Appendix for Choose Your Moments: Peer Review and Scientific Risk Taking

A Randomization Checks

Table A1: Randomization Balance: Omnibus F-test

	(1) Study 1 indicator	(2) Study 1 indicator
Fraction BDCN	-0.018	-0.078
	(0.019)	(0.061)
Fraction CB	0.024	0.014
	(0.020)	(0.063)
Fraction OBT/OTC	-0.0034	-0.057
	(0.018)	(0.061)
Total funding	0.00061	-0.0030
	(0.0010)	(0.0045)
Unique projects	-0.0021	0.014
	(0.0030)	(0.012)
Total projects	-0.00028	-0.0013
	(0.00063)	(0.0030)
F-stat	1.21	0.74
p-value	0.30	0.61
Observations	6678	563

This table shows randomization checks for the randomization into Study 1 versus Study 2. Column 1 shows the check for the total sample of potential participants. Column 2 shows the check for the sample that completed the study. Heteroskedasticity robust standard errors are in parentheses: $p < .10$, ** $p < .05$, *** $p < .01$.

B Robustness and Sensitivity Results

Table A2: Additional Preferences for Score Statistics

	(1) Project choice	(2) Project choice	(3) Project choice	(4) Project choice
Avg. score	0.82***	0.92***	0.67***	0.80***
	(0.041)	(0.057)	(0.061)	(0.044)
Score variance	0.074**	0.11***	0.086	0.074**
	(0.029)	(0.035)	(0.060)	(0.034)
Avg. score × Score variance	0.029			
	(0.026)			
Score skew		0.14**		
		(0.068)		
Score kurtosis		0.031		
		(0.055)		
Count of lowest scores			-0.13**	
			(0.052)	
Count of highest scores			0.11***	
			(0.044)	
At least one bottom score				-0.026
				(0.050)
At least one top score				0.094
				(0.077)
Clusters	313	313	313	313
N	11276	11094	11276	11276

This table shows results from estimating Equation (1) on the baseline sample. The dependent variable is an indicator for whether the participant chose a given project. All models include participant fixed effects. Standard errors are clustered at the participant level: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A3: Robustness: Main Results Using Multinomial Logit and Project Rank

	(1) rank	(2) rank	(3) rank
1			
Avg. score	1.79***	1.79***	2.27***
	(0.11)	(0.11)	(0.16)
Score variance	0.16***	0.16***	0.23***
	(0.060)	(0.060)	(0.075)
MeSH match		0.050	0.093
		(0.042)	(0.059)
2			
Avg. score	1.13***	1.13***	1.41***
	(0.081)	(0.081)	(0.11)
Score variance	0.11**	0.11**	0.17***
	(0.053)	(0.053)	(0.064)
MeSH match		0.057	0.095*
		(0.036)	(0.051)
3			
Avg. score	0.58***	0.58***	0.74***
	(0.057)	(0.057)	(0.077)
Score variance	-0.037	-0.038	-0.011
	(0.049)	(0.049)	(0.057)
MeSH match		0.057	0.091*
		(0.036)	(0.051)
Subject FEs	No	No	Yes
Clusters	313	313	313
N	5008	5008	5008

This table shows results from estimating a multinomial logit model corresponding to Equation (1) on the baseline sample. The dependent variable is the rank the participant gave to a given project (lower is better and the excluded category is rank 4 out of 4). Standard errors are clustered at the participant level: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A4: Robustness: Sensitivity to Fixed Effects

	(1) Project choice	(2) Project choice
Avg. score	0.81***	1.10***
	(0.040)	(0.071)
Score variance	0.080***	0.12***
	(0.027)	(0.034)
Sample	All obs.	Granual FEs
Clusters	356	313
N	12078	11228

This table shows results from estimating Equation (1) on the baseline sample. The dependent variable is an indicator for whether the participant chose a given project. All models include participant fixed effects. Standard errors are clustered at the participant level: * $p < .10$, ** $p < .05$, *** $p < .01$.

