Science
AAAS

Supplementary Materials for

Public views on polygenic screening of embryos

Michelle N. Meyer *et al.*

Corresponding author: Michelle N. Meyer, michellenmeyer@gmail.com

Science **379**, 541 (2023)
DOI: 10.1126/science.ade1083

The PDF file includes:

Materials and Methods

Survey methods and materials

Our survey was fielded in the January 2022 wave of the Understanding America Study (UAS). The UAS, based at the University of Southern California, is a panel of households that is nationally representative of the United States (Table S1). The UAS survey was approved by the BRANY SBER IRB (#194604). The study is an "Internet Panel," meaning that respondents answer the surveys on a computer, tablet, or smartphone, wherever and whenever they wish to participate. Our survey was fielded among English speakers during December 17, 2021–February 10, 2022, and among Spanish speakers during January 4, 2022–February 10, 2022.

In our component of the UAS survey, each respondent was first asked about their view of the morality of IVF (IVF morality) and then randomly assigned with equal probability to one of three services: SAT preparation courses ("SAT prep"), polygenic embryo selection ("PGT-P"), or gene editing. Within each of these groups, they were asked two questions: one about their view of the moral acceptability of the service ("service moral acceptability") and one about their likelihood of using the service themselves, given certain assumptions ("willingness to use"). Similar to previous research (*16*), participants could give one of four responses to service moral acceptability: morally acceptable, not a moral issue, morally wrong, or unsure.

For the willingness to use question, participants were asked to assume that using the free service would increase their chances of having a child who attends a top-100 college by two percentage points (from 3% to 5%). Willingness-to-use was then measured using a slider via which participants indicated the percentage chance they would use the service, from 0% to 100%. The order of the service, ethical acceptability, and willingness-to-use questions was randomized with equal probability for each possible order. Within the willingness-to-use question, respondents were randomly assigned with equal probability to a condition in which they are asked to assume that 90% of the relevant population uses the service ("willingness to use: 90% social uptake") or to a condition in which they are asked to assume that 10% of the population uses the service ("willingness to use: 10% social uptake"). Full survey materials, preregistration, and code are available at https://osf.io/krj8t/.

Data analyses

Our preregistration contains our power calculations. We anticipated a sample size of 7600, which is 8% larger than the actual number of respondents (n = 7024). Furthermore, in addition to differential response rates from various groups, the UAS oversamples Los Angeles County households and Native American individuals. Therefore, in order to obtain results that are based on a nationally representative sample, we used the sampling weights (also called probability weights) provided by the UAS in all of the analyses described in the main text and through Table S8, below. As part of the UAS weighting strategy, 201 individuals are given zero weight, meaning that the main results of this paper are based on 6823 individuals (SAT prep, n = 2294; PGT-P, n = 2216; gene editing n = 2313). Using these weights further reduced our *effective* sample size to 3805. (Effective sample size is the number of unweighted individuals that would be required to obtain equally powered analyses. To calculate effective sample size, we use the formula

$$N_{eff} = \frac{(\sum w_i)^2}{\sum (w_i)^2}$$

where w_i is the weight assigned to individual i in our sample.)

Using these sampling weights is a deviation from our preregistered analysis plan (we wrote the analysis plan before we received the data, which we thought at the time was already representative); the main results based on the unweighted sample can be found in Tables S1 and Tables S9-13. The results for the unweighted sample have similar point estimates compared to the results based on the weighted sample, although which of these estimates are statistically significant at $p < 0.05$ changes in a few cases. Other than giving some participants a weight of zero (as described above), we did not exclude any participants from any of our analyses.

For our descriptive analyses, for each service, we calculated the proportion of participants who responded that the service was either "morally acceptable" or "not a moral issue," the average reported likelihood of using the service, and the proportion of individuals who responded that their likelihood of using the service was greater than 50% (Table S2). Full distributions of participants' likelihood of using each service, from 0-100%, are shown in Figure S2.

We preregistered one hypothesis: in all three service domains, we would find a social norm effect such that the mean willingness to use among those randomized to the 90% social uptake condition would be greater than the mean willingness to use among those randomized to the 10% social uptake condition. To determine whether respondents' likelihood to use each service varied significantly between the 90% and 10% social uptake groups, we created a dummy variable which was coded as 1 if the respondent was randomized into the 90% social uptake group and 0 if the 10% social uptake group. We split the sample by service and regressed the likelihood to use the service on this dummy variable, for each of the three services. The resulting coefficient on the dummy variable represents the difference between the average likelihood to use the service in the 90% social uptake group and the average likelihood to use the service in the 10% social uptake group, which we refer to as the "90-10 difference" for each service. For this preregistered hypothesis, we used a one-tailed test (Table S3).

We conducted two exploratory analyses, one about heterogeneity by age and one about heterogeneity by educational attainment. Specifically, for each service, we used two-tailed tests to determine whether the mean reported likelihood of using the service and/or the proportion of individuals who responded that the service was either "morally acceptable" or "not a moral issue" differed by age (Table S4) and/or educational attainment (Table S5). We next describe these two analyses in more detail.

To determine how service moral acceptability and willingness to use each service varied with participants' age, we split the sample into individuals under and at least 35 years old. We chose 35 as the age cutoff in order to facilitate comparison between our results and those found in a recent study commissioned by the Progress Educational Trust in the UK, which used this age cutoff (*6*). Within each service grouping, we calculated the proportion of individuals who responded that the service was either "morally acceptable" or "not a moral issue" for the under-35 age group and the 35-and-over age group. We also calculated the mean reported likelihood of using the the service for the under-35 age group and compared this to the result for the full sample. Because these two statistics are based on partially overlapping samples, this sample overlap must be taken into account in a test comparing the full sample to the under-35 sample. In the subsequent section of this document, we prove that a test comparing the under-35 sample to the full sample is statistically equivalent to a test comparing the under-35 sample to the 35-and-older sample. We therefore report p values corresponding to the latter test based on nonoverlapping groups. Statistics for the full sample and for the age-stratified subsamples are reported in Table S4.

Following the suggestion of a reviewer, we also conducted a tertiary analysis in which we considered a finer partition of our age groups. Specifically, we were interested in whether the moral attitudes towards or willingness to use each service declines or increases monotonically with age or if it perhaps peaks at a time when Americans are most likely to have children. We therefore split the sample into four groups: under 30 (13% of the sample), 30-34 (11%), 35-55 (39%), and 56 and older (37%). This grouping was chosen by splitting the "Under 35" and "35 and Older" groups each approximately in half. We then evaluated the moral attitudes and willingness to use as described above in each of these groups. The results of these analyses are found in Figures S3 and S4 and Table S7.

With respect to moral acceptability, as was seen when comparing those under 35 to the full sample, the differences observed at this finer partition are also statistically indistinguishable. With respect to willingness to use, at a high level, those in the "30 to 34" group are slightly more likely to use each service than those in the "Under 30" group. However, the increase in willingness to use each service between the under 30 group and the 30-34 group is statistically very weak for all three services: the p value is 0.61 for gene editing, 0.65 for PGT-P, and 0.59 for SAT prep. Given the weak statistical evidence here, we do not believe any conclusions can be drawn about age-cohort differences at this fine a scale.

