



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2023

Sollten Förderentscheidungen partiell durch Lotterien ersetzt werden?

Mutz, Rüdiger ; Wolbring, Tobias ; Daniel, Hans-Dieter

DOI: <https://doi.org/10.31244/zfe.2023.02.04>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-256209>

Journal Article

Published Version

Originally published at:

Mutz, Rüdiger; Wolbring, Tobias; Daniel, Hans-Dieter (2023). Sollten Förderentscheidungen partiell durch Lotterien ersetzt werden? *Zeitschrift für Evaluation*, 22(2):243-274.

DOI: <https://doi.org/10.31244/zfe.2023.02.04>

DIGITALER SONDERDRUCK

Zeitschrift für
Evaluation,
22. Jahrgang,
Heft 2/2023
ISSN 1619-5515

© Waxmann
Verlag GmbH,
2023

2 · 2023
22. Jahrgang

Zeitschrift für Evaluation

Originalbeiträge

Hammersley: Was the UK Government's Policymaking ›Evidence-Based‹ During the Pandemic? Reflections on Science and Politics

Mutz/Wolbring/Daniel: Sollten Förderentscheidungen partiell durch Lotterien ersetzt werden? Grundlagen, Konzeption und Evaluationsdesign am Beispiel der Förderinitiative ›Experiment!‹ der VolkswagenStiftung

Bender/Diller: Business-Improvement Districts. Ein Beitrag der Evaluationsforschung zur Bilanz eines Instruments privatwirtschaftlich orientierter Stadtentwicklungspolitik

Praxisberichte

Hense/Rädiker: Künstliche Intelligenz in der Evaluation. Praxisbericht zum Online-Seminar ›ChatGPT & Co.‹

Meyer/Walprecht/Petersen/Gerls/Kühnhenrich: Kooperationsprojekt zur Evaluierung der Flexirente

WAXMANN



WAXMANN

Steinfurter Str. 555
48159 Münster

Fon 02 51 – 2 65 04-0
Fax 02 51 – 2 65 04-26

info@waxmann.com
www.waxmann.com

Mehr zur ZfEval
finden Sie [hier](#).

Sollten Förderentscheidungen partiell durch Lotterien ersetzt werden?

Grundlagen, Konzeption und Evaluationsdesign am Beispiel der Förderinitiative „Experiment!“ der VolkswagenStiftung

Rüdiger Mutz,¹ Tobias Wolbring,² Hans-Dieter Daniel³

Zusammenfassung: Angesichts der Kritik am Peer-Review-Verfahren haben einige Forschungsförderungsorganisationen ein Lotterie-Element in die Auswahlverfahren für ausgewählte Förderinitiativen eingeführt. Ziel dieses konzeptuellen Beitrags ist es, das Lotterie-Verfahren als Element eines Entscheidungsprozesses in der Forschungsförderung am Beispiel der Förderinitiative „Experiment!“ der VolkswagenStiftung (VWS) zu diskutieren. Dabei sollen zuerst die allgemeinen Grundlagen von Lotterien dargelegt werden, und zwar aus vier Perspektiven: des Verfahrens, der Statistik, der Ökonomie und der Sozialpsychologie. Diese generellen Überlegungen zu Lotterien sollen als Folie dienen, die Fachliteratur aufzuarbeiten. Beispielfhaft sollen die Förderinitiative „Experiment!“ der VWS dargestellt und ein Design für die Ex-post-Evaluation der Wirkungen des Lotterieelements im Auswahlverfahren entwickelt werden. Abschließend werden Empfehlungen formuliert für die Wirkungsevaluation derart modifizierter Auswahlverfahren.

Schlüsselwörter: Lotterie, Forschungsförderung, Programmevaluation, doppelt-randomisierter Präferenz-Testlauf

Abstract: In view of the criticism of the peer review process, some research funding organizations have introduced a lottery element into the selection procedures for selected funding initiatives. The aim of this conceptual paper is to discuss the lottery procedure as an element of a decision-making process in research funding, using the funding initiative “Experiment!“ of the Volkswagen Foundation (VWS) as an example. First, the general principles of lotteries will be presented, from four perspectives: the procedure, statistics, economics, and social psychology. These general considerations on lotteries will serve as a foil to review the literature on

1 Competence Center for Higher Education and Science Studies (CHESS), Universität Zürich

2 Fachbereich Wirtschafts- und Sozialwissenschaften, Friedrich-Alexander-Universität Erlangen-Nürnberg

3 Psychologisches Institut, Universität Zürich

'lotteries in research funding'. As an example, the funding initiative "Experiment!" of the VWS will be presented and a design for the ex-post evaluation of the effects of the lottery element in the selection process will be developed. Finally, recommendations for the impact evaluation of such modified selection procedures will be formulated.

Keywords: Lottery, Research Funding, Program Evaluation, Doubly Randomized Preference Trial

1. Einleitung

Das auf Begutachtungen von Fachkolleginnen und Fachkollegen beruhende System der Qualitätssicherung in der Forschung wird in den letzten Jahren mehr und mehr infrage gestellt (z.B. Gould, 2019; Roy, 1985; Sandström & Hällsten, 2008; Smith, 2006; van der Linden & Wise, 2014). So seien die Begutachtungen durch Fachkolleginnen und -kollegen wenig reliabel, d.h., sie würden in nicht ausreichendem Maße in ihren gutachterlichen Urteilen übereinstimmen. Darüber hinaus seien die Urteile nicht prädiktiv valide, d.h., sie können nicht in ausreichendem Maße den späteren tatsächlichen Erfolg eines Forschungsprojekts vorhersagen, und die Urteile seien in vielfacher Hinsicht verzerrt (z.B. Geschlechter-Bias). Dieser Kritik versuchen Forschungsförderungsorganisationen und wissenschaftliche Zeitschriften in unterschiedlicher Weise zu begegnen, zum einen durch inhouse durchgeführte Peer-Review-Forschung (z.B. der Schweizerische Nationalfonds), zum anderen durch Kanalisierung und Fokussierung der unabhängigen Peer-Review-Forschung (z.B. PEERE, RoRI) und schließlich durch neue Formen von Auswahlverfahren (z.B. „lottery“, „citizen review“).

Die Tendenz zur stärkeren *Kanalisierung und Fokussierung der Peer-Review-Forschung* wurde in den letzten Jahren in zwei Initiativen deutlich. So konnte im Bereich des Begutachtungswesens von wissenschaftlichen Zeitschriften im Rahmen des Projekts PEERE (COST Action) von 2014 bis 2018 eine Plattform mit anonymisierten Datensätzen zur Gutachtertätigkeit einer Vielzahl von Zeitschriften geschaffen werden, die Forschungsgruppen für vollumfängliche Datenanalysen nutzen können (Bravo et al., 2019; Squazzoni & Gandelli, 2012) und Ergebnisse auf eine wesentlich breitere Datengrundlage stellen als es bisher mit der Analyse von Daten einzelner Zeitschriften möglich war (z.B. Bornmann & Daniel, 2008b; Daniel, 2018). Im Bereich der Forschungsförderung muss sich die Analyse meist noch auf die Gesamtheit aller Dokumente beschränken, die im Rahmen des Begutachtungsprozesses einer einzelnen Forschungsförderungsorganisation in einer Zeitperiode angefallen sind, z.B. des Wissenschaftsfonds Österreichs (FWF) (Mutz et al., 2012a, 2012b, 2015, 2016). Daher unterstützen führende nationale Forschungsförderungsorganisationen (z.B. Wellcom Trust, Schweizerischer Nationalfonds (SNF), European Molecular

Biology Organization (EMBO)) das Research on Research Institute (RoRI) in Sheffield (UK), mit dem Ziel „bringing together people and organisations that care about research, gathering information and developing tools to inform and improve how research is funded, practised, communicated and evaluated“ (<https://researchonresearch.org/about>; Bendiscioli et al., 2021).

Mit der „fokalen Randomisierung“ schlägt das RoRI eine *neue Form von Auswahlverfahren für Forschungsanträge* vor (Bendiscioli et al., 2021; Buckley-Woods & Wilsdon, 2021), die als Antwort auf die mangelnde Güte des Peer-Review-Verfahrens „... as a way of bypassing the recognized difficulties of peer review...“ (Bedessem, 2020, S. 151) formuliert wird. An einer bestimmten Stufe des Entscheidungsprozesses über die Bewilligung eines Antrags wird eine Lotterie eingebaut, in dem der Zufall bestimmt, ob ein Antrag bewilligt wird oder nicht. Es gibt bisher mindestens vier Forschungsförderungsorganisationen, die für einen kleinen Teil ihrer Fördermittel Lotterieverfahren einsetzen (Buckley-Woods & Wilsdon, 2021, S. 15; Avin, 2019, S. 21): die Förderinitiative „Experiment!“ der VolkswagenStiftung (VolkswagenStiftung, 2019a), das „Explorer Grants“-Programm des Health Research Council of New Zealand (Health Research Council of New Zealand, 2018), und das „SEED Projects-Programm“ der neuseeländischen Science for Technological Innovation (Science for Technological Innovation, 2013). Der Schweizerische Nationalfonds hat ein Lotterie-Element für das Förderprogramm Mobility.Doc implementiert (Heyard et al., 2022; Schweizer Nationalfonds, 2018; Weber, 2022). Im Jahr 2021 wurde die Möglichkeit von Verlosungen auf alle Förderinstrumente ausgeweitet.⁴

Zum Thema „Lotterien in der Wissenschaft“ ist zwar inzwischen eine Vielzahl von Literatur veröffentlicht worden, jedoch handelt es sich meist um theoretische Beiträge mit einem spezifischen Fokus. Es mangelt an Übersichtsarbeiten, die versuchen, den Forschungsstand in ihrer ganzen Breite wiederzugeben (z.B. Bedessem, 2020; Buckley-Woods & Wilsdon, 2021). Des Weiteren fehlt es an empirischen Studien (z.B. Barlösius & Philipps, 2021; Liu et al., 2020; Röbecke & Simon, 2020), insbesondere an Studien, welche die Wirkung von Förderprogrammen mit Lotterie-Elementen im Auswahlverfahren untersuchen bzw. eine Wirkungsevaluation durchführen (Daniel et al., 2021; Jaffe, 2002; Treischl & Wolbring, 2020). So lassen sich Maßnahmen als Treatments in einem klassischen Experiment auffassen, dessen Wirksamkeit gegenüber einer Kontrollgruppe ohne Treatment geprüft werden kann. Beispielsweise kann geprüft werden, welche Wirkungen von einem „Blinding“ von Anträgen oder Manuskripten ausgehen. Experimentelle Peer-Review-Forschung hat eine lange Tradition und lässt bei einem geeigneten Versuchsdesign kausale Aussagen über die Wirksamkeit einer Maßnahme zu (z.B. Leibbrandt et al., 2018; Mahoney, 1977).

Ziel dieses Beitrags ist es, auf der Grundlage einer breiten Literaturübersicht für das Auswahlverfahren der Förderinitiative „Experiment!“ der VolkswagenStiftung ein methodisches Design für die Wirkungsevaluation vorzuschlagen, das

4 Siehe <https://www.snf.ch/de/JyifP2I9SUo8CPxI/news/news-210331-das-los-kann-entscheiden> [11.04.2023].

auch für die Evaluation anderer Förderprogramme mit Lotterie-Elementen angewendet werden kann. Zunächst soll auf der Grundlage einer Literaturrecherche die Frage beantwortet werden, warum Lotterien in Entscheidungsprozessen, insbesondere bei Förderentscheidungen von Vorteil sind und welche Probleme entstehen. Die Literaturrecherche erfolgte mittels der Suche in verschiedenen Datenbanken (Elsevier Scopus, Google Scholar, Clarivate Web of Science Core Collection) mit Schlüsselwörtern („funding“, „lottery“, „focal randomization“, „grant peer review“ ...) inklusive der zitierten und zitierenden Arbeiten. Das Ergebnis der Recherche wird in Form von vier verschiedenen Perspektiven (verfahrenstechnisch, statistisch, ökonomisch, sozialpsychologisch) systematisiert, um Bedingungen herauszuarbeiten, die erfüllt sein müssen, damit Lotterien sinnvoll angewendet werden können (Abschnitt 2). Diese Ausführungen dienen als Folien, um die themenspezifische Literatur zur Forschungsförderung mittels Lotterien aufzuarbeiten und einzuordnen (Abschnitt 3). Im Anschluss soll die Förderinitiative „Experiment!“ der VolkswagenStiftung (VWS) vorgestellt werden als bereits implementiertes Förderprogramm mit einem Lotterieelement im Auswahlverfahren (Abschnitt 4). Diesem schließt sich ein Vorschlag für ein Evaluationsdesign an, das der experimentellen bzw. quasi-experimentellen Forschung in den Sozialwissenschaften verpflichtet ist (Abschnitt 5). Zum Abschluss sollen Empfehlungen für die Konzeption und Evaluation von Auswahlverfahren mit Lotterieelementen gegeben werden.

2. Lotterien als Element von Auswahlentscheidungen

2.1 Übersicht

Im Folgenden soll die Logik von Lotterien im Zusammenhang mit Auswahlentscheidungen analysiert werden, um Bedingungen zu formulieren, wann eine Lotterie sinnvoll ist und wann nicht. Das Lotterie-Verfahren soll unter vier verschiedenen Perspektiven analysiert werden, die letztlich schon Boyle (1998) in einer der ersten Arbeit zu Lotterien in Entscheidungsprozessen nannte:

- *Verfahrensperspektive*: Lotterien als Element eines mehrstufigen Entscheidungsprozesses,
- *statistische Perspektive*: Lotterien als statistischer Selektionsmechanismus der zufälligen Zuweisung zu Gruppen,
- *ökonomische Perspektive*: Lotterien als Verfahren der Allokation von Gütern unter der Bedingung von Knappheit,
- *sozialpsychologische Perspektive*: Lotterien mit Wirkungen auf die Motivation und Emotion der Betroffenen in Leistungssituationen.