A. Generalized Multinomial Logistic Regression

Fiebig et al. (2010) generalize the logistic regression to permit estimation of heterogeneous preferences across individuals, while also relaxing scale assumptions in the standard conditional logit model. They call this model the generalized multinomial logit model (G-MNL). Here we report G-MNL estimates for our main results in Table 2. The estimating equation is

$$\Pr(choice_{it} = j|\beta_i) = \frac{\exp(\beta_i' \mathbf{x}_{itj})}{\sum_{k=1}^{J} \exp(\beta_i' \mathbf{x}_{itk})} \tag{A-1}$$

where i indexes subject, t indexes the choice scenario, and j and k index the project attributes. The variable x_{itj} is a vector containing the mean and variance of project scores and β_i is the vector of individual-specific coefficients associated with these project attributes. The coefficients are defined by

$$\beta_i = \sigma_i \beta + \{\gamma + \sigma_i(1-\gamma)\}\eta_i$$

The coefficients in this equation are a vector *beta* that is constant across individuals and measures the average utility weights across the sample for the different variables in x; a single parameter for the scale of the individual-level idiosyncratic error (σ_i), which captures overall scaling of the individual's tastes; and a random vector η_i distributed multivariate normal with mean zero and variance-covariance matrix Σ, which captures latent taste heterogeneity. The parameter γ determines how much of the variance is explained by these latter two components. The setup and parameterization of these parameters follows Fiebig et al. (2010). In particular, we assume that σ_i is distributed log normal with mean $\bar{sigma} + \theta' z_i$ and standard deviation τ. The parameter $\bar{sigma}$ is a normalizing constant, and z_i is a vector of subject fixed effects in this application.

The results of fitting this model are shown in Table A5. The coefficient estimates are qualitatively consistent with the baseline results in the sense that on average, participants prefer both higher mean and higher variance projects. The subject-level utility weights are predicted based on the average utility weights and heterogeneity in weights using the `gmnlpred` command from Gu, Hole, and Knox (2013). Based on these predictions, the average weight placed on project score variance is 8% as large as the weight that is placed on average project score, comparable to the 10% value derived from Table 2.

Table A5: G-MNL estimates of preferences over average and variance of project scores

	(1) Project choice
Average utility weight	
Average project score	2.86***
	(0.36)
Project score variance	0.016**
	(0.07)
Utility weight heterogeneity	
Average project score	0.96***
	(0.13)
Project score variance	0.65***
	(0.09)
Additional parameters	
τ	1.29***
	(0.14)
γ	0.25***
	(0.06)
Clusters	313
N	11228

This table shows results from estimating Equation (A-1) on the baseline sample. The dependent variable is an indicator for whether the participant chose a given project. All models include participant fixed effects. Standard errors are clustered at the participant level: * $p < .10$, ** $p < .05$, *** $p < .01$.

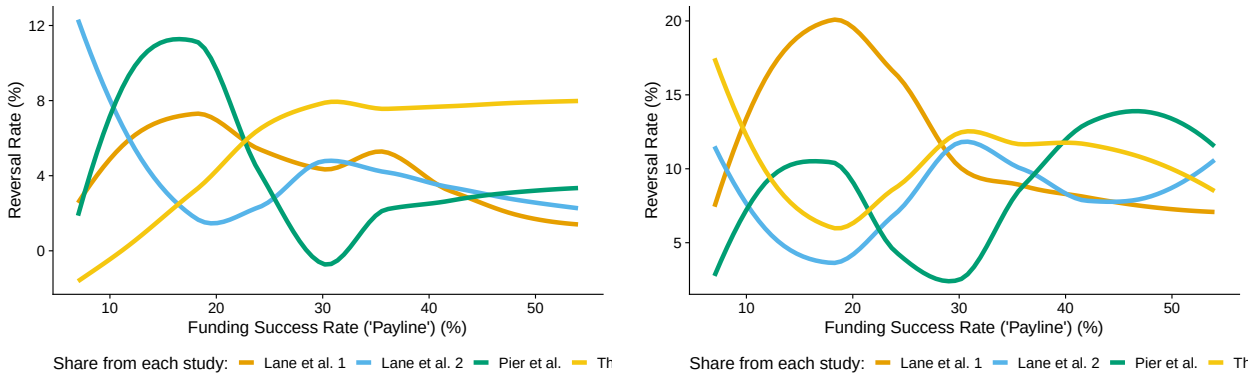
C Additional Figures and Tables

Table A6: Effects of Positive and Negative Skewness

	(1) Project choice	(2) Project choice
Avg. score	0.95***	1.02***
	(0.067)	(0.19)
Score variance	0.094***	0.054
	(0.035)	(0.12)
Score skew	0.16***	0.73
	(0.050)	(1.17)
Skew sample	Neg. skew	Pos. skew
Clusters	313	230
N	8887	2071

