Seeing Racial Avoidance on City Streets

Bryce J. Dietrich* Melissa Sands†

This Draft: August 4, 2021

Abstract

Using publicly-available traffic camera feeds in combination with a real-world field experiment we examine how pedestrians of different races behave in the presence of racial out-group members. Across two different neighborhoods and 3,552 pedestrians we generate an unobtrusive, large-scale measure of inter-group racial avoidance by measuring the distance individuals maintain between themselves and other racial groups. We find that, on average, pedestrians give a wider berth to Black confederates, as compared to white non-Hispanic confederates.

*Department of Political Science, University of Iowa, 341 Schaeffer Hall, Iowa City, IA 52242 (bryce-dietrich@uiowa.edu, www.brycejdietrich.com).

†Department of Political Science, University of California, Merced, 5200 North Lake Road, Merced, CA 95343 (msands2@ucmerced.edu, melissasands.org).

Introduction

In the United States and elsewhere, race is a salient feature of everyday social interactions. Though places of residence, work, and study remain highly segregated along racial lines, with white Americans especially isolated from Black Americans, some degree of racial contact is particularly unavoidable in cities [48]. Whether they occur on the street, on public transit, or in other public spaces, these encounters need not involve conversation or verbal exchange to be significant. Social scientists have long argued that Americans behave differently in the presence of racial out-group members than they do in encounters with their in-group, and in a way that reflects a torrid history of institutionalized racism and segregation. Even as the majority of Americans now reject explicitly racist messages [31, 27], and as individuals' willingness to express racist attitudes on surveys has generally declined over time [41], white Americans' *behavior* belies persistent racial biases. This behavior need not be obviously discriminatory; racial bias may in fact be more apparent in subtle, nonverbal spontaneous behaviors than in more explicit behaviors [12, 13, 53].

One such behavior is racial avoidance. Researchers have observed nonverbal, physical manifestations of racial bias in the context of social conversations [51], seating distance [24], and shared space invasion [6]. Prior studies have, however, largely relied on laboratory or relatively small scale field studies that employ qualitative or coarse measurement techniques, and none have systematically studied pedestrian racial avoidance. This study contributes a new, large scale field experiment-based approach to studying racial bias in pedestrian interactions in highly naturalistic scenarios and settings. Using publicly-available, real-time video feeds from New York City Department of Transportation (NYC DOT) traffic cameras (`https://nyctmc.org//`), combined with confederates of different races, we address a central social science question: How do members of different racial groups behave toward each other in commonplace public encounters? In particular, we examine how members of different groups navigate one another on the sidewalks of a major city, one of the few places

where interracial encounters regularly occur.

Experimentally, we manipulate the presence of phenotypically Black and phenotypically white young adult males on sidewalks in Manhattan, and measure the behavior of pedestrians who pass by our confederates. Our measurement technique uses NYC DOT traffic cameras to passively record pedestrian movements vis-à-vis the confederates. Our findings support our hypothesis that, on average, pedestrians will give a wider berth to phenotypically Black males than they do to phenotypically white non-Hispanic males.

This study offers several major contributions. Substantively, we document that pedestrians in the United States actively avoid Black Americans on sidewalks which likely imposes a psychological toll on a population already heavily burdened with historical and institutionalized racism. Although we are not the first to suggest such a relationship, we do so using a large field experiment in a common and important naturalistic setting. Not only do our design and measurement strategy make our findings more generalizable than previous studies, but they also allow us to uncover subtle changes in walking trajectories, which underscores one of the many ways implicit biases are physically manifested in everyday behavior. We developed a number of new methodological tools which will help scholars study such behavior from afar without costly surveys. Video camera feeds are available in many cities throughout the country, and researchers may even set up their own cameras to systematically measure behavior. By combining a field experiment with publicly available "big data," we demonstrate the virtue of blending careful research design with widely-available high-frequency data.

Social scientists have long described social behavior in terms of physical space [5]. People utilize space in their environment in ways that reflect their attitudes towards others [22]. A large body of research – most prominently work on construal level theory [34] – describes a powerful link between social or psychological distance and physical distance [4, 37]. Individuals tend to describe friends and in-group members as "close" and strangers and out-group members as "distant" [26] – a phenomenon so pervasive that it has even been observed in

young children [33]. Physical distance is not only central to how individuals speak about others, it affects how people reason about and perceive physical distances between themselves and out-groups [18, 29, 35, 15]. The tendency to equate out-groups with "distant" manifests in nonverbal behaviors, most notably physical avoidance. For example, white Americans put more space between themselves and Black Americans in conversation than they do when speaking with fellow whites [51, 2]. Another form of avoidance is seating distance, which has been used by social psychologists as a measure of racial bias in laboratory settings [28, 20].

Interracial encounters have been a topic of academic interrogation since at least the middle of the twentieth century [30]. Neighborhood racial composition influences individuals' behavior [36, 47]. In the now vast body of literature that includes observational and experimental studies of racial contact and threat [3, 16, 14], less attention has been paid to one of the most basic and common types of encounters – those that regularly occur on sidewalks and street corners. These experiences are significant because they are quintessentially public and nearly universal.

Here we focus on pedestrians. Early work on spatial displacement on sidewalks and in public spaces emphasizes dominance behavior [9, 32, 25] and its evolutionary roots [49], personal space [17], and "gallantry" [21, 52]. Scholars have focused on gender differences as a determinant of both power and gallantry in pedestrian encounters [42, 9, 43, 52]. Other features that determine how pedestrians are treated include group size, occupational uniform, age, physical weakness or disability, attractiveness, and cultural differences [46]. Though researchers have described various interpersonal behaviors between Black and white individuals and used physical avoidance as both an indicator of prejudice and as a measure of discrimination, studies of interracial pedestrian interaction are rare. Nearly forty years ago, researchers examined responses to breaches in spatial etiquette by studying what happens when Black confederates overtly violate social norms by blocking pedestrians' way [6]. We consciously depart from this paradigm with a more subtle and naturalistic intervention that is designed to reflect an everyday scenario: confederates conform to social norms by standing

out of the way of pedestrian traffic, behaving as unimposing bystanders on public sidewalks.

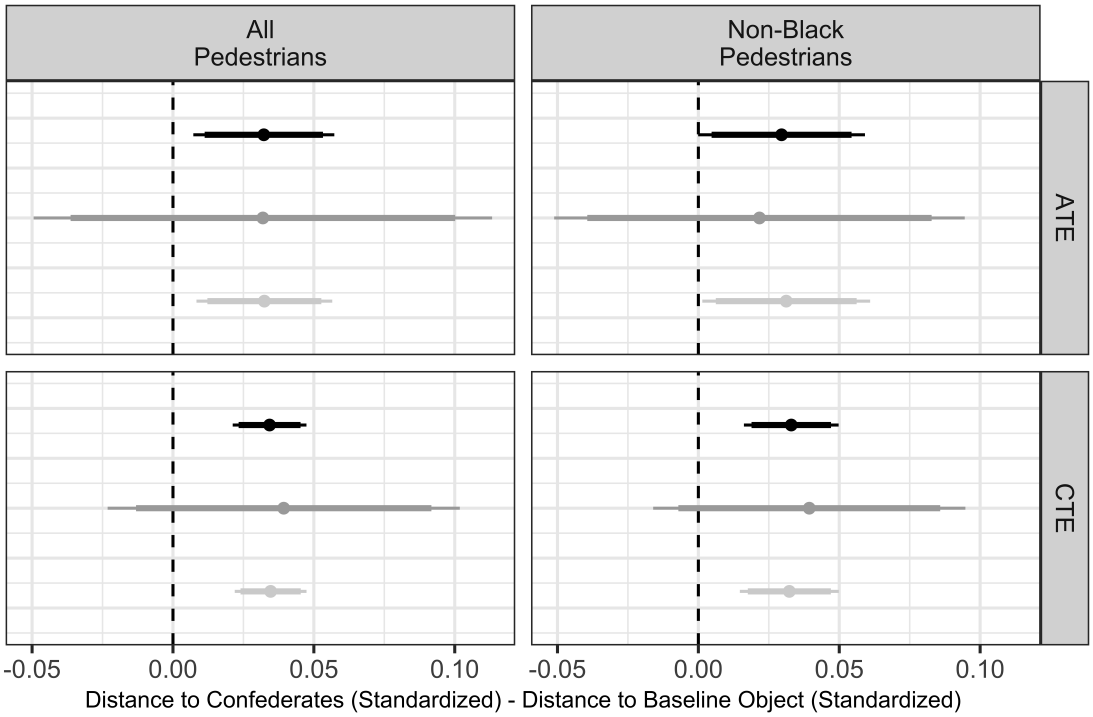
Our study leverages an unobtrusive measure of individual-level behavior – live public camera feeds from the city of New York – to understand interactions that are not typically systematically observed. Images from these cameras, which allow NYC DOT staff and the public to monitor live traffic conditions throughout NYC, were recorded for later processing. The research sites are situated in two different Manhattan neighborhoods: the more residential Upper East Side and more commercial Midtown East.

Three experimental conditions were implemented and recorded at each site. In the first condition, two young phenotypically white adult male confederates stood facing each other, in conversation, in a designated spot within camera view for 15 minutes. In the second condition, a pair of young phenotypically Black adult male confederates took their place – standing in the same spots, such that the distance between them was equidistant across race conditions – for the same period of time. In the third (baseline) condition, no confederates were present for the same period of time, recorded the same way as in the first two conditions.

Results

Figure 1 presents the average treatment effect (ATE) of the presence of Black confederates in both locations (black), on the Upper East Side (dark grey), and in Midtown (light grey), respectively. Our main outcome measure ("standardized pedestrian deviation", or SPD) reflects the deviation of a pedestrian from the confederate location as a proportion of the sidewalk width, as detailed in Appendix S3.3. The leftmost panel includes all pedestrians, while the rightmost includes only those identified as non-Black. In the top panels, each ATE reflects a simple difference-in-means, calculated by estimating a bivariate OLS regression of SPD on an indicator for whether the confederates present are phenotypically Black or not. The bottom panels report covariate-adjusted average treatment effects (CTE), with controls for pedestrian race and gender, 45-minute time block, and indicators for pedestrians traveling in groups or pairs on the Upper East Side and for pedestrian race and time

Figure 1: Pedestrians Give a Wider Berth to Black Confederates



Note: Treatment effects from OLS regressions of standardized pedestrian deviation on an indicator for whether the confederates present are Black (versus white). The top panels (ATE) reflect simple differences-in-means while the bottom panels (CTE) include controls for pedestrian characteristics and time block fixed effects. Positive values indicate deviation from Black confederates relative to white confederates as a proportion of total sidewalk width. Black (■) denotes both locations ($N_{all} = 3419; N_{non-Black} = 3208$), while dark grey (■) and light grey (□) correspond to the Upper East Side ($N_{all} = 516; N_{non-Black} = 448$) and Midtown ($N_{all} = 2903; N_{non-Black} = 2758$), respectively. For the Upper East Side, all variables other than the treatment indicator were coded by a graduate research assistant at a later date (see Appendix S5.1). There were also six pedestrians in the Upper East Side who were identified as outliers by the same RA. Although these are excluded in this figure, their inclusion does not change our substantive results (see Appendix S5.4). Thicker (▬) and thinner (—) lines represent 90 and 95-percent confidence intervals, bootstrapped to account for dependence within 15-minute clusters. All reported statistical tests are two-tailed.

block in Midtown. Thicker and thinner lines reflect 90% and 95% confidence intervals, respectively, bootstrapped to account for dependence within 15-minute clusters (the unit at which randomization occurred) using a wild block bootstrap [7]. Note that this approach is highly conservative: though clustered standard errors are appropriate when clusters of units, rather than individual units, are assigned to treatment [1], in our setup it is implausible that pedestrians receiving treatment towards the end of a 15-minute time block are dependent on those treated at the beginning.

Across both locations, estimated treatment effects are generally consistent with pedestrian avoidance of Black confederates relative to white. In the left-hand panel of Figure 1, pooling across the two neighborhoods, pedestrians deviate by, on average, 3.43% of the sidewalk width ($t_{3417} = 2.527, p = 0.012, \beta = 0.032, CI_{95\%} = [0.007, 0.057]$, without controls, $t_{3322} = 5.136, p = 0.000000296, \beta = 0.0334, CI_{95\%} = [0.021, 0.047]$ with controls), or around 4 inches, in the presence of Black confederates. On the Upper East side, estimated effects do not attain statistical significance at conventional levels; point estimates reflect that pedestrians move away between 3.19% and 3.93% of the sidewalk width ($t_{514} = 0.770, p = 0.441, \beta = 0.032, CI_{95\%} = [-0.049, 0.113]$ without controls, $t_{423} = 1.236, p = 0.217, \beta = 0.039, CI_{95\%} = [-0.023, 0.102]$ with controls), or between 3.1 and 3.8 inches, but are too noisy to reject the null hypothesis of no racial avoidance. In Midtown, pedestrians deviate between 3.24% and 3.47% of the sidewalk width ($t_{2901} = 2.640, p = 0.008337402, \beta = 0.032, CI_{95\%} = [0.008, 0.057]$ without controls, $t_{2895} = 5.338, p = 0.000000101, \beta = 0.035, CI_{95\%} = [0.018, 0.044]$ with controls), amounting to between 4.1 and 4.4 inches. When subsetting to non-Black pedestrians, who make up 93% of our sample, our results do not meaningfully change.

Discussion

Our findings are indicative of racial avoidance. Even in a highly-stimulating, densely populated, and politically left-leaning locale pedestrians systematically change their behavior in

observable ways in the presence of racial out-groups. The effects we measure may in fact represent a lower bound; if such behavior is detectable in NYC, pedestrian racial avoidance is likely even more pervasive in more racially-conservative environments. Moreover, we isolate this effect both with and without controls, suggesting our results are not dependent on specific model corrections. And, while our treatment effect estimates at the Upper East Side location do not reach statistical significance, the direction of the relationship is similar, providing some evidence that the effect we isolated persists across divergent neighborhoods.

These findings comport with our pre-registered hypotheses and are consistent with well-established theories of outgroup bias and threat, as well as evidence that young Black men in particular are stereotyped as threatening [8, 23, 19]. They are also consistent with evidence from across the social sciences that who and what we encounter as we move through space matters for a wide range of political and social outcomes. An experience as seemingly trivial as passing someone of another race or social class on a city sidewalk can have meaningful implications for decision-making [40].

Links between non-verbal measures of prejudice and more explicit behavioral indicators of racism are well-established. Physical avoidance in particular has been shown to track racial bias in the laboratory. For example, researchers have used immersive virtual environments to demonstrate that implicit prejudice against an ethnic out-group (as measured by an implicit attitude test, or IAT) predicts avoidance behavior [11], and that racial avoidance behavior (interpersonal distance and head orientation in a virtual encounter) predicts aggressiveness and hostility towards Black males in a subsequent virtual gunfight [39]. This body of work suggests that what is at stake is much more than a few inches of sidewalk space. One alternative interpretation of our findings is that pedestrians are "giving people space" rather than avoiding them. We reject this interpretation, and point to the literature on proxemics, which focuses on how people use space during social interactions, and demonstrates physical avoidance of stigmatized groups in a variety of contexts, linking patterns of avoidance to implicit measures of prejudice (see [38] for a review).

Our study remains subject to several limitations. As is the case in most field studies, research subjects do not represent a random sample of the population. Pedestrians on our selected Manhattan street corners are likely to be wealthier, better educated, and more politically Left-leaning than the average American. These traits may, on the whole, make subjects less prone to racial avoidance. Technical and feasibility constraints limit our ability to estimate pedestrian race and perfectly measure actual physical distances in the camera feeds. Future research might make use of high-definition video feeds now available in other locales, or researchers might employ their own cameras.

Our study also carries important implications. First, we document systematic racial avoidance in the real world, a finding that is highly consistent with the narratives voiced by people of color, for whom stereotype threat [44] and 'micro-aggressions' carry pernicious consequences [45]. Avoidance experienced by Black individuals, day in and day out, likely imposes a psychological toll on a population that already carries the extra burden of historical and institutionalized racism, eroding mental and physical health. Moreover, pedestrian racial avoidance can be both a cause and consequence of misattributions of threat. The very public, widespread, and chronic occurrence of pedestrian racial avoidance may have spillover effects that influence others' behavior in subtle but destructive ways [50]. For example, law enforcement – even if trained to recognize their own implicit biases – may implicitly or explicitly detect bystander behavior and overestimate the level of threat posed by Black Americans in ambiguous situations.

Second, we develop a method that allows researchers to study mass pedestrian behavior in the real world, and which can be used anywhere that cameras capture pedestrians in their feeds. Openly accessible video feeds are increasingly ubiquitous, yet under-utilized as a research tool [10]. By combining our measurement strategy with field experiments future scholars can replicate our study in different neighborhoods, and with different confederate characteristics, such as socioeconomic status, gender, ethnicity, and age.

Methods

Traffic cameras in the Upper East Side and Midtown neighborhoods were selected based on several factors: visibility of a large swath of unobstructed sidewalk, camera angle and image quality. These neighborhoods and the camera selection process are described in Appendix S2.

Confederates in each pair were dressed similarly to those in the other pair, and were similar in height, weight, and age, such that between-pair differences were minimized. Steps were taken to ensure that confederates behaved identically across conditions. They were instructed to talk about the same, non-political topic and monitored to make sure that they did not deviate from this protocol. As shown in Figure 2, confederates were positioned such that they created a slight bottleneck which funnels pedestrians into the treatment area without obstructing their movement.

The conditions were block randomized to control for natural fluctuations in pedestrian flow that occur throughout the day, such as lunchtime and commuting hours. That is, each day was divided into 45-minute blocks, and the order in which the three conditions were implemented within a block was randomized. The experiment took place over the course of two weekdays. On the first day, five 45-minute blocks were implemented in immediate succession at a pre-determined location in Midtown. On the second day, the same procedure was repeated – and the order re-randomized – at the Upper East Side location.

As the experiment occurred, a research assistant (RA) on site unobtrusively identified phenotypically Black pedestrians as they passed the fixed confederate location, a task that was independently validated for one site using the recorded video, as detailed in Appendix S5.1. Based on a power analysis informed by a pilot study conducted at a different intersection in Manhattan, we established and pre-registered a target sample size of $1,350$ pedestrians over 225 minutes of video. Ultimately we recorded 225 minutes at each location, capturing 776 pedestrians on the Upper East Side and $4,632$ in Midtown. Of the 776 pedestrians

identified on the Upper East Side (by condition: $N_{\text{Black}} = 253; N_{\text{white}} = 263; N_{\text{control}} = 260$), 48 (by condition: $N_{\text{Black}} = 9; N_{\text{white}} = 17; N_{\text{control}} = 22$) were said to be Black or African American by our RA on the day of the experiment. This same RA identified 230 phenotypically Black pedestrians in Midtown (by condition: $N_{\text{Black}} = 57; N_{\text{white}} = 88; N_{\text{control}} = 85$), representing 4.97% of the 4,632 pedestrians tracked at that location (by condition: $N_{\text{Black}} = 1447; N_{\text{white}} = 1456; N_{\text{control}} = 1729$). In Appendix S5.2 we show balance in the proportion of Black pedestrians across experimental treatments. In Appendix S5.1.2 we present analyses of inter-coder reliability with respect to the pedestrian race coding. All confederates and RAs were blind to the researcher hypotheses.

Recovering measurements from a photograph, or from any two-dimensional representation of a three-dimensional world, is difficult due to well-known features of perspective: objects appear smaller as distance from the observer increases, and perceived distance is distorted by the angle of vision. Thus, in a given frame of video, on the portion of sidewalk towards the top of the image (further away from the camera) a single pixel represents a greater distance than a pixel toward the bottom of an image (closer to the camera). As a pedestrian moves along the sidewalk, the relationship between pixels and actual distance on the ground changes as they progress from background to foreground. The nature of this dynamic relationship, furthermore, is dependent on the camera angle and zoom, which differs by location. Most relevant for our study, measuring the actual distance between individuals on a sidewalk is complicated by distortions due to the cameras' angled position, rather than it being located directly overhead. Although we cannot solve this problem, we take steps to show that it does not invalidate our experimental findings.

As preregistered, we consider the relationship between the pedestrian and a neutral, baseline object, located on the other edge of the sidewalk, directly adjacent and parallel to the confederates. For each frame of video, we subtract the distance (in pixels) from a given pedestrian to that object ("baseline distance") from the distance from that pedestrian to the confederate ("confederate distance"). This value is a rough proxy for how far a pedestrian is

from the confederate location relative to the baseline object when the three points appear on the same visual plane.

We provide evidence of the validity of this approach in Appendix S3.4. This validation exercise is based on the idea that a camera located directly above the confederates, providing a "bird's eye view", yields the equivalent of a two-dimensional representation of the scene; from this view, one could directly measure distances from the confederates to pedestrians without distortions due to camera angle or perspective. As described in more detail in Appendix S3, we use agent-based modeling and three-dimensional models of our experimental locations to demonstrate that the SPD yields the same results regardless of whether images reflect the original camera angle or a "bird's eye view" in an animated simulation. The average treatment effects calculated using these two camera perspectives are nearly identical, suggesting that our measurement approach is appropriate in the context of a randomized experiment.

Since distortions due to perspective are minimized when a pedestrian is on the same visual plane as the confederates, we use only the measurement for each pedestrian that is closest to the confederate location in our main model specifications. We preregistered our intent to use only the closest observation, and anticipated using pixel distance to identify the closest observation.. However, we discovered when implementing a validation exercise suggested by a reviewer that using pixel distance for this purpose does not always reveal the pedestrian observation that is closest in actual distance. Thus, we use "closest point" frames identified by RAs, as detailed in Appendix S3.5. The accuracy and reliability of these visual assessments is demonstrated in Appendix S5.1.3. We also show in Appendix S4 that our findings are not sensitive to the decision to use a single observation for each pedestrian; models utilizing all observations for each pedestrian yield the same substantive conclusions.