2.2 Verfahrensperspektive

Bevor das Lotterieverfahren als explizites Element einer Auswahlentscheidung diskutiert wird, muss geklärt werden, was eine Lotterie eigentlich ist. Das Werfen einer Münze, das Würfeln, oder das Ziehen eines Strohhalms sind bekannte Beispiele für eine Lotterie. Eine Definition von Lotterie gibt Stone (2011):

Eine Lotterie ist ein Prozess, der in der Lage ist, eine Menge von Ergebnissen zu generieren, wobei ein spezifisches Ergebnis, wann immer der Prozess stattfindet, aufgrund der verfügbaren Informationen nicht vorhersagbar ist (Stone, 2011, S. 20, von den Autoren dieses Beitrages ins Deutsche übersetzt).

Zwei wichtige Elemente dieser Definition erfordern weitere Erläuterungen. So handelt es sich bei einer Lotterie um einen Prozess und nicht um einen Gegenstand oder Apparat. Nicht die Münze macht eine Lotterie aus, sondern das Werfen einer Münze, d.h., es werden bestimmte klar definierte diskrete Ergebnisse mit endlicher Zahl durch eine Lotterie generiert (z.B. Kopf oder Zahl). Das zweite wichtige Element ist das der Unvorhersagbarkeit unter der Voraussetzung der verfügbaren Information. Eine Vorhersage des Lotterieverganges aus der vorhandenen Information ist rational nicht möglich. Zwei weitere Bedingungen im Hinblick auf den formalen Prozess sollten erfüllt sein: So sollte von einer Gleichwahrscheinlichkeit ausgegangen werden, d.h., bei einem Münzwurf sollte die Wahrscheinlichkeit ‚Kopf‘ zu erhalten 0.5 sein. Münzwürfe mit einer ungleichen Wahrscheinlichkeit z.B. von 0.6 wären zwar auch zufällig, dennoch würde ein Ergebnis besser vorhersagbar sein als ein anderes Ergebnis (gewichtete Lotterie). Des Weiteren sollten Muster z.B. ZZKKZZKK keine Vorhersage ex-ante erlauben können, beispielsweise die Regel „Auf zweimal Kopf folgt zweimal Zahl“. Das schließt nicht aus, dass in einer Ex-post-Betrachtung solche Muster per Zufall aufgetreten sind. Je weniger ein Ergebnis vorhersagbar ist, desto fairer ist die Lotterie.⁵ Dieses Vorgehen impliziert darüber hinaus, dass eine Lotterie umso fairer ist, je weniger Möglichkeiten zu deren Manipulation bestehen. Das Lotterie-Verfahren ist umso stärker gegenüber Manipulation geschützt, je mehr Informationen über den Lotterie-Prozess vorliegen. So werden Lotterien häufig öffentlich durchgeführt, um damit Manipulation möglichst auszuschließen.

Lotterien stehen nicht für sich, sondern werden im Rahmen eines mehrstufigen Entscheidungsprozesses eingesetzt. In einem ersten Schritt wird aus den vorhandenen Optionen ein Set von Optionen entsprechend bestimmter Kriterien herausgefiltert. Für die Optionen des finalen Sets wird eine Vielzahl von Gründen für ihre Auswahl bestimmt, die zusammengefasst ihre Stärke bestimmen. Eine Rangordnung der Optionen nach der Stärke erlaubt die finale Auswahl der Option, die dann auch akzeptiert wird. Die übrigen Optionen werden abgelehnt. Es kann aber der Fall eintreten, dass sich die Stärke der Gründe für ein kleines Set von Optionen nicht be-

5 Fairness wird hier implizit im Hinblick auf die „Equality“-Norm interpretiert (Deutsch, 1975). Alle Ergebnisse werden gleichbehandelt, unabhängig vom Einsatz. Unter Maßgabe der „Equity“-Norm würde zusätzlich der Einsatz einbezogen werden.

deutsam unterscheidet (Gleichstand). Dieses Set hebt sich darüber hinaus von den übrigen Optionen deutlich ab. In diesem Fall kann nach Stone (2011, S. 32) wie folgt ein Lotterie-Element in den Entscheidungsprozess eingebaut werden:

1. Ein Filter erzeugt eine Liste von Optionen, die Gegenstand der Lotterie sind.
2. Es wird ein Lotterieverfahren gewählt, wobei die Anzahl der diskreten Ausgänge dieser Lotterie der Zahl der Optionen entspricht.
3. Jedem Ausgang der Lotterie wird eine Option zugeordnet.
4. Die Lotterie wird ausgeführt und ein Ausgang generiert.
5. Es werden alle Optionen ausgefiltert außer der, die dem Ausgang der Lotterie entspricht. Diese Option wird ausgewählt.

Berger, Osterloh und Rost (2020) sprechen auf der Grundlage der Ergebnisse einer solchen Zufallsauswahl von „focal random selection“, einer Vorselektion mit anschließender Randomisierung von Forschungsvorhaben oder Manuskripten. Jaffe (2002) verwendet hierfür den Begriff „randomization at the margin“.

Lotterien schaffen Surrogat-Gründe für Entscheidungen. In Lotterien mangelt es den Akteuren an jeglicher Information über den Ausgang der Entscheidung. Diese Unvorhersagbarkeit kann prinzipiell auch für andere Entscheidungen wie den Ausgang politischer Wahlen gelten. Dennoch wird in diesen Fällen nicht auf Lotterien zurückgegriffen, da wichtige Informationen und Gründe für oder gegen eine Option vorliegen: Legitimität, Bedürftigkeit, Verdienst oder Bereitschaft zu bezahlen (z.B. Auktionen). Prinzipiell sind drei Arten von Gründen zu unterscheiden: Erstens positive Gründe, d.h. Gründe, die für eine Entscheidungsoption sprechen. Zweitens neutrale Gründe, die für die Entscheidung irrelevant sind, z.B. das Wetter für die Wahl des qualitativ besten Restaurants, und drittens negative Gründe. Damit sind Gründe gemeint, die eigentlich den Entscheidungsprozess nicht beeinflussen sollten (z.B. Nationalität, Geschlecht, ethnische Zugehörigkeit ...). Lotterien erlauben Entscheidungen ohne jegliche Referenz auf Gründe (Surrogat-Gründe). Damit sind diese Entscheidungen unabhängig von positiven, aber eben auch von negativen Gründen. Nach Boyle (1998, S. 295–296) sollte ein Selektionsprozess erstens effizient sein. Es sollten nicht unnötig Zeit und Ressourcen seitens der betroffenen Individuen aufgewendet werden müssen. Zweitens sollten klare Cut-offs (z.B. Ablehnung bei mittlerer Note der Gutachtenden von 80) vermieden werden zugunsten von Zonen der Unbestimmtheit der Entscheidung, wo keine Unterschiede zwischen den Personen oder hier Forschungsanträgen mehr bestehen („not significantly different“, Boyle, 1998, S. 295). Drittens sollten Verzerrungen („biases“) vermieden werden, d.h., Personen oder Anträge sollten nicht aufgrund des Geschlechts, der ethnischen Zugehörigkeit oder anderer Merkmale der antragstellenden Person unterschiedlich behandelt werden, und viertens sollte der Auswahlprozess frei von Korruption sein (Boyle, 1998, S. 296). Lotterien erfüllen nach Boyle diese Kriterien.

Zusammenfassend ist eine Lotterie dann sinnvoll, wenn Entscheidungen erstens unbestimmt sind, d.h. mehrere gleichwertige Optionen vorliegen, zweitens keine

weiteren positiven Gründe mehr bestehen, eine Option einer anderen Option vorzuziehen, und drittens, wenn die Gefahr besteht, dass negative Gründe ausschlaggebend werden. Stone spricht von einem reinigenden Effekt der Lotterie (Stone, 2011, S. 36).

2.3 Statistische Perspektive

Bei der auf das Auswahlverfahren abhebenden Sichtweise bleibt das rein statistische Moment der Zufallsauswahl unbeachtet. Zufallsauswahlen haben im Bereich der Statistik, insbesondere im Bereich der kausalen Inferenz, eine zentrale Bedeutung. Im Potential-Outcome-Modell (Rubin, 2007, 2011) ist ein randomisiertes Kontrollgruppenexperiment dadurch definiert, dass für ein Individuum sowohl ein (potentielles) Ergebnis unter der Bedingung des Treatments als auch ein (potentielles) Ergebnis unter der Kontrollbedingung anzunehmen ist. Dieser auch als Rubin Causal Model bekannte Ansatz hat sich insbesondere für Evaluationsfragestellungen als fruchtbar erwiesen (Caspari, 2009; Wolbring, 2014). Ein individueller kausaler Effekt ist definiert als Unterschied zwischen den zwei Ergebnissen für ein Individuum. Das fundamentale Problem der kausalen Inferenz besteht darin, dass in einem konkreten Experiment letztlich für ein Individuum nur ein Ergebnis realisiert wird, und zwar dadurch, dass ein Individuum einer der beiden Gruppen (Kontroll- oder Treatmentbedingung) zugewiesen wird. Ein individueller kausaler Effekt ist auf dieser Basis zwar nicht identifizierbar (Holland, 1986, S. 947), aber unter bestimmten Umständen kann zumindest der durchschnittliche kausale Effekt zwischen Treatment und Kontrollgruppe über alle Individuen hinweg bestimmt werden. Diese Inferenz ist möglich unter der Bedingung der Zufallszuweisung der Individuen zu den Gruppen.

Mit der Zufallszuweisung wird erreicht, dass bei ausreichender Stichprobengröße Treatment- und Kontrollgruppe ausbalanciert, d.h. im Erwartungswert hinsichtlich aller beobachteten und nicht beobachteten Variablen vergleichbar sind (siehe Mutz & Daniel, 2012; Mutz et al., 2017). Letztlich ist ein Lotterie-Verfahren dahingehend fair, als es unterschiedliche Personengruppen (Männer und Frauen, junge und ältere Wissenschaftlerinnen und Wissenschaftler, Deutsche und Nicht-Deutsche) in gleicher Art und Weise behandelt.

2.4 Ökonomische Perspektive

Bei Auswahlentscheidungen, die ein Lotterie-Verfahren verwenden, geht es häufig um die Verteilung von knappen Gütern an diejenigen, die einen Anspruch auf das Gut erheben, beispielsweise die Vergabe von Wohnungen an eine Gruppe von Wohnungsinteressierten. Die Tatsache, dass es sich bei vielen Lotterie-Entscheidungen um ökonomische Entscheidungen handelt, hat weitere Implikationen für die Organisation der Lotterie und für die Gestaltung des Auswahlprozesses.

Stone (2011, S. 50, 2013, S. 581) nennt folgende vier Bedingungen für eine „gerechte Lotterie“ (von den Autoren dieses Beitrags ins Deutsche übersetzt):

- B1: *Das Gut muss stückweise vorliegen.* Das Gut muss in diskreten Einheiten vorliegen, die nicht weiter unterteilt werden können, ohne den Wert des Gutes zu mindern. Wohnungen oder Stipendien sind typische Beispiele für solche Güter.
- B2: *Das Gut muss homogen sein.* Einheiten desselben Gutes müssen austauschbar sein. Eine Einheit des Gutes darf nicht gegenüber einer anderen Einheit bevorzugt werden. Das ist mitunter schwierig zu erfüllen. Bestimmte Wohnungen sind für bestimmte Wohnungsinteressierte nicht geeignet, z.B. Wohnungen für Ältere im obersten Stock ohne Aufzug.
- B3: *Die Ansprüche auf das Gut müssen homogen sein.* Jede Partei hat Anspruch genau auf eine Einheit des Gutes und nicht mehr. Damit werden Entscheidungen ausgeschlossen, zwei oder mehr Einheiten eines Gutes einer Partei, oder jeweils eine Einheit mehreren Parteien zuzuteilen. So können nicht zwei gleichwertige Stipendien an eine Person vergeben werden.
- B4: *Die Ansprüche variieren allein in ihrer Stärke.* Die Ansprüche sind zwar qualitativ homogen, aber können quantitativ in ihrer Stärke variieren. Was Stärke bedeutet, bestimmt der jeweilige Kontext. Für die Vergabe eines Stipendiums können die vorherigen Leistungen im Studium oder die Qualität eines Projektantrags die Stärke des Anspruchs bestimmen.

Wenn die obigen vier Bedingungen erfüllt sind, wäre ein Lotterie-Verfahren strenggenommen nicht erforderlich, da die finalen Parteien im ersten Schritt aufgrund bestimmter Kriterien ausgewählt werden, und im zweiten Schritt die Partei ausgewählt wird, die den stärksten Anspruch hat. Lotterien sind dann angezeigt, wenn sich Unbestimmtheit einstellt: Es gibt eine Untermenge von Parteien, deren Ansprüche in ihrer Stärke gleich sind (Gleichstand).

Darüber hinaus stellen sich aus ökonomischer Perspektive Fragen nach den Anreizwirkungen einer fokal randomisierten Zufallsauswahl. Unter der Annahme eines rationalen Akteurs könnte im Falle einer fokal randomisierten Zufallsauswahl eine attraktive Strategie darin bestehen, eine größere Anzahl von Anträgen einzureichen, die weniger hochwertig, jedoch hinreichend gut sind, um die Stufe der Lotterie-Auswahl zu erreichen.