This table shows results from estimating Equation (1) on the baseline sample. The dependent variable is an indicator for whether the participant chose a given project. All models include participant fixed effects. Standard errors are clustered at the participant level: * $p < .10$, ** $p < .05$, *** $p < .01$.

Figure A1: Funding Reversals Under NIH and Scientist Ranking Procedures: Separate Estimates by Study



(a) Overall sample preferences

(b) Expert preferences

Notes: This figure shows the reversal rate for project funding as a function of the payline (fraction of projects that get funded) for four different sets of project scores (Study 1 from this paper, the two studies from Lane et al. (2022b), and Pier et al. (2018)). The reversal rate is the fraction of studies that change when they are funded under a mean-only ranking versus under the mean and variance-based ranking. Panel (a) shows the reversal rate when using the estimated preferences from the baseline results using the full sample of scientists. Panel (b) shows the reversal rate when using the preferences of scientists who are relative experts (MeSH match 1 standard deviation higher than average).

Table A7: Preferences for Project Scores

	(1) Project choice
Score of 1	-0.20***
	(0.026)
Score of 2	-0.18***
	(0.022)
Score of 3	-0.093***
	(0.021)
Score of 4	0.015
	(0.019)
Score of 6	0.068***
	(0.024)
Score of 7	0.074***
	(0.011)
Score of 8	0.14***
	(0.014)
Score of 9	0.23***
	(0.016)
Clusters	313
N	11228

This table shows results from estimating Equation (1) on the baseline sample where the right-hand-side variables (project attributes) are each level of possible project scores. The dependent variable is an indicator for whether the participant chose a given project. All models include participant fixed effects. Standard errors are clustered at the participant level: * $p < .10$, ** $p < .05$, *** $p < .01$.

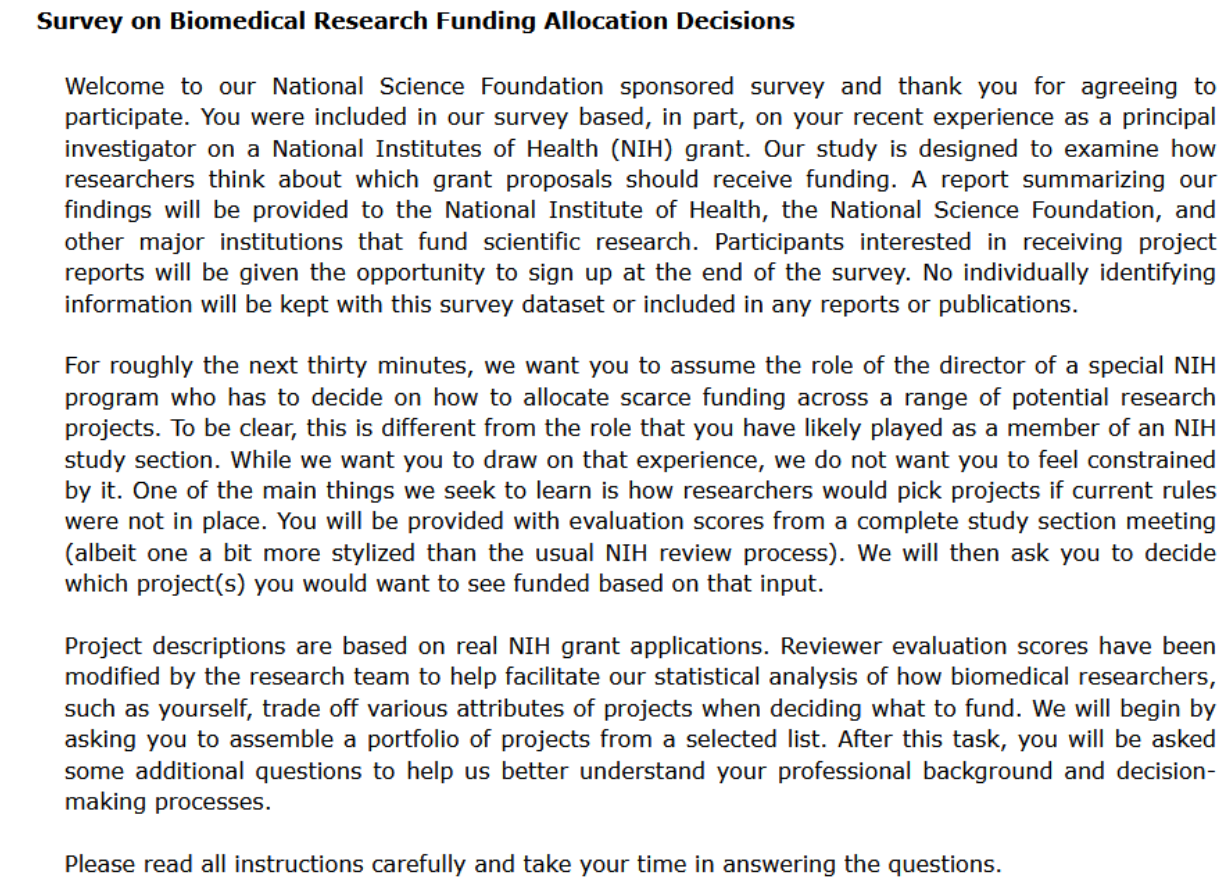
Table A8: Heterogeneity: Demographics, Preferences, and Experience

