Supplemental Information

Simulation-Based Behavioral Experiments on Active Representation

Section 1: Experimental Task

Webpages used for experiment (white male client condition)

You have just been hired as the new Director of the Valley Men's Health which is part of Men's Health Network. As is apparent in the testimonials of our patients, our system of 5 clinics provides high quality health services for disadvantaged men throughout the city. Each clinic provides services to their local communities, but the network considers over all system-wide demand for service.

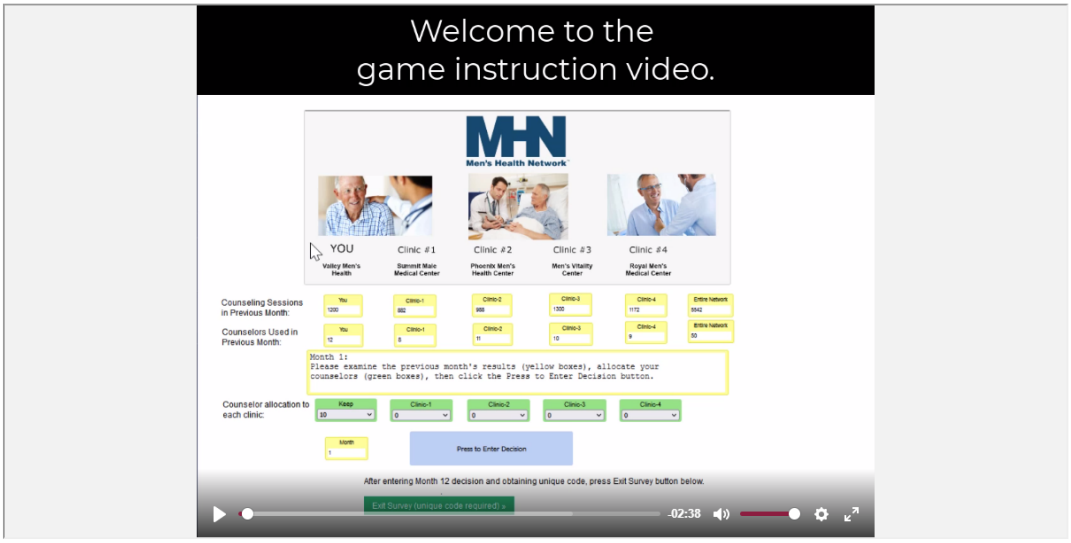
The services provided by the network include mental health counseling. As Director, each month you will be required to allocate your 10 full time counselors to work either at your facility or other facilities in the network **Your objective is to use your 10 counselors to generate as many services as possible for the network as a whole, while providing at least some services for your unit.** The directors of the other facilities in the network will also face the same decision for their 10 counselors.

Press the Next button below to continue.

« Previous Next »

In each month of the game you will be given information on the number of counselors used and sessions conducted by each health center and the entire network in the previous month. You can use that information to make your own allocation decision for the current month. Once you have completed 12 monthly decisions, you will be asked to complete a short questionnaire.

The video below illustrates how to allocate your 10 counselors to your own and other clinics each month, as well as how to view the previous month's counselor and session information. **Press play on the video to watch it. When it is complete, you will be able to start the game by pressing the Next button below the video.**



« Previous Next »

Counseling Sessions in Previous Month:

You	Clinic-1	Clinic-2	Clinic-3	Clinic-4	Entire Network
1102	1690	720	1559	657	5728

Counselors Used in Previous Month:

You	Clinic-1	Clinic-2	Clinic-3	Clinic-4	Entire Network
11	13	8	12	6	50

Month 1:
Please examine the previous month's results (yellow boxes), allocate your counselors (green boxes), then click the Press to Enter Decision button.

Counselor allocation to each clinic:

Keep	Clinic-1	Clinic-2	Clinic-3	Clinic-4
10	0	0	0	0

Month: 1

Press to Enter Decision

After entering Month 12 decision and obtaining unique code, press Exit Survey button below.

Exit Survey (unique code required) »

Section 2: Sample Size and Data Collection

As the basis of our target sample size calculation, we used the maximum number of identity groups required for testing our hypotheses. We estimated the error variance to use for our calculations with a two-way ANOVA of the sum gives variable using a sample of 174 valid cases from pre-experiment pilots, and assuming 8 cells of equal size (4 identity groups X matched/unmatched condition). Further assuming 80% statistical power, a 5% level of statistical significance, and an effect size of 0.1 -- which Cohen (1988) characterizes as "small" -- resulted in a target sample size of 792. Assuming 80% statistical power and a smaller effect size of 0.0625 approximated from the pilot data resulted in a target sample size of 2016. Using the more conservative assumption of 90% power and an effect size of 0.0625 resulted in target sample size of 2688. We inflated this number by 20% to account for rejected cases due to data integrity checks, leading us to a recruitment target of 3226 participants.