To determine whether service moral acceptability and willingness to use each service varied with participants' educational attainment (EA), we explored how both responses differed between participants with a bachelor's degree or above ("high EA") and participants with an associate degree or below ("low EA"). We chose associate degree as the education level cutoff for the groupings because this was the median education level of the sample. Within each service grouping, we split the sample by educational attainment and calculated the mean reported likelihood of using each service, and we calculated the proportion of participants who responded that the service was either "morally acceptable" or "not a moral issue."

Following the suggestion of a reviewer, we also conducted a tertiary analysis in which we considered a finer partition of our education groups. In these analyses, we split the data into four groups: high school graduate or less (38% of the sample), some college or Associates' degree (27%), Bachelor's degree (20%), or greater than a Bachelor's degree (15%). These divisions were meant to create groups of approximately equal size. We then evaluated the moral attitudes and willingness to use as described above in each of these groups. The results of these analyses are found in Figures S5 and S6 and Table S8.

From these figures, we still see a broad pattern of more educated groups reporting that PGT-P is morally acceptable or not a moral issue. This is largely driven by a decreasing number of respondents replying that they are "not sure" about the moral acceptability of PGT-P in more educated groups (the fraction reporting that it is morally unacceptable is roughly constant across all groups). With respect to willingness to use PGT-P, we observe a jump between the "some college/Associate's degree" group and the "Bachelor's degree" group in the mean reported willingness-to-use PGT-P under the assumed circumstances. There is no statistically significant difference between the two lower-education groups ($p = 0.17$) or between the two higher-education groups ($p = 0.77$).

<u>Equivalence of tests for age-stratified samples</u>

In some of our analyses, we compare the mean responses of the full sample to the subsample of individuals under 35 years of age. Because these samples partially overlap, this overlap must be accounted for in tests comparing the two groups. Here we show that a test

comparing these groups is equivalent to a test comparing the subsample of individuals under 35 and the subsample of individuals 35 and older.

First, we define $\hat{\mu}_{\text{full}}$, $\hat{\mu}_{\text{under}}$, and $\hat{\mu}_{\text{over}}$ as the estimated weighted-mean response to some survey question for the full sample, the under-35 sample, and the 35-and-older sample, respectively. This means the test statistic for comparing the full sample to the under-35 sample is

$$\frac{(\hat{\mu}_{\text{full}} - \hat{\mu}_{\text{under}})^2}{\text{Var}(\hat{\mu}_{\text{full}} - \hat{\mu}_{\text{under}})} \sim \chi^2(1).$$

The test statistics for comparing the under-35 sample to the 35-and-older sample is

$$\frac{(\hat{\mu}_{\text{over}} - \hat{\mu}_{\text{under}})^2}{\text{Var}(\hat{\mu}_{\text{over}} - \hat{\mu}_{\text{under}})} \sim \chi^2(1).$$

Notice that, because the under-35 sample and the 35-and-older sample are nonoverlapping, the square-root of this latter test corresponds to the conventional two-sample *t* test.

Next, we show that we can express $\hat{\mu}_{\text{full}}$ as a weighted average of $\hat{\mu}_{\text{under}}$ and $\hat{\mu}_{\text{over}}$. We define $x_{\text{under},i}$ as the survey response of individual i from the under-35 sample and define $w_{\text{under},i}$ as that person's survey weight. We define $x_{\text{over},i}$ and $w_{\text{over},i}$ similarly for a person from the 35-and-older sample. Then,

$$\begin{aligned}
\hat{\mu}_{\text{full}} &= \frac{\sum_i w_{\text{under},i} x_{\text{under},i} + \sum_i w_{\text{over},i} x_{\text{over},i}}{\sum_i w_{\text{under},i} + \sum_i w_{\text{over},i}} \\
&= \frac{\frac{\sum_i w_{\text{under},i} x_{\text{under},i}}{\sum_i w_{\text{under},i}} \sum_i w_{\text{under},i} + \frac{\sum_i w_{\text{over},i} x_{\text{over},i}}{\sum_i w_{\text{over},i}} \sum_i w_{\text{over},i}}{\sum_i w_{\text{under},i} + \sum_i w_{\text{over},i}} \\
&= \frac{\theta_{\text{under}} \hat{\mu}_{\text{under}} + \theta_{\text{over}} \hat{\mu}_{\text{over}}}{\theta_{\text{under}} + \theta_{\text{over}}},
\end{aligned}$$

where $\theta_{\text{under}} \equiv \sum_i w_{\text{under},i} / \left(\sum_i w_{\text{under},i} + \sum_i w_{\text{over},i}\right)$ and $\theta_{\text{over}} \equiv \sum_i w_{\text{over},i} / \left(\sum_i w_{\text{under},i} + \sum_i w_{\text{over},i}\right)$ are the weights that define the average.

Next, substituting this expression for $\hat{\mu}_{\text{full}}$ into $(\hat{\mu}_{\text{full}} - \hat{\mu}_{\text{under}})$, we obtain

$$\begin{aligned}
\hat{\mu}_{\text{full}} - \hat{\mu}_{\text{under}} &= \frac{\theta_{\text{under}} \hat{\mu}_{\text{under}} + \theta_{\text{over}} \hat{\mu}_{\text{over}}}{\theta_{\text{under}} + \theta_{\text{over}}} - \hat{\mu}_{\text{under}} \\
&= \frac{\theta_{\text{over}} \hat{\mu}_{\text{over}} - \theta_{\text{over}} \hat{\mu}_{\text{under}}}{\theta_{\text{under}} + \theta_{\text{over}}} \\
&= \frac{\theta_{\text{over}}}{\theta_{\text{under}} + \theta_{\text{over}}} (\hat{\mu}_{\text{over}} - \hat{\mu}_{\text{under}}).
\end{aligned}$$

Finally, substituting this expression into the test statistics for the full vs. under-35 samples, we obtain

$$\begin{aligned}
\frac{(\hat{\mu}_{\text{full}} - \hat{\mu}_{\text{under}})^2}{\text{Var}(\hat{\mu}_{\text{full}} - \hat{\mu}_{\text{under}})} &= \frac{\left[\frac{\theta_{\text{over}}}{\theta_{\text{under}} + \theta_{\text{over}}} (\hat{\mu}_{\text{over}} - \hat{\mu}_{\text{under}})\right]^2}{\text{Var}\left[\frac{\theta_{\text{over}}}{\theta_{\text{under}} + \theta_{\text{over}}} (\hat{\mu}_{\text{over}} - \hat{\mu}_{\text{under}})\right]} \\
&= \frac{\left(\frac{\theta_{\text{over}}}{\theta_{\text{under}} + \theta_{\text{over}}}\right)^2 (\hat{\mu}_{\text{over}} - \hat{\mu}_{\text{under}})^2}{\left(\frac{\theta_{\text{over}}}{\theta_{\text{under}} + \theta_{\text{over}}}\right)^2 \text{Var}(\hat{\mu}_{\text{over}} - \hat{\mu}_{\text{under}})} \\
&= \frac{(\hat{\mu}_{\text{over}} - \hat{\mu}_{\text{under}})^2}{\text{Var}(\hat{\mu}_{\text{over}} - \hat{\mu}_{\text{under}})},
\end{aligned}$$

which is the same as the test statistic for the under-35 vs 35-and-older comparison. Therefore, the two tests are statistically equivalent.