	(1) Project choice	(2) Project choice	(3) Project choice	(4) Project choice	(5) Project choice	(6) Project choice
Avg. score	0.80***	0.82***	0.78***	0.91***	0.74***	0.83***
	(0.069)	(0.050)	(0.051)	(0.066)	(0.073)	(0.067)
Score variance	0.022	0.11***	0.079**	0.088*	0.012	0.12***
	(0.048)	(0.033)	(0.034)	(0.046)	(0.045)	(0.047)
Subgroup	Female	Male	Risk averse	Not risk averse	Least experience	Most experience
Clusters	91	222	211	102	103	105
N	3277	7999	7604	3672	3710	3783

This table shows results from estimating Equation (1) on the baseline sample. The dependent variable is an indicator for whether the participant chose a given project. All right-hand-side variables are standardized. The models include participant fixed effects. Standard errors are clustered at the participant level: * $p < .10$, ** $p < .05$, *** $p < .01$.

D Experiment Instructions and Instruments

Figure A2: Experimental Instrument: Welcome Screen



Survey on Biomedical Research Funding Allocation Decisions

Welcome to our National Science Foundation sponsored survey and thank you for agreeing to participate. You were included in our survey based, in part, on your recent experience as a principal investigator on a National Institutes of Health (NIH) grant. Our study is designed to examine how researchers think about which grant proposals should receive funding. A report summarizing our findings will be provided to the National Institute of Health, the National Science Foundation, and other major institutions that fund scientific research. Participants interested in receiving project reports will be given the opportunity to sign up at the end of the survey. No individually identifying information will be kept with this survey dataset or included in any reports or publications.

