Supplementary Information

Title: Rethink funding: Putting the lottery first.

Subtitle: To increase access to research funding for historically marginalized communities, it is necessary to decrease biases in the funding process.

Authors: Finn Luebber, [1,2,3] Sören Krach, [1,3,*] Marina Martinez Mateo, [4] Frieder M. Paulus, [1,3] Lena Rademacher, [1,3] Rima-Maria Rahal, [5] Jule Specht. [6]

Affiliations:

1: Department of Psychiatry and Psychotherapy, University of Lübeck, Germany

2: Department of Rheumatology and Clinical Immunology, University of Lübeck, Germany

3: Open Science Initiative, University of Lübeck, Germany

4: Academy of Fine Arts, Munich, Germany

5: Behavioral Law & Economics, Max Planck Institute for Research on Collective Goods, Bonn, Germany

6: Department of Psychology, Humboldt University, Berlin, Germany

Email:

Finn Luebber f.luebber@uni-luebeck.de

Sören Krach soeren.krach@uni-luebeck.de

Marina Martinez Mateo martinezmateo@adbk.mhn.de

Frieder M. Paulus frieder.paulus@uni-luebeck.de

Lena Rademacher lena.rademacher@uni-luebeck.de

Rima-Maria Rahal rahal@coll.mpg.de

Jule Specht jule.specht@hu-berlin.de

***Corresponding Author:**

Professor Dr Sören Krach

Department of Psychiatry and Psychotherapy, University of Lübeck, Germany

University of Lübeck, Ratzeburger Allee 160, D-23538 Lübeck, Germany

Phone: +49 45131017526

Table of contents Supplementary Information

Supplementary Information

Brief manual and background

Introduction and R packages

We here introduce an app-based simulation approach to model and analyze the scientific grant-application process. This approach can help to compare different grant-application scenarios with regards to costs, diversity, and quality to inform the debate on how to improve the process. The Shiny app "GrantInq" was built using the R packages shiny (Chang et al. 2021), shinyBS (Bailey 2022), shinyjs (Attali 2021), shinydashboard (Chang and Borges Ribeiro 2021), spsComps (Zhang 2022), and shinyWidgets (Perrier, Meyer, and Granjon 2022) in R (R Core Team 2020) using RStudio (RStudio Team 2021). Additionally, packages from the tidyverse (Wickham et al. 2019) were used. In the following text, user customizable parameter names or outputs found in the Shiny app are printed in **bold** for referencing between the GrantInq and the additional material.

Example funding scenarios

The user can choose from a set of four **example funding scenarios** which are inspired by established funding scenarios (which are further described in the Supplementary Material – Scenarios & Simulations). They consist of a one-stage scenario, a two-stage-scenario, the tiebreaker lottery, and the pre-lottery scenario.

General settings

The user chooses how many applications there are initially (**Number of applications**) to compete for a **target volume per application** from the overall **funding budget** in a process consisting of one to up to three stages (**Nr. of process stages**). A stage is defined as a step in the selection process where applications are potentially rejected and only a subset of applicants advances to the next stage (or are finally accepted to receive funding). The whole process is simulated in a given number of samples (**Number of simulated samples**) with larger samples resulting in more robust estimates.

Distribution of Idea Qualities

Each application has a given quality (idea quality) which is drawn from a distribution with **mean** (fixed to 100), **standard deviation** and **skew** (**parameters of idea distribution**) according to user input using the "rsnorm"-function from the fGarch package (Wuertz et al. 2022). The resulting distribution of idea qualities is plotted above the parameter settings as empirical histogram per individual sample together with a theoretically derived density curve.

Group proportions

Each application comes from a member of either a **disadvantaged** or an **advantaged group**. Labels for both groups can be provided by the user and are initialized with "Women" and "Men", respectively. The group proportions in the target population can be selected using the slider "**Population proportion of [Disadvantaged group]**". The disadvantages of the respective group are characterized by an **entry bias** and bias in the review process (see **Diversity & Quality** below). The **entry bias** defines the proportion of the disadvantaged group which eventually does not apply and enter the applicant pool so that the actual proportion of the disadvantaged group in the simulation is defined as "**Population proportion** * **Entry Bias**" (**Initial Proportions**). Using this proportion, a group indicator is sampled for each idea quality so that the applications of both groups have on average the same idea qualities and the initial group proportion in the data will on average be equal to the **Initial Proportions**.