Calibration of the effect of PGT-P

In the survey administered for this research, respondents were told to assume that using gene editing, PGT-P, or SAT prep courses could raise the chances that their child would attend a top-100 ranked college or university from 3% to 5%. This assumption was based on the back-of-the-envelope approximation below.

We first looked up the number of children born in the US each year. In the mid 2000s—when today's high school students were born—there were approximately 4.1 million children born each year. Next, we approximated that the average entering class at a top-100 ranked college or university has an entering class of 1200 students, for a total of 120,000 students across all 100 schools. Taking the ratio of the total number of students and the total size of the birth cohort, we calculate that a random student from that birth cohort would have a 3% chance of attending a top-100 ranked college or university.

Next, we estimated the effect of PGT-P for a family selecting among 10 embryos using a PGI with a within-family R^2 of 4%. This estimate is consistent with the within-family R^2 estimated for a PGI predicting years of complete education in Okbay et al (2022). Applying this estimate to the approximation of Karavani et al. (2019), this would lead to an increase in the liability-scale for top-100 college attendance by 0.23 standard deviations. Transforming this to the observed scale, this implies that PGT-P would raise the probability of attending a top-100 ranked college or university from 3% to 5%. These are the numbers that we asked survey respondents to assume for the effect of PGT-P.

We are not aware of any data on the effect of SAT prep courses on attending a top-100 ranked college or university. We also do not have data on the effect of germline gene editing since such a service does not currently exist, and calculating an effect would be sensitive to speculative assumptions. However, we believe that the effect of PGT-P calculated above is not an unreasonable approximation for the effect of SAT prep, and we were interested in holding the effect of each service constant across participants for ease of interpreting and comparing results.

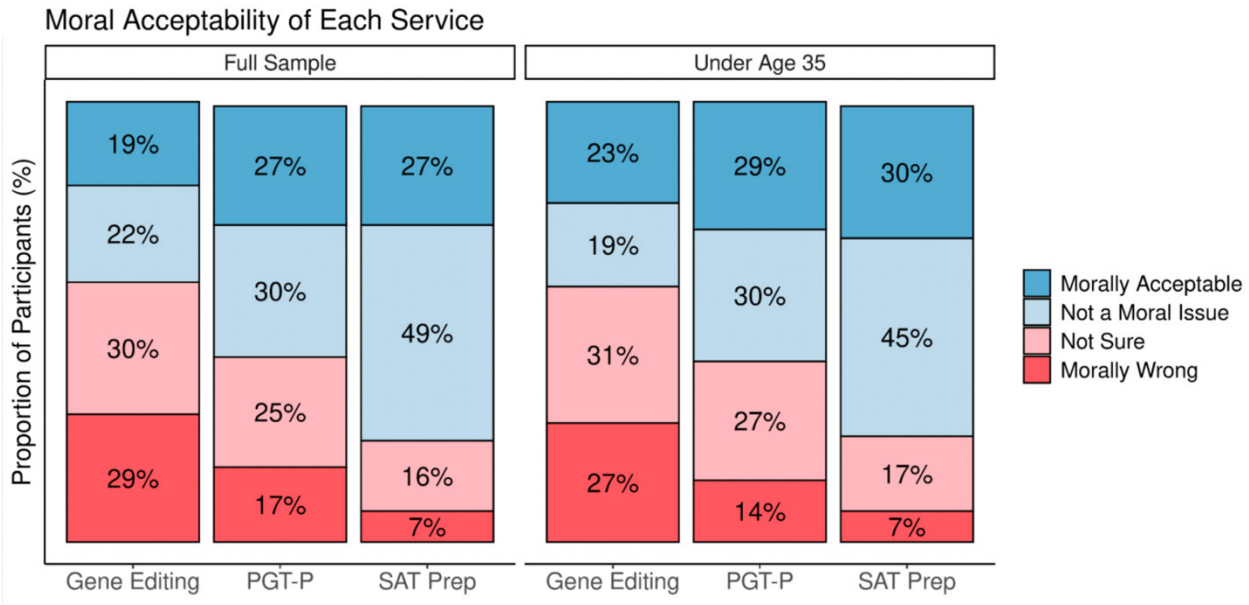

Figure S1. Moral acceptability of each service. Proportion of participants in the full sample (left) and among those under 35 years of age (right) who responded that each service was morally acceptable," "not a moral issue," or "morally wrong," or about which they were "not sure." Gene editing and PGT-P were described as being available for "medical and non-medical traits"; SAT prep was described as being used "to improve [high school students'] results on the SAT." Some bars do not sum to 100% due to rounding.