2.5 Sozialpsychologische Perspektive

Lotterien können im Erleben der Betroffenen eine ganz andere Bedeutung haben, als es ökonomische Rationalitätskonzepte implizieren. Ein wichtiger sozialpsychologischer Ansatz, der den Zufall als Ergebnis von Handlungen einbezieht, ist die Attributionstheorie (Försterling & Stiensmeier-Pelster, 1994). Personen haben das Bedürfnis, akzeptable Erklärungen für Ereignisse zu finden, die für sie bedeutsam sind. Akzeptabel sind ihnen Erklärungen, die das Selbstwertgefühl aufrechterhalten und ihrem Bedürfnis nach Wertschätzung gerecht werden – unabhängig davon, ob sie ökonomisch rational sind oder nicht (Schnotz, 2011, S. 103–104.). Bereits Elster (1989) wies im Zusammenhang mit Lotterien in Auswahlverfahren darauf hin, dass

Betroffene nach einer Erklärung für eine Entscheidung suchen, selbst wenn diese nach dem Zufall gefällt wurde.

Die unterschiedliche Ursachenzuschreibung von Ereignissen hat auch unterschiedliche Konsequenzen auf Erleben und Verhalten der Personen. Ein wichtiger Bereich betrifft die Leistungsmotivation. Nach Weiner (1985) lassen sich die Möglichkeiten, in Leistungssituationen Erfolg und Misserfolg zu attribuieren, nach drei Dimensionen ordnen: a) internal oder external, b) stabil oder variabel und c) kontrollierbar oder nicht kontrollierbar. Internal wird attribuiert, wenn eine Person die Erklärung von Leistungen bei sich selbst sieht (z.B. mangelnde Anstrengung), external, wenn eine Person die Erklärung von Leistungen in der Umwelt sieht (z.B. Schwierigkeit der Prüfung). Stabil ist eine Ursache, wenn sie über Zeit und Situation konstant ist (z.B. Fähigkeiten), variabel, wenn sie von Situation zu Situation sich ändert (z.B. Pech, Anstrengung). Kontrollierbar sind Leistungen dann, wenn Personen subjektiv meinen, die Kontrolle über Ereignisse zu haben (z.B. Fähigkeit). Aus diesen drei binären Dimensionen lässt sich kombinatorisch eine Tabelle mit $2 \times 2 \times 2 = 8$ Zellen generieren, wobei in jeder Zelle ein Ergebnis einer Attribution steht. So ist die ‚Fähigkeit‘, die sich eine Person zuschreibt, beispielsweise eine interne, stabile, nicht kontrollierbare Ursache. Die Attribution einer Ursache hat unmittelbare Auswirkungen auf die Emotionen und die Leistungsbereitschaft der attribuierenden Person. Erfolgsorientierte Personen attribuieren ihren Erfolg überwiegend internal (z.B. auf eigene Fähigkeiten, Anstrengung), Misserfolge dagegen external (z.B. Schwierigkeit der Prüfung). Misserfolgsorientierte Personen attribuieren ihren Misserfolg internal (z.B. mangelnde Fähigkeit), ihren Erfolg dagegen external (z.B. Glück).

In dieser Klassifikation von Ursachen ist der ‚Zufall‘ eine externale, instabile, und nicht kontrollierbare Ursache. Personen, deren Erfolg von Lotterien abhängt, sehen keinen Zusammenhang mehr zwischen ihren Leistungen, Anstrengungen und Fähigkeiten einerseits und dem Ausgang der Entscheidung andererseits. Bei einer Lotterie zu gewinnen leistet nichts zur Aufrechterhaltung des Selbstwertgefühls, zur wahrgenommenen Selbstwirksamkeit und zur Steigerung der Leistungsmotivation. Bei einer Lotterie zu verlieren kann zu einem Gefühl führen, eine Situation nicht kontrollieren zu können, was im Wiederholungsfall „erlernte Hilflosigkeit“ zur Folge haben kann. Gegebenenfalls wird dieses Unbehagen durch Versuche gemindert, wieder Kontrolle über die Situation zu erhalten, beispielsweise durch Verfahrenskritik („Lotterien sind nicht akzeptabel“) oder interne Attribution („Ich habe immer Pech bei Lotterien“).

Unter bestimmten Bedingungen können Lotterien aber auch positive Wirkungen haben. So gehen Frauen tendenziell Wettbewerbssituationen eher aus dem Wege. Berger et al. (2020) können in einem Laborexperiment zeigen, dass Frauen fokale Randomisierung auf der Grundlage einer Vorauswahl gegenüber Wettbewerbssituationen präferieren, ohne dass damit die Qualifikation tangiert wird. Mit fokaler Randomisierung könne der Gender-Gap im Bereich von Führungspositionen reduziert werden. Auch im Hinblick auf wissenschaftliche Zeitschriften kann foka-

le Randomisierung vorteilhaft sein (siehe Brezis, 2007; Osterloh & Frey, 2019, 2020; Oswald, 2020; Wooding, 2020; Yaqub, 2020). Im Hinblick auf die Förderinitiative „Experiment!“ der VolkswagenStiftung, die fokale Randomisierung verwendet, war jedoch der Frauenanteil unter den Geförderten relativ gering: ~16% in den Jahren 2013 bis 2016, 21% im Jahr 2017 (Röbecke & Simon, 2020, S. 10).

2.6 Zusammenfassung

Aus den obigen Überlegungen lassen sich zusammenfassend folgende sieben Bedingungen formulieren, unter denen ein Lotterieverfahren sinnvoll ist:

1. *Mehrstufiger Entscheidungsprozess*: Das Auswahlverfahren ist mehrstufig. In einer Vorauswahl werden Optionen herausgefiltert, die später in eine Endauswahl kommen. In der Endauswahl werden die Optionen entsprechend der Stärke ihrer Gründe in Rangordnung gebracht und die Option mit dem höchsten Rang ausgewählt.
2. *Unbestimmtheit*: Es gibt eine Untermenge von Optionen in der Endauswahl, die sich (1) in der Stärke ihrer Gründe gleichen und (2) die sich in der Stärke der Gründe von allen anderen Optionen, die nicht in der Untermenge sind, unterscheiden.
3. *Reinigender Effekt*: Es gibt (1) für die Untermenge von Optionen mit gleicher Stärke keine positiven Gründe (Legitimität, Bedürftigkeit, Verdienst, Zahlungsbereitschaft) mehr, eine Option gegenüber einer anderen zu bevorzugen, und (2) es muss mit negativen Gründen gerechnet werden, die eine mögliche Auswahl bedingen könnten. In dem Fall hat die Lotterie einen reinigenden Effekt.
4. *Allokation von knappen Gütern*: Im Falle der Allokation eines knappen Guts durch Lotterie an Personen, die einen Anspruch auf das Gut erheben, müssen neben der Unbestimmtheit der Entscheidung, folgende Bedingungen zutreffen:
 - a. Das zu verteilende Gut als Option muss stückweise vorliegen und darf nicht teilbar sein.
 - b. Das zu verteilende Gut als Option muss homogen und austauschbar sein.
 - c. Die Ansprüche an das Gut müssen homogen sein.
 - d. Die Ansprüche an das Gut dürfen nur hinsichtlich ihrer Stärke variieren.
5. *Negative psychologische Nebeneffekte*: Wird das Lotterieverfahren im Entscheidungsprozess in leistungsaffinen Situationen eingesetzt, muss gewährleistet sein, dass keine negativen psychologischen Nebeneffekte (Attributionen) auftreten, die das Selbstwertgefühl, die wahrgenommene Selbstwirksamkeit und die Leistungsmotivation der Betroffenen stark beeinflussen. Das Auswahlverfahren sollte von den betroffenen Personen nicht infrage gestellt werden (Akzeptanz). Mögliche Geschlechter-Effekte sind zu beachten.
6. *Balancierung*: Durch wiederholte Anwendung der Lotterie wird die Gruppe der Akzeptierten sich von der Gruppe der Abgelehnten in keinem Merkmal zufällig unterscheiden. Die zwei Gruppen sind ausbalanciert, was sich bei einer hinreichend großen Zahl an Wiederholungen empirisch prüfen lässt. Dabei un-

terscheiden sich die Gruppen der akzeptierten und der abgelehnten Anträge in keinem Merkmal mehr, sondern sie bilden im Durchschnitt die Heterogenität in der Population möglichst gut ab.

7. *Fairness des Lotterieverfahrens*: Das Lotterieverfahren ist fair, wenn den Ausgängen der Lotterie Wahrscheinlichkeiten zugeordnet werden können und alle Ausgänge die gleiche Wahrscheinlichkeit haben. Es ist keine Ex-ante-Vorhersage des Ausgangs aufgrund von einfachen Regeln (z.B. nach dreimal Kopf kommt zweimal Zahl) möglich. Je mehr Information über den Lotterienprozess vorliegt, desto eher ist das Verfahren gegenüber Manipulationen gewappnet, d.h., desto weniger besteht die Möglichkeit, den Ausgang der Lotterie vorherzusagen.

3. Lotterien als Element von Förderentscheidungen

3.1 Lotteriekonzepte in der Forschungsförderung

Während Lotterien im Hinblick auf politische Entscheidungen eine lange Geschichte haben (Stone, 2011, S. 6–10), werden sie im Bereich der Forschungsförderung erst seit wenigen Jahren diskutiert. Die Ersten, die Lotterien im Zusammenhang mit der Forschungsförderung diskutiert haben, waren Horton (1996) und Greenberg (1998). Sie schlugen vor, einen Teil der Fördersumme (15–20% über 5 Jahre) beiseite zu legen. Dieser Betrag wird dann über eine Lotterie an Bewerberinnen und Bewerber vergeben, deren Qualifikationen und Projektanträge als respektabel beurteilt wurden (Greenberg, 1998, S. 686). Aus einer statistischen Perspektive diskutierte Boyle (1998) die Bedeutung von Lotterien für die Auswahl von Personen in Organisationen. Inzwischen ist eine beachtliche Anzahl von einschlägigen Veröffentlichungen erschienen.⁶

Im Folgenden soll die Literatur zu Lotterien in der Forschungsförderung kurz umrissen werden, um dann diese Argumentationen im Lichte der Forschung zum „grant peer review“ und vor dem Hintergrund der Ausführungen zu Lotterien (Abschnitt 3.2) zu diskutieren.

Der Einsatz von Lotterien in der Forschungsförderung wird in der Literatur fast überwiegend mit Problemen des Peer-Review-Verfahrens begründet (z.B. Avin, 2015, 2018; Barnett, 2016; Barnett et al., 2015; de Peuter & Conix, 2021; Frey & Osterloh, 2015; Osterloh, 2017; Osterloh & Frey, 2015; Roumbanis, 2019). Das Peer-Review-Verfahren in der Forschungsförderung sei nicht reliabel in dem Sinne, dass die Gutachterinnen und Gutachter nicht ausreichend in ihren Urteilen über die Förderungswürdigkeit eines Forschungsantrags übereinstimmen (Inter-Rater-Reliabilität). Gut-

⁶ Siehe z.B. Adam, 2019; Avin, 2014; Avin, 2015, 2018, 2019; Barnett, 2016; Barnett et al., 2015; Bedessem, 2020; Berger et al., 2020; Brezis, 2007; Clarke et al., 2016; Daniel, 2018; de Peuter & Conix, 2021; Fang & Casadevall, 2016; Frey & Osterloh, 2015; Gillies, 2014; Graves et al., 2011; Harnagel, 2019; Heyard et al., 2022; Neff & Olden, 2006; Osterloh, 2017; Osterloh & Frey, 2015, 2019, 2020; Reinhart & Schendzielorz, 2020; Roumbanis, 2019.

achterurteile seien nicht prädiktiv valide, d.h., sie können den späteren Erfolg eines beantragten Projektes nicht vorhersagen, und sie seien darüber hinaus nicht ausreichend fair. Beispielsweise könnten Anträge von Männern eine höhere Wahrscheinlichkeit haben, akzeptiert zu werden, als Anträge, die von Frauen gestellt werden. Fang und Casadevall (2016, S. 1) argumentieren „... scientific peer review is no longer effective, because review panels cannot accurately stratify proposals to identify the most meritorious ones. Bias has a major influence on funding decisions, and the impact of reviewer bias is magnified by low funding paylines.“ Die Unsicherheiten im Peer-Review-Verfahren führten bei manchen Autoren zu der Schlussfolgerung, dass die auf den gutachterlichen Urteilen beruhenden Entscheidungen selbst reiner Zufall seien (Greenberg, 1998; Neff & Olden, 2006; Scholten et al., 2018; Yokoyama, 1996).

Neben den Gütekriterien des Peer-Reviews wird häufig auch auf die immensen Kosten des Verfahrens hingewiesen. „The peer reviewing itself both by the referees and by the committee members is time consuming and hence expensive“ (Gillies, 2014, S. 6). Graves, Barnett und Clarke (2011, S. 2) rechneten aus Zahlen einer Umfrage unter 41 Forschenden in Australien (Herbert et al., 2013) folgendes Ergebnis hoch. So wand jedes Mitglied des gutachtenden Panels pro Antrag 20 Minuten zum Lesen und 46 Stunden Arbeitszeit insgesamt in der Woche des Panelmeetings auf. Smith (2006), ehemaliger Chief-Editor des British Medical Journal (BMJ), kommt zu folgendem vernichtenden Urteil über das Peer-Review-Verfahren insgesamt: „So we have little evidence on the effectiveness of peer review, but we have considerable evidence on its defects. In addition to being poor at detecting gross defects and almost useless for detecting fraud it is slow, expensive, profligate of academic time, highly subjective, something of a lottery, prone to bias, and easily abused“ (Smith, 2006, S. 179). Reinhart und Schendzielorz (2020) argumentieren, dass mit der Fokussierung auf Gütekriterien und Objektivität einer eher dogmatischen Epistemologie das Wort geredet wird. Im Falle von partieller Validität (z.B. Bias) der gutachterlichen Voten sei eine Lotterie die bessere Alternative (Reinhart & Schendzielorz, 2020, S. 27).