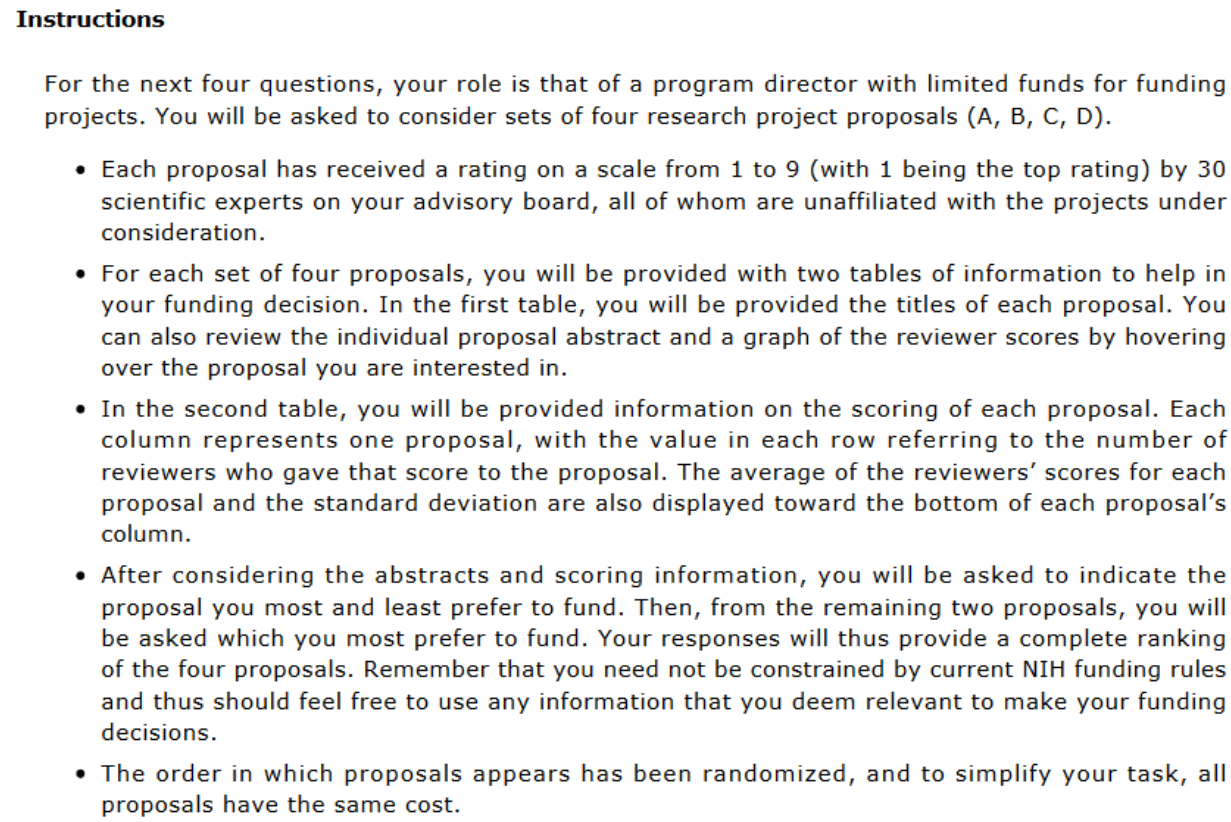
For roughly the next thirty minutes, we want you to assume the role of the director of a special NIH program who has to decide on how to allocate scarce funding across a range of potential research projects. To be clear, this is different from the role that you have likely played as a member of an NIH study section. While we want you to draw on that experience, we do not want you to feel constrained by it. One of the main things we seek to learn is how researchers would pick projects if current rules were not in place. You will be provided with evaluation scores from a complete study section meeting (albeit one a bit more stylized than the usual NIH review process). We will then ask you to decide which project(s) you would want to see funded based on that input.

Project descriptions are based on real NIH grant applications. Reviewer evaluation scores have been modified by the research team to help facilitate our statistical analysis of how biomedical researchers, such as yourself, trade off various attributes of projects when deciding what to fund. We will begin by asking you to assemble a portfolio of projects from a selected list. After this task, you will be asked some additional questions to help us better understand your professional background and decision-making processes.

Please read all instructions carefully and take your time in answering the questions.

Notes: This figure shows the welcome screen that greeted participants.

Figure A3: Experimental Instrument: Main Instructions



Instructions

For the next four questions, your role is that of a program director with limited funds for funding projects. You will be asked to consider sets of four research project proposals (A, B, C, D).

