Trust Radius In An Experimental Design: To What Extent Are Those "Most People" Out-Groups?

Wahideh Achbari

The proposed study is an experimental design aiming a) to examine the validity of the Generalized Trust (GT) question as a measure of outgroup attitudes and behavior in an unprecedentedly broad manner, and b) compare the results with 3 existing large-scale studies: WVS; ESS; and Project Implicit, which employs the Race Implicit Association Test (hereafter IAT) and many of the attitudinal constructs mentioned below as well as GT. WVS and ESS data contain only the explicit attitudes and GT.

Background and hypotheses

Generalized trust or trust in 'most people' is a conspicuous indicator in observational studies on intergroup relations. While prior research has debated the negative link between ethnic diversity indices and GT, very few have so far focused on measurement issues. A British study using think aloud protocols demonstrated that the majority of respondents high in GT think 'most people' refers to people they know, whereas a high proportion of those who are low in GT think about strangers. Clearly, this questions the ecological validity of previous research that assumes GT measures out-group trust. Following the homophily principle – the tendency of individuals to associate and bond with similar others – we can expect that people known to the respondents are likely to be in-groups, and strangers are more likely to be out-groups, although formally we do not know this. This study examines the construct validity of GT and its relation to out-group trust by comparing explicit self-reported attitudes with two unobtrusive measures of out-group bias: Implicit Association Test scores and social distance towards out-groups in an experimental design. If 'most people' refers to out-groups GT should also have an affective, more automatic, component that would relate well to implicit prejudice. In contrast, if an evaluation of 'most people' does not entail assessing the trustworthiness of out-groups, it should not relate well to the implicit measure and preferred social distance, and GT would be better conceptualized as confidence in one's social environment without implying any preference for an in-group over an out-group.

Based on previous findings – about the relative low correlation between implicit and explicit attitudes – I hypothesize three scenarios.

(1) High variance in predicting behavior by IAT scores in comparison to explicit measures, especially in a spontaneous context, would be evidence of social desirability since the sensitivity of the prejudice topic affects the likelihood of impression management.

(2) But if the relation between the IAT and explicit measures are relatively equal, this could be attributed to a dual mental process or dissociation. Each criterion would have a unique effect on predicted behavior. This would suggest that self-reports are valid indicators of prejudice.

(3) It may also be that implicit and explicit prejudice both explain greater variance in predicting behavior relative to GT, if 'most people' would not extend to out-groups.

Participants/sample

I will focus on the majority population (White Dutch) and their attitudes towards the Moroccan Dutch since they are highly stigmatized and this enhances the likelihood of finding stronger effect sizes with a small sample of Dutch students. In previous research GT has been found to be higher among people with a university degree compared to those without. Therefore, we will recruit participants outside the UvA premises in order to increase educational variation. We will approach passers-by at the Albert Heijn (Jodenbreestraat), Dappermarkt, and at the HvA. If this strategy does not yield enough participants, we will then approach adult MBO-students via schools. The student subsample is recruited via the psychology subject pool. Participation is rewarded with 7.5 euros. Outside sample will get 15 to compensate their travel time.

Cover story

Participants are told they are part of a study about cognitive abilities.

They are told that they will be:

- Filling out a questionnaire
- Completing a few tasks in which their reaction time will be measured. Through these tests they are asked to associate pictures to words. These images and words will appear in the middle of the computer screen. These tasks will be repeated several times.
- Reading a passage from someone's diary. Afterwards they are told they will meet this person, and together they will be asked questions about the text they have read by researchers conducting this study.

Overview of the protocol

The experiment will be conducted at the UvA psychology LAB. Research participants will get a unique (memoizable) ID when they turn up for the session: a combination of digits of week number, first two letters of weekday, followed by a three digit sequence (001, 002, etc.). This will enable us to merge the different parts of the study and track participants. Throughout the experiment, participants will carry a sticker with this info, and only when we ensure all data has been recorded, we will shred the sticker.