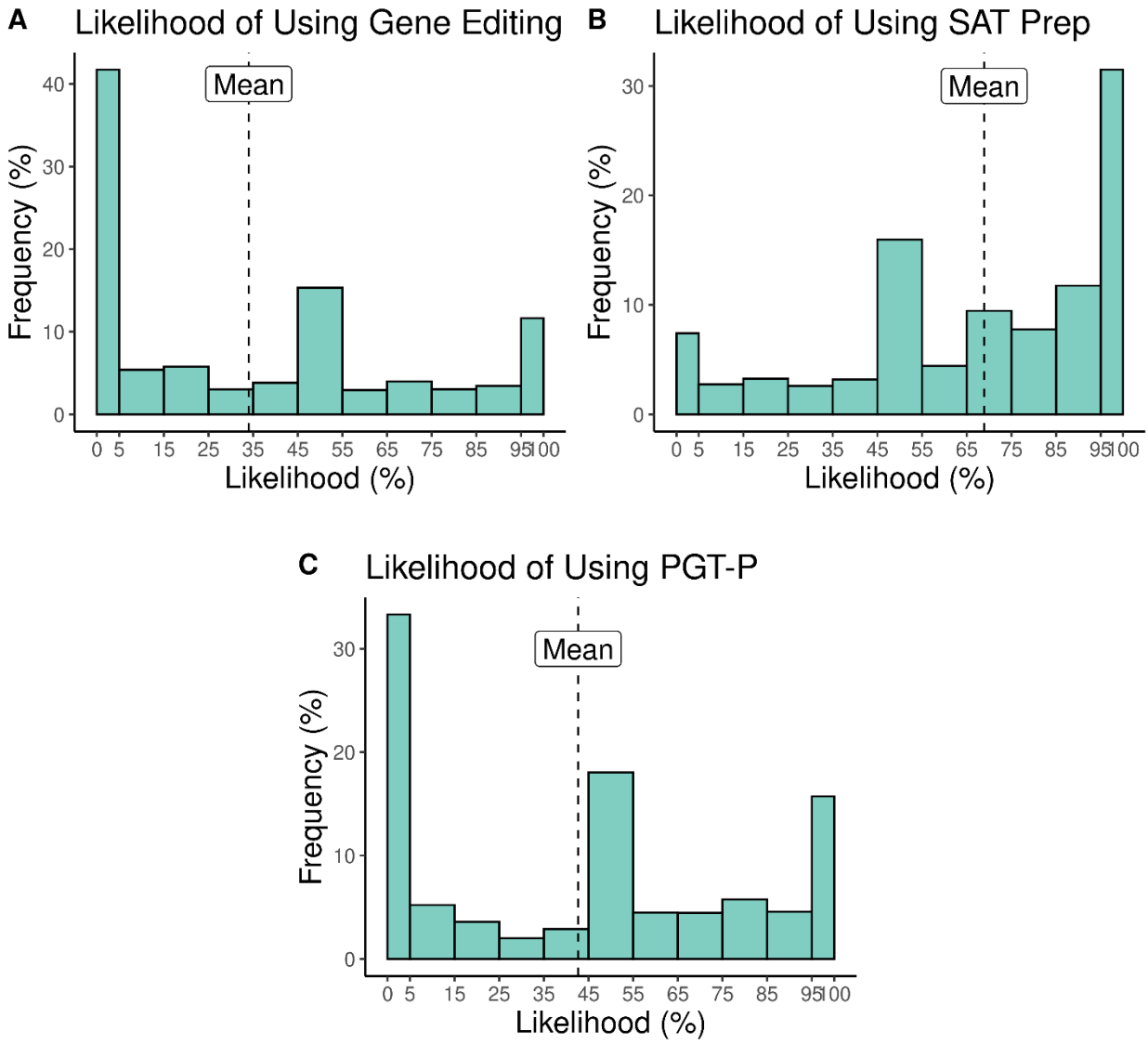

Figure S2. Distribution of participants' likelihood of using each service. Participants reported their likelihood of using each service as a probability from 0% to 100% chance. Dotted lines show the mean likelihood. Estimates in this figure are based on the US-representative sample using the sampling weights provided by the UAS (N = 6823, Effective N = 3805).

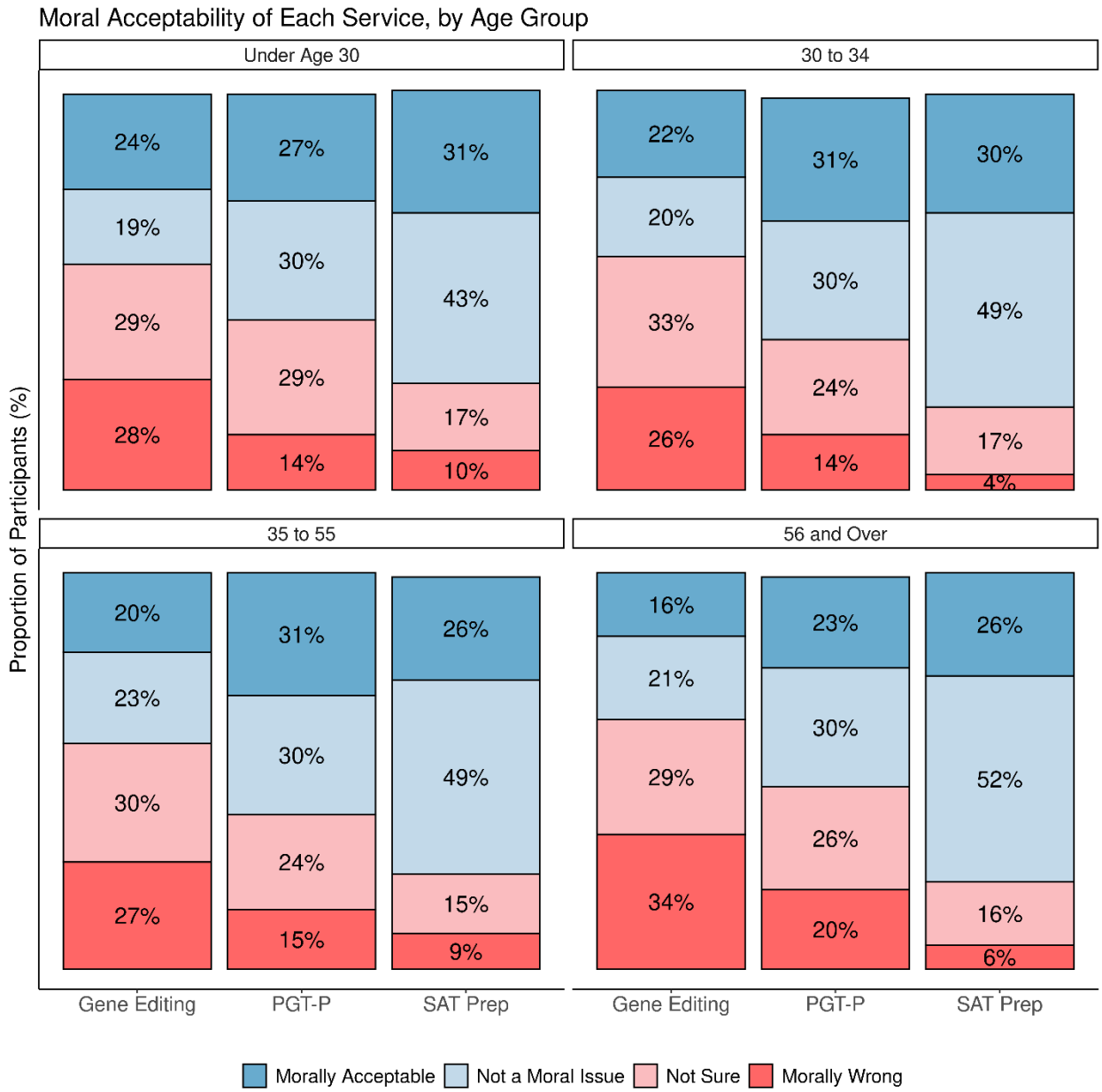

Figure S3. Moral acceptability of each service, by age group. Proportions of participants giving each of four responses to the moral acceptability question for each service. Standard errors for the estimates reported in this table are found in Table S7. Estimates in this figure are based on the US-representative sample using the sampling weights provided by the UAS (N = 6823, Effective N = 3805).