- Each proposal has received a rating on a scale from 1 to 9 (with 1 being the top rating) by 30 scientific experts on your advisory board, all of whom are unaffiliated with the projects under consideration.
- For each set of four proposals, you will be provided with two tables of information to help in your funding decision. In the first table, you will be provided the titles of each proposal. You can also review the individual proposal abstract and a graph of the reviewer scores by hovering over the proposal you are interested in.
- In the second table, you will be provided information on the scoring of each proposal. Each column represents one proposal, with the value in each row referring to the number of reviewers who gave that score to the proposal. The average of the reviewers' scores for each proposal and the standard deviation are also displayed toward the bottom of each proposal's column.
- After considering the abstracts and scoring information, you will be asked to indicate the proposal you most and least prefer to fund. Then, from the remaining two proposals, you will be asked which you most prefer to fund. Your responses will thus provide a complete ranking of the four proposals. Remember that you need not be constrained by current NIH funding rules and thus should feel free to use any information that you deem relevant to make your funding decisions.
- The order in which proposals appears has been randomized, and to simplify your task, all proposals have the same cost.

Notes: This figure shows the main instructions that participants were shown prior to making any choices.

Figure A4: Experimental Instrument: Choice Scenario Example

Please indicate which of the four proposals you would **most** and **least** like to fund. The tables below show the proposal titles and the number of reviewers (out of your panel of 30) who gave the proposals each score, 1 (best) through 9 (worst), along with the average score and standard deviation of the scores for each proposal. Hover over each proposal in the first table to view its abstract and see a graph of reviewer scores.

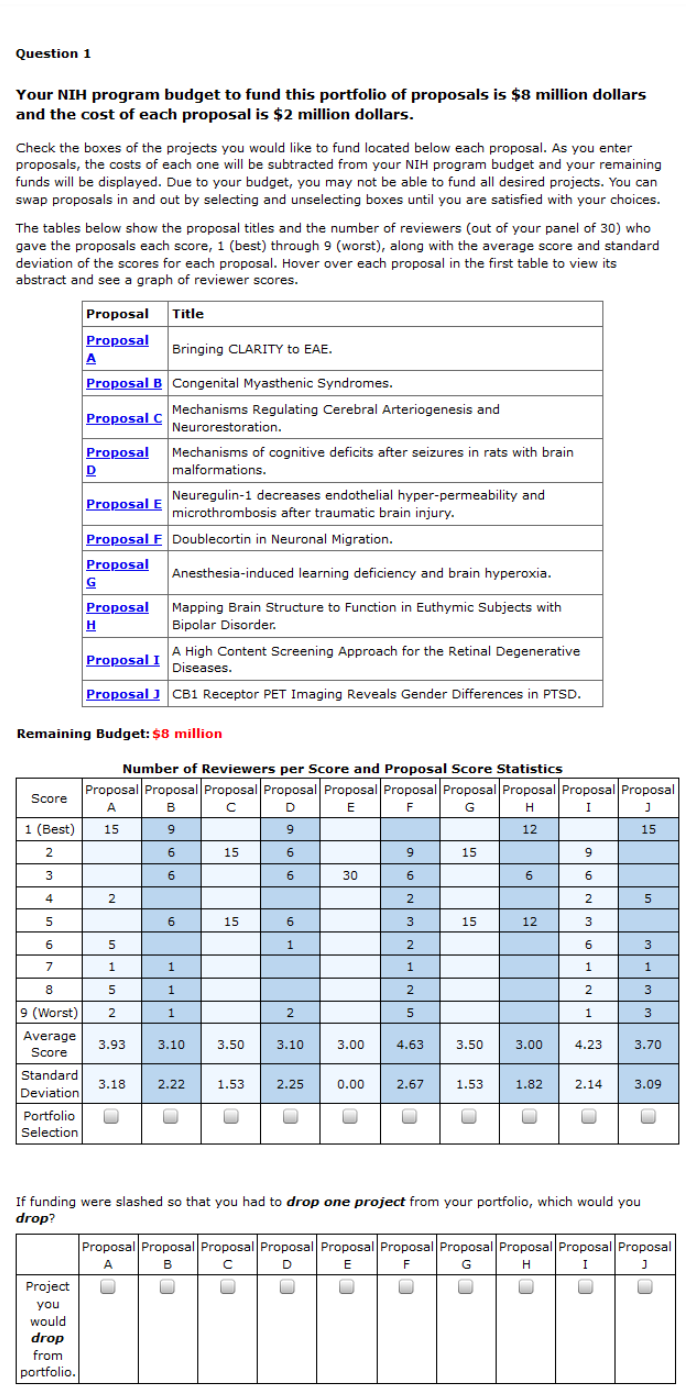
Proposal	Title
Proposal A	Bringing CLARITY to EAE.
Proposal B	Congenital Myasthenic Syndromes.
Proposal C	Mechanisms Regulating Cerebral Arteriogenesis and Neurorestoration.
Proposal D	Mechanisms of cognitive deficits after seizures in rats with brain malformations.