Pedestrians were manually tracked on the recorded camera images using the Fiji distribution of *ImageJ*. We consider only pedestrians who crossed between the confederates and the baseline object, and focus on those who are within close proximity to, and are not separated

by a physical barrier from, the confederates. The setup, measurement strategy, and manual tracking protocol are further detailed in Appendix S3.

Our main outcome measure (SPD) is explained in greater detail in Appendix S3.3 and reflects the deviation of a pedestrian from the confederate location as a proportion of the sidewalk width. We standardize by sidewalk width – 126 and 96.5 inches wide at the Midtown and Upper East Side locations, respectively – to make results from the two sites more comparable, as pedestrian behavior is in part dependent on the amount of room one has to maneuver. To obtain estimates of pedestrian avoidance in inches, we divide the pedestrian deviation in pixels by the pixel width of the sidewalk, and then use the actual sidewalk width at the plane at which the confederates are standing to convert that percentage to inches. Note that the main specifications involve comparisons between the Black and white conditions only; data from the pure control (no confederate) condition are not used. In Appendix S5.5.1 we discuss these results, which demonstrate how pedestrian behavior changes in presence of confederate of either race relative to a baseline object.

Protocol registration: Study design and hypotheses were registered with EGAP: http://egap.org/des registrations (ID # 20170616AA) on June 16, 2017. Our preregistration plan can be found at this URL: https://osf.io/vtqez (Accessed on August 11, 2020).

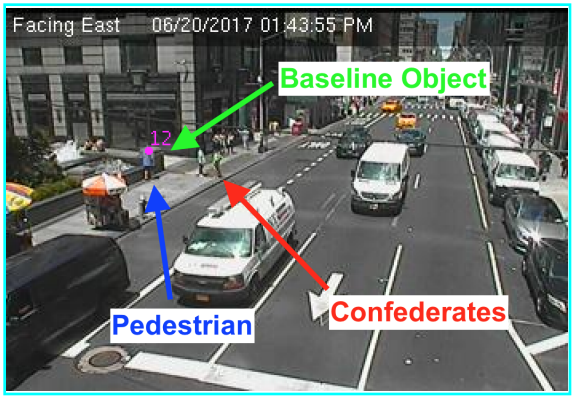
Data and Code Availability

Data availability: All data necessary to replicate the analyses and figures in this paper and supporting information are available at [DATAVERSE REPLICATION DOI HERE].

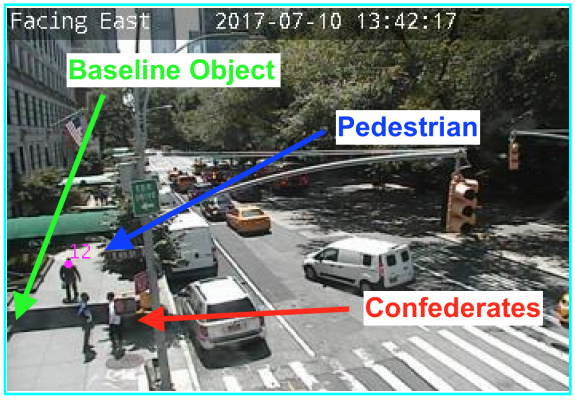
Code availability: All code necessary to replicate the analyses and figures in this paper and supporting information are available at [DATAVERSE REPLICATION DOI HERE]. R (open source, version 3.6.2) was used for data analysis.

Figure 2: Experiment Set-Up and Example Tracking Images

(a) Midtown Set-Up



(b) Upper East Side Set-Up



(c) Midtown Example

(d) Upper East Side Example

Note: Camera images from Midtown (panels A and C), and Upper East Side (panels B and D) locations. In the top row, confederates are referenced in red, pedestrians in blue, and the baseline object in green. Pedestrians in all images are numbered for the purpose of manual tracking.

This research complies with all relevant ethical regulations, and received approval from the University of Iowa Institutional Review Board (protocol # 201706768), with a reliance issued by the University of California, Merced Institutional Review Board. As explained in Appendix S1, waiver of informed consent was obtained.

References

[1] Alberto Abadie, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research, 2017.

[2] John R Aiello and Stanley E Jones. Field study of the proxemic behavior of young school children in three subcultural groups. *Journal of Personality and social psychology*, 19(3):351, 1971.

[3] Gordon Willard Allport. *The nature of prejudice*. Basic Books, 1979.

[4] Yoav Bar-Anan, Nira Liberman, Yaacov Trope, and Daniel Algom. Automatic processing of psychological distance: evidence from a stroop task. *Journal of Experimental Psychology: General*, 136(4):610, 2007.

[5] Emory Stephen Bogardus. A social distance scale. *Sociology & Social Research*, 1933.

[6] Clifford E Brown. Shared space invasion and race. *Personality and Social Psychology Bulletin*, 7(1):103–108, 1981.

[7] A Colin Cameron, Jonah B Gelbach, and Douglas L Miller. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427, 2008.

[8] Catherine A Cottrell and Steven L Neuberg. Different emotional reactions to different groups: a sociofunctional threat-based approach to 'prejudice'. *Journal of Personality and Social Psychology*, 88(5):770, 2005.

[9] James M Dabbs Jr and Neil A Stokes III. Beauty is power: The use of space on the sidewalk. *Sociometry*, pages 551–557, 1975.

[10] Bryce J Dietrich. If a picture is worth a thousand words, what is a video worth? In Robert X Browning, editor, *Exploring the C-SPAN Archives: Advancing the Research Agenda*. Purdue University Press, 2015.

[11] Ron Dotsch and Daniël HJ Wigboldus. Virtual prejudice. *Journal of experimental social psychology*, 44(4):1194–1198, 2008.

[12] John F Dovidio, Kerry Kawakami, and Samuel L Gaertner. Implicit and explicit prejudice and interracial interaction. *Journal of personality and social psychology*, 82(1):62, 2002.

[13] John F Dovidio, Kerry Kawakami, Craig Johnson, Brenda Johnson, and Adaiah Howard. On the nature of prejudice: Automatic and controlled processes. *Journal of experimental social psychology*, 33(5):510–540, 1997.

[14] Ryan D Enos. Causal effect of intergroup contact on exclusionary attitudes. *Proceedings of the National Academy of Sciences*, 111(10):3699–3704, 2014.

[15] Ryan D. Enos. *The Space Between Us: Social Geography and Politics*. Book Manuscript, Harvard University, 2016.

[16] Ryan D Enos. What the demolition of public housing teaches us about the impact of racial threat on political behavior. *American Journal of Political Science*, 60(1):123–142, 2016.

[17] Gary W Evans and Roger B Howard. Personal space. *Psychological bulletin*, 80(4):334, 1973.

[18] Anthony C Gatrell. *Distance and space: a geographical perspective*. Oxford University Press, USA, 1983.

[19] Phillip Atiba Goff, Matthew Christian Jackson, Brooke Allison Lewis Di Leone, Carmen Marie Culotta, and Natalie Ann DiTomasso. The essence of innocence: Consequences of dehumanizing black children. *Journal of Personality and Social Psychology*, 106(4):526, 2014.

[20] Phillip Atiba Goff, Claude M Steele, and Paul G Davies. The space between us: stereotype threat and distance in interracial contexts. *Journal of personality and social psychology*, 94(1):91, 2008.

[21] Erving Goffman. Relations in public: Microstudies of the social order. *London: Allen Lane*, 1971.

[22] Edward Twitchell Hall and T Hall. *The silent language*, volume 948. Anchor books, 1959.

[23] Melissa Harris-Lacewell. No place to rest: African american political attitudes and the myth of black women's strength. *Women & Politics*, 23(3):1–33, 2001.

[24] Michael Hendricks and Richard Bootzin. Race and sex as stimuli for negative affect and physical avoidance. *The Journal of Social Psychology*, 98(1):111–120, 1976.

[25] Nancy Henley. *Body politics: Power, sex, and nonverbal communication*. Prentice Hall, 1977.

[26] A Lee Hoxter and David Lester. Social distance evaluations in white and african-american students. *Perceptual and Motor Skills*, 80(2):478, 1995.

[27] Vincent L Hutchings, Nicholas A Valentino, Tasha S Philpot, and Ismail K White. The compassion strategy: Race and the gender gap in campaign 2000. *Public Opinion Quarterly*, 68(4):512–541, 2004.

[28] Kerry Kawakami, Curtis E Phills, Jennifer R Steele, and John F Dovidio. (close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of personality and social psychology*, 92(6):957, 2007.

[29] Dennis D Kerkman, David Stea, Karen Norris, and Jennifer L Rice. Social attitudes predict biases in geographic knowledge. *The Professional Geographer*, 56(2):258–269, 2004.

[30] V.O. Key, Jr. *Southern Politics in State and Nation*. Knopf, New York, 1949.

[31] Donald R Kinder and Lynn M Sanders. *Divided by color: Racial politics and democratic ideals*. University of Chicago Press, 1996.

[32] Eric S Knowles. Boundaries around group interaction: The effect of group size and member status on boundary permeability. *Journal of Personality and Social Psychology*, 26(3):327, 1973.

[33] Sandra Cohen Koslin, Marianne Amarel, and Nancy Ames. A distance measure of racial attitudes in primary grade children: An exploratory study. *Psychology in the Schools*, 6(4):382–385, 1969.

[34] Nira Liberman, Michael D Sagristano, and Yaacov Trope. The effect of temporal distance on level of mental construal. *Journal of experimental social psychology*, 38(6):523–534, 2002.

[35] Keith B Maddox, David N Rapp, Sebastien Brion, and Holly A Taylor. Social influences on spatial memory. *Memory & Cognition*, 36(3):479–494, 2008.

[36] Douglas S. Massey and Nancy A. Denton. *American Apartheid: Segregation and the Making of the Underclass*. Havard University Press, Cambridge, MA, 1993.

[37] Justin L Matthews and Teenie Matlock. Understanding the link between spatial distance and social distance. *Social Psychology*, 42(3):185, 2011.

[38] Cade McCall. Mapping social interactions: the science of proxemics. In *Social Behavior from Rodents to Humans*, pages 295–308. Springer, 2015.

[39] Cade McCall, Jim Blascovich, Ariana Young, and Susan Persky. Proxemic behaviors as predictors of aggression towards black (but not white) males in an immersive virtual environment. *Social Influence*, 2009.

[40] Melissa L Sands. Exposure to inequality affects support for redistribution. *Proceedings of the National Academy of Sciences*, 114(4):663–668, 2017.

[41] Howard Schuman, Charlotte Steeh, Lawrence Bobo, and Maria Krysan. *Racial attitudes in America: Trends and interpretations*. Harvard University Press, 1997.

[42] Jeanette Silveira. Thoughts on the politics of touch. *Women's Press*, 1(13):13, 1972.

[43] Robert S Sobel and Nancy Lillith. Determinants of nonstationary personal space invasion. *The Journal of Social Psychology*, 97(1):39–45, 1975.

[44] Claude M Steele. *Whistling Vivaldi: How stereotypes affect us and what we can do.* WW Norton & Company, 2011.

[45] Derald Wing Sue, Christina M Capodilupo, Gina C Torino, Jennifer M Bucceri, Aisha Holder, Kevin L Nadal, and Marta Esquilin. Racial microaggressions in everyday life: implications for clinical practice. *American psychologist*, 62(4):271, 2007.

[46] Stephen Thayer and Lewis Alban. A field experiment on the effect of political and cultural factors on the use of personal space. *The Journal of Social Psychology*, 88(2):267–272, 1972.

[47] Jessica Trounstine. Segregation and inequality in public goods. *American Journal of Political Science*, 60, 2015.

[48] Qi Wang, Nolan Edward Phillips, Mario L. Small, and Robert J. Sampson. Urban mobility and neighborhood isolation in america's 50 largest cities. *Proceedings of the National Academy of Sciences*, 115(30):7735–7740, 2018.

[49] Sherwood Larned Washburn and Irven DeVore. The social life of baboons. *Scientific American*, 204(6):62–71, 1961.

[50] Max Weisbuch, Kristin Pauker, and Nalini Ambady. The subtle transmission of race bias via televised nonverbal behavior. *Science*, 326(5960):1711–1714, 2009.

[51] Frank N Willis. Initial speaking distance as a function of the speakers' relationship. *Psychonomic Science*, 5(6):221–222, 1966.

[52] Frank N Willis, Joseph A Gier, and David E Smith. Stepping aside: Correlates of displacement in pedestrians. *Journal of communication*, 29(4):34–39, 1979.

[53] Carl O Word, Mark P Zanna, Joel Cooper, et al. The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of experimental social psychology*, 10(2):109–120, 1974.

Acknowledgments

We are grateful to Jason Anastasopoulos, Dan Carpenter, Maureen Craig, Daniel de Kadt, Ryan Enos, Justin Grimmer, Matt Hibbing, Hakeem Jefferson, Gary King, Alex Theodoridis,

Cara Wong, and audiences at the American Political Science Association annual meeting, the Midwest Political Science Association annual meeting, the New York University Center for Experimental Social Science conference, Stanford University, the University of California, Riverside, the University of California, Berkeley, and the University of Iowa for helpful feedback.

Author contributions

B.D. and M.S. contributed equally to the authorship of this manuscript.

Competing interests

The authors declare no competing interests.

Supporting Information (to go online) for: Seeing Racial Avoidance on City Streets

Contents

S1 Pre-registration and IRB

Our experimental design was pre-registered with EGAP (Evidence in Governance and Politics) and registered on June 16, 2017 (ID# 20170616AA). The full form is located in Section S6 and can be accessed at this URL: https://osf.io/vtqez (Accessed on August 11, 2020). Although we followed our pre-analysis plan, we did reserve some analyses for the Supplemental Information (SI) due to space limitations in the main text. Please see Section S5.5 found on page S37 in the SI for a detailed comparison of our pre-analysis plan to what is reported below and in the main text.

The University of Iowa Institutional Review Board (IRB) approved our proposed design on August 3, 2017 (IRB ID# 201706768). In doing so, we asked for informed consent to be waived. This was done for three reasons. First, we use publicly available camera feeds which record pedestrians every day without obtaining informed consent, meaning there is not the same expectation of privacy in this setting as opposed to in settings such as research laboratories. Second, the facial features of specific pedestrians are nearly impossible to discern from camera images. This in addition to the way the results are reported – in the aggregate – means there is limited risk for the pedestrians involved. Finally, obtaining consent would have required stopping pedestrians, which is invasive and likely infeasible.

Although the New York City Department of Transportation (NYCDOT) traffic cameras are fully public, we entered into a data-use agreement between NYCDOT and the University of California, Merced and the University of Iowa, dated October 19, 2020.

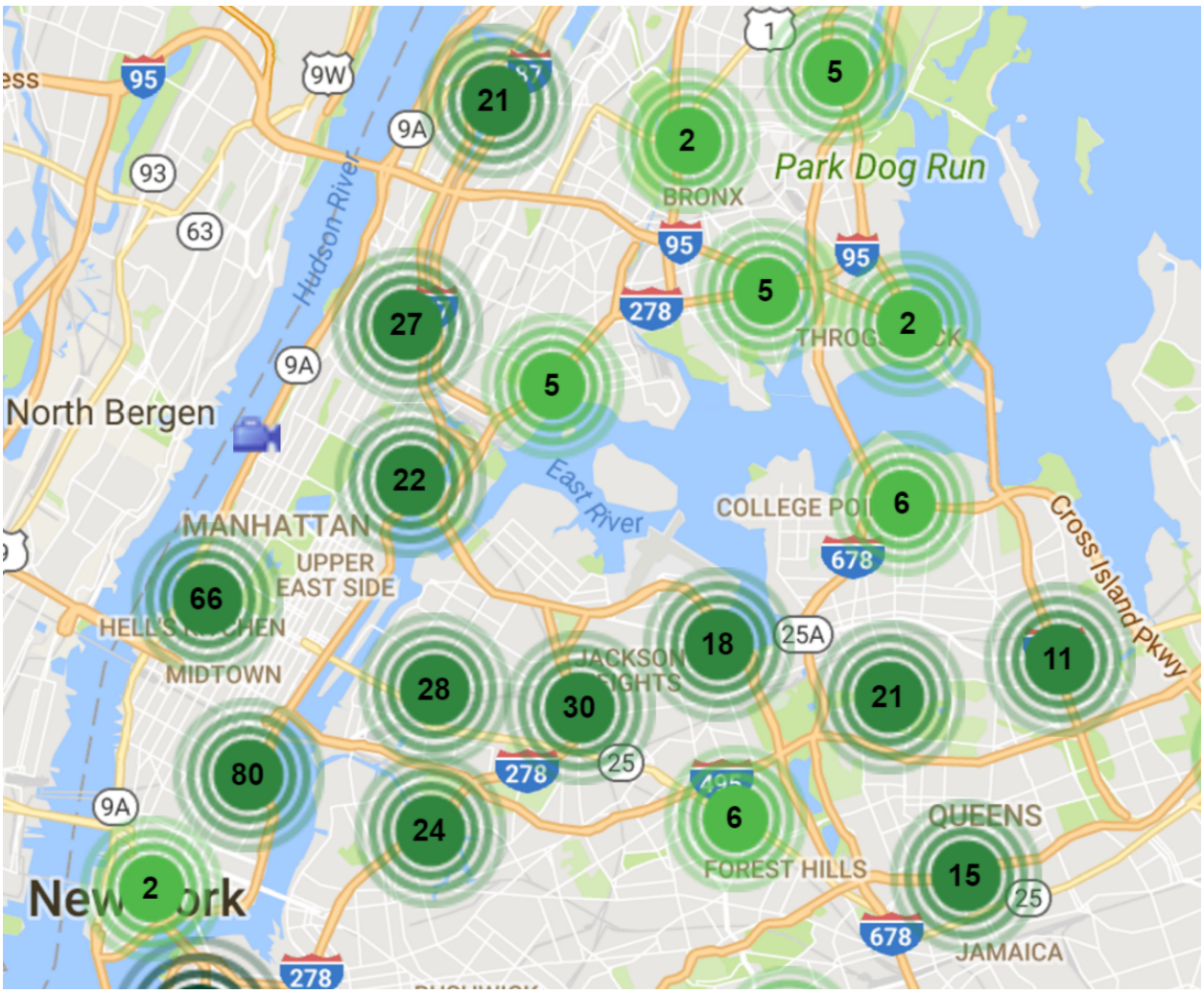
S2 Camera and Neighborhood Selection

At the onset of our study, the New York City Department of Transportation (NYCDOT) traffic camera system included 573 cameras (see Figure S1). These were distributed across all five boroughs, with the highest density of cameras in Manhattan (223) followed by Queens (175), Brooklyn (98), Staten Island (40), and the Bronx (37). Because these cameras are intended to capture vehicular rather than pedestrian traffic, most were unusable for our purposes, particularly outside of Manhattan. After eliminating cameras that focus on vehicular bridges or highways, we identified 123 cameras which partially overlooked city sidewalks.

We elected to implement the experiment in two distinct neighborhoods: the Upper East Side and Midtown East. Table S1 presents summary statistics from the Census tract in which each of the study cameras is situated, using data from the NYC Planning Population FactFinder,[S1] which reports demographic information from the 2012-2016 American Community Survey (ACS). Overall, the two sites differ significantly in terms of the demographics of their residents, but not median income or rent. In particular, Upper East Side residents are, on average, older and whiter. Most relevant for our study, almost all residents of the Upper East Side tract identify as white (98%), while none identify as Black or African American (0.00%). Meanwhile, residents of the Midtown tract are more mixed, with 77.2% identifying as white.

[S1] `https://popfactfinder.planning.nyc.gov/`

Figure S1: NYC Cameras



Note: Map of NYCDOT camera placement. The highest density of cameras is located in Midtown and lower Manhattan. Source: `https://webcams.nyctmc.org/`

There are a number of other differences between the two experiment sites. First, average weekday ridership for the subway station closest to our Midtown location is substantially higher than the comparable ridership for our Upper East Side location (69,332 passengers per weekday in Midtown versus 19,245 on the Upper East Side). Second, the number of businesses proximate to the camera locations varies notably across the two sites. At the time of the study, there were no commercial businesses within one block in each direction of the Upper East Side camera, compared to at least 15 businesses within one block of the Midtown camera. According to zoning information for the Community Districts in which

Table S1: Sidewalk Demographics (Census Tracts)

Variable	Midtown	Upper East Side
Percent Female	59.95	54.94***
	(57.12, 62.72)	(53.18, 56.70)
Percent White	77.20	97.95***
	(74.70, 79.52)	(97.38, 98.41)
Percent Black	2.65	0.00***
	(1.85, 3.77)	(0.00, 0.15)
Percent Latino	4.48	6.24***
	(3.41, 5.84)	(5.43, 7.16)
Percent Asian	18.24	2.05***
	(16.13, 20.56)	(1.59, 2.62)
Median Age	40.10	55.90***
	(33.31, 46.89)	(49.35, 62.45)
Median Income	$116,250	$151,346
	($85,030.60, $147,469.40)	($105,660.72, $197,031.28)
Median Rent	$2,778	$2,376
	($2,402.68, $3,153.32)	($1,980.43, $2,771.57)

Note: The two neighborhoods differ significantly in terms of their demographic composition. In total, 1,206 people resided in the Midtown census tract which encompasses 4 city blocks North-to-South and 2 city blocks East to West. The census tract in which the Upper East Side camera is located is home to 3,125 people, and encompasses 8 city block North-to-South and 2 city blocks East-to-West. Levels of significance are reported as follows: $^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$. 95 percent confidence intervals reported in parentheses.

the cameras are situated,[S2] 64 percent of the lots associated with the Midtown camera are used for commercial purposes whereas only 20 percent of the lots are residential. Conversely, in the Community District associated with our Upper East Side location 4 and 74 percent of the lots are used for commercial and residential purposes, respectively.