Als Lösungsansatz wird von den meisten Autoren ein mehrstufiges Entscheidungsverfahren empfohlen. Nach einer Vorselektion von Anträgen, in der das Peer-Review-Verfahren eingesetzt wird, erfolgt aus der Gruppe der förderungswürdigen Anträge die Auswahl der Anträge, die tatsächlich gefördert werden, über eine Lotterie.

Im Hinblick auf die Forschungsförderung wird von zusammenfassenden Peer-Review-Urteilen ausgegangen, auf deren Grundlage eine Gruppierung vorgenommen wird. So schlägt Brezis (2007) vor, anhand der Gutachterurteile drei Gruppen zu bilden, eine Gruppe von Anträgen, die so exzellent bewertet wurde, dass sie auf jeden Fall bewilligt werden sollte; eine Gruppe von Anträgen, die auf jeden Fall abgelehnt werden sollte; und eine Gruppe von Anträgen, über die Unsicherheit besteht, ob sie abgelehnt oder akzeptiert werden sollte. Fang und Casadevall (2016) und Avin (2014, 2015, 2018, 2019) favorisieren ein Verfahren mit zwei Gruppen. Aufgrund der Gutachterurteile wird eine Gruppe von Anträgen gebildet, die eine Förderung ver-

dienen, und eine Gruppe von Anträgen, die keine Förderung verdienen und daher abgelehnt werden. Diese Gruppe sollte aber die Möglichkeit erhalten, eine überarbeitete Version des Antrags wieder einzureichen. Bei den Anträgen, die eine Förderung eigentlich verdienen, wird das Lotterie-Verfahren angewendet, um die tatsächlich zu fördernden Anträge auszuwählen. Avin (2019) nennt dieses Verfahren „Triage“ und setzt dies in Kontrast zu einem reinen Lotterie-Verfahren, in dem alle Anträge in die Lostrommel kommen. Er kann in einer Simulation nachweisen, dass das Triage-Verfahren im Hinblick auf Effizienz und Breite unterschiedlicher Themen dem reinen Lotterie-Verfahren überlegen ist. Heyard et al. (2022) verwenden statistisch geschätzte Ränge von Anträgen („Empirical Bayes Ranking Method“) in Kombination mit Randomisierung als Grundlage der Auswahl der zu fördernden Anträge.

Barlösius und Philipps (2021) befragten im Hinblick auf Losverfahren eine Stichprobe von 50 Wissenschaftlerinnen und Wissenschaftlern, die nach Fachgebieten (Biologie, Physik, Geschichte), akademischer Position (Professor, Postdoc, PreDoc) und Geschlecht stratifiziert wurde. Die Verteilung von Forschungsmitteln allein dem Zufall zu überlassen, wird von den meisten Befragten zugunsten von Varianten abgelehnt, die Qualitätssicherung und Randomisierung kombinieren. Barlösius und Philipps (2020) sowie Philipps (2021) führten problemzentrierte Interviews mit 32 Naturwissenschaftlerinnen und Naturwissenschaftlern in Deutschland durch. Losverfahren als Alternative zu den bisherigen Verfahren sollten nach den Befragten eine Chance bekommen. Insbesondere wurde die Fairness des Losverfahrens betont, da alle Anträge unabhängig von Merkmalen der antragstellenden Person (Geschlecht, akademische Position) die gleichen Chancen hätten, bewilligt zu werden. Risikoreiche, spekulative Forschungsprojekte hätten eine größere Chance, es würde Zeit gespart werden und die Entscheidungsprozesse würden beschleunigt. Reserviert stehen die Befragten allerdings einer reinen Zufallsauswahl gegenüber, da sie von der Überlegenheit des Peer-Review-Verfahrens überzeugt sind. In einer Befragung von Antragstellerinnen und Antragstellern des Health Research Council of New Zealand (Liu et al., 2020) waren 63% der 126 Personen, die an der Befragung teilnahmen (39% Rücklaufquote), für eine Randomisierung im Rahmen des Förderprogramms „Explorer Grant“, 25% waren dagegen. Für andere Förderprogramme lag die Zustimmung nur bei 40%.⁷

3.2 Diskussion der Kritik im Lichte der Peer-Review-Forschung

Wird die Diskussion um das Peer-Review-Verfahren als Form eines juristischen Prozesses gesehen mit dem Peer-Review-Verfahren als Angeklagten, so wird schnell

⁷ Santana (2022) schlägt in seiner Publikation „Why citizen review might beat peer review at identifying pursuitworthy scientific research“ vor, statt Losverfahren einzuführen, Laien stärker in die Beurteilung von Anträgen und Manuskripten einzubeziehen. Dies steht in Einklang mit der „Open Science“-Bewegung und der Idee von „Citizen Science“: „Open Science also aims to encourage the inclusion of non-institutional participants, in other words the general public, in the scientific process“ (European Commission Directorate-General for Research and Innovation, 2016, S. 53).

klar, dass die Position der Anklage in der Literatur sehr viel stärker vertreten ist als die Position der Verteidigung des Verfahrens. Statt nun positive Argumente der Verteidigung ins Feld zu führen, sollen im Folgenden zumindest begründete Zweifel an der angeführten „Anklage“ formuliert werden. So werden die Ergebnisse der „grant peer review“-Forschung (z.B. mangelnde Gutachterübereinstimmung) häufig zitiert, ohne die Frage aufzuwerfen, wie diese Ergebnisse methodisch zustande gekommen sind, und ob die Ergebnisse die Schlussfolgerungen (z.B. Einführung eines Lotterieverfahrens) überhaupt rechtfertigen. Drei dieser möglichen methodischen Probleme seien beispielhaft kurz skizziert.

In der ‚grant peer review‘-Forschung dominieren *Fallanalysen* einzelner Forschungsförderorganisationen, welche die Generalisierbarkeit auf andere Organisationen der Forschungsförderung insgesamt aufgrund unterschiedlicher Bedingungen (Förderprogramme, Förderjahre, Höhe der zur Verfügung stehenden Fördermittel ...) schwierig machen. Häufig werden Daten verwendet, die in laufenden Peer-Review-Verfahren anfallen (z.B. Peer-Review-Ratings). Das Vorgehen, die Auswahl der Variablen und letztlich auch die Ergebnisse sind damit nicht per se unabhängig von den Hypothesen und Präferenzen der Forschenden selbst. Dies ist bei einem randomisierten Kontrollgruppenexperiment anders, da die Hypothesen vor der Durchführung vorliegen. In der Sozialpsychologie spricht man auch von „präferenzkonsistenter Informationsverarbeitung“ (Greitemeyer & Schulz-Hardt, 2003). Damit wird nicht unterstellt, Forscherinnen und Forscher würden in der „grant peer review“-Forschung bewusst täuschen oder manipulieren, aber ihre Entscheidungen im Analyse- und Veröffentlichungsprozess können implizit beeinflusst sein durch ihre Präferenzen (z.B. Ablehnung des Peer-Review-Verfahrens).

Am ehesten geben Metaanalysen als Zusammenfassung der Ergebnisse aus einer Vielzahl von Primärstudien einen Hinweis auf generalisierbare Ergebnisse zum Peer-Review. So erbrachte beispielsweise eine Metaanalyse zum „gender bias“ von Bornmann et al. (2010) folgendes Ergebnis: „Even though the estimates of the gender effect vary substantially from study to study, the model estimation shows that all in all, among grant applicants men have statistically significant greater odds of receiving grants than women by about 7%“ (S. 226). Marsh et al. (2009) schließen in ihrem Synthesebericht sogar die Nullhypothese von einem fehlendem Effekt nicht aus, obwohl sich in Einzelstudien – wie für den Wissenschaftsfonds Österreich (FWF) – Geschlechter-Effekte zeigen (Mutz et al., 2012a). Darüber hinaus deuten Metaanalysen auf einen publication bias hin (Rothstein et al., 2005). So werden Ergebnisse mit großen Effekten auch dann publiziert, wenn kleine Stichproben vorliegen.

Ein weiteres methodisches Problem ist der verwendete *Reliabilitätskoeffizient*, der im Hinblick auf Peer-Review-Urteile eine Interrater-Reliabilität darstellt. Häufig wird die Intraklassenkorrelation (single-rater reliability) verwendet, die Auskunft darüber gibt, wie reliabel – sprich austauschbar – einzelne gutachterliche Urteile sind. Bei Förderentscheidungen spielt aber häufig der Mittelwert von gutachterlichen Urteilen eine Rolle, deren Reliabilität infolge der Mittelwertbildung höher ist als die „single-rater reliability“ (Marsh et al., 2008, S. 166; Mutz et al., 2012b).

In Untersuchungen zur *prädiktiven Validität* von Peer-Review-Urteilen wird schließlich häufig eine Korrelation berechnet zwischen den Peer-Review-Urteilen und externen Kriterien, z.B. dem Forschungsoutput der geförderten Projekte. So berechneten Scheiner und Bouchie (2013) einen Zusammenhang zwischen Umfang der Förderung durch die National Science Foundation (USA) und der Anzahl Publikationen und kamen zu dem Schluss: „On average, although the projects were quite productive (Web Table 1) the reviewers and panelists were unable to predict productivity Reviewer scores and panel rankings were moderately, but non-significantly, correlated with project outcomes ($r=0.12-0.29$), ...“ (Scheiner & Bouchie, 2013, S. 406). In die Analyse gingen aber nur die Anträge ein, die tatsächlich auch gefördert wurden. Gutachterinnen und Gutachter geben Empfehlungen ab, Anträge entweder zu bewilligen oder nicht zu bewilligen. Um tatsächlich die prädiktive Validität der Gutachterurteile zu untersuchen, müssten prinzipiell auch die Anträge im Hinblick auf ihren Forschungsoutput untersucht werden, die abgelehnt und anderweitig gefördert wurden. Ansonsten werden Korrelationen infolge einer Restriktion des Wertebereichs unterschätzt. Einen solchen Ansatz haben Bornmann und Daniel (2008a) im Hinblick auf das Begutachtungsverfahren von Zeitschriften verwendet. „Here we examine the predictive validity of the peer review process at *Angewandte Chemie*, on the basis of the mean citation counts of manuscripts accepted for publication and manuscripts rejected by the journal but then published elsewhere and on the Journal Impact Factors (JIFs) of the journals in which the rejected *Communications* were later published.“ (Bornmann & Daniel, 2008a, S. 7175) Im Bereich der Forschungsförderung ist ein solches Vorgehen kaum realisierbar. Häufig werden Anträge ganz aufgegeben oder stark umformuliert. Mutz et al. (2015) schlugen daher einen statistischen Missing-Value-Imputation-Ansatz vor, um auch für die abgelehnten Forschungsanträge am Beispiel des FWF einen Wert im Forschungsoutput zu generieren, auch wenn er nur statistisch geschätzt ist. Die prädiktive Validität allein für die geförderten Anträge war mit $r=.26$ erwartungsgemäß gering ausgeprägt. Wenn die Korrelation jedoch für die geförderten und (sic!) die abgelehnten Anträge berechnet wurde, war die prädiktive Validität deutlich stärker ausgeprägt ($r=.49$).

Die methodischen Ausführungen und geäußerten Zweifel legen auf der einen Seite nahe, das Peer-Review-Verfahren nicht kategorisch abzulehnen, auf der anderen Seite auch, Alternativen zum Peer-Review-Verfahren in Betracht zu ziehen. Im Hinblick auf die in Abschnitt 2.5 formulierten Bedingungen wird deutlich, dass mangelnde Reliabilität des Peer Reviews zu Unbestimmtheit (Bedingung 2) führt. Infolge von mangelnder Reliabilität und Messfehlern unterscheiden sich Anträge trotz unterschiedlichen Gutachterurteilen nicht in ihrer Qualität (Cole et al., 1981). Wird als Maß der Reliabilität die Interrater-Reliabilität zugrunde gelegt, so erfordern gerade Anträge mit extremen Bewertungen (exzellente, nicht förderungswürdig) eine hohe Reliabilität, da nur die Übereinstimmung in den Urteilen (z.B. dreimal exzellente) eine solche exzellente Gesamtbewertung möglich macht. Eine Bewertung im mittleren Bereich kann jedoch zu einem dadurch zustande

kommen, dass die Gutachterinnen und Gutachter in ihrer mäßigen Einschätzung eines Antrags perfekt übereinstimmen. Sie kann jedoch auch dadurch zustande kommen, dass die Bewertungen unter den Gutachterinnen und Gutachern stark divergieren (hohe und niedrige Bewertungen), was ein Hinweis auf kontroverse Anträge sein kann. Im ersten Fall wäre eine hohe Interrater-Reliabilität die Folge, Unbestimmtheit wäre jedoch nicht das Problem. In diesem Fall könnten Anträge auf der ganzen Breite der gutachterlichen Skala trennscharf unterschieden werden. Im letzteren Fall wäre gerade im mittleren Bereich mit hoher Unbestimmtheit zu rechnen, sodass für diesen Bereich die Voraussetzungen für eine Lotterie gegeben wären.

Die Beantwortung der Frage, ob Lotterien einen „reinigenden Effekt“ haben, bedarf des Nachweises von „negativen Gründen“. Im Peer-Review-Verfahren sind negative Gründe Verzerrungen von Gutachterurteilen z.B. „gender bias“, „affiliation bias“ und „prestige bias“, die Gegenstand einer Vielzahl von Studien zum Peer-Review-Verfahren waren (für einen Überblick siehe Lee et al., 2013). Mutz et al. (2015) unterscheiden zwischen potentiellen und realen Verzerrungen. So können sich die Gutachterurteile zu Anträgen aus zwei verschiedenen Nationen (z.B. China, USA) im Mittel unterscheiden und damit verzerrt sein (Nationalitätsbias). Dies ist aber nur eine potentielle Verzerrung, denn es kann selbstverständlich sein, dass beide Nationen sich im Forschungoutput der geförderten Projekte tatsächlich unterscheiden. In diesem Falle würde man nicht unbedingt von einer realen Verzerrung sprechen. Die Frage von Verzerrungen bedarf empirischer Analysen im Hinblick auf das Antragsverfahren der einzelnen Forschungsförderungsorganisation (z.B. Mutz et al., 2015).