Review Parameters

Here, the user can control details of the selection process. In each stage there is a number of reviewers (**Number Reviewers**) who rate the applications according to idea quality but with a given accuracy and/or bias (see **Diversity & Quality**), as long as **Competition Mode** is set to “Competitive” or “Normative”. When **Competition Mode** is set to “Lottery”, the reviewers have no effect but the user can still enter a number greater than zero (that is mostly so that a given setting here does not get lost when one chooses “Lottery” by mistake, but can also be used to model a process where there is a given amount of administrative cost on the side of the funding agency for a “Lottery” setting). In a “Competitive” or “Lottery” setting, in each stage a given number of applications advance (**Number Accepted Applications**) which is based on their ranking: **Review Mode** controls how applications are ranked: based on the mean of all reviewers, based on the maximum score of all reviewers per application or based on the minimum score. When **Competition Mode** is set to "Normative”, the reviewer rating works as in the "Competitive” setting, although all applications below a given **Cut-Off** in rated quality are rejected.

Diversity & Quality

In this tab, users can control the details of the evaluation of applications by reviewers for “Competitive” and “Normative” modes. In each stage, reviewers will show a given degree of inaccuracy and biases which specifics are subject to user choice. Conceptually, inaccuracies and biases are modeled as error scores which are added to the true idea quality score, once for each reviewer. This leads to an “observed idea quality” per reviewer which is the variable the applications are ranked or evaluated by. These error scores are drawn from distributions which can have different shapes, which can depend on the true idea quality, and which can vary between groups (which would constitute a bias). There are three ways to define the error distributions:

1. Buttons with predefined biases on the left-hand side (“**Constant Bias**”, “**Brilliance Bias**”, “**Error Bias**”, “**Combined Bias**”)
2. **Dropdown menus** between these buttons and plots
3. Manual curve modelling in “**Advanced settings**”

Since the buttons and dropdown-menus depend on the advanced settings, the latter shall be explained first. In each stage there are three **plots**. Each control one parameter of the error distribution: the **mean** (left), the **standard deviation** (SD, center), and the **skew** (right). By clicking inside a plot, a point for the selected group (**Which group do you want to model for stage x?**, below the plots) is added at the position of the cursor, which triggers a linear model fit (lm function of stats package in base R). The algorithm will fit a polynomial to a degree given by the **terms for the best fit** the user selects above the plot. In case there a not enough points to fit a uniquely defined curve, the highest degree possible will be fit. When no term is selected, there will only be an intercept term present creating a value that is constant across all values of idea qualities. Thus, the user can customize the dependency of idea quality and the individual parameters of the error distribution, but also the overall size of the parameters independent of idea quality.

The **buttons** with predefined biases control the **dropdown menus** which in turn control the advanced settings via the addition of individual points to the plots, thus triggering the model fitting as well. These options are supposed to facilitate bias modeling without going into the details of how the modeling works.

Based on the so created error distributions, a single error-term per reviewer and application is drawn from the respective distribution. These distributions are visualized in the first plot to the right of the reviewer bias options. Next, the true idea quality and the errors per reviewer are added to form the observed idea quality for each reviewer. This is displayed in the rightmost plot within each stage which shows all applications' true idea qualities against their observed idea qualities (one for each application, respecting the **Review Mode** setting as either mean, max, or min of all reviewers). Depending on the **Competition Mode**, the applications are then either ranked based on the observed score and a fixed **Number Accepted Applications** advance ("Competitive") or all applications with observed scores above a certain **Cut-Off** advance ("Normative"). In case of a selected "Lottery", the promoted applications are selected by a random draw.

There are three main output plots at the bottom: Left, there is a scatter plot showing the different idea qualities of applications for each stage split by group, together with the mean line of initial applications (this does not need to equal 100, as it can vary due to sampling variability). Next, there is a line graph showing the average idea quality for each stage together with the grand mean of all samples and its standard deviation. Last, the bar graph shows the proportions of the disadvantaged group by stage for the selected sample (left) and an average across all samples (right). With the slider below the plots the user can scroll through all samples that were simulated (and which also applies to the idea distribution in the **General Settings** box).