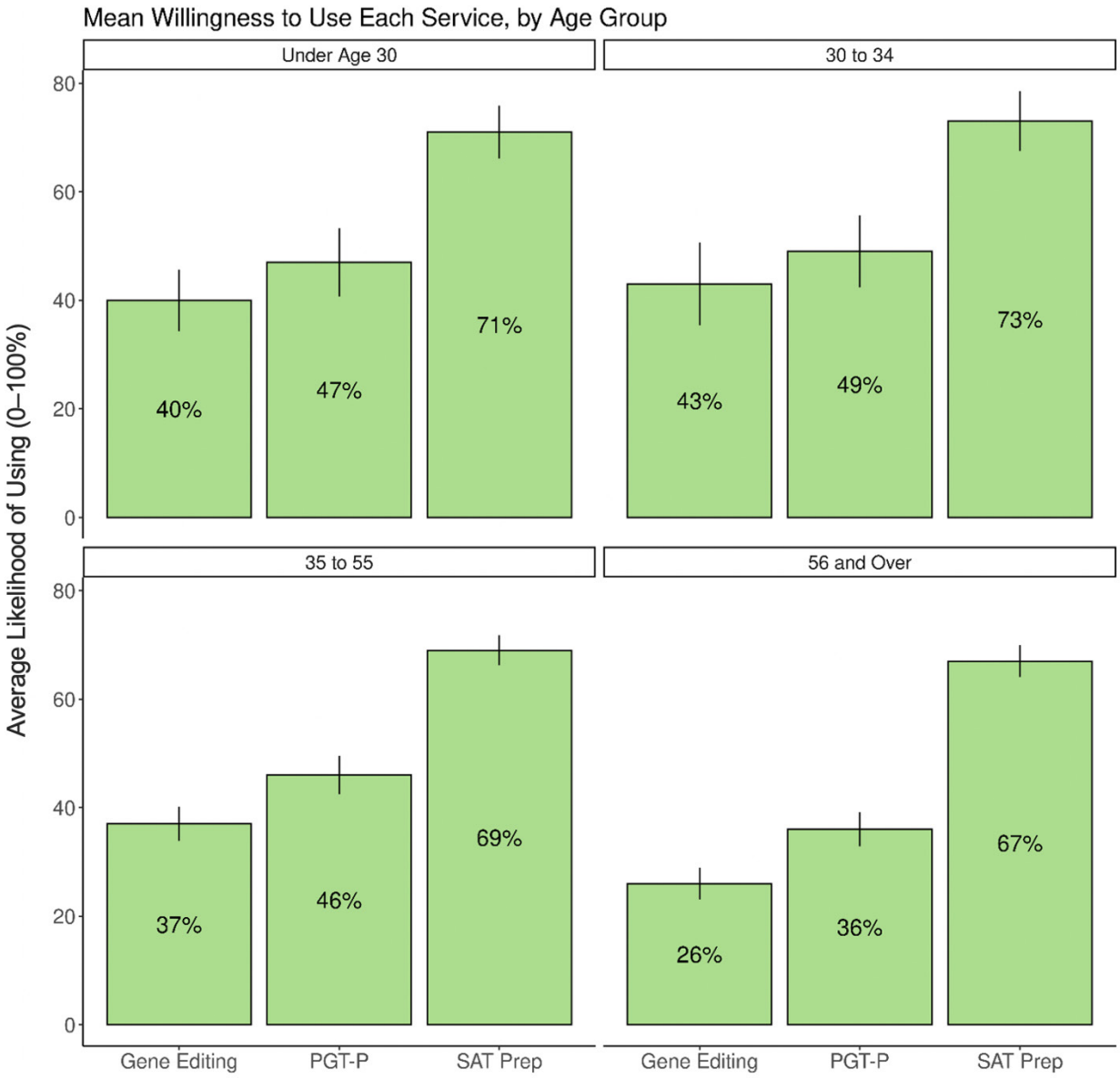

Figure S4. Mean willingness to use, by age group. Mean willingness to use each service. Standard errors for the estimates reported in this table are found in Table S7. Estimates in this table are based on the US-representative sample using the sampling weights provided by the UAS (N = 6823, Effective N = 3805).

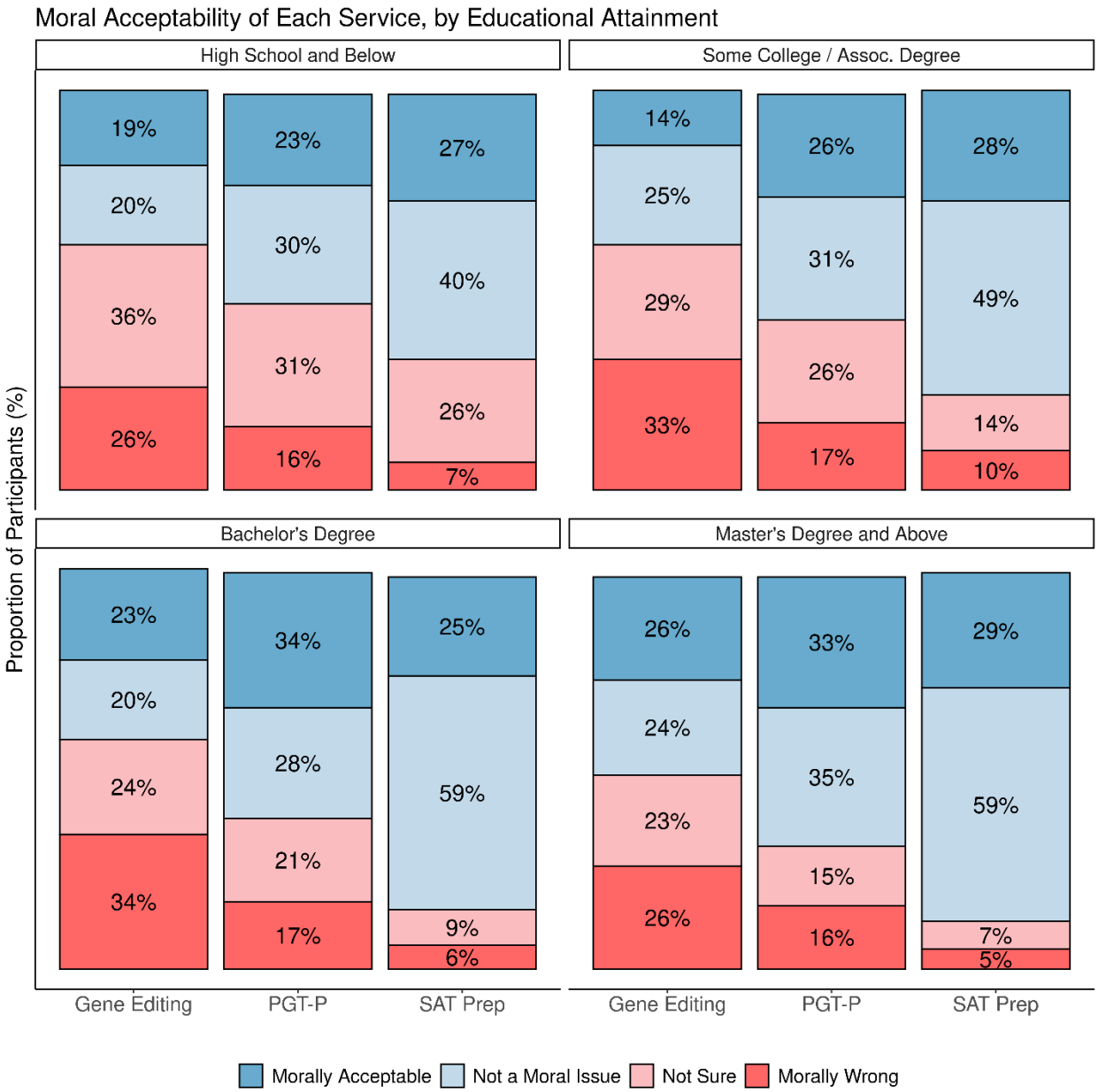

Figure S5. Moral acceptability of each service, by educational attainment. Proportions of participants giving each of four responses to the moral acceptability question for each service. Standard errors for the estimates reported in this table are found in Table S8. Estimates in this figure are based on the US-representative sample using the sampling weights provided by the UAS (N = 6823, Effective N = 3805).