Im Hinblick auf die ökonomischen Bedingungen für Lotterien (Abschnitt 2.4) besteht das grundsätzliche Problem, inwieweit Anträge kommensurabel sind, d.h. auch stückweise vorliegen und homogen sind. So variieren Anträge in vielfältigen Dimensionen: Fachgebiet, Höhe der Antragssumme, Art der Förderung (Sachmittel, Mitarbeitende), Forschungsgegenstand, Forschungsansatz, Abhängigkeit der Teilprojekte, anwendungsbezogene Forschung oder Grundlagenforschung. In einem Lotterieverfahren müssten Gruppen von Anträgen gebildet werden, die homogen in den zentralen Dimensionen sind. Innerhalb dieser Gruppen fände dann die Zufallsauswahl statt.

Schließlich sind die psychologischen Nebeneffekte zu diskutieren. Loveday (2017) führte qualitative Interviews mit 44 Wissenschaftlerinnen und Wissenschaftlern (18 Männer, 26 Frauen) an Hochschulen in Großbritannien durch, die am Beginn ihrer Karriere standen. Ziel der Interviews war es, das subjektive Verständnis ihrer Anstellung zu erforschen. Die Beteiligung am Hochschulleben wird von den Befragten als stark individualisiert erlebt. „The demands of the so-called ‘neoliberal university’ have created ‘enterprising’ academic subjects who understand their own participation as being increasingly individualized“ (Loveday, 2017, S. 775). Eine wichtige Bedeutung kommt dem Zufall zu: „Arguing that precarious employment situations precipitate the feeling of being ‘out of control’, the majority of the participants’ nar-

ratives were characterized by a distinct lack of agency. The paper explores the recourse to notions of chance and the consolidation of ‘luck’ as an explanatory factor in accounting for why good things happen“ (Loveday, 2017, S. 758). Zwar ist die Datengrundlage für weiterreichende Schlussfolgerungen mit 44 Befragten zu gering, dennoch wird deutlich, dass unabhängig von Lotterien Wissenschaftlerinnen und Wissenschaftler dazu neigen, ihre aktuelle Situation auf den Zufall zu attribuieren, und dass sie Symptome „erlernter Hilflosigkeit“ („lack of agency“) zeigen. Ein Lotterieverfahren würde in dieser Hinsicht den Status quo eher bestätigen. Mit der olympischen Idee, dass die Teilnahme an den Spielen wichtiger sei als der Sieg, fassen Ayoubi et al. (2019) ihr Ergebnis zu einer Untersuchung zum Schweizerischen Nationalfonds (SNF) zusammen. Nach ihren Analysen profitieren Wissenschaftlerinnen und Wissenschaftler von einer Teilnahme in einem klassischen wettbewerblichen Antragsverfahren des SNF im Hinblick auf die Ausweitung des wissenschaftlichen Netzwerks und der Erhöhung der wissenschaftlichen Produktivität, und zwar unabhängig davon, ob der Antrag erfolgreich war oder nicht.

3.3 Zusammenfassung

Der Einsatz des Lotterieverfahrens wird meist mit Problemen des Peer-Review-Verfahrens begründet. Das Verfahren sei wenig reliabel, könne das Ausmaß des Forschungserfolgs nicht vorhersagen, und sei z.T. verzerrt (z.B. gender bias). Daher wird in der Regel ein zweistufiges Verfahren vorgeschlagen. Im ersten Schritt wird aufgrund des Peer-Reviews eine Auswahl getroffen, bei der VolkswagenStiftung erfolgt die Vorselektion allerdings durch die Geschäftsstelle. Im zweiten Schritt werden idealtypisch exzellente Anträge angenommen, wenig überzeugende Anträge abgelehnt und Anträge mit mäßiger oder unsicherer Bewertung einer Lotterie unterzogen. Schwierig bleibt es, Gruppen von Anträgen zu finden, die kommensurabel sind, d.h., für die keine Gründe mehr vorhanden sind, den einen Antrag gegenüber einem anderen zu bevorzugen. Im Hinblick auf die subjektive Wirkung von Förderungen zeigt die qualitative Studie von Loveday (2017) auf, dass Wissenschaftlerinnen und Wissenschaftler schon in einer frühen Phase der Karriere Glück und Zufall als Narrativ für ihre bisherige Karriere sehen.

4. Die Förderinitiative „Experiment!“ der VWS mit einem teilrandomisierten Auswahlverfahren

Im November 2012 führte die VolkswagenStiftung (VWS) die Förderinitiative „Experiment!“ ein, um wissenschaftliche Projekte mit gewagten und riskanten Forschungsideen im Bereich der Natur-, Ingenieur- und Lebenswissenschaften zu unterstützen (Bendisoli et al., 2021, S. 78–87; VolkswagenStiftung, 2019a).

„Vorrangig werden Ansätze unterstützt, die

– radikal neu sind und über eine kontinuierliche Verbesserung hinauszielen;

- akzeptiertes Fachwissen grundlegend herausfordern;
- unkonventionelle Hypothesen, Methodiken oder Technologien etablieren wollen oder
- ganz neue Forschungsrichtungen stimulieren“. (VolkswagenStiftung, 2019b, S. 1)

In den Jahren 2013 bis 2016 wurden 67 Projekte gefördert (Röbbecke & Simon, 2020, S. 9). Im Jahre 2017 führte die VWS zwei Änderungen ein: Erstens sollten doppelt so viele Projektanträge pro Auswahlrunde (30–40 Anträge) bewilligt werden, zweitens wurde für eine Erprobungsphase bis 2021 ein Lotterie-Element eingeführt („teil-randomisiertes Verfahren“). Die Fördersumme war auf maximal 120.000 Euro und die Laufzeit auf die Dauer von maximal 18 Monaten begrenzt. Wissenschaftlerinnen und Wissenschaftler in Deutschland aller Karrierestufen nach der Promotion konnten sich bewerben. Insgesamt wurden im neuen Verfahren von 2013 bis 2021 183 Bewilligungen mit jeweils 120.000 Euro nach dem teilrandomisierten Verfahren ausgesprochen (VolkswagenStiftung, 2023). Das Verfahren wird von der VWS wie folgt beschrieben:

„An der teil-randomisierten, anonymisierten Auswahl sind drei Akteure beteiligt: die Stiftung, eine externe Wissenschaftsjury und – in einer Erprobungsphase von 2017 bis 2020 – der Zufall. Der gesamte Prozess ist ohne Einholung schriftlicher Fachgutachten konzipiert. Zuerst werden alle Anträge in der Förderabteilung gemäß den Programmkriterien, der Selbsteinschätzung und ihrer Schlüssigkeit geprüft und vorselektiert. Die resultierende Shortlist (ca. 120–140 Anträge) bewertet eine international und interdisziplinär zusammengesetzte Jury doppel-blind, d. h. anhand anonymisierter Unterlagen. So zählt alleine die gewagte, potenziell bahnbrechende Idee und nicht die Reputation der dahinterstehenden Person. Die Jury sondert qualitativ ungenügende Anträge aus und wählt aus den verbleibenden Anträgen die überzeugendsten Ideen (ca. 15–20 Bewilligungen) aus. Jedes Jurymitglied verfügt über genau einen „Förderjoker“, um optional ein Vorhaben, über das kein Konsens erzielt wird, zur Förderung vorzuschlagen. Testweise werden aus demselben Pool von Anträgen, die den Programm- und den Qualitätskriterien genügen, ebenso viele Bewilligungen (15–20) unter Aufsicht der Stiftungsjustitiarin ausgelost. Auf diese Weise wird Bias vermieden und es erhalten auch Ideen, die ansonsten leicht übersehen werden, eine Chance.“ (VolkswagenStiftung, 2019a, S. 2)

Ein sogenanntes „Logic-Model“ (Daniel et al., 2021, S. 56–57; Jordan, 2013), das explizit die Programmkriterien nennt und ein Wirkungsmodell aufzeigt, wie diese erreicht werden können, ist nicht formuliert worden. Als Auswahlkriterien werden Folgende genannt: a) Originalität, Unkonventionalität, Vision der Idee, b) vorweggenommene Wirkung des Forschungsbeitrags, c) möglicher Lerneffekt, wenn das

Projekt scheitert, d) Wagnis und e) ob sich eine explorative Phase eignet, um eine Idee wirklich voranzubringen (vgl. VolkswagenStiftung, 2019b, S. 2–3). Vor dem Hintergrund der Ausführungen in den Abschnitten 2 und 3 kann das Auswahlverfahren der VWS-Förderinitiative „Experiment!“ wie folgt rekonstruiert werden:

Es handelt sich um ein mehrstufiges Entscheidungsverfahren. Die Vorauswahl wird jedoch nicht auf der Grundlage eines externen Gutachtervotums getroffen, sondern allein von Personen der Geschäftsstelle der Stiftung nach formalen Antragskriterien (Programmkriterien, Selbsteinschätzung, Schlüssigkeit). Mehr als 75% der eingereichten Anträge werden schon in der Vorauswahl abgelehnt. Die Frage bleibt, inwieweit es den Personen der Stiftung gelingt, ausreichend intersubjektiv vergleichbar und reliabel nicht geeignete Anträge auszusortieren. Die eigentliche Förderentscheidung wird einer wissenschaftlichen Fachjury überlassen, der sowohl die Aufgabe der Begutachtung als auch die Auswahlentscheidung zukommt. In einem ersten Schritt werden Anträge, die aufgrund mangelnder Qualität auf jeden Fall nicht gefördert werden sollen, von Anträgen getrennt, die prinzipiell gefördert werden können. Die ‚Unbestimmtheit‘ der Optionen (Abschnitt 2.5) drückt sich letztlich darin aus, dass diese grundsätzlich förderungswürdigen Anträge sowohl die Grundlage der finalen Entscheidung der Jury als auch die Grundlage der Lotterie bilden. Im Grunde genommen liegt damit auch eine Dreiteilung oder Triage vor: 1) Anträge, die auf jeden Fall abgelehnt werden (70–80%), 2) Anträge, die auf jeden Fall per Jury-Entscheidung bewilligt werden (10–15%), und 3) potentiell förderungswürdige Anträge („meritorius“), die per Zufall bewilligt werden (10–15%). Doppelungen in der Auswahl sind dabei nicht ausgeschlossen (Zufallsauswahl und Jury-Entscheidung). Die Verwendung des Lotterieverfahrens wird letztlich ebenso mit dem Ausschluss „negativer Gründe“ gerechtfertigt, d.h. aufgrund möglicher Verzerrungen (z.B. gender bias), wie sie in Abschnitt 3.2 diskutiert wurden.

Die VWS erhofft sich von der Einführung des Lotterie-Elements eine höhere Fächerbreite und Geschlechtergerechtigkeit. Damit wird auf die Eigenschaft von Randomisierungen Bezug genommen, im Erwartungswert eine Heterogenität der Gruppen herzustellen vergleichbar der Heterogenität in der Population der Antragstellenden und ihrer Anträge. Bei perfekter Randomisierung und hinreichender Fallzahl liegt Heterogenität nämlich nicht nur bezüglich der Fächer und des Geschlechts vor, sondern bezüglich aller Variablen (z.B. Alter, Forschungseinrichtung, Position). In einer Online-Befragung, an der 25 Geförderte teilnahmen, wurde die Förderinitiative durchweg positiv beurteilt, insbesondere der Aufwand für die Antragsstellung (Röbecke & Simon, 2020, S. 11). Die Förderinitiative würde aus Sicht der Geförderten die Chancengleichheit fördern (92% der Befragten), zu Anträgen mit riskanter Forschung ermutigen (84%), bessere Chancen für riskante Forschung bieten (80%) und die Chancen für im Verfahren schwach vertretene Fächer erhöhen (84%). Informationen zu einem potentiellen Non-Response-Bias und einer subjektiven Bewertung des Verfahrens durch Nicht-Geförderte liegen jedoch nicht vor.

5. Evaluationskonzepte

Bei der Entwicklung neuer Auswahlverfahren für Forschungsförderungsinitiativen stellt sich unmittelbar die Frage, ob sie tatsächlich einen kausalen Effekt auf den Forschungsoutput der Geförderten haben, der höher liegt als bei den bisherigen Auswahlverfahren. In Ermangelung eines Logic-Models, das Erfolgskriterien für die Förderinitiative „Experiment!“ vorgibt, wird im Folgenden nur allgemein von Forschungsoutput gesprochen als Substitut möglicher spezifischer Output- und Outcome-Maße.

Im Kontext der Evaluation von politischen Programmen zur Förderung von Forschung und Entwicklung (R&D) im privaten Sektor wird die Frage nach Effekten und Wirkungen von Förderungen schon lange diskutiert, da hier der Return of Investment von besonderer Bedeutung ist. Daher stammen die Konzepte zur Evaluation solcher Programme meist aus der Ökonomik (z.B. Benavente et al., 2012; Clarke et al., 2016; Jacob & Lefgren, 2011; Jaffe, 2002; Klette et al., 2000). Jaffe (2002) regt an, eine Ex-post-Evaluation schon in der Konzeption solcher Programme anzulegen, anstatt erst nach Abschluss eines Förderprogramms Vorschläge für eine Wirkungsevaluation einzuholen. Die Stringenz von Ex-post-Evaluationen könnte dadurch wesentlich erhöht werden.