Cameras were selected in each neighborhood which had the best camera angle and image quality for tracking pedestrians. The former was determined using data collected from a small pilot study. In that study, the camera angle was similar to the angles of the cameras shown in Figure S2 in that they face east and the sidewalks tend to run in a similar trajectory. Outside of that, we also examined each camera feed for noticeable shadowing and glaring. The cameras we ultimately used in the study (see Figure S2) had neither problem which made them ideal for collecting our tracking data.

[S2] `communityprofiles.planning.nyc.gov`

Figure S2: Study Cameras

(a) Midtown (b) Upper East Side

Note: This figure shows examples of the cameras we used in this study. The colored numbers were generated by the researchers for tracking purposes.

S3 Measurement Strategy and Tracking Pedestrians

S3.1 Using the "Manual Tracking" Tool From *ImageJ*

Research Assistants (RAs) tasked with manually tracking pedestrians were trained as follows. In order to participate, they were required to watch three video tutorials and answer questions about the tracking task. They then were asked to track all the pedestrians in a five-minute sample clip obtained from the more difficult Midtown location. Seven RAs successfully completed the task on their first try. Three RAs missed between 3 and 5 pedestrians out of a possible 50. These individuals received additional training, and on their second attempt successfully tracked all of the pedestrians in the sample data. Finally, the tracking data was reviewed independently by a graduate RA to assure quality.

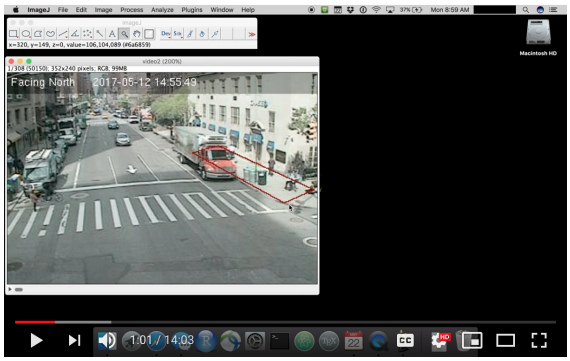
All the manually tracking was conducted using the "Manual Tracking" tool in the Fiji distribution of *ImageJ* (`https://imagej.net/Welcome`). The basic procedure is described in a *YouTube* video which was sent to each RA.[S3] Additional videos were distributed describing some basic post-processing steps[S4] and how to code the gender[S5] and race[S6] of pedestrians in the Upper East Side location. Finally, we also labeled pedestrian groups at this location which we also described on a private *YouTube* video.[S7]

Figure S3 outlines the main components of our training video which used frames from our pilot study. RAs were first presented video images over which the outline of a box had been superimposed to indicate the area of interest, and were asked to track pedestrians who enter that portion of the sidewalk (see Panel A). This box was drawn to indicate the area

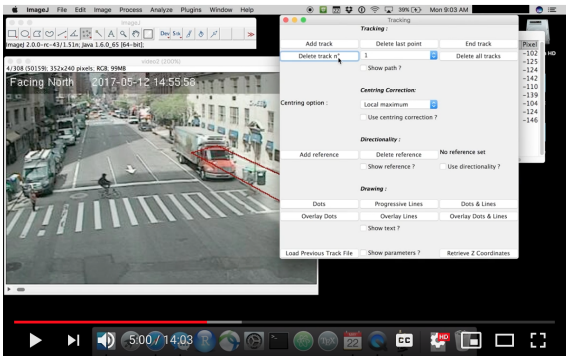
[S3] `https://www.youtube.com/watch?v=49nek32FbMc&feature=youtu.be`
[S4] `https://www.youtube.com/watch?v=6dCqjFC1Xns&feature=youtu.be`
[S5] `https://www.youtube.com/watch?v=pXiUv8EDtDw&feature=youtu.be`
[S6] `https://www.youtube.com/watch?v=T8G-7jQD3QU&feature=youtu.be`
[S7] `https://www.youtube.com/watch?v=LxzzS20pWW0&feature=youtu.be`

Figure S3: Training Video

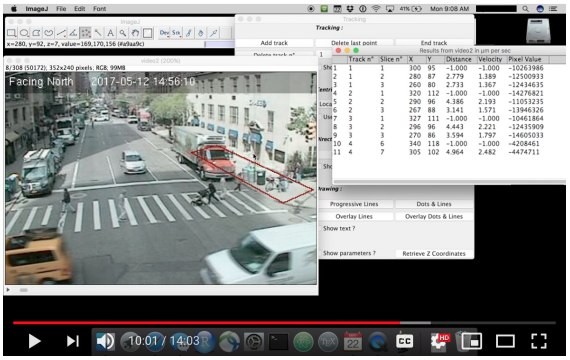
(a) 1 Minute



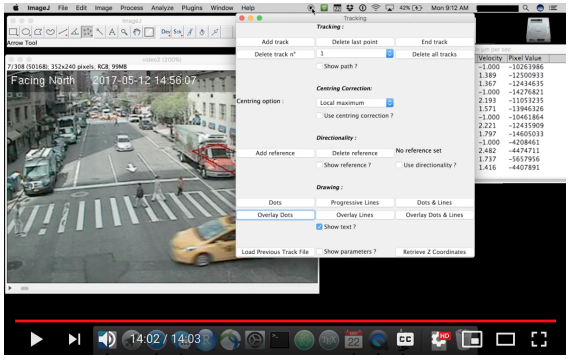
(b) 5 Minutes



(c) 10 Minutes



(d) 14 Minutes



Note: This figure shows frames from our training video and exercise described on page S5. The video was from our pilot data but looks similar to the videos we used for our paper.

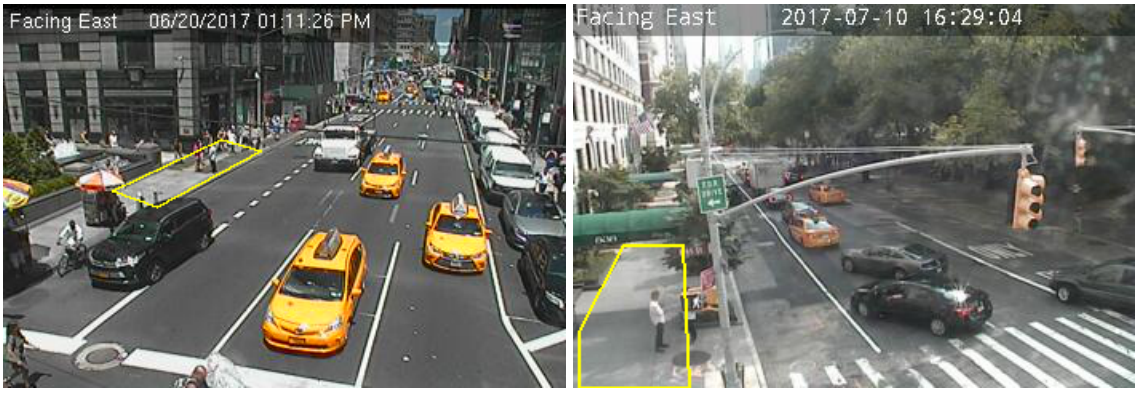
in which the confederates would have been clearly visible to approaching pedestrians, and was defined in part by 'natural barriers' in the physical environment. RAs tracked only pedestrians moving up and down the sidewalk (see Panel B), and excluded those whose trajectories do not cross the confederates' location. Results were saved to a CSV (see Panel C) and checked using the "overlay" function in the "Manual Tracking" tool (see Panel D).

To define the areas in which pedestrians are likely to be "treated" by the presence, or absence, of confederates, we defined measurement areas on the video images using "natural breaks" in the Midtown and Upper East Side sidewalks. Figure S4 outlines both areas in yellow. In Midtown, the length of the measurement area was defined by an entrance to a public courtyard in front of a condominium building, while at the Upper East Side site it was defined by the entrance to condominium building. The width of the measurement area in both locations extended from the edge of the entrance to the confederate location.

RAs were asked to imagine a line on the sidewalk from the confederates to the opposing wall or barrier, and to track only those pedestrians who crossed the imaginary line. They were instructed not to track children or bicycles. Consequently, our pedestrian data is comprised of adults walking up and down the sidewalk (see Figure S5, Panel A), which in practice constituted almost all of the foot traffic observed. Pedestrians who crossed the

Figure S4: Identifying "Treated" Pedestrians

(a) Midtown (b) Upper East Side



Note: Panels A and B represent the "Midtown" and "Upper East Side" locations, respectively. In both panels, the yellow box outlines the measurement area in which pedestrians were tracked.

Figure S5: Tracked Pedestrians

(a) Tracked Pedestrians (b) Did Not Track These Pedestrians

Note: This figure shows examples of some of the pedestrians we included in the study and those who were not. As described on page S7, we restricted our study to pedestrians moving up-and-down the sidewalk which made the data more comparable across the treatment and control groups.

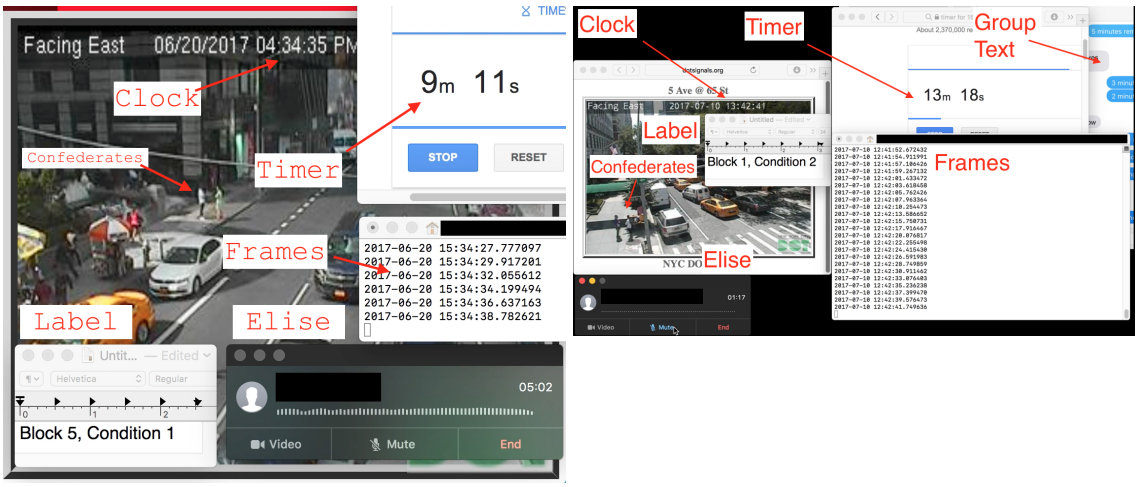
sidewalk horizontally (see Figure S5, Panel B) were excluded, as per the pre-analysis plan. Finally, our RA identified nine other pedestrians who appeared to walk in irregular zigzag patterns, six of which were in the Black and white confederate treatments. These individuals tended to linger in the frame for long periods of time, or appear and re-appear as they crossed the sidewalk repeatedly (e.g., as if making multiple deliveries). These pedestrians were flagged as outliers and excluded in our final models. In Figure S18 we show that the results are unchanged by our decision to exclude them.

S3.2 Identifying the Race of Pedestrians

During the experiment, RAs reported whether each pedestrian passing the confederates was, in their "best guess", Black or African American.[S8] They were also instructed to give additional details about those individuals (e.g., "wearing a red hat") where feasible, to assist with remote identification. This audio was recorded in conjunction with the video, so that the race description could be paired with the visual data.

Figure S6: Counting Phenotypically Black Pedestrians

(a) Midtown (b) Upper East Side



Note: This figure provides examples of the recording setups we used for our Midtown and Upper East Side Locations. Each label is described on pages S8–S9.

The recording setup is shown in Figure S6. *Quicktime* was first used to create a screen recording of the experiment. The audio for the screen recording was provided by a video call with one of the RAs on the ground (see "Elise"). The actual frames used for the manual tracking described in Section S3.1 were scraped from the camera URLs provided by the New York Department of Transportation at two-second intervals (see "Frames"). Since the frames had to be downloaded individually, one researcher watched the experiment remotely using a stable internet connection. Another researcher was at the camera site implementing the experimental protocol. Research team members communicated via a group text message (see "Group Text") and the timing for each block and condition was kept remotely using a Google Timer (see "Timer"). It was only feasible for RAs to track pedestrian race during 12 out of the total 15 blocks (each were 15-minutes in length) in our Midtown location. Race was recorded in all 15 blocks in our Upper East Side location.

[S8] Though we pre-registered our intent to identify more detailed information about pedestrian characteristics, this proved unfeasible during the implementation of the experiment. The density of pedestrians, combined with the difficulty of surreptitiously reporting this information without subjects overhearing, led us to this decision.

Once the experiment was complete another RA merged the narrated screen recordings with the manual tracking information, allowing us to create an indicator for whether a pedestrian was phenotypically Black. Ultimately, this imputation process was feasible for 3,113 pedestrians in the Black and White treatments, 176 (or 5.65%) of which were identified in real-time as phenotypically Black. This includes 516 pedestrians at the Upper East Side site, of which 31 (6.01%) were identified as Black. At the Midtown location, 5.58% (or 145) of 2,597 pedestrians were identified as Black, but as explained in the previous paragraph we were only able to code race for 12 out of 15 blocks at this location. To leverage all of the data acquired from our Midtown location, pedestrians who we were unable to identify in real-time are assume to be non-Black. Since imputing race on the ground of our Midtown location was significantly more challenging, we expect more error in the racial coding at that location.[S9]

Finally, we attempted to validate the racial information by employing another RA to independently code pedestrian characteristics based only on the recorded video. While we found imputation of pedestrian characteristics in the Midtown video to be impossible, due again to the density of pedestrian and the camera angle, this validation exercise was successfully completed for the Upper East Side recording. The RA coded her best guess at the race (African American / Black, white, Hispanic, Asian) and gender (man or woman) for each pedestrian, largely confirming the race statistics reported above. Additional analyses utilizing this coding are reported in Section S5.

S3.3 Measuring Three-Dimensional Effects Using Two-Dimensional Images

Recovering measurements from a photograph, or from any two-dimensional representation of a three-dimensional world, requires understanding the relationship between distance in a three-dimensional space and the pixels observed in a two-dimensional image. This is due to well-known features of perspective: objects appear smaller as distance from the observer increases, and perceived distance is distorted by the angle of vision (a phenomenon known as foreshortening). Thus, in a given frame of video, on the portion of sidewalk towards the top of the image (further away from the camera) a single pixel represents a greater distance than a pixel toward the bottom of an image (closer to the camera). As a pedestrian moves along the sidewalk, the relationship between pixels and actual distance on the ground changes as they progress from background to foreground. The nature of this dynamic relationship, furthermore, is dependent on the camera angle and zoom, which differs by location.

To help address this problem, we consider not just the relationship between the pedestrian and confederate, but also the relationship between the pedestrian and a neutral, baseline object, located on the other edge of the sidewalk, directly adjacent and parallel to the confederates. In each frame of video, we measure the distance (in pixels) from a given pedestrian to that object ("baseline distance"), and from that pedestrian to the confederate

[S9] Please note this also explains why these numbers are slightly different from those reported in Figure 1 in the main text. In the main text, 306 pedestrians from our Midtown location are assumed to be non-Black.

("confederate distance"). We use these two distances to calculate a measure we call "distance difference", which reflects how far a pedestrian is from the confederate location *relative to the baseline object*. As described in the Methods section, distance difference is equal to confederate distance minus baseline distance, such that more positive values reflect more avoidance (pedestrian is closer to the baseline object than to the confederate location), negative values reflect less avoidance (pedestrian is further from the baseline object than to the confederate location), and zero reflects equidistance.

This approach is intended to address some of the distortions created by the camera angle by focusing on the pedestrian's position in relation to two different fixed points which represent the extent of the available sidewalk when the pedestrian is passing the confederate. In our main analyses, we take the measurement for each pedestrian that is closest to the confederate as identified by human coders. See Section S3.5 for an explanation of how the closet point is identified.

Figure S7 shows two pedestrians, #1 (see red dot) and #2 (see blue dot). In the Panel A, Pedestrian #1 is around 22 pixels (or 37 inches) from the wall and around 30 pixels (or 50 inches) from the confederates. Thus the distance difference is 8 pixels (or 13 inches), meaning that Pedestrian #1 is a little over a foot further from the confederates as compared to the wall. In Panel B, Pedestrian #2 is around 12 pixels (or 20 inches) from the wall and around 38 pixels (or 64 inches) from the confederates. The distance difference is 26 pixels (or 44 inches), meaning that Pedestrian #2 is a little over three and half feet further away from the confederates as compared to the wall. The width of the sidewalk from the place where the confederates are standing to the opposite wall is 96 inches, which means that Pedestrian #2 is 32.29 percent ($\frac{44-13}{96} \times 100 = 32.29$) further away from the confederates as compared to Pedestrian #1.

These scalar distance estimates are derived from dividing the pedestrian deviation in pixels by the pixel width of the sidewalk, and then using the actual sidewalk width at the plane at which the confederates are standing to convert that percentage to inches.

Figure S7: Explaining Distance Difference Measure (Example)

(a) Pedestrian #1

(b) Pedestrian #2

Note: Panels A and B depict two different pedestrians at the Upper East Side location. Using the distance difference calculation, Pedestrian #2 is 32.29 percent further away from the confederates as compared to Pedestrian #1.

S3.4 Validating Distance Difference Measure Using 3D Simulations

As explained in the previous section, well-known features of perspective make it nearly impossible to precisely measure pedestrian distance without knowing certain features of the camera and/or objects that appear in the image. Problems associated with perspective are further complicated when the camera is at an unknown angle; if we knew the focal length of the camera we could use the height of the people to approximate the distances. Unfortunately, when using public camera feeds this information is unknown. Instead we rely on the assumption that the measurement error associated with confederate distance is roughly equivalent to the measurement error associated with the distance to a baseline object, which may not always be true given the refresh rates of the cameras. However, since the baseline object is constant across the experimental conditions, we are able to estimate an effect of the randomized treatment, if not precise measurements of the relevant distances.

While there are distortions in any projection of a three-dimensional space into a two-dimensional image, distance distortion are minimized when the camera provides a "bird's eye" view. If the cameras were located directly overhead, we could simply measure the distance between confederates and pedestrians. In fact, previous scholars have often relied on overhead shots when measuring pedestrian movement (for a review, see Haghani and Sarvi, 2018). Unlike other camera angles, these bird's eye views yield images akin to two-dimensional representations. In this section we describe a simulation exercise that relies on this logic to provide evidence in support of the validity of our measures. We thank an anonymous reviewer for this suggestion.

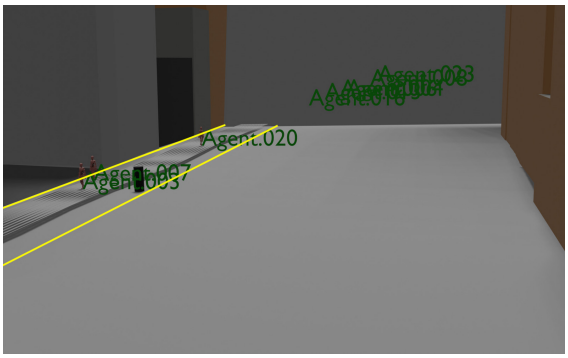
In this validation exercise, we calculated the distance difference using images from computer simulations, using both the original camera angle and a bird's eye view (placed immediately above the confederates). As shown in Figure S8, the same three-dimensional simulation was recorded using the original camera angles and, for comparison, "bird's eye" shots. The basic idea is that if our measure yields similar results for both camera angles, this suggests that the treatment effects outlined in the main text cannot be easily attributed to measurement error associated with the camera position vis-á-vis the pedestrians and confederates. This process is described in detail below.

We created two three-dimensional models of our experimental locations in Blender (`https://www.blender.org/`), an open source 3D graphics software. Pedestrians were then simulated using the CrowdSim3D (`https://crowdsim3d.com/`) plug-in. In total, 250 pedestrians were randomly assigned to the Black and white treatments at each of the two locations (Upper East Side and Midtown), for a total of 500 simulated pedestrians. To simulate the within-cluster dependence theoretically present in our original experiment, we created 10 batches of 25 pedestrians. Each batch was generated using a different random seed, resulting in ten unique distributions of 25 pedestrians each, allowing us to simulate experimental clusters. Each pair of batches was said to be in a "block", meaning the first two pedestrian batches were said to be in block 1 and the next two were said to be in block 2, etc. This ultimately yielded 5 blocks for each treatment, the same number as in our experiment.