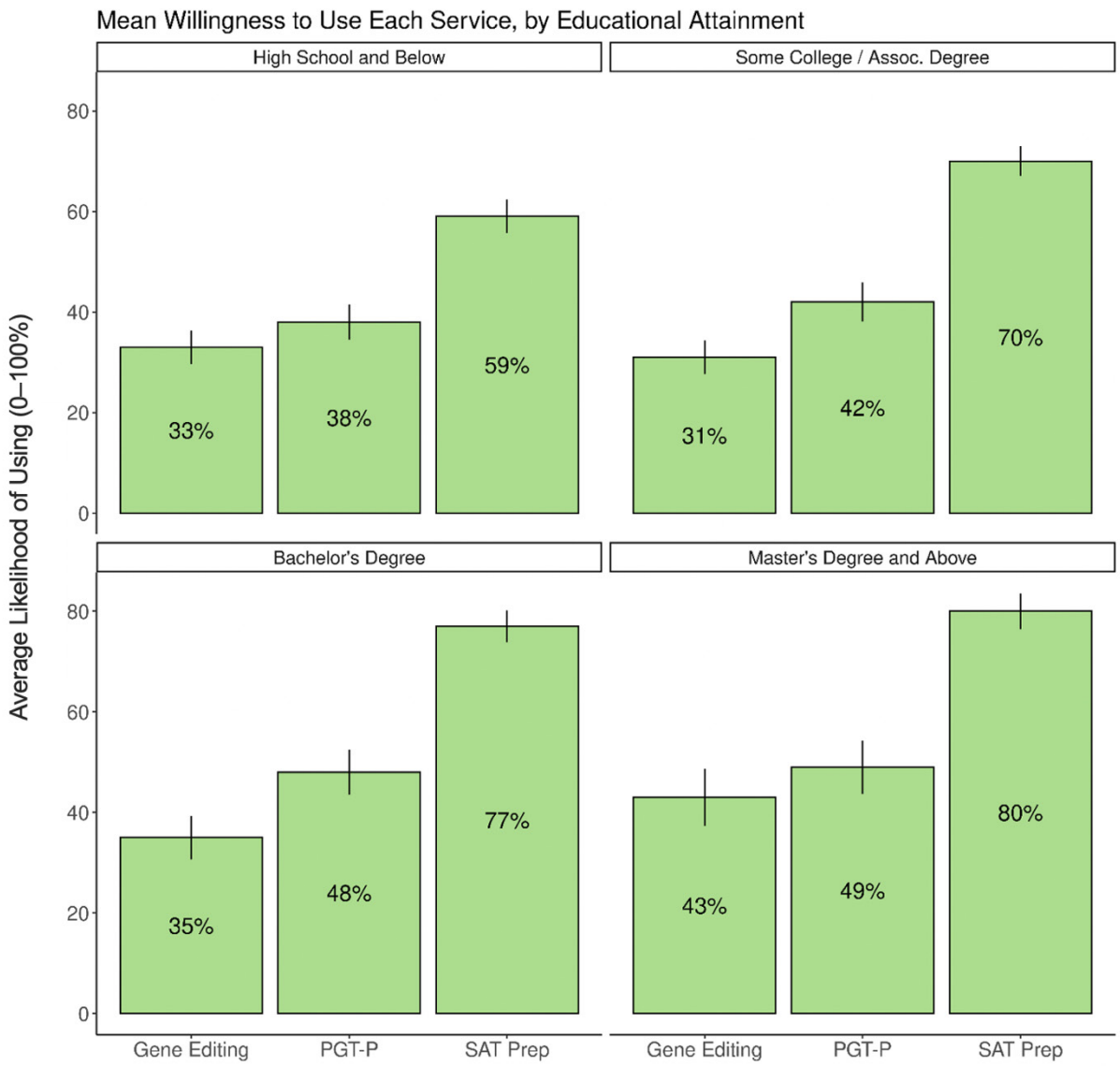

Figure S6. Mean willingness to use, by educational attainment. Mean willingness to use each service. Standard errors for the estimates reported in this table are found in Table S8. Estimates in this table are based on the US-representative sample using the sampling weights provided by the UAS (N = 6823, Effective N = 3805).

		Statistic (%)* (Representative)	Statistic (%)* (Unweighted)
Age	Minimum (years)	18	18
	Maximum (years)	111	111
	Mean (years)	48.9	51.8
	SD (years)	16.7	16.1
Sex	Male	48.3	40.5
	Female	51.7	59.5
Race	White	77.2	78.3
	Black or African American	12.8	8.1
	American Indian or Alaska Native	0.8	2.2
	Asian	5.1	5.5
	Native Hawaiian or Other Pacific Islander	0.2	0.6
	Two or more races	3.9	5.2
Ethnicity	Spanish, Hispanic, or Latino	16.9	15.8
	Not Spanish, Hispanic, or Latino	83.1	84.2
Marital Status	Married	54.4	55.9
	Separated	2.1	1.8
	Divorced	12.9	14.6
	Widowed	4.6	5.1
	Never Married	26.0	22.5
Household Income	Less than $5,000	5.2	3.8
	$5,000 to $7,499	2.0	1.5
	$7,500 to $9,999	2.0	1.7
	$10,000 to $12,499	2.7	2.3
	$12,500 to $14,999	2.6	2.1
	$15,000 to $19,999	3.9	3.4
	$20,000 to $24,999	4.9	4.4
	$25,000 to $29,999	4.7	4.2
	$30,000 to $34,999	4.9	4.7
	$35,000 to $39,999	4.8	4.6
	$40,000 to $49,999	6.9	6.8
	$50,000 to $59,999	7.7	7.9
	$60,000 to $74,999	11.3	11.0
	$75,000 to $99,999	12.5	13.8
	$100,000 to $149,000	13.3	14.9
	$150,000 or more	10.8	13.0
US Region	Midwest	20.6	22.4
	Northeast	17.2	10.8
	South	38.3	27.7
	West	23.9	39.1
Education	Less than high school graduate	8.2	4.7
	High school graduate or GED	30.2	16.0
	Some college - no degree	16.5	22.4
	Associate college degree	10.2	13.6
	Bachelor's degree	20.4	25.1
	Master's degree	10.8	13.9
	Professional school degree or Doctorate	3.7	4.3

** Units in % unless otherwise denoted.*

Table S1. Sample characteristics. All categories replicate nomenclature used in the demographics portion of the UAS survey. Representative sample size: $N = 6{,}823$. Unweighted sample size: $N = 7024$.

	"Morally acceptable"	"Not a moral issue"	"Not sure"	"Morally wrong"	Willingness to Use > 50%	Mean Willingness to Use
Gene Editing	19% *(1.1)*	22% *(1.1)*	30% *(1.3)*	29% *(1.2)*	28% *(1.3)*	34% *(1.0)*
PGT-P	27% *(1.3)*	30% *(1.3)*	25% *(1.3)*	17% *(1.1)*	38% *(1.4)*	43% *(1.1)*
SAT Prep	27% *(1.3)*	49% *(1.4)*	16% *(1.1)*	7% *(0.8)*	68% *(1.3)*	69% *(0.9)*

Table S2. Moral acceptability and willingness to use each service. Proportions of participants (*standard errors* in parentheses) giving each of four responses to the moral acceptability question for each service (white); among the three services, all three differences between the proportion of participants who have no moral objections to the service ("morally acceptable" or "not a moral issue") and those who do and might ("morally wrong" or "not sure") were significant at $p < 0.01$. Proportions of participants reporting that they are more than 50% likely to use each service (light gray); among the three services, all three differences were significant at $p < 0.01$. Mean willingness to use each service (darker gray); among the three services, all three differences were significant at $p < 0.01$. Estimates in this table are based on the US-representative sample using the sampling weights provided by the UAS ($N = 6823$, Effective $N = 3805$).