Ausgangspunkt von Programmevaluationen in der Ökonomik sind häufig Paneldaten. So lässt sich im Hinblick auf die Forschungsförderung im öffentlichen Bereich für eine Wissenschaftlerin oder einen Wissenschaftler i jährlich der Forschungsoutput Y erheben. Für ein Jahr t lässt sich dann folgende einfache Regressionsgleichung erstellen (Jaffe, 2002, S. 25; Klette et al., 2000, S. 479; Pion & Cordray, 2008):

$$Y_{it} = \alpha_i + u_t + \beta_1 D_i + \sum_j \gamma_j X_{jit} + e_{it}$$

wobei D_i eine Dummy-Variable darstellt, die ab dem Zeitpunkt t „1“ ist, wenn eine Person i durch das betreffende Programm gefördert wurde, und „0“, wenn nicht. Der Parameter β_1 gibt dann den durchschnittlichen Effekt der Förderung für den Geförderten an. Der Parameter α_i bezeichnet den fixen Effekt einer Person, der das durchschnittliche Niveau des Forschungsoutputs des Individuums angibt. Der Schätzer für den Effekt der Förderung im Rahmen dieser „fixed effects regression“ (Allison, 2009) beruht damit auf dem mit der Förderung einhergehenden fixen Effekt im Forschungsoutput, wodurch die Einflüsse zeitinvarianter Kovariaten mitkontrolliert werden. Der Parameter u_t spiegelt jährliche Schocks, die alle Geförderten in einem Jahr gleichermaßen betreffen. Zur Schätzung dieser zeitlichen Trends wird eine (Kontroll-)Gruppe von Personen miteinbezogen, die nicht gefördert wurde. Gleiches gilt für die Bestimmung der Effekte γ_j der j zeitveränderlichen Kovariaten X_{jit} . Mit e_{it} sind die Zufallsfehler bezeichnet, die über die Zeit hinweg auch (auto-)korreliert sein können.

Nach Klette et al. (2000) wurde über viele Jahre der Effekt einer Förderung mit derartigen Regressionsmodellen geschätzt. Diese Modelle verlangen jedoch, dass nach Kontrolle auf die fixen Effekte und zeitveränderlichen Kovariaten die Zuweisung zur Gruppe der Geförderten und zur Gruppe der Nicht-Geförderten rein zufällig erfolgt.⁸ Diese Annahme ist jedoch in der Forschungsförderung kaum erfüllt. Es ist von einem Selektionsbias auszugehen (Heckman et al., 1998). Die Zuweisung zu den Gruppen erfolgt aufgrund bestimmter Merkmale des Antrags oder der antragstellenden Person. Die geschätzten Programmeffekte sind damit nur begrenzt auf Nicht-Geförderte übertragbar und möglicherweise auch verzerrt.

Jaffe (2002) und Klette et al. (2000) führen verschiedene Untersuchungsdesigns an, um den durchschnittlichen kausalen Effekt der Förderung zu schätzen (z.B. „regression discontinuity design“, „difference in difference“). Ein Modell, das sich besonders für Forschungsförderung eignet, ist das „regression discontinuity design“ (Thistlethwaite & Campbell, 1960), das aber eine kontinuierliche Bewertungsskala voraussetzt, die bei der VWS nicht vorliegt. Der Zuweisungsentscheidung muss eine kontinuierliche Bewertungsskala der Anträge zugrunde liegen, die sich beispielsweise aus den Urteilen der Gutachtenden ableitet. Ein Schwellenwert auf der Skala bestimmt die Förderentscheidung als zentraler Prädiktor. Für die Schätzung der kausalen Effekte werden gerade die Anträge verwendet, die um den Schwellenwert liegen („gerade abgelehnt“, „gerade noch bewilligt“), was vergleichbar dem Konzept des „selection at the margin“ von Jaffe (2002, S. 28) ist. Forschungsanträge lassen sich entsprechend der gutachterlichen Urteile nach Jaffe prinzipiell in drei Subgruppen einteilen: „... ‘high priority’, ‘marginal’ und ‘rejects’, with the probability of receiving funding higher for the ‘high priority’ group than the ‘marginal’ group, and no funds awarded to the ‘rejects.’“ (Jaffe, 2002, S. 28) Für jede dieser Subgruppen erfolgt eine zufällige Zuteilung in Treatment und Kontrollgruppe. Die Unterschiede im Mittelwert im Output zwischen den beiden Gruppen („high priority“, „marginal“) sind dann die durchschnittlichen kausalen Effekte einer Förderung.

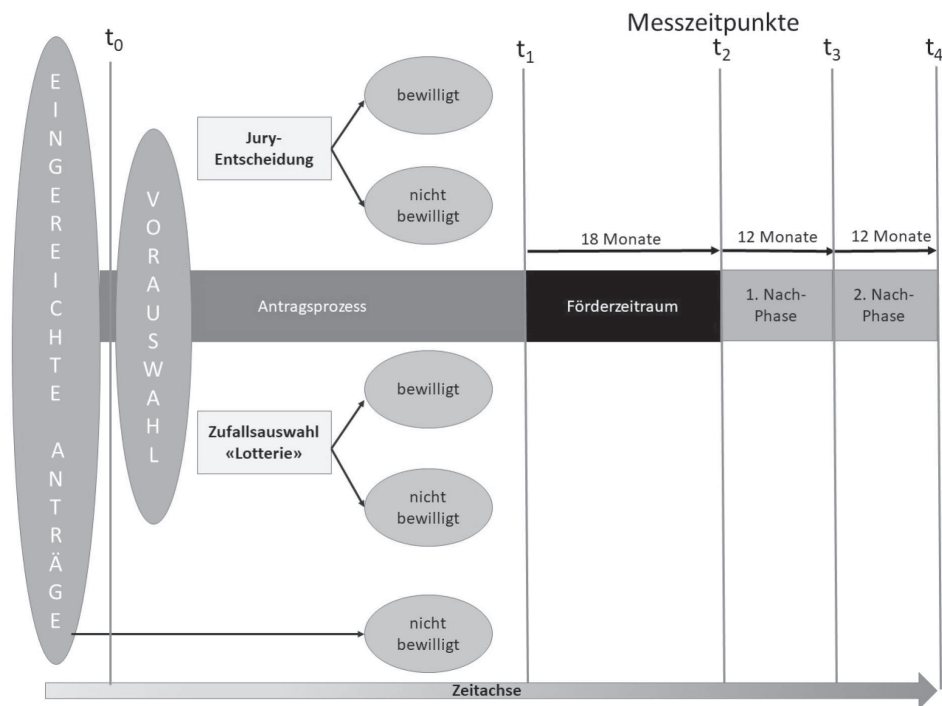
Aus forschungsethischen und wissenschaftspolitischen Gründen ist eine Ablehnung von „high-priority“-Projektanträgen bei ausreichenden Fördermitteln kaum möglich, sodass sich die Analyse dann auf die „marginal“-Gruppe konzentriert, mit einer Unterschätzung des kausalen Effekts („randomization at the margin“) im Vergleich zu der „high-priority“-Gruppe (Daniel, 2018). Bol, de Vaan und van de Rijt (2018) nutzten beispielsweise diese Überlegung, um einen Matthäus-Effekt in der Förderung des wissenschaftlichen Nachwuchses nachzuweisen. Mit der „marginal“-Gruppe wird eine Gruppe von Forschungsanträgen bezeichnet, die prinzipiell förderungswürdig sind, für die allerdings infolge der begrenzten Haushaltsmittel einer Forschungsförderungsorganisation keine Fördermittel mehr zu Verfügung stehen.

8 Genauer verlangen diese Modelle, dass u.a. die sogenannte Conditional-Independence-Annahme (CIA) (Morgan & Winship, 2015; Treischl & Wolbring, 2020) bzw. „strong ignorability condition“ erfüllt ist (Mutz & Daniel, 2012; Rubin, 2011). Vereinfacht gesagt, muss die Zuweisung zur Gruppe der Geförderten und zur Gruppe der Nicht-Geförderten nach Kontrolle der Kovariaten statistisch unabhängig von den erwarteten Effekten einer Förderung sein.

So wird beispielsweise beim European Research Council (ERC) der Score A für die besten Anträge unter Vorbehalt vergeben: „A. fully meets the ERC’s excellence criterion and *is recommended for funding, if sufficient funds are available ...*“ (European Research Council, 2021, S. 40, Fettdruck statt Kursivdruck im Original).

Für die Evaluation von Förderprogrammen wie „Experiment!“ der VWS kann ein spezifisches Design von Interesse sein, das als „doubly randomized preference trial (DRPT)“ bezeichnet wird (Marcus et al., 2012) und in dem ein echtes randomisiertes Experiment mit einem Quasi-Experiment mit Selbstselektion der Gruppen verglichen wird im Hinblick auf dasselbe Treatment, hier das Förderprogramm (siehe auch Shadish et al., 2008, 2011; Steiner & Wong, 2018). Statt Selbstselektion werden bei der VWS-Förderinitiative „Experiment!“ die Antragstellenden aufgrund der Entscheidung der Jury der Gruppe der Geförderten bzw. der Gruppe der Nicht-Geförderten zugewiesen.

Abbildung 1: Innerhalb-Gruppen-Versuchsdesign



Mit einem Messwiederholungsplan (Abbildung 1), der vier verschiedene Messzeitpunkte („within-group design“) vorsieht, lassen sich verschiedene Fragestellungen beantworten:

- Hat die Tatsache, gefördert oder nicht gefördert zu werden, einen Effekt auf den späteren Forschungsoutput der Antragstellenden, unabhängig davon, wie die Entscheidung getroffen wurde (Jury-Entscheidung oder Zufallszuweisung)?

- Unterscheiden sich Jury-Entscheidung und Zufallsentscheidung in ihren Effekten auf den Forschungsoutput der Antragstellenden?
- Sind Zufallsentscheidungen tatsächlich nach dem Zufall erfolgt bzw. Jury-Entscheidungen tatsächlich nicht zufällig, wenn eine Vielzahl von Variablen (z.B. Alter, Geschlecht der antragstellenden Person, Merkmale des Antrags) in Betracht gezogen werden?
- Kommen Jury-Entscheidungen und Lotterie-Entscheidungen zu demselben Ergebnis, wenn die Gruppen der Bewilligten und Nicht-Bewilligten mit einem Propensity-Score-Matching-Verfahren vergleichbar gemacht wurden (Mutz & Daniel, 2012; Mutz et al., 2017)?
- Gibt es kurzfristige und langfristige Effekte auf den Forschungsoutput abhängig von der Art der Förderentscheidung? Lässt sich die Vorauswahl der Geschäftsstelle aufgrund von Merkmalen des Antrags bzw. der Antragstellenden vorher-sagen?

Mit diesem Versuchsdesign ist eine umfangreiche Evaluation der Effekte eines Förderprogramms, wie die Förderinitiative „Experiment!“ der VWS möglich. Ein Problem ist sicherlich, dass die Anzahl der Geförderten bei der VWS pro Jahr mit 30–40 Anträgen vermutlich zu klein ist, um statistisch verlässliche Analysen durchzuführen.

6. Diskussion und Empfehlungen

Forschungsförderungsorganisationen geraten immer wieder in die Kritik, vor allem wegen der mangelnden Güte ihrer Peer-Review-Verfahren. Sie versuchen dieser Kritik mit geeigneten Maßnahmen zu begegnen, wie beispielsweise die VWS, die eine Lotterie-Komponente in das Auswahlverfahren einer ihrer Förderinitiativen eingebaut hat, um damit deren Fairness zu erhöhen.

Ziel dieses Beitrags war es, die allgemeinen Grundlagen von Lotterieverfahren aus verfahrenstechnischer, statistischer, ökonomischer und sozialpsychologischer Perspektive darzustellen, deren Anwendung auf die Forschungsförderung am Beispiel der VWS-Förderinitiative „Experiment“ zu diskutieren und ein mögliches Design für eine ex-post durchgeführte Wirkungsevaluation des Lotterieelements in der Auswahlentscheidung vorzuschlagen.

Insgesamt zeigt die Literaturanalyse, dass für die Implementation von Lotterien besondere Bedingungen erfüllt sein müssen. So muss es sich um einen mehrstufigen Entscheidungsprozess handeln, in dessen Zuge anhand einer Vorauswahl ein finales Set von Optionen bestimmt wird. Dieses finale Set darf sich in der Stärke der Argumente nicht mehr unterscheiden („Unbestimmtheit“), d.h., es dürfen keine positiven Gründe mehr vorhanden sein, eine Option, sprich einen Forschungsantrag, einer anderen Option vorzuziehen. Des Weiteren muss mit negativen Gründen (z.B. „gender bias“) gerechnet werden, die Einfluss auf die finale Entscheidung haben können und

durch die Lotterie beseitigt werden („reinigender Effekt“). Auch im Hinblick auf die Gestaltung der Förderung muss beachtet werden, dass das Gut, hier die Förderung, stückweise vorliegt und weitgehend homogen ist (annähernd gleicher Förderbetrag), die Ansprüche der Antragstellenden homogen sein müssen und nur hinsichtlich ihrer Stärke (z.B. Projektumfang) variieren dürfen. Wird das Lotterieverfahren in leistungsaffinen Situationen eingesetzt, muss gewährleistet sein, dass keine negativen psychologischen Nebeneffekte auftreten, die das Selbstwertgefühl, die wahrgenommene Selbstwirksamkeit, und die Leistungsmotivation der Betroffenen stark beeinflussen.