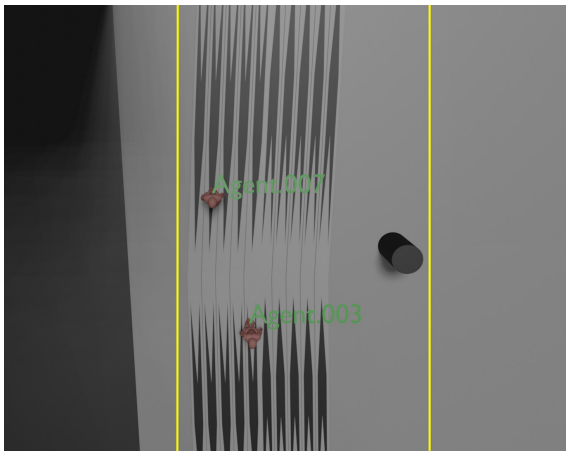
Pedestrian speed and acceleration were held constant at the program defaults of 1 meter (39.37 inches) per second and 1 meter per second squared, respectively. Each pedestrian's

Figure S8: Example Frames from 3D Simulations Used for Measurement Validation

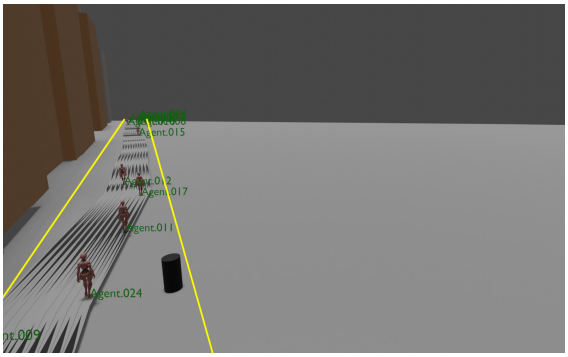
(a) Midtown Original View



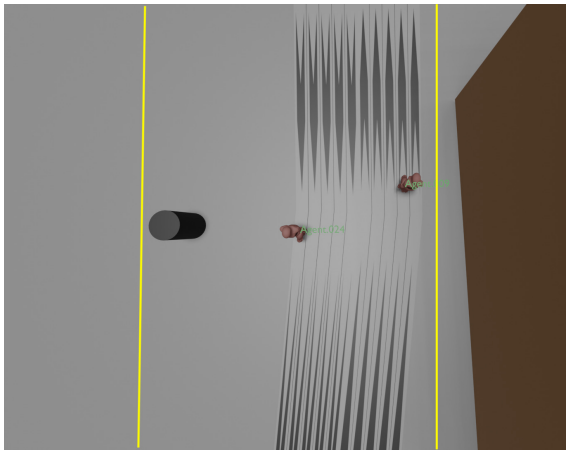
(b) Midtown Bird's Eye View



(c) Upper East Side Original View



(d) Upper East Side Bird's Eye View



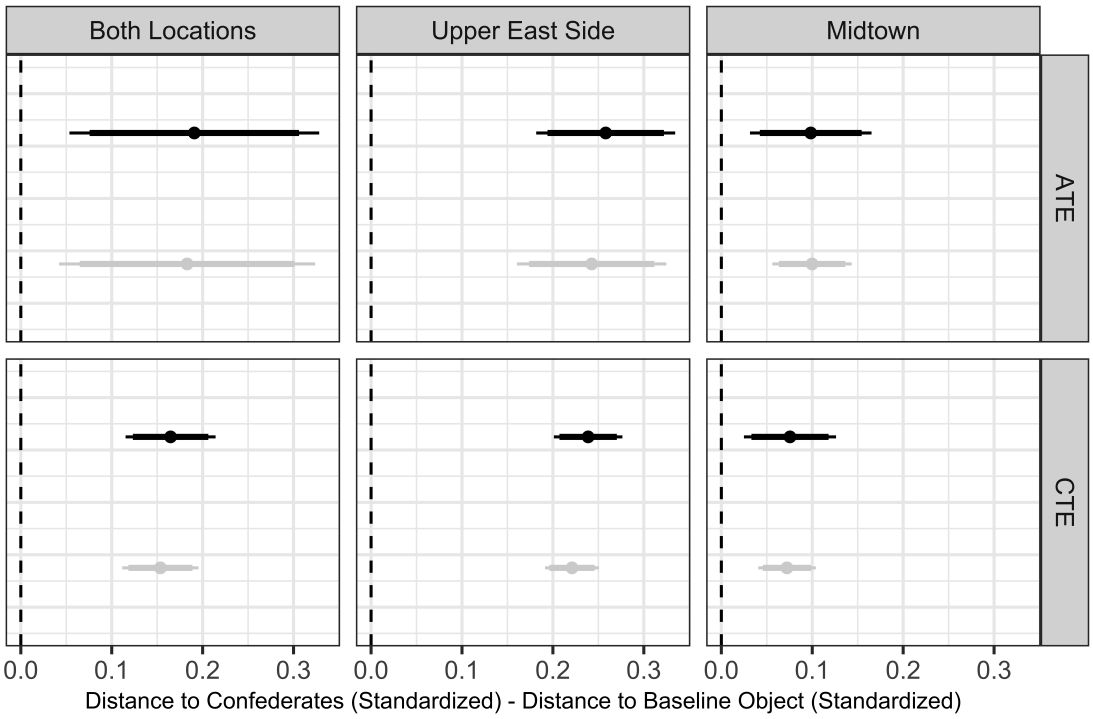
Note: Images from Midtown (panels A and B) and Upper East Side (panels C and D) simulations. In the first column, the original camera angles are used, whereas in the second column an overhead ("bird's eye") camera angle is used. Given that the latter eliminates distortions in distance due to perspective, we compare measurements taken using the two different perspectives to provide evidence in support of the validity of our experimental findings.

height was also held constant at 1.75 meters (69 inches) which was the default of the pedestrian model we obtained from Adobe Mixamo (`https://www.mixamo.com`). In order to increase the likelihood pedestrians were exposed to the treatments we only allowed pedestrians to be placed plus or minus three city blocks from where the confederates – represented as pegs in the simulation – were standing. The locations of these pegs mirrored the place where the confederates stood in our original experiment. We again used the program default for the simulation length, which was 250 frames. Ultimately, 135 pedestrians passed the confederates in the Upper East Side simulation. At our Midtown location, 163 simulated pedestrians passed the confederates. To simulate our Black and white treatments, we created walking trajectories for all pedestrians. These consisted of five "lanes" heading in opposite directions, yielding ten lanes total, representing the various paths that pedestrians might follow as they proceed along the sidewalk. Each pedestrian was randomly assigned to a lane; the simulated pedestrian walks in the same direction and lane for the duration of the simulation.

The simulated treatments were calibrated using two small pilot simulations in which we approximated the results originally reported in the main text. In the Upper East Side simulation, the distance between the confederate peg and the first lane was 1.623 meters (64 inches) and 1.301 meters (51 inches) under the Black and white treatments, respectively. Note, the simulated Upper East Side sidewalk is 8.63 meters (340 inches) wide, meaning the difference between the Black and White confederate pegs (0.322 meters or 13 inches) is 3.73% of the simulated sidewalk width. The Upper East Side estimate reported in the main text were 3.93% and 3.19% of the width of the sidewalk with and without controls, respectively. In the Midtown simulation, the Black treatment distance was 1.539 meters (61 inches) and the white treatment distance was 1.347 meters (53 inches), with the simulated sidewalk equaling 8.81 meters (347 inches) in width. For this location, the estimates reported in the main text were 3.47% and 3.24% with and without controls, respectively. The difference between the simulated treatments (0.192 meters or 8 inches) is 2.18% of the simulated sidewalk which is smaller than what we report in the main text, but is consistent with our theoretical expectations (see Section S5.5.3).

Figure S9 presents the simulated average treatment effect (ATE) of the presence of Black confederates using the original (black) and bird's eye (light grey) camera angles in both locations (first column), on the Upper East Side (second column), and in Midtown (third column), respectively. Our main outcome measure ("standardized pedestrian deviation", or SPD) reflects the deviation of a pedestrian from the confederate location as a proportion of the sidewalk width, as detailed in Appendix S3.3. Similar to the main text, simulated pedestrians were manually tracked using the Fiji distribution of *ImageJ*, as described in Section S3. In the top panels, each ATE reflects a simple difference-in-means, calculated by estimating a bivariate OLS regression of SPD on an indicator for whether the simulated Black confederates were present or not. The bottom panels report covariate-adjusted average treatment effects (CTE), with controls for the direction in which the pedestrian was walking and the two-batch blocks (see page S12 above). Thicker and thinner lines reflect 90% and 95% confidence intervals, respectively, bootstrapped to account for dependence within each pedestrian batch (see page S12 above) using a wild block bootstrap.

Figure S9: Pedestrians Give A Wider Berth to Black Confederates (Simulated)



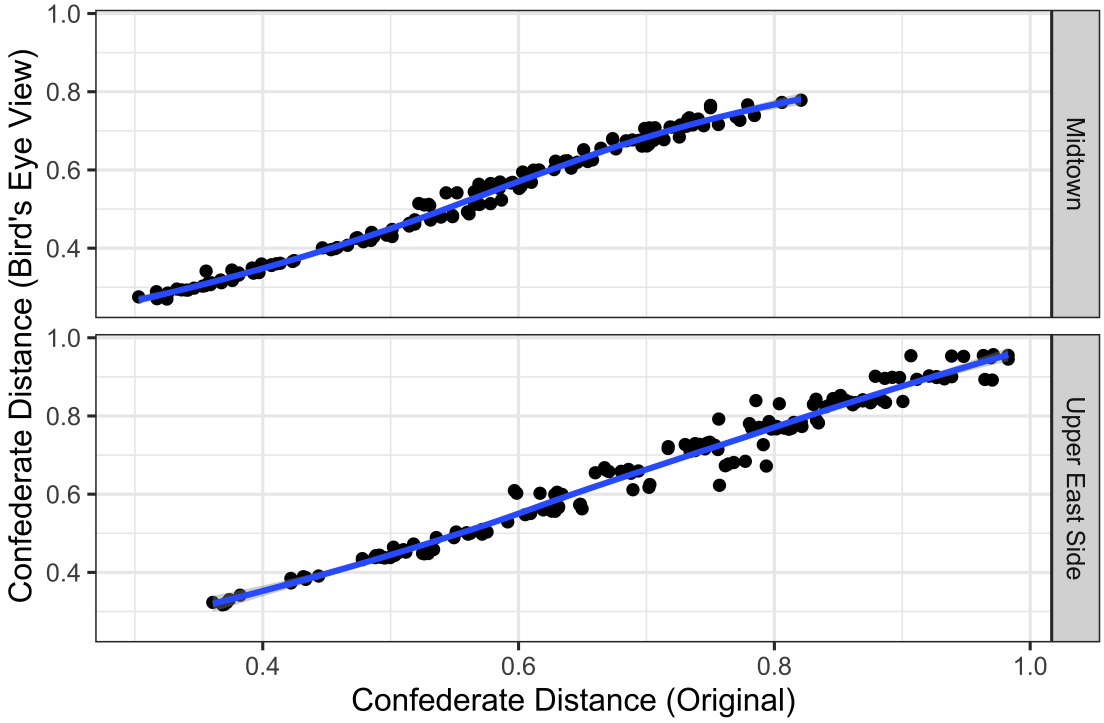
Note: Treatment effects from OLS regressions of standardized pedestrian deviation on an indicator for whether the confederates present are Black (versus white). The top panels (ATE) reflect simple differences-in-means while the bottom panels (CTE) include controls for pedestrian characteristics and time block fixed effects. Positive values indicate deviation from Black confederates relative to white confederates as a proportion of total sidewalk width. Black (■) denotes estimates were obtained using the original camera angles, while light grey (□) corresponds to the "bird's eye" view. The columns correspond to the specific simulation with the first column using all pedestrians from both simulations (N = 298), the second only using pedestrians from the Upper East Side simulation (N = 135), and the third only using pedestrians from the Midtown simulation (N = 163). Thicker (▬) and thinner (—) lines represent 90 and 95-percent confidence intervals, bootstrapped to account for dependence within simulated pedestrians clusters. All reported statistical tests are two-tailed.

Across both locations, the estimated treatment effects for the original and bird's eye camera views are essentially identical. In the left-hand panel of Figure S9, pooling across the two simulations, pedestrians deviate between 16.48% and 19.09% of the width of the simulated sidewalk ($t_{296} = 2.693, p = 0.007, \beta = 0.191, CI_{95\%} = [0.051, 0.33]$, without controls, $t_{286} = 6.747, p = 0.0000000000835, \beta = 0.165, CI_{95\%} = [0.117, 0.213]$ with controls) in the presences of Black confederates when the original camera angles are used. When the bird's eye view is used, pedestrians deviate between 15.35% and 18.29% of the width of the simulated sidewalk ($t_{296} = 2.542, p = 0.012, \beta = 0.183, CI_{95\%} = [0.041, 0.325]$, without controls, $t_{286} = 7.319, p = 0.00000000000254, \beta = 0.154, CI_{95\%} = [0.112, 0.195]$ with controls), suggesting our measure yields similar results regardless of whether camera angles correspond to a two-dimensional ("bird's eye") view or the three-dimensional view provided by public traffic cameras. In the Upper East Side simulation, we found similar results. More specifically, in this simulation pedestrians deviate between 23.88% and 25.81% of the sidewalk width ($t_{161} = 6.593, p = 0.000000000586, \beta = 0.258, CI_{95\%} = [0.181, 0.335]$ without controls, $t_{156} = 12.734, p = 0.000000000000000000000000000642, \beta = 0.239, CI_{95\%} = [0.202, 0.276]$ with controls) when the original camera angle is used. When the bird's eye view is used, pedestrians deviate between 22.10% and 24.26% of the width of the simulated sidewalk ($t_{161} = 5.584, p = 0.0000000979, \beta = 0.243, CI_{95\%} = [0.157, 0.328]$, without controls, $t_{156} = 14.404, p = 0.000000000000000000000000000000187, \beta = 0.221, CI_{95\%} = [0.191, 0.251]$ with controls). Finally, in our Midtown simulation nearly identical results were again found when the original camera and "bird's eye" views were used. Here, when using the former pedestrians deviate between 7.55% and 9.84% of the sidewalk width ($t_{133} = 2.962, p = 0.004, \beta = 0.098, CI_{95\%} = [0.033, 0.164]$ without controls, $t_{128} = 2.931, p = 0.004, \beta = 0.076, CI_{95\%} = [0.025, 0.127]$ with controls). Similarly, when the bird's eye view is used, pedestrians deviate between 7.22% and 9.98% of the width of the simulated sidewalk ($t_{133} = 4.64, p = 0.00000825, \beta = 0.1, CI_{95\%} = [0.057, 0.142]$, without controls, $t_{128} = 4.611, p = 0.00000957, \beta = 0.072, CI_{95\%} = [0.041, 0.103]$ with controls).

As an additional test, we calculated the linear correlation between the confederate distance at the closest point when the original and bird's eye camera angles are used. A simple scatter plot of the distances between the simulated agents and confederates can be found in Figure S10. In order to make the figure panels comparable, we divided by the sidewalk width for the Midtown and Upper East Side simulations. On the x-axis, we measure the confederate distance using the original camera angle, whereas on the y-axis the confederate distance is measured using the bird's eye view. The figure shows a nearly perfect linear correlation between the two measures. For the Midtown and Upper East Side simulations, the correlation between the confederate distance when the original and bird's eye camera angles are used are 0.991 ($t = 83.302, df = 133, p < 0.0000001$) and 0.987 ($t = 77.178, df = 161, p < 0.0000001$), respectively.

Collectively, these results show that our distance difference measure can capture the underlying treatment effect even when three-dimensional images are used. This is unsurprising given the nature of the measurement error discussed in Sections S3.3 and S3.4. Since we are using a randomized design, the estimated treatment effects should not be biased by con-

Figure S10: Correlation Between Confederate Distance When the Original and Bird's Eye Camera Angles are Used



Note: Simple scatter plots showing the correlation between the confederate distance at the closest point when the original (x-axis) and bird's eye (y-axis) camera angles are used. All distances are shown as proportions of the sidewalk width.

founding due to distortions of perspective, since these features are constant across conditions (for each camera).

That said, the camera angles in the experiment still produce measurement error in the dependent variable that makes it difficult to map pixel distances in the image to actual distances on the sidewalk. While we attempt to convert treatment effects to scalar distance using the limited on-the-ground measurements that we have available, we cannot be sure that these distances (estimated to be around 4 inches) are precise. Scalar distance estimates come from dividing the pedestrian deviation in pixels by the pixel width of the sidewalk, and then using the actual sidewalk width at the plane at which the confederates are standing to convert that percentage to inches. This approach is imperfect because it relies on a single on-the-ground measurement at each location. Fortunately, this error should be the same in each of the experimental conditions captured with the same camera; the estimated difference *between* the conditions should not be biased by this type of error. The goal of this study is not to estimate exactly how many inches of racial avoidance are demonstrated by the average pedestrian, but rather to document the phenomenon unobtrusively in a real world everyday setting.

S3.5 Identifying the Closest Point

In our pre-analysis plan (see Section S6), we indicated that we would identify the point at which pedestrians were closest to the confederates and use only that measurement for each pedestrian in our main analysis. Because of distortions due to the camera angle, the closest point as measured by pixel distance is not necessarily the closest point in real distance. Our simulation-based validation exercise demonstrates the extent of this distortion. In our Upper East Side simulation when the closest frame was identified as the frame in which the simulated pedestrians were closest in pixels to the confederates, the frame was, on average, nine frames from the actual frame where the simulated pedestrians were the closest. Since the default frame rate for Blender is 25 frames per second, this is the equivalent to an error of around 9 twenty-fifths of a second (0.36 seconds). When pixel distance was used to identify the frames where the pedestrians were closest to the confederates in our Midtown simulation, those frames were, on average, ten frames from the frames where the simulated pedestrians were actually closest. Given the frame rate, this is around two-fifths of a second (0.40 seconds).

To test whether human coders perform better at this task, we asked a graduate research assistant to view the simulations and identify the closest frame by watching the trajectory of each simulated pedestrian and determining the point where the pedestrian seemed closest to the confederates. They also identified any pedestrian who did not actually walk by our confederates. This process resulted in substantial error reduction. Here, the frames selected by our graduate research assistant were, on average, one and two frames (or between 0.04 and 0.08 seconds) away from the actual frames where the pedestrians were closest to the confederates in our Midtown and Upper East Side simulations, respectively.

To determine how this error impacts estimates of the treatment effect using pixel distances to identify the closest pedestrians, we randomly assigned a constant error to the actual closest

frame ranging from -5 to 5, meaning for a given pedestrian if the closest frame was 100, then an assigned error of -5 means the frame we used to estimate the treatment effect for that pedestrian was frame 95. Again, since the default frame rate for is 25 frames per second, this is the equivalent of varying the error by around a $\pm$ fifth of a second. When this was done for the Upper East Side, the ATE ranged between 19.97% and 27.86% and the CTE ranged between 20.04% and 25.70% of the width of the simulated sidewalk when the original camera angle was used. The ATE and CTE reported in Figure S9 for the Upper East Side simulation were 26.44% and 23.80%, respectively. When the same was done with the original camera angle for the Midtown simulation, the ATE ranged between 6.26% and 15.75% and the CTE ranged between 7.49% and 9.20% of the width of the simulated sidewalk. The ATE and CTE reported in Figure S9 for the Midtown simulation were 9.84% and 7.55%, respectively. These results suggest an error rate of 1 to 2 frames (or between 0.04 and 0.08 seconds) when identifying the closest frame likely does not significantly impact our results. However, we do note the frame rate for the cameras used in the actual experiment (two frames per second) is significantly lower, so these results should be applied with some degree of caution.

Given what we observed in the simulation, we asked the same research assistant to identify the closest frame at the Upper East Side location. Since the number of pedestrians is substantially higher at the Midtown location, we asked three additional research assistants to identify the closest frames at this location. Inter-coder reliability results are reported for this latter task in Section S5.1.3. Ultimately, we found high agreement across our coders, suggesting the closest frames identified at this location are reliable.

S4 Robustness to Different Measure Specifications

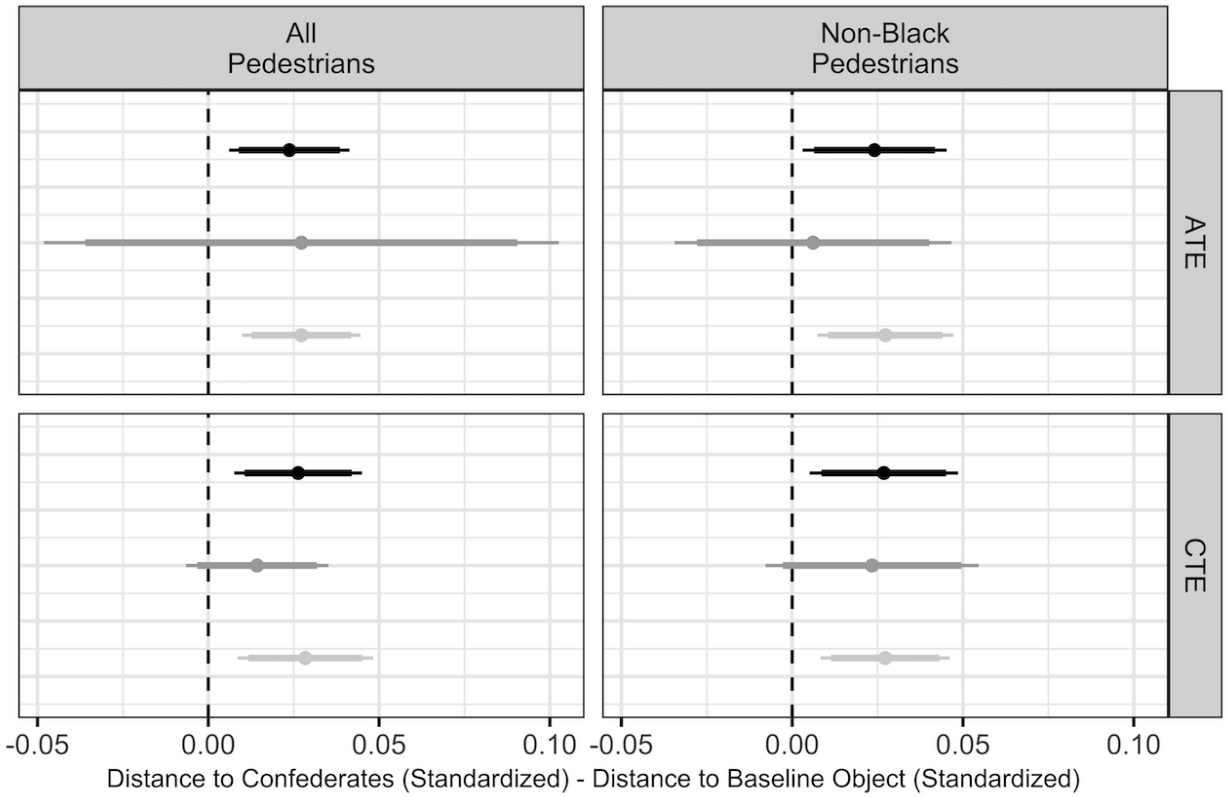
Our raw data structure is akin to an unbalanced panel with spatial dependence. As they proceed along the sidewalk, pedestrians are captured in multiple frames, and are thus observed more than once. Our main specifications collapse each pedestrian into a single observation – the individual's closest point to the confederate. Our reasoning for this decision is two-fold. First, the number of frames in which an individual appears depends on their walking speed and where they enter into the camera view, and thus varies from person to person. If speed is correlated with racial avoidance, then over-weighting pedestrians who move slowly and under-weighting those who move more quickly may bias our estimates of racial avoidance. Second, as discussed above, we use only the closest observation to the confederate to reduce the error generated by tracking pedestrians on different visual planes. Moreover, as pedestrians walk between our confederates and the baseline object, they tend to angle towards the midpoint between the confederates and the adjacent wall which means as pedestrians move towards the confederates the error structure changes. Thus, taking the closest observation ensures that pedestrians are located in approximately the same location on the sidewalk.