We initiated the recruitment of an equal number of white and black participants using the online crowdsourcing platform Mechanical Turk (MTurk). To facilitate recruitment, we used MTurk Toolkit, a platform provided by a third-party vendor CloudResearch that assists in conducting research on MTurk (Litman et al., 2017). MTurk Toolkit grants greater control over the recruitment of subjects on MTurk, such as exclusion of participants from previous tasks, managing bonus payments, and making changes to the study while it is running. MTurk Toolkit also provides access to CloudResearch's "approved group", which is an aggregated list of MTurk participants over time that have been recognized to be high quality participants in studies using MTurk (Hauser et al., 2022). In addition to the use of CloudResearch's "approved group", we filtered out participants with unfavorable histories on MTurk Human Intelligence Tasks and also limited participation to individuals residing in the United States (Aguinis et al., 2021; Goodman et al., 2013; Paolacci et al.,2010).[1]

After the data were collected, we conducted several integrity checks. First, we verified that the IP addresses of the participants were not virtual addresses or outside the US. Second, we removed participants associated with a duplicate IP address, treating the duplication as evidence of an individual attempting to participate twice from the same computer. Third, we included an attention checker in the survey that asked, "Please indicate how often you do each of the following activities", one of which was "how often do you eat cement?" Any participant responding with an answer other than the 'never' option was excluded. Fourth, we excluded any participant that completed the first round of the simulation in less than 20 seconds (less than 5% of participants). We deemed completion in that time as impossible and took such a fast response as evidence of a participant just clicking through to completion as quickly as possible. Fifth, we excluded 15 cases where the participant reported difficulties in understanding or executing the task in response to an open-ended survey question about their decision-making. Finally, we excluded any participant that did not self-identify as male or female and either white or black. Applying this final filter was not a data integrity or quality issue, but instead necessary given the importance in our experiment of having an unambiguous match between the race and gender of the participant and the clients in the experimental condition.

Obtaining participation by black participants proved to be much more difficult than collection of white participants, despite extending the period of collection for black participants by multiple weeks and increasing the level of base compensation. Consequently, we were not able to obtain our target number of black participants. After applying all the data suitability filters above, our final sample was comprised of 404 black women (13.6%), 204 black men (6.9%), 1405 white women (47.3%), and 955 white men (32.2%).

Since the number of black participants was smaller than we targeted, to help put the power of our analysis into further context, we calculated the minimum effect that we would have been able to detect given the final sample we collected. Table S1 shows the estimates assuming a 5% of significance level and 80% power, broken out by hypothesis. Our calculations indicate that for Hypotheses 1, which test for matching effects for either race or gender across all participants, our analysis could have detected an effect as small as 8% of the mean sum gives (5.31 sum gives for gender, 5.25 sum gives for race, equivalent to sharing a little less than 1

[1] When workers perform a Human Intelligence Task (HIT) on MTurk (HIT), the recruiters are given the option to approve the task performed by the worker to ensure payment is made to only workers that carried out the task demanded by the recruiter. Number of HITs that have been approved as well as the approval rate of HITs provides a track record of the quality of the participant in terms of the participant's ability to accurately understand and perform the task given. We filtered for participants with 50 or more total approved HITs and an approval rate of 70% or greater.

person every other round). For Hypothesis 4, which tests the effect of matching on both race and gender across all participants, the minimum detectable effect is 2 times as large (equivalent to sharing a little less than 1 person every round). For Hypothesis 2, which tests whether such matching effects differ in intensity between black and white participants, the minimum detectable effect is 2.5 times as large (equivalent to sharing a little more than 1 person every round). For Hypothesis 3, which tests whether such matching effects differ in intensity between male and female participants, the minimum detectable effect is 2 times as large (equivalent to sharing a little less than 1 person every round).