Allerdings bestehen in Forschungsförderorganisationen ganz unterschiedliche Auffassungen der Implementation von Randomisierung. So unterscheiden Feliciani und Shankar (2022, S. 3) auf einem Kontinuum von Peer-Review-Verfahren ohne Lotterie (Typ 0) zu voller Lotterie (Typ 4) drei weitere Typen: Typ 1 ist eine „tie-breaking mixed lottery“ mit fokaler Randomisierung, die wie beim SNF nur dann in Aktion tritt, wenn Anträge nach einer Begutachtung im Peer-Review-Verfahren die gleiche Rangposition in der gutachterlichen Bewertung („ties“) einnehmen. Randomisierung spielt bei diesem Typ nur eine untergeordnete Rolle. Beim Typ 2 „mixed lottery with bypass“ wird die fokale Randomisierung durch eine Bypass-Option erreicht. Ein Beispiel ist die VWS. Das gutachtende Panel erstellt den Pool der Anträge für die Lotterie, kann exzellente Anträge aber auch zur direkten Förderung vorschlagen und damit die Lotterie umgehen („bypass“). Im Typ 3 „mixed lottery of fundable proposal“ wird ganz auf fokale Randomisierung verzichtet und alle finanzierbaren Anträge, die Mindestkriterien erfüllen, werden einer Lotterie zugeführt wie beim Health Research Council of New Zealand.

Für die Einrichtung und Evaluation von Lotterielelementen in Auswahlverfahren, lassen sich folgende Empfehlungen ableiten (siehe auch Daniel, 2018, S. 25):

- Es muss geklärt werden, wie „bold research ideas“ und „riskiness“ von Forschungsanträgen zu messen sind.
- Es muss nachgewiesen werden, dass das Peer-Review-Verfahren tatsächlich Mängel im Hinblick auf die Inter-Rater-Reliabilität der Gutachterurteile und Verzerrungen (z.B. „gender bias“, „Milde-Effekt“) aufweist, sodass für ein Set von Anträgen keine inhaltlichen Gründe mehr bestehen, einen Antrag gegenüber einem anderen Antrag vorzuziehen.
- Auch die Vorauswahl der Anträge muss bestimmten Gütekriterien (z.B. prädiktive Validität, Fairness, Übereinstimmungsreliabilität) genügen.
- Es muss überprüft werden, ob das Lotterieverfahren bei kleinen Sets von Anträgen überhaupt sinnvoll ist und funktioniert, d.h. zufällig ist, und über einen größeren Zeitraum hinweg zu einer Balancierung der Gruppen im Hinblick auf Merkmale der Antragstellenden und des Antrags führt. Die Lotterie müsste ausreichend transparent sein.
- Es muss geprüft werden, ob dieses Verfahren auf Akzeptanz bei den Antragstellenden und der ‚scientific community‘ insgesamt stößt, und nicht das Selbstwertgefühl, die Leistungsmotivation und die wahrgenommene Selbstwirksamkeit der Antragstellenden negativ beeinflusst.

- Die Förderung sollte so gestaltet werden, dass an der Lotterie lediglich Projektanträge teilnehmen, die im Hinblick auf Größe, Antragssumme, Forschungsfeld etc. vergleichbar sind.
- Es muss sichergestellt werden, dass der Kreis der Antragsstellenden sich nicht durch die Art des Auswahlverfahrens unbeabsichtigt verändert.
- Die Möglichkeit von Wirkungsevaluationen sollte schon bei der Konzeption solcher Förderprogramme berücksichtigt werden und die für Ex-post-Evaluationen benötigten Daten sollten bereits während der Durchführung des Programms prozessbegleitend erhoben werden.
- In Wirkungsevaluationen sollte geprüft werden, ob das Lotterieverfahren tatsächlich zu besseren Ergebnissen im Forschungsoutput führt als die klassische Jury-Entscheidung. Es müssten in die Wirkungsevaluation nicht nur die „Marginals“ einbezogen werden, sondern alle Anträge, um eine Beschränkung des Wertebereichs durch Selektionseffekte („restriction of range“) zu vermeiden, die zu einer Unterschätzung der bewirkten Effekte führen würde.
- Im Hinblick auf den Datenschutz muss gewährleistet sein, dass alle, die einen Antrag stellen (bewilligt oder nicht bewilligt), sich schon zum Zeitpunkt der Antragsstellung verpflichten, sich später an Ex-post-Evaluationen zu beteiligen bzw. Daten beispielsweise über ihren Forschungsoutput zur Verfügung zu stellen.
- Ferner stellt sich die Frage, inwieweit Antragstellende über den Status der Förderentscheidung (Losglück oder Jury-Empfehlung) informiert werden sollten.

Angesichts der angeführten Bedingungen und Empfehlungen waren die Forschungsförderorganisationen in der Vergangenheit sicherlich gut beraten, ein Lotterieverfahren nur in begrenztem Umfang und als ein Element in einem mehrstufigen Auswahlverfahren einzusetzen.

Insgesamt ist die Diskussion über das Lotterie-Element in den Auswahlverfahren der Forschungsförderungsorganisationen bisher wenig evidenzbasiert. Es bedarf daher unbedingt empirischer Wirkungsevaluationen, auch wenn dies vielfach wegen der vergleichsweise geringen Fallzahlen und der disziplinären Heterogenität der Forschungsprojekte eine wissenschaftliche Herausforderung darstellt.

Literatur

- Adam, D. (2019). Science Funders Gamble on Grant Lotteries. *Nature*, 575(7784), 574–575. <https://doi.org/10.1038/d41586-019-03572-7>
- Allison, P. D. (2009). *Fixed Effects Regression Models*. Sage. <https://doi.org/10.4135/9781412993869>
- Avin, S. (2014). *Breaking the Grant Cycle: On the Rational Allocation of Public Resources to Scientific Research Projects*. (PhD). University of Cambridge.

- Avin, S. (2015). Funding Science by Lottery. In U. Maki, I. Votsis, S. Ruphy, & G. Schurz (Hrsg.), *Recent Developments in the Philosophy of Science: EPSA13 Helsinki* (S. 111–126). Springer. https://doi.org/10.1007/978-3-319-23015-3_9
- Avin, S. (2018). Policy Considerations for Random Allocation of Research Funds. *Journal on Research Policy & Evaluation*, 6(1), 1–27. <https://doi.org/10.13130/2282-5398/8626>
- Avin, S. (2019). Mavericks and Lotteries. *Studies in History and Philosophy of Science*, 76, 13–23. <https://doi.org/10.1016/j.shpsa.2018.11.006>
- Ayoubi, C., Pezzoni, M., & Visentin, F. (2019). The Important Thing is not to Win, it is to Take Part: What if Scientists Benefit from Participating in Research Grant Competitions? *Research Policy*, 48(1), 84–97. <https://doi.org/10.1016/j.respol.2018.07.021>
- Barlösius, E., Philipps, A. (2020). *Lotto in der Wissenschaft: Feldinterne und -externe Widersprüche durch nichtwissenschaftliche Entscheidungsverfahren auflösen?* <https://www.repo.uni-hannover.de/handle/123456789/10277>
- Barlösius, E., Philipps, A. (2021). Verlosung von Forschungsgeldern: Welche Losverfahren können sich Wissenschaftlerinnen und Wissenschaftler vorstellen? *Qualität in der Wissenschaft*, 15 (3+4), 67–82.
- Barnett, A. G. (2016). Funding by Lottery: Political Problems and Research Opportunities. *mBio*, 7(4), e01369-01316. <https://doi.org/10.1128/mBio.01369-16>
- Barnett, A. G., Graves, N., Clarke, P., & Blakely, T. (2015). *What is the Impact of Research Funding on Research Productivity?* Queensland University of Technology. <https://eprints.qut.edu.au/83127/>
- Bedessem, B. (2020). Should we fund research randomly? An epistemological criticism of the lottery model as an alternative to peer review for the funding of science. *Research Evaluation*, 29(2), 150–157. <https://doi.org/10.1093/reseval/rvz034>
- Benavente, J. M., Crespi, G., Garone, L. F., & Maffioli, A. (2012). The Impact of National Research Funds: A Regression Discontinuity Approach to the Chilean FONDECYT. *Research Policy*, 41(8), 1461–1475. <https://doi.org/10.1016/j.respol.2012.04.007>
- Bendiscioli, S., Firpo, T., Bravo-Biosca, A., Czibor, E., Garfinkel, M., Stafford, T., Wildson, J., & Buckley-Woods, H. (2021). *The Experimental Research Funder's Handbook* (RoRI Working Paper No.6). https://rori.figshare.com/articles/report/The_experimental_research_funder_s_handbook_final_version_/19459328
- Berger, J., Osterloh, M., & Rost, K. (2020). Focal Random Selection Closes the Gender Gap in Competitiveness. *Science Advances*, 6(47), eabb2142. <https://doi.org/10.1126/sciadv.abb2142>
- Bol, T., de Vaan, M., & van de Rijt, A. (2018). The Matthew Effect in Science Funding. *Proceedings of the National Academy of Sciences of the United States of America*, 115(19), 4887–4890. <https://doi.org/10.1073/pnas.1719557115>
- Bornmann, L., & Daniel, H.-D. (2008a). The Effectiveness of the Peer Review Process: Inter-Referee Agreement and Predictive Validity of Manuscript Refereeing at *Angewandte Chemie. Angewandte Chemie-International Edition*, 47(38), 7173–7178. <http://dx.doi.org/10.1002/anie.200800513>
- Bornmann, L., & Daniel, H.-D. (2008b). Selecting Manuscripts for a High-Impact Journal through Peer Review: A Citation Analysis of Communications that were Accepted by *Angewandte Chemie International Edition*, or Rejected but Published Elsewhere. *Journal of the American Society for Information Science and Technology*, 59(11), 1841–1852. <http://dx.doi.org/10.1002/asi.v59:11>
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2010). A Reliability-Generalization Study of Journal Peer Reviews: A Multilevel Meta-Analysis of Inter-Rater Reliability and its Determinants. *Plos One*, 5(12), e14331. <http://dx.doi.org/10.1371/journal.pone.0014331>

- Boyle, C. (1998). How the Process Could be Made Fairer by the Appropriate Use of Lotteries. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 47(2), 291–321. <https://www.jstor.org/stable/2988669>
- Bravo, G., Grimaldo, F., López-Iñesta, E., Mehmani, B., & Squazzoni, F. (2019). The Effect of Publishing Peer Review Reports on Referee Behavior in Five Scholarly journals. *Nature Communications*, 10(322), 1–8. <https://doi.org/10.1038/s41467-018-08250-2>
- Brezis, E. S. (2007). Focal Randomisation: An Optimal Mechanism for the Evaluation of R&D Projects. *Science and Public Policy*, 34(10), 691–698. <https://doi.org/10.3152/030234207X265394>
- Buckley-Woods, H., & Wilsdon, J. (2021): *Experiments with Randomisation in Research Funding: Scoping and Workshop Report*. https://rori.figshare.com/articles/report/Experiments_with_randomisation_in_research_funding_scoping_and_workshop_report/16553067
- Caspari, A. (2009). ‚Rigorose‘ Wirkungsevaluation – methodische und konzeptionelle Ansätze der Wirkungsmessung in der Entwicklungszusammenarbeit. *Zeitschrift für Evaluation*, 8(2), 183–213.
- Clarke, P., Herbert, D., Graves, N., & Barnett, A. G. (2016). A Randomized Trial of Fellowships for Early Career Researchers Finds a High Reliability in Funding Decisions. *Journal of Clinical Epidemiology*, 69, 147–151. <https://doi.org/10.1016/j.jclinepi.2015.04.010>
- Cole, S., Cole, J. R., & Simon, G. A. (1981). Chance and consensus in peer review. *Science*, 214(4523), 881–886. <https://doi.org/10.1126/science.7302566>
- Daniel, H.-D. (2018). *Peer Review and Beyond: Randomisation at the Margin in the Selection of Research Grant Proposals*. Invited paper presented at the 2nd PEERE International Conference on Peer Review, Rome, 7.–9. März 2018. <http://www.peere.org/wp-content/uploads/2018/03/Daniel.pdf>
- Daniel, H.-D., Bobokova, J., & Mutz, R. (2021). *Funding Programmes and Initiatives for Internationally Mobile Postdocs. Perceived Impacts on Individuals, Institutions and Society – An Exploratory Study*. Lemmens.
- Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues*, 31(3), 137–149. <https://doi.org/10.1111/j.1540-4560.1975.tb01000.x>
- de Peuter, S., & Conix, S. (2021). The Modified Lottery: Formalizing the Intrinsic Randomness of Research Funding. *Accountability in Research*, 29(5), 1–22. <https://doi.org/10.1080/08989621.2021.1927727>
- Elster, J. (1989). *Solomonic Judgements: Studies in the Limitations of Rationality*. Cambridge University Press.
- European Commission Directorate-General for Research and Innovation (2016). *Open Innovation, Open Science, Open to the World: A Vision for Europe*. <https://op.europa.eu/en/publication-detail/-/publication/3213b335-1cbc-11e6-ba9a-01aa75ed71a1>
- European Research Council (2021). *ERC Work Programme 2022*. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2022/wp_horizon-erc-2022_en.pdf
- Fang, F. C., & Casadevall, A. (2016). Research Funding: The Case for a Modified Lottery. *mBio*, 7(2), e00422-16. <https://doi.org/10.1128/mBio.00422-16>
- Feliciani, T., & Shankar, K. (2022). *Fairness in Funding Research by Lottery*. Paper presented at the STI conference 2022, Granada.
- Försterling, F. & Stiensmeier-Pelster, J. (1994). *Attributionstheorie*. Hogrefe.
- Frey, B. S., & Osterloh, M. (2015). *Zufallsauswahl in der Wissenschaft?* <https://www.bsfrey.ch/wp-content/uploads/2021/08/zufallsauswahl-in-der-wissenschaft.pdf>