I. *Survey* (20 minutes max.) in *Qualtrics* with:
   1. Detailed demographics (2 minutes)
      a. Gender
      b. Year of birth
      c. Country of birth (cob)
      d. Mother's cob
      e. Father's cob
      f. Marital/partner status

g. Highest educational degree
h. No. people in household
i. Household income
j. Religiosity
k. Religious identity

2. Complete the following individual differences measures (83 items – 6 point Likert Scale – 8.5 minutes):
    a. Balanced Inventory of Desirable Responding – Impression Management (18 items)
    b. Big 5 – Agreeableness, Openness to experience (19 items) All items start with: I see myself as …..
    c. GT-question (6-point scale)
    d. Need for cognitive closure – close-mindedness (8 items)
    e. Right-Wing Authoritarianism (original scale 20 items) (or Social Dominance Orientation? 12 items – less old fashioned terms)
    f. Spheres of Control Battery – Personal efficacy (10 items) (or Optimism? 9 items – positive correlation with GTQ is shown in previous research, but these are not included in Race IAT.)

3. Then complete out-group attitudes and trust items (39 (or 59) items different scales – 5 minutes or 7 minutes)
    a. Out-group and in-group trust (Moroccan-Dutch, Turkish-Dutch, Surinamese-Dutch, Antillean-Dutch, Dutch without migration background – 5 items).
    b. Rosenberg faith in people (including GTQ 3 items, dichotomous)
    c. Trust items WVS (neighborhood, known personally, people met for first time, another religion, nationality -- 5 items, 6-point or original 5-points)
    d. Social Distance (preferred neighbors: other races, immigrants/foreign workers, other religions – 3 items, dichotomous)
    e. Attitudes towards Moroccan-Dutch, adapted from Attitudes towards Blacks (20 items, 6-points)
    f. Attitudes towards White-Dutch? (20 items)
    g. Feeling thermometer Moroccan-Dutch (2 items)
    h. Preference Moroccan-Dutch (1 item)

2 minute break. Participants don't leave the cubicle.

II. From Qualtrics participants are automatically forwarded to *Inquisit* to complete an adapted *Race IAT* (5.5 minute)

1. Participants are *randomly divided in 2 groups*. The purpose of the grouping is to mitigate order effects, and has no other conditional purpose. This will become clear below.

2. Participants are given *instructions* to complete the following task:

Participants are asked to categorize attributes (e.g. "joyful"; "tragic") and target items (e.g. pictures of Moroccan Dutch and White Dutch) into predetermined categories via keystroke presses. The basic task is to press a left key (E) if an item (e.g. "joyful") belongs to the category presented on the left (e.g. "Good") and to press the right key (I) if the word (e.g. "tragic") belongs to the category ("Bad") presented on the right.
For practice, participants sort items into the target categories "Moroccan Dutch vs. White Americans" and the attribute categories "Good vs. Bad".
For the test, participants are asked to sort categories into the paired/combined categories (e.g. "White Dutch OR Good" on the left vs. "Moroccan Dutch OR Bad" on the right). Pairings are reversed for a second test (e.g. "Moroccan Dutch OR Good" on the left vs. " White Dutch OR Bad" on the right). Order is counterbalanced by group number.

3. Block Sequence (7 blocks):
   a. Target Category sorting training
   b. Attribute sorting training
   c. 1. Test Block of hypothesis-consistent* pairings with 20 trials (half the participant start with inconsistent pairings)
   d. 2. Test Block of hypothesis-consistent pairings with 40 trials
   e. Target Category sorting training with targets switching sides
   f. 1. Test Block of hypothesis-inconsistent pairings with 20 trials
   g. 2. Test Block of hypothesis-inconsistent pairings with 40 trials

In all Test Blocks:
- attributes and targets alternate
- attributes as well as targets are randomly selected without replacement

List of attributes:
Good words are: "Joy", "Happy", "Laughter", "Love", "Glorious", "Pleasure", "Peace", and "Wonderful";
Bad words are: "Evil", "Agony", "Awful", "Nasty", "Terrible", "Horrible", "Failure", and "Hurt".

The strength of an association between words and pictures is measured by the standardized mean difference score of the 'hypothesis-inconsistent' pairings and 'hypothesis-consistent' pairings (d-score).
Positive d-scores: support a stronger association between 'White Dutch-Good' and 'Moroccan Dutch-Bad' than for the opposite pairings.
Negative d-scores: support a stronger association between Moroccan Dutch-Good' and 'White Dutch-Bad' than for the opposite pairings.