Below we show that our results are robust to alternative measurements and specifications. We explain different ways to combine the pedestrian tracks into different measures and show that the substantive results reported in the main text hold in each of these instances.

S4.1 Average Distance

Figure S11 replicates the analysis in the main text, replacing the closest pedestrian observation with the average of all pedestrian observations. That is, we subtract the average of distance (in pixels) from the baseline object ("baseline distance") from the average distance from the confederates ("confederate distance") for all available data points. As in the main analysis, we find that pedestrians move significantly further away from the Black confederates as compared to the white ($t_{3407} = 2.634, p = 0.008, \beta = 0.024, CI_{95\%} = [0.006, 0.041]$, without controls, $t_{3322} = 2.755, p = 0.006, \beta = 0.026, CI_{95\%} = [0.008, 0.045]$ with controls).

Figure S11: Pedestrians Give A Wider Berth to Black Confederates (Average Distance)

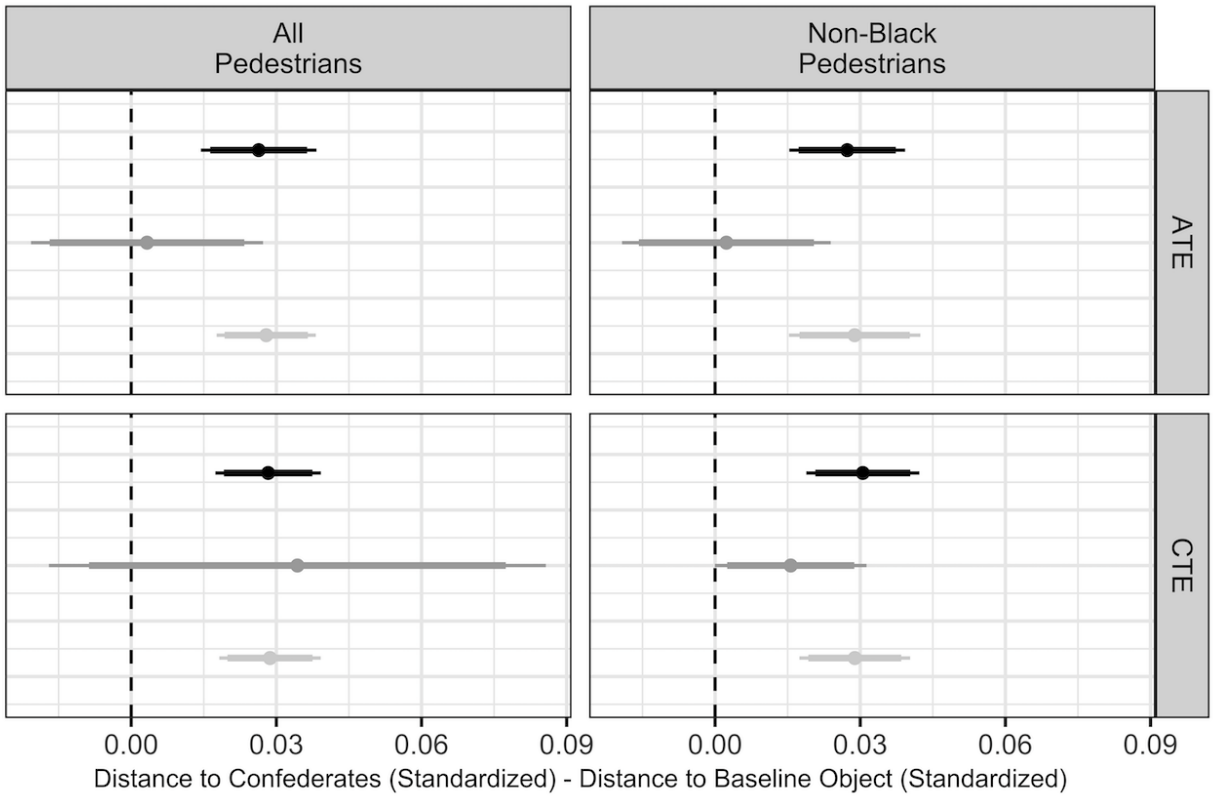


Note: Treatment effects from OLS regressions of standardized pedestrian deviation on an indicator for whether the confederates present are Black (versus white). Measures were created by averaging over all tracking data. Top panels (ATE) reflect simple differences-in-means while the bottom panels (CTE) include controls for pedestrian characteristics and time block fixed effects. Positive values indicate deviation from Black confederates relative to white confederates as a proportion of total sidewalk width. Black (■) denotes both locations ($N_{all} = 3419; N_{non-Black} = 3208$), while dark grey (■) and light grey (□) correspond to the Upper East Side ($N_{all} = 516; N_{non-Black} = 448$) and Midtown ($N_{all} = 2903; N_{non-Black} = 2758$), respectively. Thicker (▬) and thinner (—) lines represent 90 and 95-percent confidence intervals, bootstrapped to account for dependence within 15-minute clusters.

S4.2 Weighted Average Distance

In Figure S12 we replicate this analysis, weighting the averages based on the proximity of the pedestrian to the confederates. As pre-registered, weights equal to the inverse of the confederate distance are applied. Again, our results do not noticeably change; we find that pedestrians move significantly further away from the Black confederates as compared to the white ($t_{3407} = 4.324, p = 0.00001579, \beta = 0.026, CI_{95\%} = [0.014, 0.038]$, without controls, $t_{3322} = 5.084, p = 0.00000039, \beta = 0.028, CI_{95\%} = [0.017, 0.039]$ with controls) and that these results tend to be more pronounced when the data is restricted to non-Black pedestrians.

Figure S12: Pedestrians Give A Wider Berth to Black Confederates (Weighted Average Distance)



Note: Treatment effects from OLS regressions of standardized pedestrian deviation on an indicator for whether the confederates present are Black (versus white). Measures were created by averaging over all tracking data giving more weight to points closer to the confederates. Top panels (ATE) reflect simple differences-in-means while the bottom panels (CTE) include controls for pedestrian characteristics and time block fixed effects. Positive values indicate deviation from Black confederates relative to white confederates as a proportion of total sidewalk width. Black (■) denotes both locations ($N_{all} = 3419; N_{non-Black} = 3208$), while dark grey (■) and light grey (□) correspond to the Upper East Side ($N_{all} = 516; N_{non-Black} = 448$) and Midtown ($N_{all} = 2903; N_{non-Black} = 2758$), respectively. Thicker (▬) and thinner (—) lines represent 90 and 95-percent confidence intervals, bootstrapped to account for dependence within 15-minute clusters.

S4.3 Line Distance

We next take a step back and imagine what the ideal scenario would be given the difficulty of measuring three-dimensional space with a two-dimensional image. As shown in Figure S13, the ideal scenario, from a measurement perspective, would be to have our confederates standing on the same visual plane as every data point we collected from each pedestrian (see Panel A). An approximation of this ideal scenario can be found in Panel B. Here, we draw a line that divides the sidewalk in half in the same way as the vanishing confederate pairs do in Panel A. This line can then be used to determine whether pedestrians are positioning themselves further away from the confederates under both the Black and white conditions. Looking back to Figure S11, in many ways this is precisely what the distance difference measure captures since a zero implies the pedestrian is just as close to the wall as they are to the confederates. The line found in Panel B achieves the same end, but does so by projecting the confederates position up the sidewalk instead of using the baseline object.

Figure S13: Comparing Pedestrians on the Same Visual Plane (Second Approach)

(a) Same Visual Plane (Hypothetical)

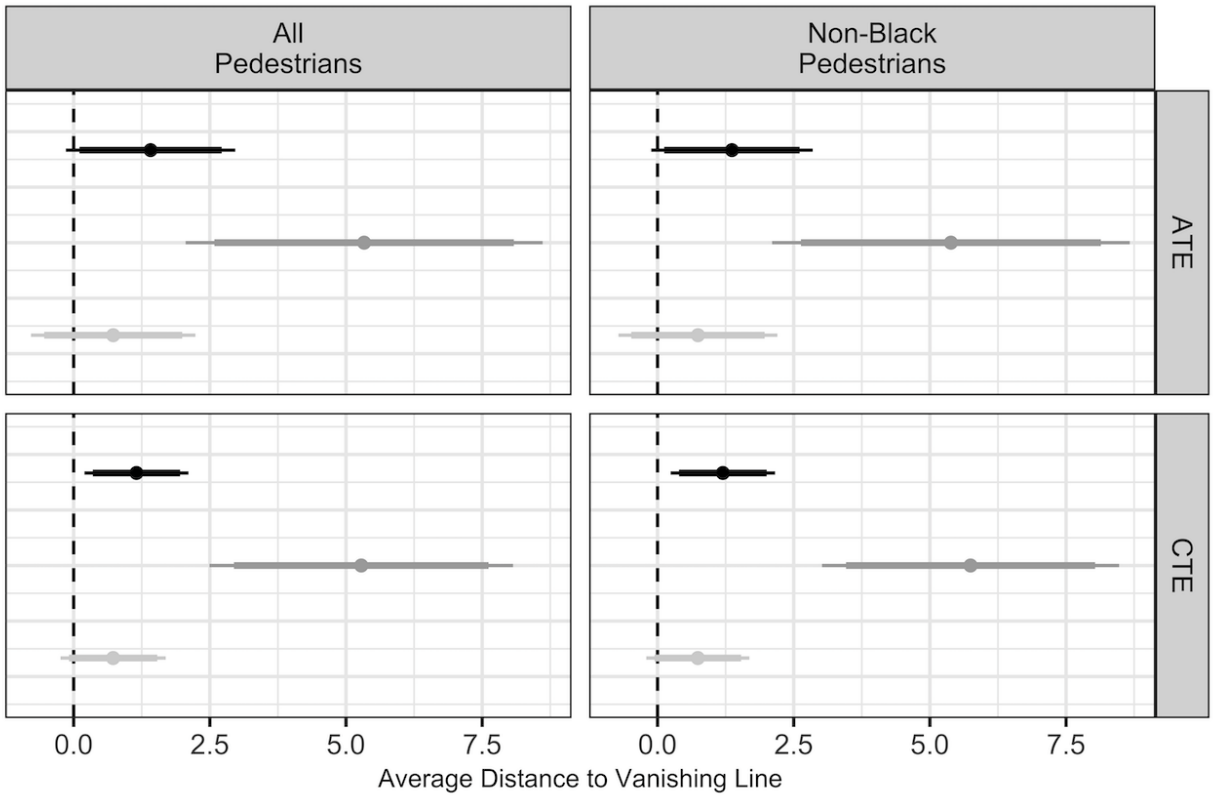
(b) Same Visual Plane (Line)

Note: Panels A reflects the ideal scenario of confederates talking with one another at each visual plane. Panel B replaces these pairs with a vanishing line. In the results below, we calculate the distance of each pedestrian (in pixels) from this line to estimate the treatment effect.

S4.3.1 Average Line Distance

Figure S14 replicates the analysis in the main text, replacing the closest pedestrian observation with the average distance of all pedestrian observations from the vanishing line explained in Figure S13. That is, we measure the distance (in pixels) from the vanishing line to each pedestrian and then – in Figure S14 – take the average. As in the main analysis, we find that pedestrians move further away from the Black confederates as compared to the white, but this effect is only statistically significant at the 0.05-level when controls are included ($t_{3416} = 1.780, p = 0.075, \beta = 1.412, CI_{95\%} = [-0.144, 2.968]$, without controls, $t_{3322} = 2.366, p = 0.018, \beta = 1.153, CI_{95\%} = [0.197, 2.109]$ with controls).

Figure S14: Pedestrians Give A Wider Berth to Black Confederates (Average Line Distance)

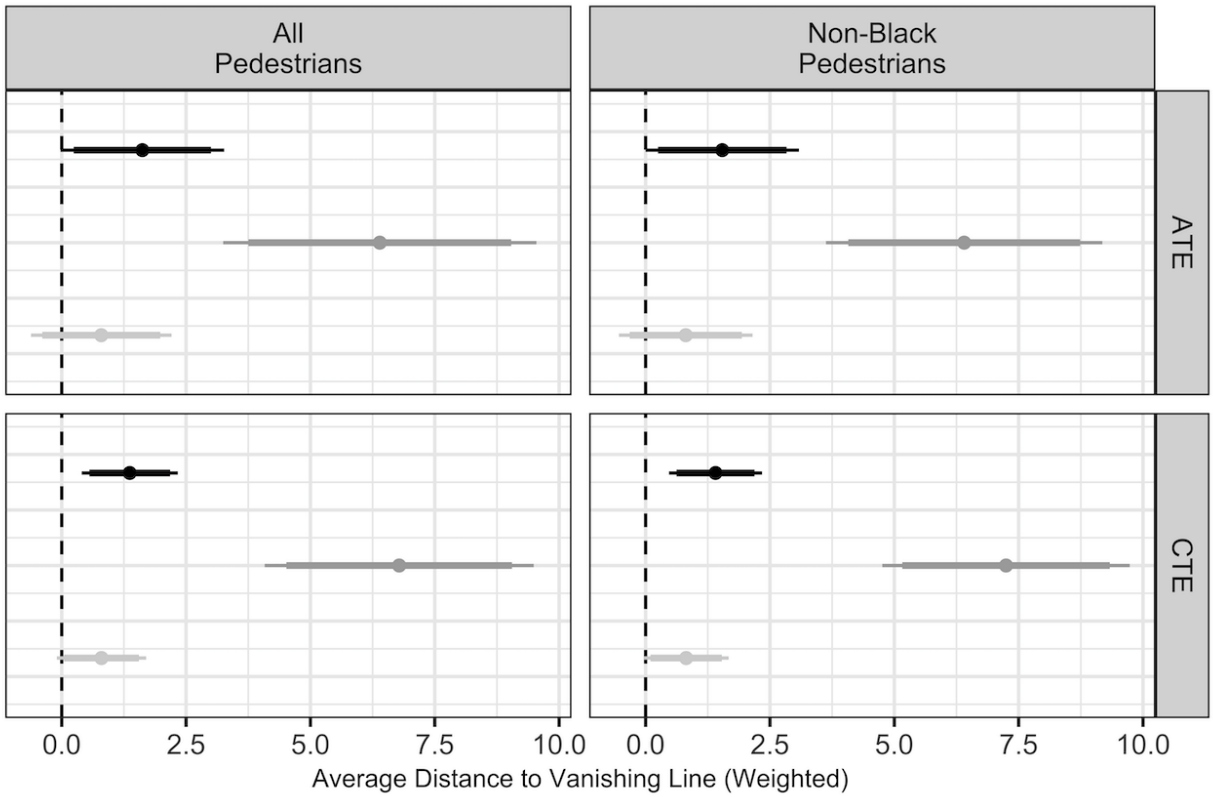


Note: Treatment effects from OLS regressions of pedestrian distance from vanishing line on an indicator for whether the confederates present are Black (versus white). Measures were created by averaging over all tracking data. Top panels (ATE) reflect simple differences-in-means while the bottom panels (CTE) include controls for pedestrian characteristics and time block fixed effects. Positive values indicate deviation from Black confederates relative to white confederates as a proportion of total sidewalk width. Black (■) denotes both locations ($N_{all} = 3419; N_{non-Black} = 3208$), while dark grey (■) and light grey (□) correspond to the Upper East Side ($N_{all} = 516; N_{non-Black} = 448$) and Midtown ($N_{all} = 2903; N_{non-Black} = 2758$), respectively. Thicker (▬) and thinner (—) lines represent 90 and 95-percent confidence intervals, bootstrapped to account for dependence within 15-minute clusters.

S4.3.2 Weighted Average Line Distance

In Figure S15 we replicate the analysis reported in Figure S14, weighting the averages based on the proximity of the data point to the confederates. Similar to what was pre-registered in Section S4.2, weights equal the inverse of the confederate distance. Again, our results are similar to those reported in Figure 1 in the main text. Specifically, we find that pedestrians move further away from the Black confederates as compared to the white, but this result is only statistically significant at the 0.05-level when controls are included ($t_{3416} = 1.931, p = 0.054, \beta = 1.621, CI_{95\%} = [-0.025, 3.266]$, without controls, $t_{3322} = 2.772, p = 0.006, \beta = 1.368, CI_{95\%} = [0.400, 2.335]$ with controls).

Figure S15: Pedestrians Give A Wider Berth to Black Confederates (Weighted Average Line Distance)



Note: Treatment effects from OLS regressions of pedestrian distance from vanishing line on an indicator for whether the confederates present are Black (versus white). Measures were created by averaging over all tracking data giving more weight to points closer to the confederates. Top panels (ATE) reflect simple differences-in-means while the bottom panels (CTE) include controls for pedestrian characteristics and time block fixed effects. Positive values indicate deviation from Black confederates relative to white confederates as a proportion of total sidewalk width. Black (■) denotes both locations ($N_{all} = 3419; N_{non-Black} = 3208$), while dark grey (■) and light grey (□) correspond to the Upper East Side ($N_{all} = 516; N_{non-Black} = 448$) and Midtown ($N_{all} = 2903; N_{non-Black} = 2758$), respectively. Thicker (▬) and thinner (—) lines represent 90 and 95-percent confidence intervals, bootstrapped to account for dependence within 15-minute clusters.

S5 Additional Analyses

S5.1 Inter-coder Reliability

S5.1.1 Pedestrian Tracking

Object-tracking does not easily yield itself to traditional measures of inter-coder reliability. This is because we must simultaneously evaluate both the pedestrian labels (i.e., which pedestrians appear in which frames) and the pedestrian tracking (i.e., the location of a given pedestrian in a given frame). If the coders' task was simply to count the number of pedestrians in each frame, then traditional inter-coder reliability statistics (e.g., Cohen's Kappa) could be used. However, the complexity of our coders' task requires a different sort of metric, which does not appear to be readily available in the literature. In the object-tracking literature we could not find a single article dedicated to inter-coder reliability for the tracking labels themselves. Instead, most studies assume "ground truth" labels are (essentially) infallible, an approach that diverges radically from the way social scientists approach coding tasks (Lombard, Snyder-Duch and Bracken, 2002).

An extensive literature exists in computer science on performance metrics (for review, see Luo et al., 2014). This literature is dedicated to comparing algorithms rather than the coders themselves. We adapt these approaches for our unique context by treating each undergraduate coder as a distinct algorithm to be evaluated against the coding performance of one of the researchers, as a proxy for the "ground truth". The inter-coder reliability metrics reported below are based coder performance in the training tracks described in Section S3.1.

To assess inter-coder reliability for our tracking data, we relied on four metrics: Multiple-Object Tracking Accuracy (MOTA), Identified Detections F1 (IDF1), Recall, and Precision. Each of these measures is intended to evaluate the performance of a multi-object tracking algorithm, as compared to a "ground truth". For all four metrics, higher scores indicate better performance, with 100% representing a perfect score.

As explained in Table S2, MOTA combines false negatives, false positives and mismatch rates. In this instance, the object tracker is the individual coder and the errors are the extent to which the coder labels tracking information in the same way as the ground truth which was created by a member of the research team. In this way, higher MOTA implies the coder tracks pedestrians in the same way as the the research team. Not only is this measure "by far the most widely accepted evaluation measure for MOT [Multiple Object Tracking]," but it provides a "fairly reasonable quantity for the overall tracking performance" (Luo et al., 2014, 11).

Identification precision is the fraction of correct detections to the total number of ground truth pedestrians. Conversely, identification recall is the ratio of ground truth pedestrians correctly identified to the total number of ground truth pedestrians. Both measures help assess the extent to which our coders consistently and correctly identify the pedestrians in the ground truth file over their entire trajectory, meaning our coders are being tested for each step in the pedestrian's path, rather than simply whether they can identify the pedestrian at all. IDF1 is simply the harmonic mean of identification precision and recall with higher

Table S2: Multiple Object Tracking Performance Measures Used to Assess our Coders

Measure	Better	Perfect	Description
MOTA	higher	100%	"The MOTA accounts for all object configuration errors made by the tracker, false positives, misses, mismatches, over all frames" (Bernardin and Stiefelhagen, 2008, 5).
Precision	higher	100%	Ratio of correct detections over the total number of detections.
Recall	higher	100%	Ratio of the total number of correct detections that were actually retrieved.
IDF1	higher	100%	"IDF_1 is the ratio of correctly identified detections over the average number of ground-truth and computed detections" (Ristani et al., 2016, 9).

Note: This table defines all of the performance metrics we used to assess our coders. In the second and third columns, we also provide information about how to interpret the measures.

values implying better performance.

Table S3: Multiple Object Tracking Performance Measures (Coders)

Coder	MOTA	IDF1	Recall	Precision
6	68.2%	73.3%	90.9%	87.0%
1	63.8%	78.4%	97.9%	83.6%
7	63.8%	78.4%	97.9%	83.6%
4	59.6%	75.5%	95.7%	88.2%
3	48.9%	76.5%	89.4%	76.4%
2	46.8%	74.8%	95.7%	75.0%
5	46.8%	71.7%	93.6%	74.6%
All	56.8%	75.5%	94.4%	81.2%

Note: This table reports the performance metrics outlined in Table S2 for the seven coders who did the vast majority of manual tracking for this study.

As explained in Section S3.1, in order to participate, RAs were asked to track all the pedestrians in a short sample clip from the more difficult Midtown location. In total, ten RAs completed the training task, but we were only able to recover the actual tracking files

from seven, meaning we only have performance metrics for 70% of the RAs who worked on the study. With that said, these RAs labeled the vast majority of the videos from both the Midtown and Upper East side locations, so we feel the results reported in Table S3 are representative of the performance of the coders in this study.