Costs

In this tab users can define the individual costs of an application which translates to the overall cost of a given grant-application scenario. For this simulation we decided to focus on two main aspects: Costs on the side of applicants, and costs on the side of reviewers, each customizable for each stage. The inputs have identical structure: First, there are **fixed costs per application**. This might for instance include travel costs or office equipment (note that this is fixed costs per application, not per applicant). Next, there are **average work hours per applicant** and **reviewer**, followed by **average costs per work hour**. Last, there is the **average number of people working on a single application** (the corresponding **average number of reviewers** is placed in the **General Settings** box). Although many grant applications call for applications of individuals, this input enables to simulate a more realistic situation where grants are developed in teams.

The first output splits the total work hours (**work hours per applicant/reviewer** * **average number of applicants/reviewers**) by applicant (left) and reviewers (right) and individual application (top) and all applications summed (bottom, values multiplied by the number of applications per stage).

Next, the cost of an individual application is calculated by multiplying the **work hours**, the **price of work hours**, and the **average number of people** separately for applicants and reviewers, and add the respective **fixed costs** to each. This number is then multiplied by the number of applications in each stage. The next output shows these costs in a way split like the work hours plot left to it.

Finally, the costs by stage are calculated by adding applicant costs and reviewer costs and split by stage (left) or by application status (right, as percentage). The application status shows the percentage of costs incurred by the ultimately winning application across all stages and its respective opposite: the costs of those who got rejected at some point in the process.

Save and compare

In the last tab, users can save individual settings to compare the main outputs with either predefined settings or a different individual setting. By naming and confirming the first setting, the respective summary plots are created. The results of any subsequently saved scenario are added to the right within each plot. For the comparison of idea qualities (center) and the group proportion, only the winning applications are considered.

Download of data

The raw data and summary tables can be downloaded in the box ”Download Raw Data”. Here, users may specify a **File prefix** consisting of characters and digits (excluding special characters). Data will be downloaded in a ZIP-compressed folder including comma separated value files on “individual_results” (reviewer ratings, rankings, and success indicators per application), “individual_costs” (raw data of application costs per application), “mean_quality” (summaries of idea quality), “costs_summarized” (summaries of the costs), “relative_costs” (costs per winning/losing) and “group_proportions” (summary of proportions of the disadvantaged group at each stage).

Scenarios & simulations

1. One-stage scenario

The one-stage scenario for funding reflects widely applied selection processes as for example those by the German Research Foundation or National Institute of Health (see e.g., research grant proposal at the Deutsche Forschungsgemeinschaft[1] or R01[2]). Here, in a call-for-submission or an open submission system, applicants submit a full proposal, including the state-of-the-art and preliminary work, the objectives and work program, as well as detailed information on the requested funds together with some general information (e.g., CV, work contract). The review process for this funding scheme is a one-step procedure. Each proposal is reviewed by several experts in the field who submit their evaluation to the council which recommends funding or not.

Simulation for process effectiveness in terms of economic burden

For the one-stage scenario we used the following configurations to simulate the costs (Table S1). All parameter settings are assumed to reflect actual investments and basic parameters such as number of applicants only roughly. The funding volume and the target grant volume were kept constant across all scenarios for comparison purposes.

Table S1. Setup of the one-stage scenario

Configuration of general selection characteristics	
Number of applications	200
Number of selection stages	1
Number of accepted grants	20
Target grant volume *(MU)*	500,000
Investments for application	
Number of people working on application	1
Average amount of work (*h*) per applicant	80
Average costs per hour of the applicant *(MU)*	50

[1] https://www.dfg.de/formulare/50_01/50_01_en.pdf
[2] https://www.niaid.nih.gov/grants-contracts/overview-r01-process

Investments for review	
Fix costs per review *(MU)*	0
Number of people reviewing the application	2
Average amount of work *(h)* per reviewer	6
Average costs per hour of the reviewer *(h)*	50

Note. MU = monetary units; multiple reviews are aggregated by the arithmetic mean (instead of the minimum or maximum being used for selection purposes); ratings are ranked for selection purposes.

In summary, adopting these parameters for the one-stage scenario yields 92 total working hours per application (equaling 4,600 *monetary units [MU]* per application). In total, across all *N*=200 applicants this one-stage scenario has an overall process cost of 920,000 *MU*, which is nearly a tenth of the in this scenario set funds of 10 Million MU. Most of the costs (90%) are sunk costs in the sense that the predominant investment of work, by applicants as well as reviewers, was on applications that were finally rejected (see Figure 2, main article).