	(A) Willingness to Use > 50%			(B) Mean Willingness to Use		
	10% Social Uptake	90% Social Uptake	*p* -value	10% Social Uptake	90% Social Uptake	*p* -value
Gene Editing	26% *(1.7)*	30% *(1.8)*	0.056	32% *(1.4)*	36% *(1.5)*	0.020
PGT-P	35% *(1.9)*	41% *(2.0)*	0.019	40% *(1.5)*	45% *(1.6)*	0.007
SAT Prep	65% *(1.9)*	72% *(1.8)*	0.004	67% *(1.3)*	71% *(1.3)*	0.022

Table S3. Willingness to use each service, by social uptake. (A) Proportion of participants who reported, on a scale from 0–100%, being more than 50% likely to use each service, by low (10%) versus high (90%) social uptake. (B) Mean response of participants of how likely they are, on a scale from 0–100%, to use each service, by low (10%) versus high (90%) social uptake. The *p*-value reported in this table corresponds to a two-sample *t*-test comparing the two percentages corresponding to the 10% Social Uptake and 90% Social Uptake columns. Estimates in this table are based on the US-representative sample using the sampling weights provided by the UAS ($N = 6823$, Effective $N = 3805$).

		"Morally acceptable" or "Not a moral issue"		Willingness to Use > 50%		Mean Willingness to Use	
		Percentage	*p* -value	Percentage	*p* -value	Mean	*p* -value
Gene Editing	Under 35	42% *(3.2)*	0.59	32% *(3.0)*	0.071	41% *(2.3)*	4.0×10^{-4}
	35 & older	40% *(1.5)*		26% *(1.4)*		32% *(1.1)*	
	Full sample	41% *(1.4)*		28% *(1.3)*		34% *(1.0)*	
PGT-P	Under 35	59% *(3.1)*	0.62	44% *(3.1)*	0.018	48% *(2.3)*	0.013
	35 & older	57% *(1.6)*		36% *(1.5)*		41% *(1.2)*	
	Full sample	58% *(1.4)*		38% *(1.4)*		43% *(1.1)*	
SAT Prep	Under 35	76% *(2.7)*	0.69	72% *(2.8)*	0.083	72% *(1.9)*	0.039
	35 & older	77% *(1.4)*		67% *(1.5)*		68% *(1.0)*	
	Full sample	76% *(1.2)*		68% *(1.3)*		69% *(0.9)*	

Table S4. Moral acceptability of and willingness to use each service, by age. Proportion of participants (*standard errors* in parentheses) who have no moral objection to each service, and proportion of participants who report being > 50% likely to use each service, by age. Estimates in this table are based on the US-representative sample using the sampling weights provided by the UAS (N = 6823, Effective N = 3805).

		"Morally acceptable" or "Not a moral issue"		Willingness to Use > 50%		Mean Willingness to Use	
		Percentage	*p* -value	Percentage	*p* -value	Mean	*p* -value
Gene Editing	High EA	46% *(2.2)*	0.012	33% *(2.1)*	0.002	38% *(1.8)*	0.008
	Low EA	39% *(1.7)*		25% *(1.5)*		32% *(1.2)*	
PGT-P	High EA	65% *(2.1)*	1.7×10^{-4}	46% *(2.2)*	1.7×10^{-5}	48% *(1.8)*	1.6×10^{-4}
	Low EA	54% *(1.9)*		33% *(1.8)*		40% *(1.4)*	
SAT Prep	High EA	86% *(1.6)*	3.9×10^{-10}	80% *(1.7)*	8.0×10^{-15}	78% *(1.2)*	4.7×10^{-17}
	Low EA	71% *(1.7)*		61% *(1.8)*		64% *(1.2)*	

Table S5. Moral acceptability of and willingness to use each service, by educational attainment. Proportion of participants (*standard errors* in parentheses) who have no moral objection to each service, and proportion of participants who report being > 50% likely to use each service, by educational attainment. The split between low (associate degree or below) and high (bachelor's degree or above) educational attainment reflects the median educational attainment of the sample. Estimates in this table are based on the US-representative sample using the sampling weights provided by the UAS (N = 6823, Effective N = 3805).

	Proportion of sample
Morally acceptable	39% *(0.8)*
Not a moral issue	39% *(0.8)*
Not sure	17% *(0.6)*
Morally wrong	6% *(0.4)*

Table S6. Moral acceptability of IVF. Proportion of sample (*standard errors* in parentheses) giving each of four responses. Estimates in this table are based on the US-representative sample using the sampling weights provided by the UAS (N = 6823, Effective N = 3805)

		"Morally acceptable"	"Not a moral issue"	"Not sure"	"Morally wrong"	Mean Willingness to Use
Gene Editing	Under 30	24% *(3.7)*	19% *(3.3)*	29% *(3.9)*	28% *(4.0)*	40% *(2.9)*
	30 to 34	22% *(4.0)*	20% *(3.9)*	33% *(4.6)*	26% *(4.1)*	43% *(3.9)*
	35 to 55	20% *(1.7)*	23% *(1.8)*	30% *(2.0)*	27% *(1.9)*	37% *(1.6)*
	56 and over	16% *(1.6)*	21% *(1.7)*	29% *(2.0)*	34% *(2.0)*	26% *(1.5)*
PGT-P	Under 30	27% *(3.9)*	30% *(4.1)*	29% *(4.0)*	14% *(3.0)*	47% *(3.2)*
	30 to 34	31% *(4.1)*	30% *(4.1)*	24% *(3.8)*	14% *(3.2)*	49% *(3.4)*
	35 to 55	31% *(2.2)*	30% *(2.2)*	24% *(2.0)*	15% *(1.6)*	46% *(1.8)*
	56 and over	23% *(1.8)*	30% *(2.0)*	26% *(2.0)*	20% *(1.7)*	36% *(1.6)*
SAT Prep	Under 30	31% *(3.9)*	43% *(4.1)*	17% *(3.2)*	10% *(2.4)*	71% *(2.5)*
	30 to 34	30% *(4.3)*	49% *(4.8)*	17% *(3.6)*	4% *(2.0)*	73% *(2.8)*
	35 to 55	26% *(2.0)*	49% *(2.3)*	15% *(1.7)*	9% *(1.4)*	69% *(1.4)*
	56 and over	26% *(1.9)*	52% *(2.2)*	16% *(1.7)*	6% *(1.0)*	67% *(1.5)*

Table S7. Moral acceptability of and willingness to use each service, by age. Proportion of participants (*standard errors* in parentheses) who have no moral objection to each service, and proportion of participants who report being > 50% likely to use each service, by age. Estimates in this table are based on the US-representative sample using the sampling weights provided by the UAS (N = 6823, Effective N = 3805).