Table S4: Multiple Object Tracking Performance Measures (Top Algorithms)

Model	MOTA	IDF1	Recall	Precision
LPC_MOT	56.3%	62.5%	58.8%	96.3%
MOT20_TBC	54.5%	50.1%	62.3%	89.5%
UnsupTrack	53.6%	50.6%	55.3%	97.8%
BD_MOT	53.5%	49.3%	55.4%	97.5%
SFS	50.8%	41.1%	61.2%	86.3%
track_bnw	50.8%	52.1%	62.7%	84.7%
HTBT	48.9%	54.6%	58.1%	86.8%
All	52.6%	51.5%	59.1%	91.3%

Note: This table reports the performance metrics outlined in Table S2 for the top seven algorithms in the most recent MOT challenge (Dendorfer et al., 2020).

For each of these metrics, we used the `py-motmetrics` module in Python, but most are difficult to directly interpret. Given that, Table S4 reports the top-7 performing algorithms from the most recent MOT challenge (Dendorfer et al., 2020). Regardless of the measure, our coders out-perform the state-of-the-art in multi-object tracking. More specifically, our coders have – on average – a 4.21% higher MOTA than the top algorithms in the field, although this difference is not statistically significant ($t_{12} = 1.178, p = 0.262, \beta = 4.214, CI_{95\%} = [-3.580, 12.009]$). When the IDF1 of our coders is compared to the IDF1 of these same algorithms, we find the former is – on average – 24.04% higher than the latter and this difference is statistically significant ($t_{12} = 9.241, p = 0.00000083, \beta = 24.043, CI_{95\%} = [18.374, 29.712]$). The same is true for recall ($t_{12} = 20.679, p = 0.0000000001, \beta = 35.329, CI_{95\%} = [31.606, 39.051]$) where are coders perform significantly better than the best multi-object tracking algorithms.

We note that the inter-coder reliability measures reported in this section represent an "upper bound" on the amount of coding error present in the data. This is because the researcher manually reviewed the tracking data for obvious errors, including pedestrians missed entirely or partially by the coder, and erroneous labeling of pedestrians.

Finally, we note that this is an area ripe for future interdisciplinary collaboration, one in which computer and social scientists can work together to determine the best way to assess the reliability of manual tracking data. We hope our study motivates this future research, but – for now – we decided to use the best tools available based on the previous literature.

S5.1.2 Pedestrian Race

As explained in Section S3.2, RAs reported whether each pedestrian passing the confederates was, in their "best guess", Black or African-American. This was done at both our Midtown and Upper East Side locations. In the latter, we attempted to validate the racial information by employing another RA to watch the videos and independently code pedestrian race. This individual was also asked whether they could fully see the pedestrian and to provide a confidence rating – low, medium or high – for each of their race labels. Tables S5 and S6 report the percent agreement and Cohen's Kappa, respectively, for different configurations of the observed and video race labels.

Table S5: Percent Agreement for Observed vs. Video Race Labels

N	Confidence	Can See	% Agreement	χ^2	df	p-value	95% CI
516	L/M/H	Y/N	90.50	336.994	1	<0.001	[0.876, 0.928]
497	L/M/H	Y	90.34	321.932	1	<0.001	[0.873, 0.927]
234	M/H	Y	96.15	197.543	1	<0.001	[0.926, 0.981]
113	H	Y	99.12	107.080	1	<0.001	[0.945, 1.000]

Note: This table reports the percent agreement between the race reported in real-time on the day of the experiment versus the race reported by an RA watching the videos at a later date. Again, "race" is only whether the pedestrian is Black or African-American. In the last two columns we provide the statistic and p-value from a χ^2-test comparing the proportion of agreement to random chance. In the first two columns, we indicate what portion of the data we use for these statistics.

Looking to the first row in Table S5, when all available data from the Black and White treatments are used we find 90.50% agreement between the observed and video race labels. Not only is this significantly different from chance ($\chi^2_1 = 336.994, p < 0.001, \beta = 0.950, CI_{95\%} = [0.876, 0.928]$), but the percent agreement increases as we use video labels in which (1) the RA can actually see the pedestrian and (2) has a high degree of confidence in the rating. More specifically, when we restrict the data to only the pedestrians which the RA coding the video data can see and has a high confidence in the rating, the percent agreement increases to 99.12%. Although this is percent is again significantly different from chance ($\chi^2_1 = 107.080, p < 0.001, \beta = 0.991, CI_{95\%} = [0.945, 1.000]$), only 21.90% of the pedestrians met both of these criteria, meaning the RA coding the video data only had high confidence and could see only 113 of the 516 pedestrians for which we have a race label from the RA on the ground during the Black and White treatments. The results are essentially the same when pedestrians from our control group – when no confederates were present – are also included which underscores the difficulty in assessing race after the fact and hopefully motivates future researchers to expand and improve upon the method we introduce in this study.

Unfortunately, percent agreement does not take into consideration the difficulty of the task. Said differently, we cannot be sure whether the coders are simply making random

Table S6: Cohen's Kappa for Observed vs. Video Race Labels

N	Confidence	Can See	κ	z	p-value
516	L/M/H	Y/N	0.091	2.065	<0.001
497	L/M/H	Y	0.092	2.071	<0.001
234	M/H	Y	0.162	2.493	<0.001
113	H	Y	0.663	7.483	<0.001

Note: This table reports Cohen's kappa for the race reported in real-time on the day of the experiment versus the race reported at a later date by an RA watching the videos. Again, "race" is only whether the pedestrian is Black or African-American. In the last three columns we provide the statistic, z-score and associated p-value.

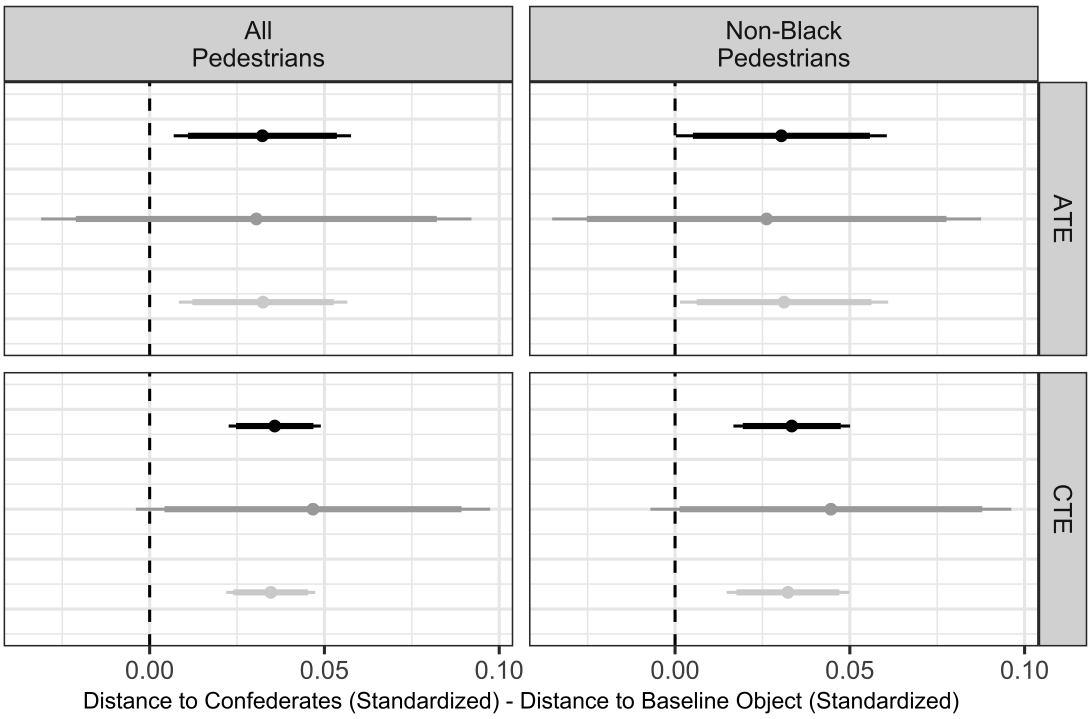
guesses and achieved high agreement by chance alone. Cohen (1960)'s kappa accounts for this problem, ranging from -1 to 1, with 0 representing the amount of agreement that can be expected by random chance. Looking to the first row of Table S6, when all available data from the Black and White treatments is used Cohen's kappa is 0.091, and is significantly different from zero ($\kappa = 0.091, z = 2.065, p < 0.001$). Cohen (1960) would say this was "slight" agreement. However, such "rules of thumb" should be taken with some qualification as there are no clear standards. With that said, when we restrict the data to only the pedestrians which the RA coding the video data can see and has a high confidence in the rating, Cohen's kappa increases substantially to 0.663 which is again significantly different from zero ($\kappa = 0.663, z = 7.483, p =< 0.001$). This result is – according to Cohen (1960) – indicative of "substantial" agreement. Again, the results are nearly identical when the control group – where no confederates were present – is included when calculating Cohen's kappa.

Finally, in Figure S16 we replicate the results we reported in Figure 1 in the main text, but we only included the 467 pedestrians at our Upper East Side location where both the on-ground and video RAs agreed on the race label in our Black and White treatments. When this is done we again find that pedestrians move significantly further away from the Black confederates as compared to the white ($t_{3368} = 2.492, p = 0.013, \beta = 0.032, CI_{95\%} = [0.007, 0.058]$, without controls, $t_{3280} = 5.321, p = 0.0000001100801, \beta = 0.036, CI_{95\%} = [0.023, 0.049]$ with controls). As before, we also find this result is more pronounced for non-Black pedestrians, suggesting the way race was coded did not substantially change the interpretation of our results.

S5.1.3 Closest Point

As explained in Section S3.5, given the high number of pedestrians at our Midtown location we used three research assistants to determine the frame in which the pedestrians were closest to our confederates. To estimate inter-coder reliability, we randomly sample twenty-five pedestrians from our simulated Midtown location, using the same camera angle as we

Figure S16: Pedestrians Give A Wider Berth to Black Confederates (Excluding Race Disagreements)



Note: Treatment effects from OLS regressions of standardized pedestrian deviation on an indicator for whether the confederates present are Black (versus white). The top panels (ATE) reflect simple differences-in-means while the bottom panels (CTE) include controls for pedestrian characteristics and time block fixed effects. Positive values indicate deviation from Black confederates relative to white confederates as a proportion of total sidewalk width. Black (■) denotes both locations ($N_{all} = 3375; N_{non-Black} = 3190$), while dark grey (■) and light grey (□) correspond to the Upper East Side ($N_{all} = 472; N_{non-Black} = 432$) and Midtown ($N_{all} = 2903; N_{non-Black} = 2758$), respectively. Thicker (▬) and thinner (—) lines represent 90 and 95-percent confidence intervals, bootstrapped to account for dependence within 15-minute clusters.

used in our experiment. Here, the frames selected by coders A, B and C were, on average, between zero and one frames away from the actual frame where the pedestrians were closest to the confederates in our Midtown simulation. Given the simulation frame rate of 25 frames per second – meaning that one frame represents 0.04 seconds – this reflects a high degree of accuracy. Moreover, the 95% confidence interval for the interclass correlation score – using two-way random effects and a single coder – ranges from 0.999 to 1, suggesting "excellent reliability" (Koo and Li, 2016, 158). Collectively, these results demonstrate that our research assistants can accurately and consistently identify the closest frame at our Midtown location.

S5.2 Balance

In Table S7 and S8 we report balance statistics for all experimental conditions. Although the results are largely the same, Table S7 uses the race labels provided by the RA who was on the ground on the day of the experiment, whereas Table S8 uses the labels that were obtained from an RA watching videos of the experiment at a later date. Given that we do not have many of the covariates we have for the Upper East Side in our Midtown location, we only assess balance using race and the total number of pedestrians.

Table S7: Number of Pedestrians in Treatment and Control Groups By Location (Race Coded by On-Ground RA)

					Black + White		*Black + Control*		*White + Control*	
Pedestrians	Black Confed.	White Confed.	No Confed.	Total	χ^2	p	χ^2	p	χ^2	p
Both Locations										
All	1700	1719	1962	5381	0.190	1.000	37.204	0.000	31.820	0.000
Black	66	105	101	272	16.889	0.001	13.844	0.005	0.087	1.000
Non-Black	1634	1614	1601	4849	0.222	1.000	0.633	1.000	0.090	1.000
Upper East Side										
All	253	263	227	743	0.314	1.000	2.604	1.000	5.000	0.684
Black	9	17	18	44	3.769	1.000	4.741	0.795	0.000	1.000
Non-Black	244	246	209	699	0.004	1.000	5.104	0.645	5.697	0.459
Midtown										
All	1447	1456	1735	4638	0.044	1.000	51.772	0.000	48.439	0.000
Black	57	88	83	228	12.414	0.012	8.929	0.076	0.187	1.000
Non-Black	1390	1368	1652	4410	0.320	1.000	44.787	0.000	53.039	0.000

Note: This table shows the number of pedestrians who were randomly assigned to either the Black, white or control conditions. Pedestrian race was coded in real-time by RA who was on the ground the day of the experiment. In the last six columns we provide the results from proportion tests with Bonferroni-adjusted p-values.

Table S8: Number of Pedestrians in Treatment and Control Groups By Location (Race Coded by RA Watching Video)

Pedestrians	Black Confed.	White Confed.	No Confed.	Total	*Black + White*		*Black + Control*		*White + Control*	
					χ^2	p	χ^2	p	χ^2	p
Both Locations										
All	1700	1719	1962	5381	0.190	1.000	37.204	0.000	31.820	0.000
Black	72	104	104	280	10.920	0.026	10.920	0.026	0.000	1.000
Non-Black	1628	1615	1598	4841	0.089	1.000	0.521	1.000	0.159	1.000
Upper East Side										
All	253	263	227	743	0.314	1.000	2.604	1.000	5.000	0.684
Black	15	16	21	52	0.000	1.000	1.389	1.000	0.865	1.000
Non-Black	238	247	206	691	0.264	1.000	4.329	1.000	7.064	0.212
Midtown										
All	1447	1456	1735	4638	0.044	1.000	51.772	0.000	48.439	0.000
Black	57	88	83	228	12.414	0.012	8.929	0.076	0.187	1.000
Non-Black	1390	1368	1652	4410	0.320	1.000	44.787	0.000	53.039	0.000

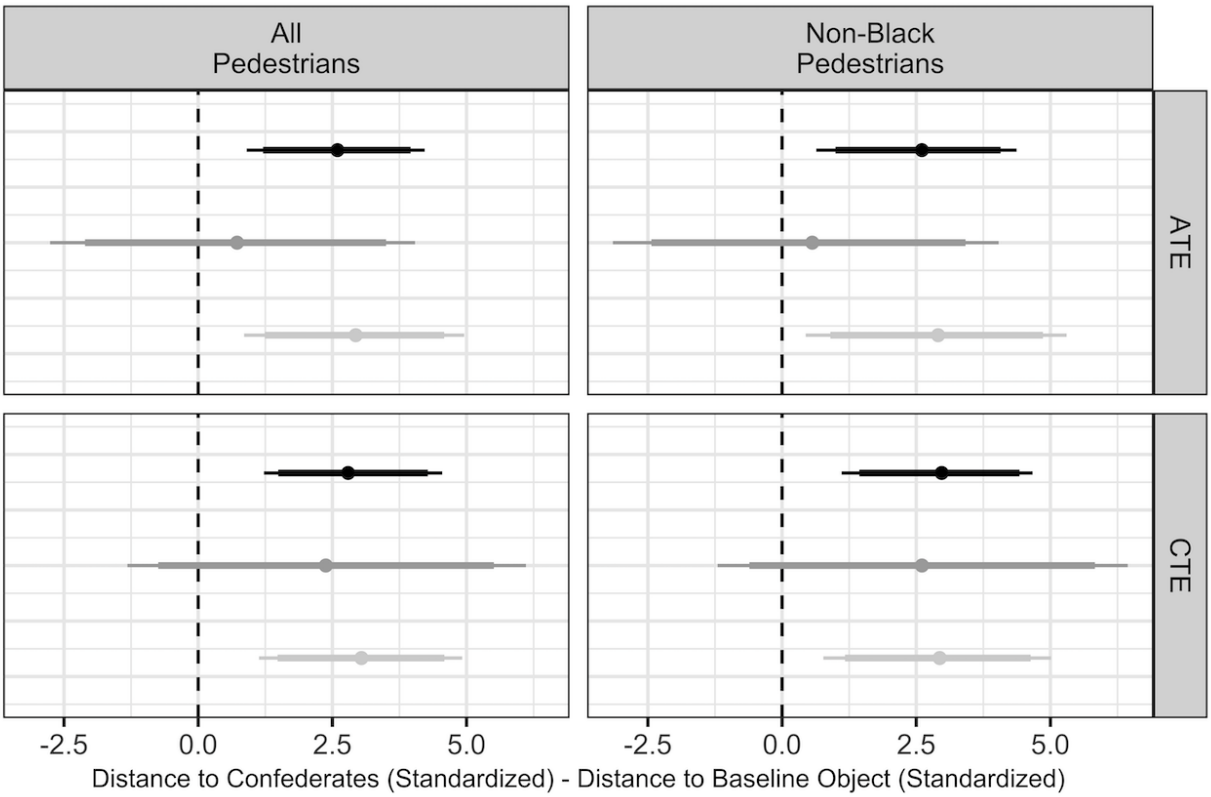
Note: This table shows the number of pedestrians who were randomly assigned to either the Black, white or control conditions. Pedestrian race was coded by a RA who watched videos of the experiment at a later date. In the last six columns we provide the results from proportion tests with Bonferroni-adjusted p-values.

S5.3 Random Intercept

Figure S17 replicates what we reported in the main text, but instead of reporting clustered standard errors we report results from a nested random intercept model. That is, we assume each measurement is nested within each pedestrian and those pedestrians are nested within within 15-minute clusters. As in the main analysis, we find that pedestrians move further away from the Black confederates as compared to the white ($t_{3267.803} = 3.129, p = 0.002, \beta = 2.595, CI_{95\%} = [0.903, 4.220]$, without controls, $t_{6.641} = 2.431, p = 0.047, \beta = 2.793, CI_{95\%} = [1.224, 4.548]$ with controls).[S10] All models were estimated using the `lme4` package in the `R` statistical software language. We also note that we received singular fit warnings for all the Upper East Side models, the Midtown model using all pedestrians with controls and the model using both locations with no controls. This suggests these models were overfitted and the random effects structure was overly complex. As such, these models are not reported in the main text despite being mentioned briefly in our pre-analysis plan.

[S10] There is considerable debate in the mixed-effects modeling literature on the best way to estimate the degrees of freedom (Pinheiro and Bates, 2006), we use the Satterthwaite approximation which is the default setting in the `lmerTest` package of the R statistical language. This is why the degrees of freedom vary and are not round numbers.

Figure S17: Pedestrians Give A Wider Berth to Black Confederates (Random Intercept)

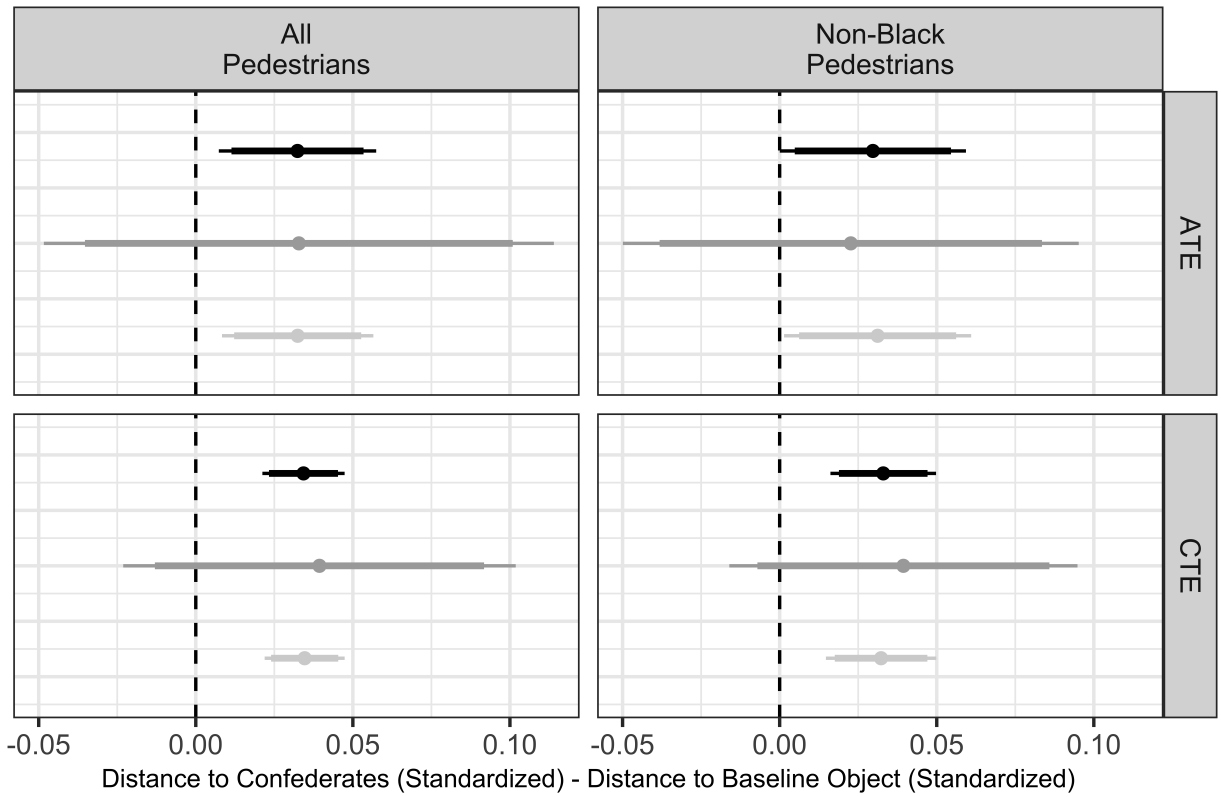


Note: Treatment effects from OLS regressions of standardized pedestrian deviation on an indicator for whether the confederates present are Black (versus white). Results are from nested random intercept model. Top panels (ATE) reflect simple differences-in-means while the bottom panels (CTE) include controls for pedestrian characteristics and time block fixed effects. Positive values indicate deviation from Black confederates relative to white confederates as a proportion of total sidewalk width. Black (■) denotes both locations ($N_{all} = 3419; N_{non-Black} = 3208$), while dark grey (■) and light grey (□) correspond to the Upper East Side ($N_{all} = 516; N_{non-Black} = 448$) and Midtown ($N_{all} = 2903; N_{non-Black} = 2758$), respectively. Thicker (▬) and thinner (—) lines represent 90 and 95-percent confidence intervals.

