and the information they provide is quite common. Across the country, fraud workers describe public reports as generally poor sources of information, and people who make these reports as underinformed and often selfishly motivated. Receiving public fraud reports missing key facts is a regular occurrence, making our pretextual reports an insignificant and unremarkable addition to a standard pattern.

Despite significant skepticism about their overall reliability, public reports nevertheless constitute a sizeable portion of the information that fraud units use to originate investigations. Moreover, fraud unit representatives in multiple states indicated that their units use email replies to seek additional necessary information in response to reports otherwise deemed promising. These circumstances provide preliminary evidence of the research design’s viability.

Pilot study and final sample
Fraud control units in all fifty US states and the District of Columbia were eligible for inclusion in our sample, setting the maximum potential sample size at fifty-one. After preliminary review, we determined that thirty-six units were eligible for inclusion in our pilot study; the other fifteen units either had no online fraud reporting option, or (more commonly) used an online form with required fields that our research design could not accommodate.

The pilot study stage tested the sensitivity of our “take action” outcome variable. It comprised sixteen pretextual reports sent to each of the thirty-six units in our initial sample over a span of eight weeks. In this stage, we found that sixteen units of the initial sample demonstrated experimental variation; the other twenty units responded to either all or none of the sixteen pilot-stage pretextual reports. Based on this result, we limited our final sample for the full data collection to the sixteen units demonstrating variation. We sent the final sample units a total of 21 reports over a span of 4 months.

Hypotheses

When rules are not explicit and transparent, bureaucrats’ default behavior is discretionary and biased (Brodkin 1997, Katzenelson 2005). Discretionary benefit provision is common even if additional provision is low or zero cost (see for example, Davis et al. 2011). Such findings give reason to suspect that such factors may also affect how cases are perceived and treated within the fraud control system, and ultimately fraud control outcomes. Apparent race/ethnicity and English proficiency may have several effects on people’s standing vis-à-vis the state in this context.

H1: There will be no significant differences in response rates to reports from different parties. 
Mechanism: Non-discrimination rules and a well-established bureaucratic profession are sufficient to minimize discrimination in responses (Einstein and Glick 2017).

H2: Reports from Latinas will have lower response rates than those from whites, independent of English proficiency. 
Mechanism: Minorities tend to receive less information and have worse quality interactions with representatives of Community Service Offices (Ernst et al. 2013). Similar processes in this context could cause Latinas to be taken less seriously as reporters.

H3A: Reports from Latinas will have higher response rates than those from whites, independent of English proficiency.
Mechanism: This effect could result from well-intentioned bureaucratic responsiveness to minority populations. Or, this difference could reflect the disproportionate targeting of Latinas for enforcement efforts, if gatekeepers infer that Latinas are likely to report other Latinas.

H3B (heterogeneous effects): Reports from Latinas will have especially elevated response rates in states with greater residential racial segregation.
Mechanism: This effect could result from conscious or unconscious assumptions that Latinas are likely to report other Latinas. Racial/ethnic minority clients are more likely to be sanctioned in public assistance programs (Born et al. 1999, Koralek and Pindus 2000, Kalil et al. 2002, Hasenfeld et al. 2004, Schram et al. 2009); accordingly, enforcement bodies may be more likely to respond to reports they infer to implicate minorities. Greater residential racial segregation may increase accuser ethnicity’s strength as a signal of accused parties’ ethnicity.

H4: Reports from lower-English proficiency individuals will have lower response rates than higher-English proficiency individuals, independent of race/ethnicity.
Mechanism: Discretion and “triage” or “creaming” (Brodkin 1993, Halushka 2017) may cause gatekeepers to deprioritize reports perceived to originate with less reliable individuals.

H5: Reports from lower-English proficiency Latinas will have particularly low response rates.
Mechanism: For ostensibly Latina reporters, low written English proficiency is a proxy for national origin and/or language nativity. Assumptions about citizenship status and therefore legitimacy and credibility as a claimant on the agency may drive gatekeepers to disproportionately disregard low-proficiency Latina reporters.

H6A: Reports from lower-English proficiency Latinas will have particularly high response rates.
Mechanism: This effect could result from well-intentioned bureaucratic responsiveness to disadvantaged populations. Or, this difference could reflect the disproportionate targeting of lower-proficiency Latinas for enforcement efforts, if gatekeepers infer that people are likely to report others similar to themselves.

H6B (heterogeneous effects): Reports from lower-proficiency Latinas will have especially elevated response rates in states with greater residential racial segregation.
Mechanism: In states with greater residential racial segregation, gatekeepers may be more likely to infer that lower-proficiency Latina reporters are reporting people like them, particularly immigrants. Suspicions about immigrants and program eligibility may intersect with desires to punish and drive agency actions on reports from low-proficiency Latinas.
Data analysis

We will create (a) two dummies (white/non-white) & (proficient/non-proficient); (b) a continuous variable measuring before response; and (c) a continuous variable measuring response length. We will first conduct simple OLS & binomial models. We will also calculate the Average Treatment Effect (ATE) using a simple difference of means approach.

Specifically, if W is white, NW is non-white, P is proficient, and NP is non-proficient, we will measure the ATEs in terms of:

\[ \phi(W) = E[Y(W) - Y(NW)] \]
\[ \phi(P) = E[Y(P) - Y(NP)] \]

In addition to average effects across the entire sample, we also posit hypotheses on heterogeneous treatment effects, in which the effects of concerns about ethnicity/proficiency vary depending on the variables discussed in the previous section. As controls we will measure segregation using Licher et al (2016).

Even with theoretically-informed hypotheses, however, there are still well-known difficulties in specifying the correct functional form among possible interaction specifications. To help overcome these difficulties, we will use Kernal Regularized Least Square (KRLS) strategies to estimate Marginal Conditional Average Treatment Effects (MCATEs) (Hainmueller and Hazlett, 2013; Grimmer, Messing and Westwood, 2017). KRLS strategies build upon machine learning techniques to learn the target function from the data. Specifically, building upon our discussion in the theory, we will run KRLS models that include our treatments and variables to explore MCATEs.

Expected results

Statistical analysis with at least 90% confidence intervals showing probability and quality of response as main dependent variables.

IRB Registration

Ref. # 1701018671 approved September 26, 2017.

References