Simulation for process effectiveness in terms of diversity and quality

Besides the costs of the one-stage scenario, funding decisions may be evaluated in terms of their potential in reducing bias (e.g., gender) and increasing quality. To simulate the quality and the impact of bias in the selection process in the one-stage scenario we adopted the following parameters (Table S2).

Table S2. Bias and quality simulations in the one-stage scenario

Sampling configuration	
Number of simulated samples	1000
Group proportions in the target population	36 % women
Entry bias	16 %
Grant characteristics	
Idea quality (mean)	100
Idea quality (standard deviation)	15
Idea quality (skew)	1.5
Review error and biases in quality assessment	
Constant underestimation *(AU)*	"medium" (-4)
Standard deviation of error *(AU)*	"large" (SD 15)

Note. *AU* = arbitrary units for idea quality; Group proportions roughly informed from the 2012 applications at the German Research foundation[3].

For the present example, we adopted the universalistic assumption that men and women provide applications of similar quality. However, many analyses have shown that women compared to men need to reach higher levels of quality to reach equal probabilities for funding (Witteman et al. 2019). To account for such bias, we assumed that applications by women were under-rated in general, so that they need to provide higher quality to achieve the identical rating by reviewers. In this example, for an above average grant proposal (+1SD) women would need to provide a quality of 119, compared to the 115 of men.

In this example, asymmetries in the target population (36% women) further increase in the pool of applicants (30% women) due to entry biases. The lack of proper representation of both groups in actual applications could have many different reasons but is thought to arise due to different opportunity costs, different resources available for grant writing, or perceived bias in the application procedure (see main article for further information). Together, including the biases in the review process, the average idea quality of accepted grants increases to 125 or 1.6 SDs, but lowers the representation of the disadvantaged group (here women) in successful grants to 23%. The average idea quality of successful grants is 125.

[3] https://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/studien/studie_gender_effects.pdf

2. Two-stage scenario

The two-stage scenario for funding reflects the grant-application process as for example by the European Research Council (ERC)[4]. Here, on a call-for-submission, applicants submit an extended synopsis and full proposal at Stage 1 (including the state-of-the-art and preliminary work, the objectives and work program, as well detailed information on the requested funds together with some general information; e.g., CV, work contract). In the ERC grant-application process, it is explicitly recommended that only "high risk, high gain" applications that are "highly competitive" should be submitted (source: https://erc.europa.eu/apply-grant/starting-grant). Selection of proposals is conducted on the "basis of excellence as the sole criterion". At Stage 1, the extended synopsis together with the applicant's track record and CV will be evaluated only. Each proposal is reviewed by several experts in the field and discussed in an in-depth panel meeting. The budget needed to fund the number of proposals invited for Stage 2 is up to three times the panel budget would all be accepted. At Stage 2, the full proposal will be evaluated. Here, selected applicants are also invited for an interview (30min) with the panel members. Only proposals that fully meet the ERC's excellence criterion are recommended for funding. However, these projects will be funded in priority order based on their rank. "This means that it is very likely that not all proposals scored 'A' will eventually be funded by the ERC."[5].

Simulation for process effectiveness in terms of economic burden

For the two-stage scenario we used the following configurations to simulate the costs (Table S3). All parameter settings are assumed to reflect actual investments and basic

[4] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/information-for-applicants_he-erc-stg-cog_en.pdf
[5] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/information-for-applicants_he-erc-stg-cog_en.pdf

parameters such as number of applicants only roughly. The funding volume and the target grant volume were kept constant for comparison purposes.

Table S3. Setup of the two-stage scenario

Configuration of general selection characteristics	
Number of applications	*N* = 200
Number of selection stages	2
Number of grants passing stage 1	60
Number of finally accepted grants	20
Target grant volume *(MU)*	500,000
Investments for application	
Stage 1 competitive review	
Number of people working on application	1
Average amount of work (*h*) per applicant	160
Average costs per hour of the applicant *(MU)*	50
Stage 2 competitive review	
Number of people working on application	1
Average amount of work (*h*) per applicant	80
Average costs per hour of the applicant *(MU)*	50
Investments for review	
Stage 1 competitive review	
Number of people reviewing the application	2
Average amount of work *(h)* per reviewer	8
Average costs per work hour of reviewer *(MU)*	50
Stage 2 competitive review	
Number of people reviewing the application	4
Average amount of work *(h)* per reviewer	16
Average costs per work hour of reviewer *(MU)*	50

Note. MU = monetary units; multiple reviews are aggregated by the arithmetic mean (instead of the minimum or maximum being used for selection purposes); ratings are ranked for selection purposes.