S5.4 Including Outliers

Figure S18 replicates the results reported in the main text including the pedestrians walking in zigzags at our Upper East Side location which was discussed above on page S7. When this is done we again find that pedestrians move further away from the Black confederates as compared to the white ($t_{3422} = 2.591, p = 0.010, \beta = 0.032, CI_{95\%} = [0.008, 0.056]$, without controls, $t_{3327} = 5.100, p = 0.00000036, \beta = 0.033, CI_{95\%} = [0.020, 0.046]$ with controls). As before, we also find this results is more pronounced for non-Black pedestrians, suggesting these nine pedestrians do not change the substantive interpretation of our results.

Figure S18: Pedestrians Give A Wider Berth to Black Confederates (Includes Outliers)



Note: Treatment effects from OLS regressions of standardized pedestrian deviation on an indicator for whether the confederates present are Black (versus white). Results include the nine pedestrians flagged as "outliers" by our RA. Six of these pedestrians were in the Black and white confederate treatments. Top panels (ATE) reflect simple differences-in-means while the bottom panels (CTE) include controls for pedestrian characteristics and time block fixed effects. Positive values indicate deviation from Black confederates relative to white confederates as a proportion of total sidewalk width. Black (■) denotes both locations ($N_{all} = 3425; N_{non-Black} = 3208$), while dark grey (■) and light grey (□) correspond to the Upper East Side ($N_{all} = 522; N_{non-Black} = 450$) and Midtown ($N_{all} = 2903; N_{non-Black} = 2758$), respectively. Thicker (▬) and thinner (—) lines represent 90 and 95-percent confidence intervals, bootstrapped to account for dependence within 15-minute clusters.

S5.5 Additional Pre-Registered Hypotheses

In our pre-analysis plan, we registered the following hypotheses:

- Obstruction avoidance hypothesis: Pedestrians (of all races) will tend to move further away from confederates of either race than the baseline object (e.g., a trashcan) as compared to pedestrians when no confederates are present.
- Racial avoidance hypothesis: Pedestrians (of all races) will tend to move further away from Black confederates than the baseline object (e.g., a trashcan) as compared to pedestrians when white confederates are present.
- Neighborhood outgroup salience hypothesis: The wider berth predicted by the racial avoidance hypothesis will be more pronounced in a predominantly white neighborhood as compared to a predominantly non-white neighborhood.
- Pedestrian outgroup salience hypothesis: The wider berth predicted by the racial avoidance hypothesis will be more pronounced in white pedestrians as compared to non-white pedestrians.
- Gender heterogeneity hypothesis: Female pedestrians (of all races) will tend to move further away from male confederates than the baseline object (e.g., a trashcan) as compared to male pedestrians.

The racial avoidance and pedestrian outgroup salience hypotheses are explicitly stated in the main text. The neighborhood outgroup salience hypothesis is examined in the main text as well, although we have de-emphasized this comparison due to the difficulty of directly comparing the two research sites. The remaining hypotheses – "Obstruction avoidance" and "Gender heterogeneity" – are discussed below.

We also note the following deviations from our pre-analysis plan. First, on page 4 of the plan we say a research assistant will be "unobtrusively observing the scene and recording characteristics of each pedestrian who passes between the trash can and confederates." As mentioned in footnote S8, we quickly found this to be intractable for every pedestrian characteristic. So, we only asked our RAs to code whether the pedestrian appeared to be Black or African-American, in their judgment.

Second, on page 8 when we describe the conditional average treatment effect, we say we would control for (1) race, (2) gender, (3) block fixed-effects, (4) presence of obstructions on the sidewalk, (5) camera lens obstructions, and (6) the degree to which the confederates where in the shade. In our analyses we did not control for camera/sidewalk obstructions and the degree to which the confederates were in the shade because none of our sidewalks ended up having obstructions and the camera lens never had anything like glare. Moreover, the confederates were never in shade which also made the last control unnecessary.

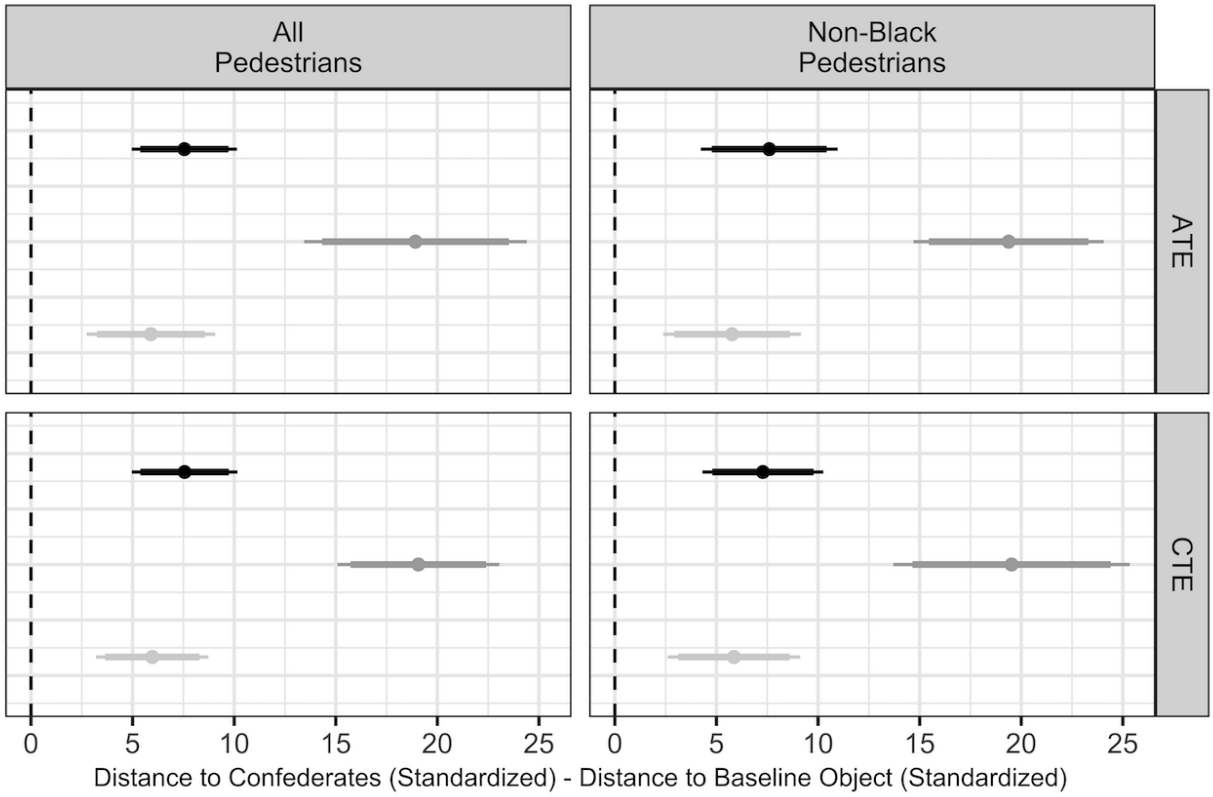
Finally, we briefly mention using a "random intercept" model where we assume there are "pedestrians w/ multiple measurements" (see pg 8). Although this model was proposed in addition to the clustered standard errors we report throughout our study, we did not

report these results in main text. Instead, these results are reported in Section S5.3 in the SI. Ultimately, the substantive results are nearly identical to those shown in Figure 1 in the main text, but we also had issues with singularity, especially for the models only using the Upper East Side data. This suggests the random effects structure of this pre-registered model is overly complex which is why it is reserved for the SI.

S5.5.1 Obstruction Avoidance

In our preanalysis plan, we specified the following obstruction avoidance hypothesis: "Pedestrians (of all races) will tend to move further away from confederates of either race than the baseline object as compared to pedestrians when no confederates are present." Given previous literature which shows pedestrians tend to adjust their gait in order to avoid obstacles and maintain personal space (e.g., Collett and Marsh, 1981; Gérin-Lajoie, Richards and McFadyen, 2005; Moussaïd, Helbing and Theraulaz, 2011), we expect pedestrians to walk closer to the baseline object when there are confederates present. Ultimately, we view this hypothesis as an important check on the validity of our measurement strategy: can we, in fact, detect pedestrian avoidance behavior?

Figure S19: Pedestrians Generally Avoid The Confederates



Note: Treatment effects from OLS regressions of standardized pedestrian deviation on an indicator for whether confederates are (either Black or white). Top panels (ATE) reflect simple differences-in-means while the bottom panels (CTE) include controls for pedestrian characteristics and time block fixed effects. Positive values indicate deviation from confederates relative to the control condition as a proportion of total sidewalk width. The later was estimated using the average position of the Black and white confederates in the same block. Black (■) denotes both locations, while dark grey (■) and light grey (□) correspond to the Upper East Side and Midtown, respectively. Thicker (▬) and thinner (—) lines represent 90 and 95-percent confidence intervals, bootstrapped to account for dependence within 15-minute clusters.

In Figure S19, the treatment effect is a simple dummy variable which equals 1 when either the Black or white confederates are present and 0 otherwise. Although we can replicate this result using any of the approaches outlined in Section S4, we report the closest point estimate which is what is included in the main text of our paper. Ultimately, we find strong evidence that pedestrians move towards the baseline object. Indeed, none of the confidence intervals in Figure S19 overlap zero which suggests when either the Black or white confederates are on the sidewalk pedestrians tend to be further away from them as compared to the baseline object.

S5.5.2 Observed Gender

In our preanalysis plan. we specified the following gender heterogeneity hypothesis: "Female pedestrians (of all races) will tend to move further away from male confederates than the baseline object (e.g., a trashcan) as compared to male pedestrians." Given previous literature which shows gender differences as a determinant of both power and gallantry in pedestrian encounters (Silveira, 1972; Dabbs Jr and Stokes III, 1975; Sobel and Lillith, 1975; Willis, Gier and Smith, 1979), we expect the treatment effect to be more pronounced in female pedestrians. Ultimately, our results are consistent with this hypothesis.

In Figure S20, we report the same models outlined in the main text, but instead of subsetting our data by race we subset it by gender. Consistent with our pre-registered hypothesis, we find that female pedestrians tend to give an even wider berth as compared to their male counterparts. However, we are only able to test this hypothesis at our Upper East Side location since we were unable to record gender in real-time and instead rely on the videos to impute gender information (it is not possible to discern gender from the Midtown videos). For these reasons, we see Figure S20 as providing preliminary evidence, and encourage future scholars to more rigorously examine this question.

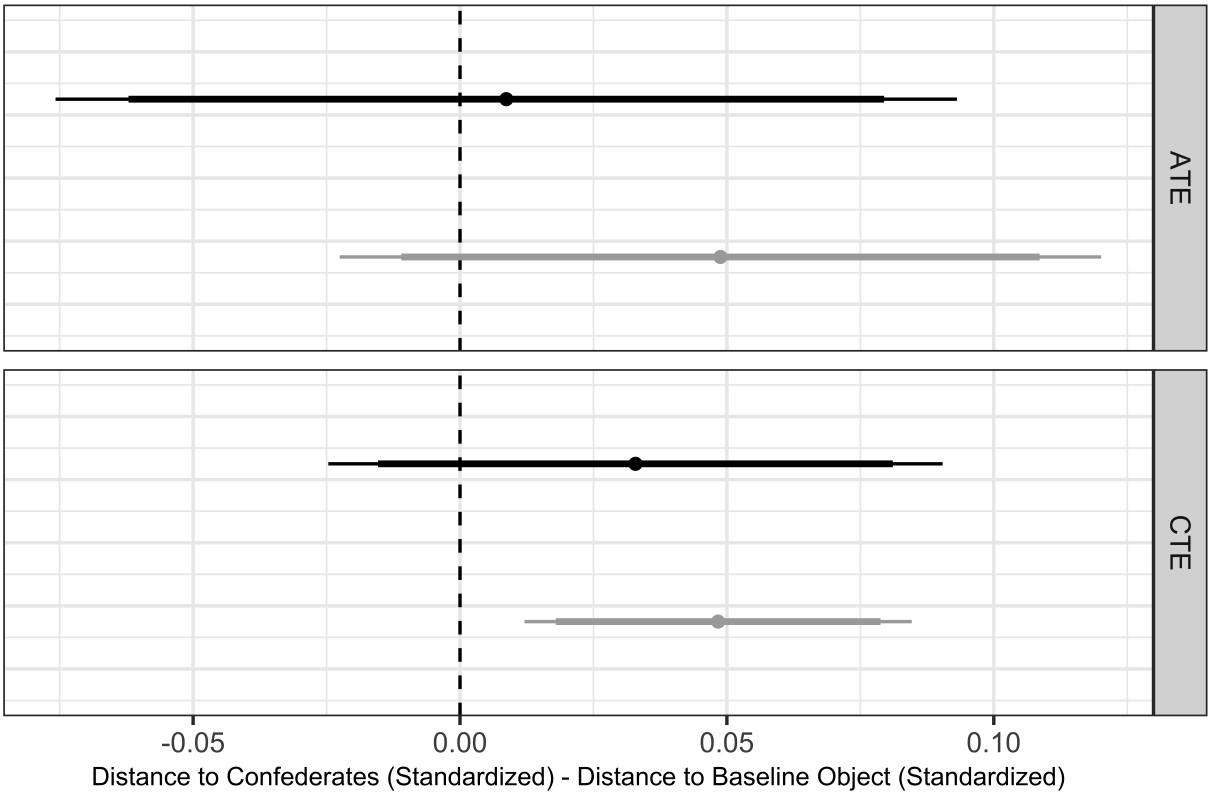
S5.5.3 Neighborhood Salience

Figure 1 in the main text shows that pedestrians move further away from our phenotypically Black as opposed to white confederates when data is restricted to our Upper East Side location. This is consistent with our neighborhood outgroup salience hypothesis, but the overlapping confidence intervals suggests there is no significant difference between the two experiment locations. When these same confidence intervals are compared in the figures reported in the SI, a similar conclusion is reached – generally pedestrians move further away from the Black confederates in the Upper East Side, but there is no significant difference between the neighborhoods.

S5.5.4 Outgroup Salience

Figure S21 replicates the analysis in the main text, but includes results for models using only pedestrians identified as phenotypically Black. That is, we use the standardized pixel distance at the closest point, but we use the following pedestrian subsets: "All Pedestrians,"

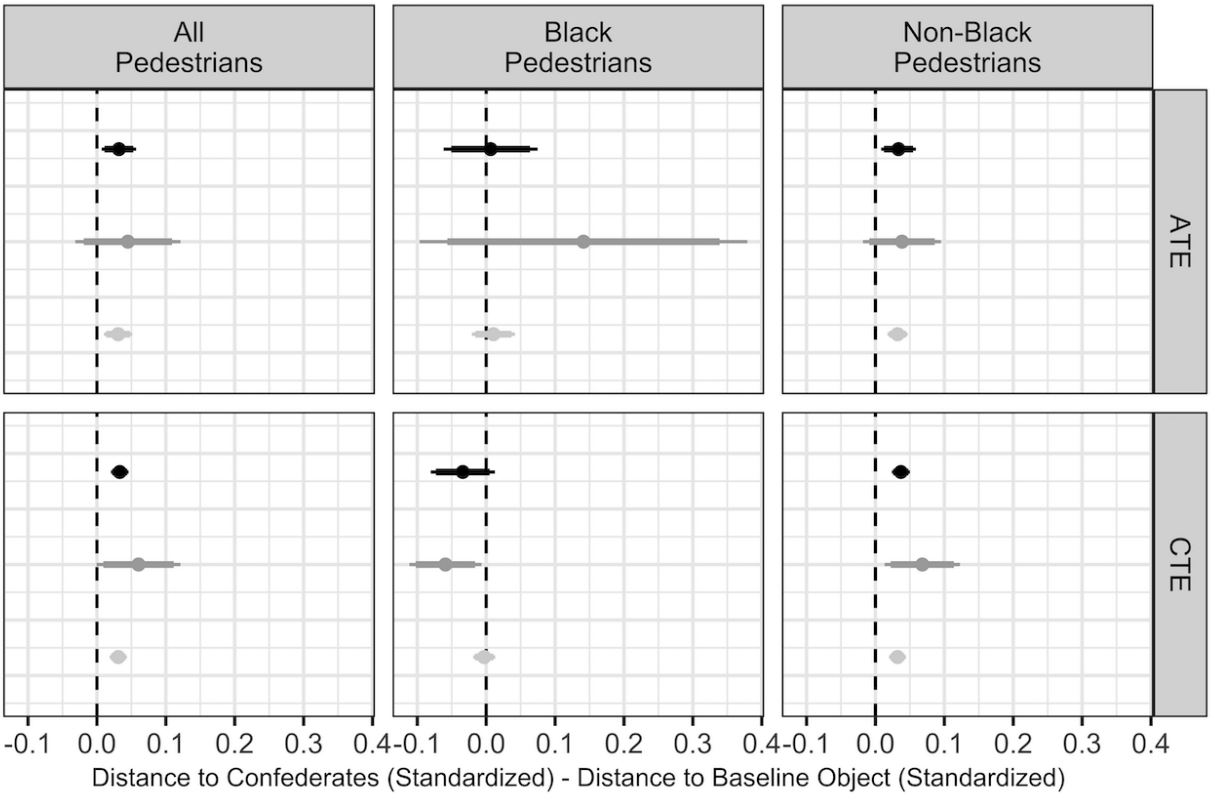
Figure S20: Female Pedestrians Give A Wider Berth to Black Confederates (Upper East Side Only)



Note: Treatment effects from OLS regressions of standardized pedestrian deviation on an indicator for whether the confederates present are Black (versus white). The top panels (ATE) reflect simple differences-in-means while the bottom panels (CTE) include controls for pedestrian characteristics and time block fixed effects. Positive values indicate deviation from Black confederates relative to white confederates as a proportion of total sidewalk width. Black (■) and dark grey (■) corresponds to male ($N = 289$) and female pedestrians ($N = 227$), respectively. Thicker (▬) and thinner (—) lines represent 90 and 95-percent confidence intervals, bootstrapped to account for dependence within 15-minute clusters.

"Black Pedestrians," and "Non-Black Pedestrians." Consistent with our pedestrian outgroup salience hypothesis, we find non-Black pedestrians tend to move further away from the Black confederates. However, due to the small number of pedestrians identified as Black in our sample, we are unable to draw definitive conclusions regarding this hypothesis, especially as it pertains to our Black Confederates.

Figure S21: Pedestrians Give A Wider Berth to Black Confederates (Includes Black Pedestrians)



Note: Treatment effects from OLS regressions of standardized pedestrian deviation on an indicator for whether the confederates present are Black (versus white). Top panels (ATE) reflect simple differences-in-means while the bottom panels (CTE) include controls for pedestrian characteristics and time block fixed effects. Positive values indicate deviation from Black confederates relative to white confederates as a proportion of total sidewalk width. Black (■) denotes both locations, while dark grey (■) and light grey (□) correspond to the Upper East Side and Midtown, respectively. Thicker (▬) and thinner (—) lines represent 90 and 95-percent confidence intervals.

S6 Pre-analysis Plan

Title: Using Public Video Feeds to Understand Intergroup Exposure on the Streets of New York

Background and explanation of rationale
Brief description of goals of project. If you are also attaching a pre-analysis plan, please refrain from simply copying and pasting a section from your plan here. If possible, please also avoid saying "see attached pre-analysis plan," as it renders the search functionality less useful. Rather, please provide a short (1-2 paragraph) summary of the project background.

Since V.O. Key (1949) scholars have argued that neighborhood racial and ethnic composition shapes individuals' behavior. Evidence from across disciplines indicates that micro-level context -- who and what we encounter as we move through space -- matters for a wide range of political and social outcomes. Yet, it is difficult to study the effect of everyday experiences on behavior in real-world settings. Using publicly-available traffic camera feeds from New York City, we examine how pedestrians of different races behave in the presence of racial outgroup members, and how these interactions vary across different types of neighborhoods. By measuring the distance that pedestrians maintain between other pedestrians of various racial groups, we generate a novel measure of intergroup behavior. In doing so we give scholars a new way to unobtrusively study implicit bias during actual social interactions, thus eliminating concerns about researcher demand effects. Our study sheds light on how racial context shapes the behavior we see on city streets, and how this effect is moderated by broader neighborhood characteristics.

What are the hypotheses to be tested/quantities of interest to be estimated?
Please list the hypotheses including hypotheses on heterogeneous effects. If you are also attaching a pre-analysis plan, please refrain from simply copying and pasting a section from your plan here. If possible, please also avoid saying "see attached pre-analysis plan," as it renders the search functionality less useful. Rather, please provide a short (1-2 paragraph) summary of project hypotheses.

H2 is our main hypothesis on race effects. H3, H4, and H5 describe hypotheses on heterogeneous effects.

1. Obstruction avoidance hypothesis: Pedestrians (of all races) will tend to move further away from confederates of either race than the baseline object (e.g., a trashcan) as compared to pedestrians when no confederates are present.