		"Morally acceptable"	"Not a moral issue"	"Not sure"	"Morally wrong"	Mean Willingness to Use
Gene Editing	High School	19% *(1.9)*	20% *(1.9)*	36% *(2.3)*	26% *(2.1)*	33% *(1.7)*
	Some College / Assoc.	14% *(1.7)*	25% *(2.1)*	29% *(2.2)*	33% *(2.3)*	31% *(1.7)*
	Bachelor's	23% *(2.5)*	20% *(2.2)*	24% *(2.3)*	34% *(2.7)*	35% *(2.2)*
	Master's and Above	26% *(3.2)*	24% *(3.0)*	23% *(2.9)*	26% *(3.1)*	43% *(2.9)*
PGT-P	High School	23% *(2.2)*	30% *(2.4)*	31% *(2.4)*	16% *(1.9)*	38% *(1.8)*
	Some College / Assoc.	26% *(2.3)*	31% *(2.3)*	26% *(2.2)*	17% *(1.9)*	42% *(2.0)*
	Bachelor's	34% *(2.8)*	28% *(2.6)*	21% *(2.4)*	17% *(2.1)*	48% *(2.3)*
	Master's and Above	33% *(3.2)*	35% *(3.4)*	15% *(2.3)*	16% *(2.5)*	49% *(2.7)*
SAT Prep	High School	27% *(2.3)*	40% *(2.6)*	26% *(2.3)*	7% *(1.4)*	59% *(1.7)*
	Some College / Assoc.	28% *(2.3)*	49% *(2.5)*	14% *(1.7)*	10% *(1.5)*	70% *(1.5)*
	Bachelor's	25% *(2.5)*	59% *(2.9)*	9% *(1.7)*	6% *(1.5)*	77% *(1.6)*
	Master's and Above	29% *(2.8)*	59% *(3.2)*	7% *(1.8)*	5% *(1.4)*	80% *(1.8)*

Table S8. Moral acceptability of and willingness to use each service, by educational attainment. Proportion of participants (*standard errors* in parentheses) who have no moral objection to each service, and proportion of participants who report being > 50% likely to use each service, by educational attainment. Estimates in this table are based on the US-representative sample using the sampling weights provided by the UAS (N = 6823, Effective N = 3805).

	"Morally acceptable"	"Not a moral issue"	"Not sure"	"Morally wrong"	Willingness to Use > 50%	Mean Willingness to Use
Gene Editing	18% *(0.8)*	23% *(0.9)*	28% *(0.9)*	31% *(1.0)*	27% *(0.9)*	33% *(0.7)*
PGT-P	28% *(0.9)*	31% *(1.0)*	24% *(0.9)*	17% *(0.8)*	37% *(1.0)*	43% *(0.8)*
SAT Prep	27% *(0.9)*	52% *(1.0)*	14% *(0.7)*	7% *(0.5)*	71% *(0.9)*	71% *(0.6)*

Table S9. Moral acceptability and willingness to use each service. Proportions of participants (*standard errors* in parentheses) giving each of four responses to the moral acceptability question for each service (white); among the three services, all three differences between the proportion of participants who have no moral objections to the service ("morally acceptable" or "not a moral issue") and those who do and might ("morally wrong" or "not sure") were significant at $p < 0.01$. Proportions of participants more than 50% likely to use each service (light gray); among the three services, all three differences were significant at $p < 0.01$. Mean willingness to use each service (darker gray); among the three services, all three differences were significant at $p < 0.01$.

	(A) Willingness to Use > 50%			(B) Mean Willingness to Use		
	10% Social Uptake	90% Social Uptake	*p* -value	10% Social Uptake	90% Social Uptake	*p* -value
Gene Editing	25% *(1.3)*	29% *(1.3)*	0.027	31% *(1.0)*	35% *(1.1)*	0.002
PGT-P	35% *(1.4)*	40% *(1.4)*	0.020	41% *(1.1)*	44% *(1.1)*	0.015
SAT Prep	67% *(1.3)*	75% *(1.3)*	2.3×10^{-5}	69% *(0.9)*	74% *(0.9)*	5.1×10^{-6}

Table S10. Willingness to use each service, by social uptake. (A) Proportion of participants who reported, on a scale from 0–100%, being more than 50% likely to use each service, by low (10%) versus high (90%) social uptake. (B) Mean response of participants of how likely they are, on a scale from 0–100%, to use each service, by low (10%) versus high (90%) social uptake. The *p*-value reported in this table corresponds to a two-sample *t*-test comparing the two percentages corresponding to the 10% Social Uptake and 90% Social Uptake columns.

		"Morally acceptable" or "Not a moral issue"		Willingness to Use > 50%		Mean Willingness to Use	
		Percentage	*p* -value	Percentage	*p* -value	Mean	*p* -value
Gene Editing	Under 35	46% *(2.6)*	0.059	34% *(2.3)*	4.4×10^{-4}	42% *(1.9)*	6.4×10^{-7}
	35 & older	40% *(1.1)*		26% *(1.0)*		31% *(0.8)*	
	Full sample	41% *(1.0)*		27% *(0.9)*		33% *(0.7)*	
PGT-P	Under 35	61% *(2.5)*	0.42	46% *(2.4)*	1.1×10^{-4}	49% *(1.9)*	7.1×10^{-5}
	35 & older	59% *(1.1)*		36% *(1.1)*		41% *(0.9)*	
	Full sample	59% *(1.0)*		37% *(1.0)*		43% *(0.8)*	
SAT Prep	Under 35	73% *(2.0)*	0.002	73% *(2.2)*	0.42	73% *(1.5)*	0.29
	35 & older	80% *(0.9)*		71% *(1.0)*		71% *(0.7)*	
	Full sample	79% *(0.8)*		71% *(0.9)*		71% *(0.6)*	

Table S11. Moral acceptability of and willingness to use each service, by age. Proportion of participants (*standard errors* in parentheses) who have no moral objection to each service, and proportion of participants who report being > 50% likely to use each service, by age.

		"Morally acceptable" or "Not a moral issue"	*p* -value	Willingness to Use > 50%	*p* -value	Mean Willingness to Use	*p* -value
Gene Editing	High EA	43% *(1.6)*	0.068	30% *(1.4)*	0.007	35% *(1.1)*	0.006
	Low EA	39% *(1.3)*		25% *(1.2)*		31% *(1.0)*	
PGT-P	High EA	64% *(1.5)*	1.8×10^{-5}	43% *(1.5)*	3.7×10^{-6}	46% *(1.2)*	1.3×10^{-5}
	Low EA	56% *(1.4)*		33% *(1.4)*		39% *(1.1)*	
SAT Prep	High EA	85% *(1.3)*	1.2×10^{-12}	81% *(1.4)*	2.9×10^{-19}	78% *(0.9)*	5.5×10^{-22}
	Low EA	73% *(1.1)*		64% *(1.2)*		66% *(0.8)*	

Table S12. Moral acceptability of and willingness to use each service, by educational attainment. Proportion of participants (*standard errors* in parentheses) who have no moral objection to each service, and proportion of participants who report being > 50% likely to use each service, by educational attainment. The split between low (associate degree or below) and high (bachelor's degree or above) educational attainment reflects the median educational attainment of the sample.

	Proportion of sample
Morally acceptable	41% *(0.6)*
Not a moral issue	40% *(0.6)*
Not sure	14% *(0.4)*
Morally wrong	5% *(0.3)*

Table S13. Moral acceptability of IVF. Proportion of full sample (*standard errors* in parentheses) giving each of four responses.