III. The final part of the experiment will record preferred social distance towards a Dutch-Moroccan out-group through a behavioral measure – *seating experiment*.

After the IAT participants will randomly be presented with a passage from a diary written by either a male or a female Dutch-Moroccan. The passage describes a typical day in his/her life

with ample references to their appearance and activities, so to clarify their ethnicity. Then they are told they are meeting the person. When entering the room, they do not see the person, but instead see a jacket and other personal items on the last chair consisting of a row of seven seats. The experimenter indicates that the out-group member has momentarily left the room, and that the items belong to him/her. The participant is requested to take a seat while waiting for his/her return. The further away they choose to sit while waiting, the more distance they prefer.

The session ends when the participants is told they haven't been selected to be included in further study (i.e. control condition). Debriefing will take place after the study is completed in order to ensure participants do not inform potential subjects.

A rough power analysis

The required sample size for testing three hypotheses for the two groups (Student and Outside) is n=128. We expect that the IAT, social distance, and attitudes correlate moderately. Hence the variance will be low. If we evaluate this with an ANOVA: Fixed effects, omnibus, one-way, with alpha = .05, power = .8, one tailed, and modest correlation between measures (r=.3), the required sample has to be a minimum of n=128 for moderate effects (f=.25). The power would increase to 0.95 with 210 participants. See Figure 1 below. If we are to incorporate an ANCOVA: Fixed effects, with main effects and interactions the required sample has to be n=269 for power = .80. See Figure 2. Below. I will aim for an n = 269 in order to be able to employ advanced tests.

Figure 1.

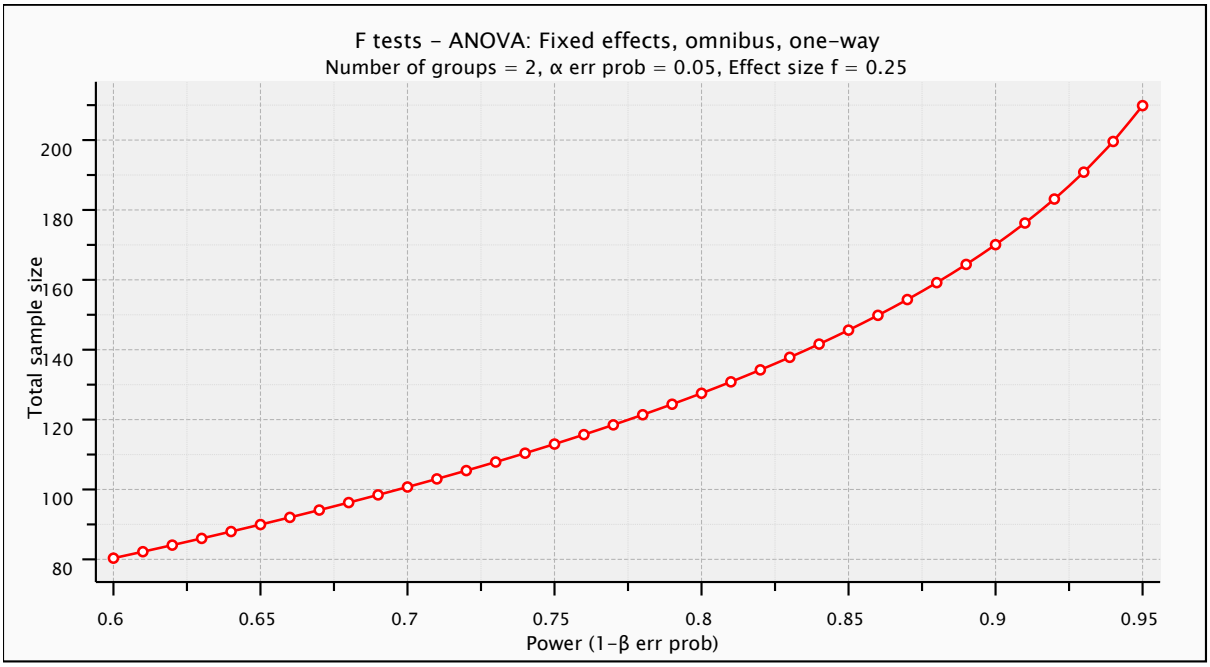
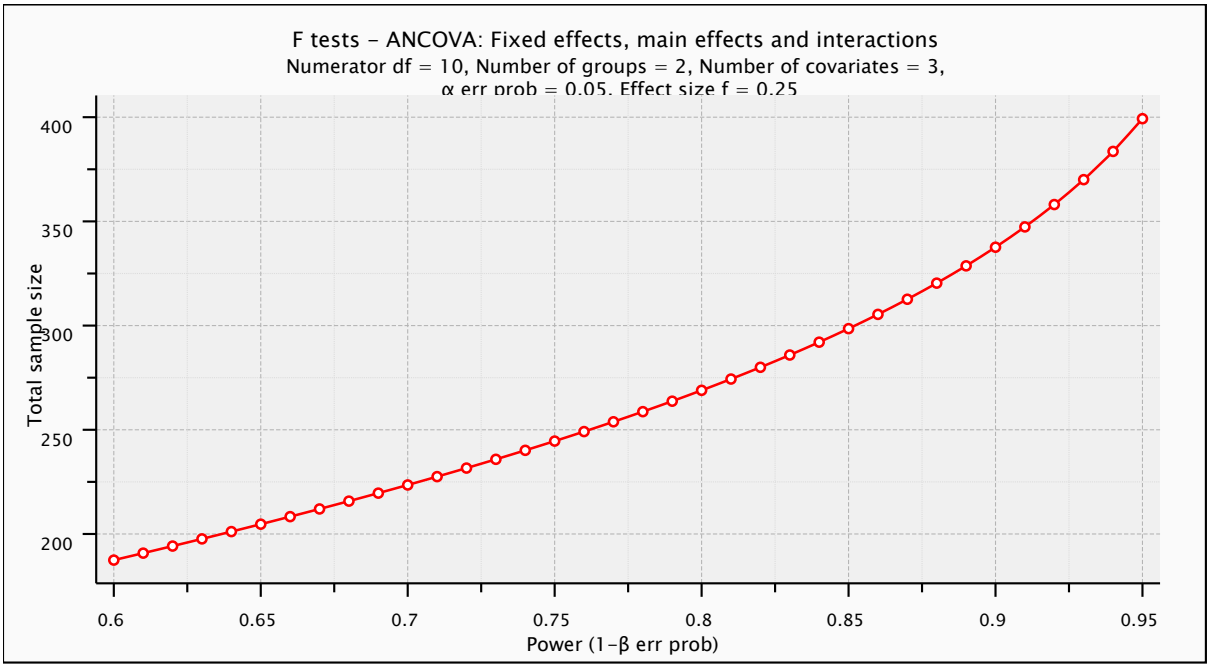