In summary, adopting these parameters for the two-stage scenario yields overall 320 working hours per application (equaling 16,000 *MU* cost per application). In total, across all *N*=200 applicants this scenario has cumulative costs of 2,192,000 *MU* which would be more than 20% of the tendered funds of 10 Million. Most of the costs (84%) are sunk costs in the sense that the predominant investment of work, by applicants as well as reviewers, was on applications that were finally rejected (see Figure 2, main article).

Simulation for process effectiveness in terms of diversity and quality

To simulate the quality and the impact of bias in the selection process in the two-stage scenario we adopted the following parameters (Table S4).

Table S4. Bias and quality simulations in the two-stage scenario

Sampling configuration	
Number of simulated samples	1000
Group proportions	36 % women
Entry bias	22 %
Grant characteristics	
Idea quality (mean)	100
Idea quality (standard deviation)	15
Idea quality (skew)	1.5
Review error and biases in quality assessment Stage 1	
Underestimation of disadvantaged group *(AU)*	“medium” (-4)
Standard deviation of error *(AU)*	“very large” (20)
Review error and biases in quality assessment Stage 2	
Constant underestimation *(AU)*	“small” (-2)
Relative underestimation of better ideas	“large” (-0.07 x idea quality)
Error spread *(AU)*	“medium” (SD 10)

Note. Group proportions: "In the past, far fewer women applied for ERC grants than men – from 2009 to 2018, on average, 26% of grant applications were from women. To put this in context, it is worth noting that the lower share of women in the ERC is likely to reflect the overall situation in terms of gender balance in science in Europe. In addition, there are major variations between applications to different research disciplines".[6]; *AU* = arbitrary units for idea quality.

Here, we assumed a different set of biases. In this example, asymmetries in the target population (36% women) further increase (28% women) in this scenario in the pool of applicants due to assumed entry biases. While the review in Stage 1 is again set to a constant underrating for the disadvantaged group, the standard deviation of the reviewer error increased to 20 points due to the shorter proposal for review. For Stage 2, we implemented a “brilliance bias”; i.e., with increasing quality the disadvantaged group is

[6] https://erc.europa.eu/news-events/magazine/woman-science-gender-balancing-erc

assumed to systematically receive lower ratings. While we assumed fewer biases in the evaluation of lower rated applications, biases increase for exceptional proposals with a weight of -.07 in idea quality. In addition, we assumed that reviewers are thought to be more accurate with an overall reduced width of error (*SD* = 10). Based on these exemplary parameters the representation of the disadvantaged group decreases from 28 % in the original applicant pool, to 24 % in the group of applicants passing Stage 1 and down to 21 % in the finally accepted applications. The average idea quality increased from 114 for those applications passing Stage 1 to 128 in the successful grants at Stage 2.

3. Tiebreaker lottery

The tiebreaker lottery, as applied by the Swiss National Science Foundation (SNSF), uses randomization to decide between the best grant-applications (Nature editorial 2022). In the tiebreaker lottery, Stage 1 is comparable to the above scenarios and based on excellence as the sole criterion for the advancement to Stage 2. However, here, at Stage 2 instead of conducting a full review, panel meeting, or interview, decision about who to fund is determined by a tiebreaker lottery. Depending on the funding agency there are some variations in who to allocate to the tiebreaker lottery. At the German Volkswagen Stiftung[7], the 20% best proposals from Stage 1 are accepted instantaneously, while another 20% of applications that advanced to Stage 2 are entered in a lottery.

Simulation for process effectiveness in terms of economic burden

For the tiebreaker lottery we used the following configurations to simulate the costs (Table S5). All parameter settings are assumed to reflect actual investments and basic parameters such as number of applicants only roughly. The funding volume and the target grant volume were kept constant for comparison purposes.