Table S1. Minimal Detectable Effect Size for Defined Hypotheses

		Minimal Detectable Effect Size (β)							
		Sum of gives (Mean = 65.5, Std. Dev. = 36.4)		Mean Gives (Mean = 5.5, Std. Dev. = 3.0)		Max Gives (Mean = 6.8, Std. Dev. = 2.7)		Last Gives (Mean = 5.6, Std. Dev. = 3.4)	
Hypothesis	Variable	Value	% of Mean	Value	% of Mean	Value	% of Mean	Value	% of Mean
H1	Gender Match (Model 1)	5.31	8%	0.44	8%	0.39	6%	0.49	9%
	Race Match (Model 1)	5.25	8%	0.44	8%	0.39	6%	0.48	9%
H4	Both Match (Model 1)	10.28	16%	0.62	11%	0.55	8%	0.69	12%
H2	Race Match * Black (Model 2)	13.60	21%	1.07	19%	0.95	14%	1.18	21%
H3	Gender Match * Female (Model 3)	10.90	17%	0.91	16%	0.80	12%	1.01	18%

Effect size is measured as model coefficients. Estimations assuming 5% of significance level and 80% of power.

Section 3: Supplemental Results for Alternative Giving Measures

Table S2. Models 1-4 from Main Text Estimated Using Different Mean Gives, Max Gives, and Last Give Measures as the Dependent Variable. (N=2968)

	Mean Gives (Mean = 5.5, Std. Dev. = 3.0)				Max Gives (Mean = 6.8, Std. Dev. = 2.7)				Last Gives (Mean = 5.6, Std. Dev. = 3.4)			
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Gender Match	0.027	0.098	-0.123		-0.100	-0.047	-0.157		0.059	0.086	-0.105	
	(0.158)	(0.177)	(0.254)		(0.140)	(0.157)	(0.225)		(0.175)	(0.196)	(0.281)	
Race Match	0.154	0.240	-0.027		0.062	0.106	-0.144		0.148	0.244	0.066	
	(0.156)	(0.176)	(0.248)		(0.138)	(0.156)	(0.220)		(0.173)	(0.194)	(0.275)	
Both Match	-0.148	-0.090	0.221	0.114	-0.005	0.009	0.201	0.018	-0.150	-0.059	0.085	0.128
	(0.223)	(0.250)	(0.356)	(0.140)	(0.197)	(0.221)	(0.316)	(0.124)	(0.246)	(0.276)	(0.394)	(0.154)
Black		0.230				0.280				0.027		
		(0.270)				(0.240)				(0.298)		
Gender Match * Black		-0.345				-0.237				-0.137		
		(0.392)				(0.348)				(0.434)		
Race Match * Black		-0.405				-0.200				-0.467		
		(0.382)				(0.339)				(0.422)		
Both Match * Black		-0.288				-0.090				-0.440		
		(0.552)				(0.490)				(0.610)		
Female			-0.412*				-0.408**				-0.417*	
			(0.227)				(0.201)				(0.251)	
Gender Match * Female			0.259				0.112				0.283	
			(0.324)				(0.287)				(0.359)	
Race Match * Female			0.306				0.347				0.144	
			(0.319)				(0.283)				(0.353)	
Both Match * Female			-0.628				-0.362				-0.407	
			(0.456)				(0.404)				(0.505)	
Both Match * Black Female				-0.976***				-0.346				-0.942**
				(0.370)				(0.328)				(0.409)
Black Female				-0.079				0.035				-0.247

				(0.188)				(0.167)				(0.208)
Constant	5.405***	5.355***	5.651***	5.477***	6.837***	6.776***	7.080***	6.820***	5.544***	5.538***	5.793***	5.647***
	(0.111)	(0.126)	(0.175)	(0.069)	(0.099)	(0.112)	(0.155)	(0.061)	(0.123)	(0.139)	(0.194)	(0.076)

* p<0.1, ** p<0.05, *** p<0.01 (two-tailed test). Standard errors in parentheses.

Section 4: Supplemental Results Controlling by Participant and Client Age

Table S3. Models 1-4 Controlling by Covariates. Results for Sum Gives and Mean Gives measures as the dependent variable. (N=2968).