Figure 2.



Issues

In the survey part

- RWA or Social Dominance Orientations?
- Spheres of Control Battery (Personal efficacy) or Optimism?
- How many points on a Likert scale for GTQ?
  Ideally I want a 6-point scale throughout the survey part to be in line with many of the constructs used in the Race IAT studies (Big 5, RWA, etc.) and to increase power by not having a mid-point (I will elaborate this below). But large scale studies that include the GT question, and with which I will compare my results, use different scales for the GT question. ESS is using an 11-point scale, while WVS and experimental IAT studies are using a dichotomous one. The 11-point scale is fine with many survey participants since it resembles normality, but the more points on a Likert scale, the more participants I would need (increased variance and loss of power). A dichotomous measure, in contrast, would introduce issues when combining different items in statistical tests due to non-linearity.
  I have opted to include the GTQ twice in this study; once as a dichotomous one, in order to be in line with at least two of the major studies (WVS and Race IAT), but instead of having an 11-point scale, I would include a 7-point scale and deviate from the ESS. At the analysis stage I will compare standardized Betas/scores to compare the results with ESS. If I were to have an 11-point scale, I would need more participants and that would become too expensive.

- Forced choice or midpoint in the scale? 6 or 7-point? It is common in surveys and experiments that many respondents take the mid-point in order to avoid making a choice between the endpoints or valence. If I remove the midpoint, some of the responses from the Race IAT and other surveys won't be comparable, but I will ensure less power loss in my sample. Since those studies are generally large-N, enough participants remain in the comparison if the answers on the mid-point are discarded.
- Recruitment outside UvA will probably be challenging. Suggestions for alternative strategies are very welcome.