Table S5. Setup of the tiebreaker lottery

Configuration of general selection characteristics	
Number of applications	N = 200
Number of selection stages	2
Number of grants passing stage 1	40
Number of finally accepted grants	20
Target grant volume *(MU)*	500,000
Investments for application	
Stage 1 competitive review	
Number of people working on application	1
Average amount of work (h) per applicant	80
Average costs per hour of the applicant *(MU)*	50
Stage 2 tiebreaker lottery	
Number of people working on application	1

[7] https://www.volkswagenstiftung.de/en/how-it-works-partially-randomized-selection-process-within-initiative-experiment

Average amount of work (*h*) per applicant	0
Average costs per hour of the applicant *(MU)*	50
Investments for review	
Stage 1 competitive review	
Number of people reviewing the application	2
Average amount of work *(h)* per reviewer	16
Average costs per work hour of reviewer *(MU)*	50
Stage 2 tiebreaker lottery	
Number of people reviewing the application	--
Average amount of work *(h)* per reviewer	--
Average costs per work hour of reviewer *(MU)*	--

Note. MU = monetary units; investments of applicants and reviewers are reflecting internal and external costs.

In summary, adopting these parameters for the tiebreaker lottery yields overall 112 working hours per application (equaling 5,600 *MU* cost per application). In total, across all *N*=200 applicants the tiebreaker lottery has an overall process cost of 1,120,000 MU, which is nearly 10% of the distributed funds. Most of the costs (90%) are sunk costs in the sense that the predominant investment of work, by applicants as well as reviewers, was on applications that were finally rejected (see Figure 2, main article).

Simulation for process effectiveness in terms of diversity and quality

To simulate the quality and the impact of bias in the selection process in the tiebreaker lottery we adopted the following parameters (Table S6).

Table S6. Bias and quality simulations in the tiebreaker lottery

Sampling configuration	
Number of simulated samples	1000
Group proportions	36 % women
Entry bias	22 %
Grant characteristics	
Idea quality (mean)	100
Idea quality (standard deviation)	15
Idea quality (skew)	1.5
Stage 1 review error and biases in quality assessment	
Underestimation of disadvantaged group	“medium” (-4)

Error spread	“medium” (SD 20)
Stage 2 tiebreaker lottery	

Note. "Nevertheless, a clear imbalance can also be seen in this area. While over 60% of applicants under the Doc.CH scheme (PhD level) are women, 40% of applicants under the Ambizione scheme (young, nearly independent researchers) are women. Under the project funding scheme (experienced researchers), the proportion of women is 34%."[8].

In this example, asymmetries in the target population (36% women) further increase (28% women) in this scenario in the pool of applicants due to assumed entry biases. In addition to the entry bias, to account for biases in the review process at Stage 1, we assumed continuous underestimation of the idea quality of the disadvantaged group by 4 units with a medium spread of reviewer error (SD = 20). These exemplary parameters led to a drop in proportions for the disadvantaged group from 36% over 28% in the applicant pool to 24% due to the cumulative effects of entry and review biases. The average idea quality of selected applications is 121.

[8] https://data.snf.ch/stories/proportion-of-women-decreasing-each-step-career-ladder-en.html

4. Pre-lottery scenario

The pre-lottery scenario reflects a "radical diversity" approach and so far, as to our knowledge, has not been applied yet. Here, the funding agency would adopt an automated selection process (e.g., all researchers with a PhD in a given area) in which all candidates are entered into the lottery at the first stage of the process without already having to spend any resources writing a grant proposal (i.e., avoiding entry bias). If a researcher is randomly selected, they move to the next phase where they have time to write up and submit their proposal (including the state-of-the-art and preliminary work, the objectives and work program, as well detailed information on the requested funds) or opt out of the process. Invited applications, within a defined amount of time (e.g., within 4 months), will then be evaluated in a normative process (i.e., similar to how the decision process is handled by journal editors).

Simulation for process effectiveness in terms of economic burden

For the pre-lottery scenario we used the following configurations to simulate the costs (Table S7). All parameter settings are assumed to reflect actual investments and basic parameters such as number of applicants only roughly. The funding volume and the target grant volume were kept constant for comparison purposes.