2. Racial avoidance hypothesis: Pedestrians (of all races) will tend to move further away from black confederates than the baseline object (e.g., a trashcan) as compared to pedestrians when white confederates are present.

3. Neighborhood outgroup salience hypothesis: The wider berth predicted by the racial avoidance hypothesis will be more pronounced in a predominantly white neighborhood

as compared to a predominantly non-white neighborhood.

4. Pedestrian outgroup salience hypothesis: The wider berth predicted by the racial avoidance hypothesis will be more pronounced in white pedestrians as compared to non-white pedestrians.

5. Gender heterogeneity hypothesis: Female pedestrians (of all races) will tend to move further away from male confederates than the baseline object (e.g., a trashcan) as compared to male pedestrians.

How will these hypotheses be tested?
Brief description of your methodology. If you are also attaching a pre-analysis plan, please refrain from simply copying and pasting a section from your plan here. If possible, please also avoid saying "see attached pre-analysis plan," as it renders the search functionality less useful. Rather, please provide a short (1-2 paragraph) summary of project methodology.

The basic measurement strategy involves comparing the distance of each pedestrian to randomly assigned confederates and a baseline object. This is shown here:

The red box is a predefined area of interest which is 20 feet long and 5 feet wide. We only track pedestrians when they are within this box since we assume that pedestrians outside of this box (i.e., those on the opposite side of the street) are essentially unaffected by the treatment.

The blue circle is the closest edge of the baseline object (a trashcan) to the first confederate. We call the red line that connects the edge of the baseline object to each confederate the “line of intersection.” Although we will track all pedestrians who appear in the area of interest, we will construct all our measures from pedestrians who cross this line. Ultimately, we will use the average distance between the pedestrians and each confederate as our primary variable of interest.

We use the baseline object to minimize the error associated with this distance measure. Since we are measuring a three dimensional space using a two dimensional image, the correlation between pixel distance and actual distance does not remain constant. By comparing the average distance from each pedestrian to both the confederates and baseline object, we assume the error is constant across both reference points to assess whether pedestrians move closer or further away from the confederates. As explained in our pre-analysis plan, we use several of these comparisons to test each of our main hypotheses.

Sample Size (# of Units):

We recorded 21 minutes and 39 seconds of video for the pilot study. In that video, 129 pedestrians crossed the line of intersection. That means we should expect to record one pedestrian of interest for approximately every 10 seconds of video. We will analyze 225 minutes of video which should yield a sample size of approximately 1350 distributed across 15 block/condition pairs, meaning we expect to have 90 pedestrians in each cell.

Was a power analysis conducted prior to data collection? *(**Y** / N / other)*

Has this research received Institutional Review Board (IRB) or ethics committee approval? *(NA / Y / N / **other**)*

Our exemption is currently under review. We will update the pre-analysis plan once it is formally approved.

IRB Number: TBD

Date of IRB Approval: TBD

Please upload your pre-analysis plan, along with any other supporting documents, such as survey instrument, research protocol, any data, etc.

Pre-analysis plan

Background and Motivation

How people relate to one another in space reflects a complex array of societal norms, biases, and power hierarchies. The study of how pedestrians interact with one another has been described by psychologists, sociologists, and urban planners. Early work on spatial displacement on sidewalks and in public spaces emphasizes dominance behavior (Dabbs and Stokes 1975, Knowles 1973, Henley 1977) and its evolutionary roots (Washborn and Devore 1961), personal space (Evans and Howard 1973), as well as “gallantry” (Goffman 1971, Willis,

Gier, and Smith 1979). Scholars have focused on gender differences as a determinant of both power and gallantry in pedestrian encounters (Silveria 1972, Dabbs and Stokes 1975, Sobel and Lilith 1975, Willis et al 1979). Other features that determine how pedestrians are treated include group size (Knowles 1973), occupational uniform, age, physical weakness or disability (Goffman 1971), attractiveness (Dabbs and Stokes 1975), and cultural differences.

Though researchers have described various interpersonal behaviors between Black and white individuals, such as speaking distance (Willis 1966), studies of interracial pedestrian interaction are virtually non-existent.[1] Nor has anyone studied how such interactions are moderated by broader neighborhood characteristics. This is despite the fact that interracial encounters have been a topic of academic interrogation since at least 1949, when V.O. Key argued that neighborhood racial and ethnic composition shapes individuals' behavior. In the now vast body of literature that includes observational and experimental studies of racial contact (Allport 1954), threat, and biases in judgements of threat, virtually no attention has been paid to one of the most basic and common types of encounters—those that occur on sidewalks and street corners.

Moreover, studies of racial contact and context tend to be plagued by a number of measurement issues that are known to generate various biases. First, individuals sort into neighborhoods of varying demographic makeups, which, due to high levels of residential segregation, allows them to a certain extent avoid individuals of other races. Second, measures of racial contact that rely on self-reports can be highly inaccurate and subject to experimenter demand effects. Third, studies that utilize aggregate measures of context are problematic because of the modifiable areal unit problem (MAUP) and other ecological inference problems.

This study avoids these problems by using an unobtrusive measure of individual-level behavior – live public camera feeds -- to eliminate demand effects, and by experimentally inducing exposure to people of other races through the use of confederates. In doing so, it provides a window into interactions that are typically unobserved, offering a new measure of implicit bias that is not subject to social desirability effects. This new measure can, in turn, be used study how revealed behaviors vary according to context; for example, how does racial context affect the actions of in-group members upon encountering members of an outgroup.

Automated pedestrian tracking has countless uses, from designing walking infrastructure and public transit timetables, to crowd management, visual surveillance and security.
Studies of pedestrian walking behavior can describe macroscopic behavior, describing the characteristics of the flow, or microscopic behavior, focused on that of individual pedestrians. Due to the ubiquity of public traffic cameras and advances in video processing technology, the method introduced in this paper can be applied in the countless locations across the globe where pedestrian traffic is captured by live video feeds.

Research Design

Two different types of neighborhoods were selected as study sites based on the demographics of their residents: the predominantly white Upper East Side of Manhattan, and the predominantly African American Harlem in upper Manhattan. Within each neighborhood type, two cameras were selected based on several factors: visibility of a large swath of unobstructed sidewalk, camera angle and image quality.

Three experimental conditions will be implemented, and recorded, at each site. In the first condition, two phenotypically white young adult male confederates will stand facing each other, in conversation, in a designated spot within the camera image for 15 minutes. In the second condition, a pair of phenotypically black young adult male confederates will take their place for the same period of time. In the third (baseline) condition, no confederates will be present for the same period of time, recorded the same way as in the first two conditions. Confederates in each pair are dressed similarly to those in the other pair, and are similar in height, weight, and age, such that between-pair differences are minimized.

The study will be block randomized to help control for natural fluctuations in pedestrian flow that occur throughout the day, such as lunchtime and commuting hours. The order in which the three conditions are implemented within a block will be randomized.

For example, the schedule of conditions might look like this:

Monday, June 19th (72nd and Lexington):
Block 1 (10am-11am): No Confederates (15min), Black (15min), White (15min)
Block 2 (11am-12pm): No Confederates (15min), White (15min), Black (15min)
Block 3 (12-1pm): White (15min), Black (15min), No Confederates (15min)
Block 4 (1-2pm): Black (15min), No Confederates (15min), White (15min)
Block 5 (2-3pm): White (15min), No Confederates (15min), Black (15min)

Confederates are positioned on the sidewalk in relation to the baseline object (e.g., trash can) that serves as a point of reference.

During this time, a research assistant will be at the site, unobtrusively observing the scene and recording characteristics of each pedestrian who passes between the trash can and confederates. The RA will also record the time and note features of pedestrians, such as the color of their clothes, to enable us to identify those individuals in the video.

Measurement

The basic measurement strategy involves comparing the distance of each pedestrian to randomly assigned confederates and a baseline object. This is shown here:

The red box is a predefined area of interest which is 20 feet long and 5 feet wide. We assume pedestrians outside of this box (i.e., those on the opposite side of the street) are essentially unaffected by the treatment. We make this assumption for both theoretical and practical reasons, since it would increase the manual tracking costs to track every pedestrian on the screen.

The “line of intersection” connects the baseline object (see blue circle) to the confederates (see green circle). We are ultimately only interested in pedestrians who cross this line. We do this to help control for the error associated with measuring a three dimensional space using a two dimensional image. Since there is not a one-to-one correspondence between pixel and actual distance, a baseline object on the same plane as the confederates must be used in order to make inferences about whether pedestrians are moving away from the confederates as they approach.

Ultimately, we will use the average distance between the pedestrians and each confederate as our primary variable of interest. This distance will then be compared to the average distance between the pedestrians and the baseline object (e.g., a trash can). Generally speaking, if pedestrians are moving away from the confederates then they should maintain a greater distance from the confederates and a closer distance to the baseline object (e.g., a trash can) as they approach.

To create this measure, we will capture frames from the traffic camera feeds outlined above with each frame representing 3 seconds in real time. To simplify the manual tracking task, we will then split the longer videos into segments that are approximately 10 frames long, meaning each segment is approximately 30 seconds long. We will then randomly assign segments to 3 RAs with a 10 percent sample being randomly assigned to all three to evaluate intercoder reliability.

Using the Manual Tracking tool found in the Fiji distribution of ImageJ (https://imagej.net/Welcome), the RAs will then complete the following tasks:

1. Record the X and Y pixel coordinates for each pedestrian appearing in the area of interest (see red box above). The basic procedure is described in this video:

   https://youtu.be/49nek32FbMc

2. Record the X and Y pixel coordinates for each confederate using the same technique outlined in the above video. This video contains the procedure using the segmented videos described above:

   https://youtu.be/v21tYN7_aCQ

3. Record the X and Y pixel coordinates for each the side of the baseline object that is closest to each confederate. The basic procedure is described in this video:

   https://youtu.be/YBkpJ9nlz8U

4. Once the X and Y pixel coordinates are recorded for each pedestrian, each RA will then watch the whole video and determine whether a pedestrian appears in multiple segments. If so, their pedestrian identification will be changed accordingly. This will ensure there are not multiple identification numbers for the same pedestrian. The results will look something like this:

   https://youtu.be/-ZRsU9MvNyg

5. Once the X and Y pixel coordinates for each pedestrian have been finalized, we will create small videos of each pedestrian. We will then ask each RA to indicate the probability the pedestrian is (a) white and (b) male. We will then use the average from all three RAs to validate the assessments of the RA we position on site (see description in paragraph immediately prior to the “Measurement” section).

Once all the pedestrians and confederates are manually tracked, then we will calculate variations of the following:

(distance from location of confederate #1 - distance from the trashcan) + (distance from location of confederate #2 - distance from the trashcan) / 2

In all conditions, we will use the Euclidian distance in pixels. For example, imagine in the 1st frame Pedestrian 1 was located at (315, 108) and Confederate 1 was located at (324, 101), then the distance between the two points would be:

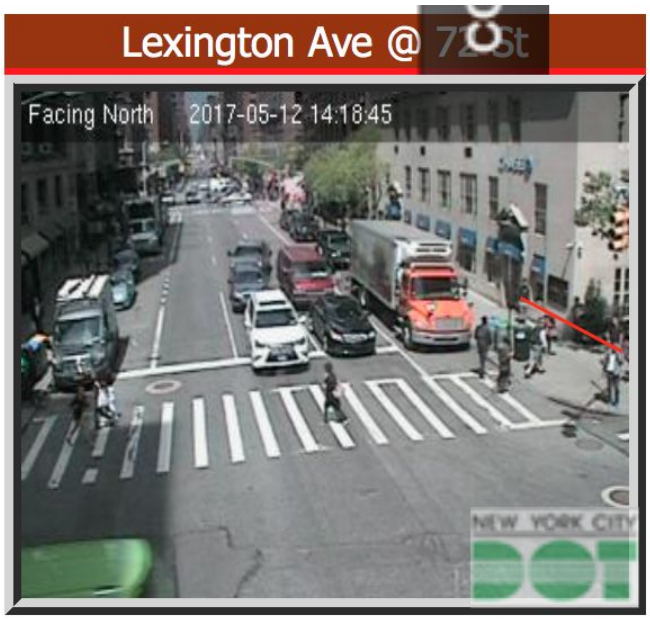
$$\sqrt{(315 - 324)^2 + (108 - 101)^2}$$

A similar measure will be calculated for the baseline object. In instances when no confederates are present (i.e., during the control condition), we will calculate distance as if the confederates *were* standing there. This will be done by using the average X and Y position of all the confederates we have in each frame.

We then take the average of those means for each condition within each block, using only the measurement that is closest to the confederates. We will also calculate a weighted average that gives more weight to measurements closer to the confederates (thus using more of the data), adjusting the standard errors for pedestrian clusters (pedestrians w/ multiple measurements) using both a random intercept model and clustered standard errors.

Conditional ATE: To increase precision, we estimate treatment effects controlling for pedestrian race and gender (as indicated by both the on-site and off-site RAs), block fixed effects, presence of obstructions (e.g., objects on the sidewalk), presence of obstructions on the camera lens (e.g., glare), and the degree to which the confederates are in the shade. This latter measure will be constructed using both pixel brightness ($\frac{R+G+B}{3}$) and luminance $\left(\sqrt{0.299R^2 + 0.587G^2 + 0.114B^2}\right)$ from the image as a whole.

Note that positive values means that the pedestrian is further away from the confederate than they are from the trash can (closer to the trash can than to the confederate). While the unit of measurement in the raw data is a pixel, these values can then be translated into (more meaningful) distance measures. To achieve this end, we will position the confederates at known distances on the street to help calibrate the pixel distance. For example, we positioned two confederates ten feet apart to help calibrate the box outlining the area of interest (see image below).



We captured this image and the others found above from a pilot study we conducted at 72nd Street and Lexington Avenue on Monday, March 27th, 2017. In this study, we implemented only the baseline (no confederate) and black confederates conditions. We found that the presence of confederates does in fact alter pedestrians' trajectories relative to the baseline, pushing them towards the trashcan and away from the confederates. We used the results of this pilot study to inform the inputs in our power analysis. If we assume that the difference between conditions is approximately half as large as the pilot then we expect to be able to detect treatment effects with 225 minutes of video. If the magnitude of the 'true' effect is less than that, we anticipate needing to collect more data in order to detect an effect.

Hypotheses

1. Obstruction avoidance hypothesis: Pedestrians (of all races) will tend to move further away from confederates of either race than the baseline object (e.g., a trashcan) as compared to pedestrians when no confederates are present.
    a. In other words: (distance from location of confederates [confederates present] - distance from the baseline object [confederates present]) > (distance from location of confederates [confederates not present] - distance from the baseline object [confederates not present])

2. Racial avoidance hypothesis: Pedestrians (of all races) will tend to move further away from black confederates than the baseline object (e.g., a trashcan) as compared to pedestrians when white confederates are present.
    a. In other words: (distance from location of confederates [black confederates present] - distance from the baseline object [black confederates present]) >

(distance from location of confederates [white confederates present] - distance from the baseline object [white confederates present])

3. Neighborhood outgroup salience hypothesis: The wider berth predicted by the racial avoidance hypothesis will be more pronounced in a predominantly white neighborhood as compared to a predominantly non-white neighborhood.
    a. In other words:
        i. $ATE_{\text{white neighborhood}}$ = ($\text{distance}_{\text{white neighborhood}}$ from location of confederates [black confederates present] - $\text{distance}_{\text{white neighborhood}}$ from the baseline object [black confederates present]) > ($\text{distance}_{\text{white neighborhood}}$ from location of confederates [white confederates present] - $\text{distance}_{\text{white neighborhood}}$ from the baseline object [white confederates present])
        ii. $ATE_{\text{non-white neighborhood}}$ = ($\text{distance}_{\text{non-white neighborhood}}$ from location of confederates [black confederates present] - $\text{distance}_{\text{non-white neighborhood}}$ from the baseline object [black confederates present]) > ($\text{distance}_{\text{non-white neighborhood}}$ from location of confederates [white confederates present] - $\text{distance}_{\text{non-white neighborhood}}$ from the baseline object [white confederates present])
        iii. $ATE_{\text{white neighborhood}} > ATE_{\text{non-white neighborhood}}$

4. Pedestrian outgroup salience hypothesis: The wider berth predicted by the racial avoidance hypothesis will be more pronounced in white pedestrians as compared to non-white pedestrians.
    a. In other words:
        i. $ATE_{\text{white pedestrians}}$ = ($\text{distance}_{\text{white pedestrians}}$ from location of confederates [black confederates present] - $\text{distance}_{\text{white pedestrians}}$ from the baseline object [black confederates present]) > ($\text{distance}_{\text{white pedestrians}}$ from location of confederates [white confederates present] - $\text{distance}_{\text{white pedestrians}}$ from the baseline object [white confederates present])
        ii. $ATE_{\text{non-white pedestrians}}$ = ($\text{distance}_{\text{non-white pedestrians}}$ from location of confederates [black confederates present] - $\text{distance}_{\text{non-white pedestrians}}$ from the baseline object [black confederates present]) > ($\text{distance}_{\text{non-white pedestrians}}$ from location of confederates [white confederates present] - $\text{distance}_{\text{non-white pedestrians}}$ from the baseline object [white confederates present])
        iii. $ATE_{\text{white pedestrians}} > ATE_{\text{non-whitepedestrians}}$

5. Gender heterogeneity hypothesis: Female pedestrians (of all races) will tend to move further away from male confederates than the baseline object (e.g., a trashcan) as compared to male pedestrians.
    a. In other words:
        i. $ATE_{\text{female pedestrians}}$ = ($\text{distance}_{\text{female pedestrians}}$ from location of confederates [confederates present] - $\text{distance}_{\text{female pedestrians}}$ from the baseline object [confederates present]) > ($\text{distance}_{\text{female pedestrians}}$ from location of

confederates [confederates not present] - $distance_{female\ pedestrians}$ from the baseline object [confederates not present])

ii. $ATE_{male\ pedestrians}$ = ($distance_{male\ pedestrians}$ from location of confederates [confederates present] - $distance_{male\ pedestrians}$ from the baseline object [confederates present]) > ($distance_{male\ pedestrians}$ from location of confederates [confederates not present] - $distance_{male\ pedestrians}$ from the baseline object [confederates not present])

iii. $ATE_{female\ pedestrians} > ATE_{male\ pedestrians}$

---

[1] One exception is Willis et al.'s (1979) observational study of pedestrian displacements, in which whites are found to be more likely to give way to Black individuals. The authors posit that this is related to prior observations that whites tend to stand back from blacks (Willis 1966), or that it may be related to socio-political factors. Another important limitation of prior studies is that they are almost always purely observational, though exceptions include Dabbs and Stokes (1975) and Sobel and Lillith (1975). In the latter, experimenters walked directly at oncoming pedestrians and observed collision avoidance behavior. In the former, Dabbs and Stokes (1975) used confederates on a public sidewalk, and observed that pedestrians more often changed their paths to avoid a male standing beside a sidewalk than a female, and changed more often for two persons than for one, and more for an attractive female than for an unattractive one.The authors conclude that pedestrian behavior is dictated by various aspects of social power. Moreover, these studies are now decades old– gender and race relations have changed, as have pedestrian behavior with the advent of smartphones.

References

Bernardin, Keni and Rainer Stiefelhagen. 2008. "Evaluating multiple object tracking performance: the CLEAR MOT metrics." *EURASIP Journal on Image and Video Processing* 2008:1–10.

Cohen, Jacob. 1960. "A coefficient of agreement for nominal scales." *Educational and psychological measurement* 20(1):37–46.

Collett, Peter and Peter Marsh. 1981. "Patterns of public behavior: Collision avoidance on a pedestrian crossing." *Nonverbal communication, interaction and gesture. Selections from Semiotica* pp. 199–217.

Dabbs Jr, James M and Neil A Stokes III. 1975. "Beauty is power: The use of space on the sidewalk." *Sociometry* pp. 551–557.

Dendorfer, P., H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler and L. Leal-Taixé. 2020. "MOT20: A benchmark for multi object tracking in crowded scenes." *arXiv:2003.09003[cs]* . arXiv: 2003.09003.
URL: *http://arxiv.org/abs/1906.04567*

Gérin-Lajoie, Martin, Carol L Richards and Bradford J McFadyen. 2005. "The negotiation of stationary and moving obstructions during walking: anticipatory locomotor adaptations and preservation of personal space." *Motor control* 9(3):242–269.

Haghani, Milad and Majid Sarvi. 2018. "Crowd behaviour and motion: Empirical methods." *Transportation research part B: methodological* 107:253–294.

Koo, Terry K and Mae Y Li. 2016. "A guideline of selecting and reporting intraclass correlation coefficients for reliability research." *Journal of chiropractic medicine* 15(2):155–163.

Lombard, Matthew, Jennifer Snyder-Duch and Cheryl Campanella Bracken. 2002. "Content analysis in mass communication: Assessment and reporting of intercoder reliability." *Human communication research* 28(4):587–604.

Luo, Wenhan, Junliang Xing, Anton Milan, Xiaoqin Zhang, Wei Liu, Xiaowei Zhao and Tae-Kyun Kim. 2014. "Multiple object tracking: A literature review." *arXiv preprint arXiv:1409.7618* .

Moussaïd, Mehdi, Dirk Helbing and Guy Theraulaz. 2011. "How simple rules determine pedestrian behavior and crowd disasters." *Proceedings of the National Academy of Sciences* 108(17):6884–6888.

Pinheiro, José and Douglas Bates. 2006. *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.

Ristani, Ergys, Francesco Solera, Roger Zou, Rita Cucchiara and Carlo Tomasi. 2016. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision*. Springer pp. 17–35.

Silveira, Jeanette. 1972. "Thoughts on the Politics of Touch." *Women's Press* 1(13):13.

Sobel, Robert S and Nancy Lillith. 1975. "Determinants of nonstationary personal space invasion." *The Journal of Social Psychology* 97(1):39–45.

Willis, Frank N, Joseph A Gier and David E Smith. 1979. "Stepping aside: Correlates of displacement in pedestrians." *Journal of communication* 29(4):34–39.