Number of Reviewers per Score and Proposal Score Statistics

Score	Proposal A	Proposal B	Proposal C	Proposal D
1 (Best)	3	9	9	3
2	6	12	6	6
3	6	6	6	6
4		3	1	2
5	6		6	6
6				2
7	3			3
8	4		2	1
9 (Worst)	2			1
Average Score	4.47	2.10	2.97	4.03
Standard Deviation	2.61	0.96	2.01	2.19
Which proposal you would **most** like to fund.	☐	☐	☐	☐
Which proposal you would **least** like to fund.	☐	☐	☐	☐

Notes: This figure shows an example of the choices participants were asked to make. In the top box, the blue, underlined proposal links would show proposal abstracts if the participant hovered over them. After choosing their top and bottom ranked projects, in a follow-up screen participants were asked to rank the remaining two projects.

Figure A5: Experimental Instrument: Portfolio Choice Scenario Example



Question 1

Your NIH program budget to fund this portfolio of proposals is $8 million dollars and the cost of each proposal is $2 million dollars.

Check the boxes of the projects you would like to fund located below each proposal. As you enter proposals, the costs of each one will be subtracted from your NIH program budget and your remaining funds will be displayed. Due to your budget, you may not be able to fund all desired projects. You can swap proposals in and out by selecting and unselecting boxes until you are satisfied with your choices.

The tables below show the proposal titles and the number of reviewers (out of your panel of 30) who gave the proposals each score, 1 (best) through 9 (worst), along with the average score and standard deviation of the scores for each proposal. Hover over each proposal in the first table to view its abstract and see a graph of reviewer scores.

Proposal	Title
Proposal A	Bringing CLARITY to EAE.
Proposal B	Congenital Myasthenic Syndromes.
Proposal C	Mechanisms Regulating Cerebral Arteriogenesis and Neurorestoration.
Proposal D	Mechanisms of cognitive deficits after seizures in rats with brain malformations.
Proposal E	Neuregulin-1 decreases endothelial hyper-permeability and microthrombosis after traumatic brain injury.
Proposal F	Doublecortin in Neuronal Migration.
Proposal G	Anesthesia-induced learning deficiency and brain hyperoxia.
Proposal H	Mapping Brain Structure to Function in Euthymic Subjects with Bipolar Disorder.
Proposal I	A High Content Screening Approach for the Retinal Degenerative Diseases.
Proposal J	CB1 Receptor PET Imaging Reveals Gender Differences in PTSD.

Remaining Budget: $8 million

Number of Reviewers per Score and Proposal Score Statistics

Score	Proposal A	Proposal B	Proposal C	Proposal D	Proposal E	Proposal F	Proposal G	Proposal H	Proposal I	Proposal J
1 (Best)	15	9		9				12		15
2		6	15	6		9	15		9	
3		6		6	30	6		6	6	
4	2					2			2	5
5		6	15	6		3	15	12	3	
6	5			1		2			6	3
7	1	1				1			1	1
8	5	1				2			2	3
9 (Worst)	2	1		2		5			1	3
Average Score	3.93	3.10	3.50	3.10	3.00	4.63	3.50	3.00	4.23	3.70
Standard Deviation	3.18	2.22	1.53	2.25	0.00	2.67	1.53	1.82	2.14	3.09
Portfolio Selection	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

If funding were slashed so that you had to ***drop one project*** from your portfolio, which would you ***drop***?

	Proposal A	Proposal B	Proposal C	Proposal D	Proposal E	Proposal F	Proposal G	Proposal H	Proposal I	Proposal J
Project you would ***drop*** from portfolio.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Notes: This figure shows an example of the choices participants were asked to make in Study 2, the portfolio choice scenario. The budget remaining changed as projects were selected. After making these choices, the participant was prompted to say which project they would add if the budget were expanded.