Table S7. Setup of the pre-lottery scenario

Configuration of general selection characteristics	
Number of applications	200
Number of selection stages	2
Number of grants passing lottery at Stage 1	50
Normative cut off in idea quality	105
Average accepted grant volume *(MU)*	500,000
Investments for application	
Stage 1 pre-lottery scenario	
Number of people working on application	1
Average amount of work (*h*) per applicant	0
Average costs per hour of the applicant *(MU)*	50
Stage 2 normative review	

Number of people working on application	1
Average amount of work (*h*) per applicant	160
Average costs per hour of the applicant *(MU)*	50
Investments for review	
Stage 1 pre-lottery scenario	
Number of people reviewing the application	--
Average amount of work *(h)* per reviewer	--
Average costs per work hour of reviewer *(MU)*	--
Stage 2 normative review	
Number of people reviewing the application	3
Average amount of work *(h)* per reviewer	8
Average costs per work hour of reviewer *(MU)*	50

Note. MU = monetary units.

yields overall 184 working hours per application (equaling 5,600 *MU* cost per application). In summary, adopting these parameters for the pre-lottery scenario yields overall 184 working hours per application (equaling 9,200 *MU* cost per application). In total, across all *N* = 200 applicants this scenario has cumulative costs of 460,000 *MU,* roughly 5% of the distributed funds. The sunk costs in this scenario are comparably smaller with 32 % of the process costs invested in finally rejected applications. Due to the normative cut-off, on average 16.25 applications were accepted across all 1000 simulated samples.

Simulation for process effectiveness in terms of diversity and quality

To simulate the quality and the impact of bias in the selection process in the pre-lottery scenario we adopted the following parameters (Table S8).

Table S8. Bias and quality simulations in the pre-lottery scenario

Sampling configuration	
Number of simulated samples	1000
Group proportions	36 % women
Entry bias	0.1 %
Grant characteristics	
Idea quality (mean)	100
Idea quality (standard deviation)	15
Idea quality (skew)	1.5

Stage 1 pre-lottery scenario	
Stage 2 review error and biases in quality assessment of normative review	
Underestimation of disadvantaged group	“small” (-2)
Error spread	“small” (SD 5)

In this scenario, asymmetries in the target population (36% women) do not increase due to assumed fairness of the lot. Thus, the proportion of women in the target population equal the proportion in the applicant pool. At Stage 2, we assumed a more thorough review process with more reviewers and better training to even further reduce biases. The better training of reviewers can be financed by the overall lower cost of this scenario. The underestimation of the quality in the disadvantaged group was 2 points and the spread of reviewer error was smaller with an *SD* = 5. Given these assumptions and a negligible entry bias the disadvantaged group was represented with 32 % while the average idea quality of accepted applications was 117.

References

Attali, Dean. 2021. "Shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds." https://CRAN.R-project.org/package=shinyjs.

Bailey, Eric. 2022. "ShinyBS: Twitter Bootstrap Components for Shiny." https://CRAN.R-project.org/package=shinyBS.

Chang, Winston, and Barbara Borges Ribeiro. 2021. "Shinydashboard: Create Dashboards with 'Shiny.'" https://CRAN.R-project.org/package=shinydashboard.

Chang, Winston, Joe Cheng, J. J. Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. 2021. "Shiny: Web Application Framework for R." https://CRAN.R-project.org/package=shiny.

Nature editorial. 2022. "The Case for Lotteries as a Tiebreaker of Quality in Research Funding." *Nature* 609 (653): 653–653. https://doi.org/10.1038/d41586-022-02959-3.

Perrier, Victor, Fanny Meyer, and David Granjon. 2022. "ShinyWidgets: Custom Inputs Widgets for Shiny." https://CRAN.R-project.org/package=shinyWidgets.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

RStudio Team. 2021. *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, PBC. http://www.rstudio.com/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software*. https://doi.org/10.21105/joss.01686.

Witteman, Holly O., Michael Hendricks, Sharon Straus, and Cara Tannenbaum. 2019. "Are Gender Gaps Due to Evaluations of the Applicant or the Science? A Natural

Experiment at a National Funding Agency." *The Lancet* 393 (10171): 531–40. https://doi.org/10.1016/S0140-6736(18)32611-4.

Wuertz, Diethelm, Yohan Chalabi, Tobias Setz, and Martin Maechler. 2022. "FGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling." https://CRAN.R-project.org/package=fGarch.

Zhang, Le. 2022. "SpsComps: 'systemPipeShiny' UI and Server Components." https://CRAN.R-project.org/package=spsComps.