	Sum Gives (Mean = 65.5, Std. Dev. = 36.4)				Mean Gives (Mean = 5.5, Std. Dev. = 3.0)			
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Gender Match	0.569	0.779	-2.788		0.047	0.065	-0.232	
	(1.904)	(2.159)	(3.137)		(0.159)	(0.180)	(0.261)	
Race Match	0.822	5.196	-2.665		0.069	0.433	-0.222	
	(2.023)	(3.322)	(3.264)		(0.169)	(0.277)	(0.272)	
Both Match	-1.982	-0.600	3.984	1.011	-0.165	-0.050	0.332	0.084
	(2.677)	(3.033)	(4.340)	(1.891)	(0.223)	(0.253)	(0.362)	(0.158)
Black		4.586				0.382		
		(4.141)				(0.345)		
Gender Match * Black		-3.411				-0.284		
		(4.743)				(0.395)		
Race Match * Black		-9.289				-0.774		
		(6.780)				(0.565)		
Both Match * Black		-4.908				-0.409		
		(6.755)				(0.563)		
Female			-6.351**				-0.529**	
			(2.838)				(0.236)	
Gender Match * Female			5.807				0.484	
			(4.184)				(0.349)	
Race Match * Female			5.249				0.437	
			(3.927)				(0.327)	
Both Match * Female			-10.157*				-0.846*	
			(5.670)				(0.473)	
Both Match * Black Female				-10.671**				-0.889**
				(5.064)				(0.422)
Black Female				-1.519				-0.127
				(2.354)				(0.196)
Participant's age	-0.049	-0.067	-0.050	-0.065	-0.004	-0.006	-0.004	-0.005
	(0.054)	(0.055)	(0.054)	(0.055)	(0.005)	(0.005)	(0.005)	(0.005)
Client age[1]	0.068	-0.090	0.093*	0.021	0.006	-0.008	0.008*	0.002
	(0.050)	(0.099)	(0.054)	(0.050)	(0.004)	(0.008)	(0.004)	(0.004)
Constant	63.7***	70.9***	66.31***	67.4***	5.3***	5.9***	5.5***	5.6***
	(3.541)	(5.065)	(3.727)	(3.516)	(0.295)	(0.422)	(0.311)	(0.293)

* p<0.1, ** p<0.05, *** p<0.01 (two-tailed test). Standard errors in parentheses.
[1]Client age is measured as the average age of the clients depicted in the testimonials in the experimental condition to which the participant was assigned.

Table S4. Models 1-4 Controlling by Covariates. Results for Max Gives, and Last Give Measures as the Dependent Variable. (N=2968)

	Max Gives (Mean = 6.8, Std. Dev. = 2.7)				Last Gives (Mean = 5.6, Std. Dev. = 3.4)			
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Gender Match	-0.093	-0.083	-0.196		0.084	0.043	-0.236	
	(0.141)	(0.160)	(0.232)		(0.175)	(0.199)	(0.289)	
Race Match	0.040	0.291	-0.213		0.043	0.499	-0.166	
	(0.149)	(0.245)	(0.241)		(0.186)	(0.306)	(0.301)	
Both Match	-0.011	0.052	0.245	-0.006	-0.170	-0.007	0.217	0.081
	(0.198)	(0.224)	(0.320)	(0.140)	(0.247)	(0.279)	(0.400)	(0.174)
Black		0.396				0.242		
		(0.306)				(0.381)		
Gender Match * Black		-0.164				-0.064		
		(0.350)				(0.437)		
Race Match * Black		-0.549				-0.957		
		(0.501)				(0.624)		
Both Match * Black		-0.222				-0.591		
		(0.499)				(0.622)		
Female			-0.451**				-0.557**	
			(0.209)				(0.262)	
Gender Match * Female			0.192				0.551	
			(0.309)				(0.386)	
Race Match * Female			0.405				0.298	
			(0.290)				(0.362)	
Both Match * Female			-0.447				-0.666	
			(0.419)				(0.523)	
Both Match * Black Female				-0.271				-0.810*
				(0.374)				(0.467)
Black Female				-0.031				-0.302
				(0.174)				(0.217)
Participant's age	-0.01**	-0.01**	-0.01**	-0.01**	-0.003	-0.005	-0.003	-0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Client age[1]	0.002	-0.007	0.002	0.001	0.007	-0.010	0.009*	0.003
	(0.004)	(0.007)	(0.004)	(0.004)	(0.005)	(0.009)	(0.005)	(0.005)

Constant	7.157***	7.493***	7.387***	7.164***	5.325***	6.184***	5.544***	5.704***
	(0.261)	(0.374)	(0.275)	(0.260)	(0.326)	(0.466)	(0.343)	(0.324)

* p<0.1, ** p<0.05, *** p<0.01 (two-tailed test). Standard errors in parentheses.

[1]Client age is measured as the average age of the clients depicted in the testimonials in the experimental condition to which the participant was assigned.

References

Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk research: Review and recommendations. *Journal of Management*, 47(4), 823-837.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.) Lawrence Erlbaum Associates, publishers.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, 26(3), 213-224.

Hauser, D. J., Moss, A. J., Rosenzweig, C., Jaffe, S. N., Robinson, J., & Litman, L. (2022). Evaluating CloudResearch's Approved Group as a solution for problematic data quality on MTurk. *Behavior Research Methods*, 1-12.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 49(2), 433-442.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 411-419.