- Gillies, D. (2014). Selecting Applications for Funding: Why Random Choice is Better than Peer Review. *Journal on Research Policy & Evaluation*, 2(1), 1–14. <https://doi.org/10.13130/2282-5398/3834>
- Gould, J. (2019, 8.2.2019): Working Scientist Podcast: The Grant Funding Lottery and How to Fix It. *Nature Careers Podcast*. <https://www.nature.com/articles/d41586-019-00525-y>
- Graves, N., Barnett, A. G., & Clarke, P. (2011). Funding Grant Proposals for Scientific Research: Retrospective Analysis of Scores by Members of Grant Review Panel. *BMJ*, 343, d4797. <https://doi.org/10.1136/bmj.d4797>
- Greenberg, D. S. (1998). Chance and Grants. *The Lancet*, 351(9103), 686. [https://doi.org/10.1016/S0140-6736\(05\)78485-3](https://doi.org/10.1016/S0140-6736(05)78485-3)
- Greitemeyer, T., & Schulz-Hardt, S. (2003). Preference-consistent Evaluation of Information in the Hidden Profile Paradigm: Beyond Group-Level Explanations for the Dominance of Shared Information in Group Decisions. *Journal of Personality and Social Psychology*, 84(2), 322–339. <http://dx.doi.org/10.1037/0022-3514.84.2.322>
- Harnagel, A. (2019). A mid-level approach to modeling scientific communities. *Studies in History and Philosophy of Science Part A*, 76, 49–59. <https://doi.org/10.1016/j.shpsa.2018.12.010>
- Health Research Council of New Zealand (2018). *Explorer Grants*. <http://www.hrc.govt.nz/funding-opportunities/researcher-initiated-proposals/explorer-grants>
- Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing Selection Bias using Experimental Data. *Econometrica*, 66(5), 1017–1098. <https://doi.org/10.2307/2999630>
- Herbert, D. L., Barnett, A. G., Clarke, P., & Graves, N. (2013). On the Time Spent Preparing Grant Proposals: An Observational Study of Australian Researchers. *BMJ Open*, 2013(3), e002800. <https://doi.org/10.1136/bmjopen-2013-002800>
- Heyard, R., Ott, M., Salanti, G., & Egger, M. (2022). Rethinking the Funding Line at the Swiss National Science Foundation: Bayesian Ranking and Lottery. *Statistics and Public Policy*, 9(1), 110–121. <https://doi.org/10.1080/2330443X.2022.2086190>
- Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396), 945–960. <https://doi.org/10.2307/2289064>
- Horton, R. (1996). Luck, Lotteries, and Loopholes of Grant Review. *The Lancet*, 348(9037), 1255–1256. [https://doi.org/10.1016/S0140-6736\(05\)65751-0](https://doi.org/10.1016/S0140-6736(05)65751-0)
- Jacob, B. A., & Lefgren, L. (2011). The Impact of Research Grant Funding on Scientific Productivity. *Journal of Public Economics*, 95(9–10), 1168–1177. <https://doi.org/10.1016/j.jpubeco.2011.05.005>
- Jaffe, A. B. (2002). Building Programme Evaluation into the Design of Public Research Support Programmes. *Oxford Review of Economic Policy*, 18(1), 22–34. <https://doi.org/10.1093/oxrep/18.1.22>
- Jordan, G. B. (2013). Logic Modeling: A Tool for Designing Program Evaluations. In A. Link, & N. Vonortas (Hrsg.), *Handbook on the Theory and Practice of Program Evaluation* (S. 143–165). Edward Elgar. <https://doi.org/10.4337/9780857932402.00012>
- Klette, T. J., Møen, J., & Griliches, Z. (2000). Do Subsidies to Commercial R & D reduce Market Failure? Microeconomic Evaluation Studies. *Research Policy*, 29(4–5), 471–495. [https://doi.org/10.1016/S0048-7333\(99\)00086-4](https://doi.org/10.1016/S0048-7333(99)00086-4)
- Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in Peer Review. *Journal of the American Society for Information Science and Technology*, 64(1), 2–17. <https://doi.org/10.1002/asi.22784>
- Leibbrandt, A., Wang, L. C., & Foo, C. (2018). Gender Quotas, Competitions, and Peer Review: Experimental Evidence on the Backlash against Women. *Management Science*, 64(8), 3501–3516. <https://doi.org/10.1287/mnsc.2017.2772>

- Liu, M., Choy, V., Clarke, P., Barnett, A., Blakely, T., & Pomeroy, L. (2020). The Acceptability of Using a Lottery to Allocate Research Funding: A Survey of Applicants. *Research Integrity and Peer Review*, 5(3), 1–7. <https://doi.org/10.1186/s41073-019-0089-z>
- Loveday, V. (2017). Luck, Chance, and Happenstance? Perceptions of Success and Failure Amongst Fixed-Term Academic Staff in UK Higher Education. *The British Journal of Sociology*, 69(3), 758–775. <https://doi.org/10.1111/1468-4446.12307>
- Mahoney, M. J. (1977). Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System. *Cognitive Therapy and Research*, 1(2), 161–175. <https://doi.org/10.1007/BF01173636>
- Marcus, S. M., Stuart, E. A., Wang, P., Shadish, W. R., & Steiner, P. M. (2012). Estimating the Causal Effect of Randomization Versus Treatment Preference in a Doubly Randomized Preference Trial. *Psychological Methods*, 17(2), 244–254. <https://doi.org/10.1037/a0028031>
- Marsh, H. W., Bornmann, L., Mutz, R., Daniel, H.-D., & O’Mara, A. (2009). Gender Effects in the Peer Reviews of Grant Proposals: A Comprehensive Meta-Analysis Comparing Traditional and Multilevel Approaches. *Review of Educational Research*, 79(3), 1290–1326. <http://dx.doi.org/10.3102/0034654309334143>
- Marsh, H. W., Jayasinghe, U. W., & Bond, N. (2008). Improving the Peer Review for Grant Applications. *American Psychologist*, 63(3), 160–168. <http://dx.doi.org/10.1037/0003-066X.63.3.160>
- Morgan, S. L., & Winship, C. (2015). *Counterfactuals and Causal Inference. Methods and Principles for Social Research*. Cambridge University Press. <https://doi.org/10.1017/CBO9781107587991>
- Mutz, R., Bornmann, L., & Daniel, H.-D. (2012a). Does Gender matter in Grant Peer Review? An Empirical Investigation using the Example of the Austrian Science Fund. *Zeitschrift für Psychologie*, 220(2), 121–129. <http://dx.doi.org/10.1027/2151-2604/a000103>
- Mutz, R., Bornmann, L., & Daniel, H.-D. (2012b). Heterogeneity of Inter-Rater Reliabilities of Grant Peer Reviews and its Determinants: A General Estimating Equations Approach. *Plos One*, 7(10), e48509. <https://doi.org/10.1371/journal.pone.0048509>
- Mutz, R., Bornmann, L., & Daniel, H.-D. (2015). Testing for the Fairness and Predictive Validity of Research Funding Decisions: A Multilevel Multiple Imputation for Missing Data Approach using Ex-Ante and Ex-Post Peer Evaluation Data from the Austrian Science Fund. *Journal of the Association for Information Science and Technology*, 66(11), 2321–2339. <http://dx.doi.org/10.1002/asi.23315>
- Mutz, R., Bornmann, L., & Daniel, H.-D. (2016). Funding Decision-Making Systems: An Empirical Comparison of Continuous and Dichotomous Approaches based on Psychometric Theory. *Research Evaluation*, 25(4), 416–426. <https://doi.org/10.1093/reseval/rvw002>
- Mutz, R., & Daniel, H.-D. (2012). The Generalized Propensity Score Methodology for Estimating Unbiased Journal Impact Factors. *Scientometrics*, 92(2), 377–390. <https://doi.org/10.1007/s11192-012-0670-4>
- Mutz, R., Wolbring, T., & Daniel, H.-D. (2017). The Effect of the “Very Important Paper” (VIP) Designation in *Angewandte Chemie International Edition* on citation impact: A Propensity Score Matching Analysis. *Journal of the Association for Information Science and Technology*, 68(9), 2139–2153. <https://doi.org/10.1002/asi.23701>
- Neff, B. D., & Olden, J. D. (2006). Is Peer Review a Game of Chance? *BioScience*, 56(4), 333–340. [https://doi.org/10.1641/0006-3568\(2006\)56\[333:IPRAGO\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2006)56[333:IPRAGO]2.0.CO;2)
- Osterloh, M. (2017). Würfelt Gott? Würfelt die Wissenschaft? *Beiträge zur Hochschulforschung*, 39(1), 30–48.
- Osterloh, M., & Frey, B. S. (2015). Würfeln in der Wissenschaft? Über die Verbindung von Expertenurteilen und Zufall. *Forschung & Lehre*, 2, 134–135.

- Osterloh, M., & Frey, B. S. (2019). Dealing with Randomness. *Management Revue*, 30(4), 331–345. <https://doi.org/10.5771/0935-9915-2019-4-331>
- Osterloh, M., & Frey, B. S. (2020). If at First an Idea is Not Absurd, Then There is No Hope For it: A Response to the Comments on our Discussion Paper on 'Borrowed Plumes'. *Research Policy*, 49(1). <https://doi.org/10.1016/j.respol.2019.103870>
- Oswald, A. J. (2020). Rational randomization by journal editors—a mathematical derivation: A response to Osterloh and Frey's discussion paper on 'Borrowed Plumes'. *Research Policy*, 49(1). <https://doi.org/10.1016/j.respol.2019.103869>
- Philipps, A. (2021). Science Rules! A Qualitative Study of Scientists' Approaches to Grant Lottery. *Research Evaluation*, 30(1), 102–111. <https://doi.org/10.1093/reseval/rvaa027>
- Pion, G. M., & Cordray, D. S. (2008). The Burroughs Wellcome Career Award in the Biomedical Sciences: Challenges to and Prospects for Estimating the Causal Effects of Career Development Programs. *Evaluation & the Health Professions*, 31(4), 335–369. <https://doi.org/10.1177/0163278708324434>
- Reinhart, M., & Schendzielorz, C. (2020). The Lottery in Babylon – On the Role of Chance in Scientific Success. *Journal Responsible Innovation*, 7(sup2), S25–S29. <https://doi.org/10.1080/23299460.2020.1806429>
- Röbecke, M., & Simon, D. (2020). Die Macht des Zufalls – Neue Wege für die Förderung risikanter Forschungsideen. *Forschung*, 2020 (1+2), 9–14.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Hrsg.). (2005). *Publication Bias in Meta-Analysis – Prevention, Assessment and Adjustment*. John Wiley. <https://doi.org/10.1002/0470870168>
- Roumbanis, L. (2019). Peer Review or Lottery? A Critical Analysis of Two Different Forms of Decision-Making Mechanisms for Allocation of Research Grants. *Science, Technology, & Human Values*, 44(6), 994–1019. <https://doi.org/10.1177/0162243918822744>
- Roy, R. (1985). Funding Science: The Real Defects of Peer Review and an Alternative to it. *Science, Technology, & Human Values*, 10(3), 73–81. <https://doi.org/10.1177/016224398501000309>
- Rubin, D. B. (2007). Statistical Inference for Causal Effects, with Emphasis on Applications in Psychometrics and Education. In C. R. Rao, & S. Sinharay (Hrsg.), *Handbook of Statistics* (Vol. 26, S. 769–800). North Holland. [https://doi.org/10.1016/S0169-7161\(06\)26024-3](https://doi.org/10.1016/S0169-7161(06)26024-3)
- Rubin, D. B. (2011). Rubin Causal Model. In M. Lovric (Hrsg.), *International Encyclopedia of Statistical Science* (S. 1263–1265). Springer. https://doi.org/10.1007/978-3-642-04898-2_64
- Sandström, U., & Hällsten, M. (2008). Persistent Nepotism in Peer-Review. *Scientometrics*, 74(2), 175–189. <https://doi.org/10.1007/s11192-008-0211-3>
- Santana, C. (2022). Why Citizen Review Might beat Peer Review at Identifying Pursuitworthy Scientific Research. *Studies in History and Philosophy of Science*, 92, 20–26. <https://doi.org/10.1016/j.shpsa.2022.01.012>
- Scheiner, S. M. & Bouchie, L. M. (2013). The Predictive Power of NSF Reviewers and Panels. *Frontiers in Ecology*, 11(8), 406–407. <https://doi.org/10.1890/13.WB.017>
- Schnotz, W. (2011). *Pädagogische Psychologie kompakt*. Beltz.
- Scholten, W., van Drooge, L., & Diederens, P. (2018). *Excellence is Extra-Ordinary – Thirty Years of Focus on Excellence in Dutch Science Policy*. <https://www.rathenau.nl/en/knowledge-ecosystems/excellence-extra-ordinary>
- Schweizer Nationalfonds (2018). *Postdoc.Mobility*. <http://www.snf.ch/de/fokusForschung/newsroom/Seiten/news-181101-ausschreibung-postdoc-mobility.aspx>

- Science for Technological Innovation (2013). *Seed project development process*. <https://www.sfti-challenge.govt.nz/for-researchers/funding-and-get-involved/seed-project-development-process/>
- Shadish, W. R., Clark, M. H., Steiner, P. M. (2008). Can Nonrandomized Experiments Yield Accurate Answers? A Randomized Experiment Comparing Random and Nonrandom Assignments. *Journal of the American Statistical Association*, 103(484), 1334–1344. <https://doi.org/10.1198/016214508000000733>
- Shadish, W. R., Galindo, R., Wong, V., Steiner, P. M., & Cook, T. D. (2011). A Randomized Experiment Comparing Random and Cutoff-based Assignment. *Psychological Methods*, 16(2), 179–191. <https://doi.org/10.1037/a0023345>
- Smith, R. (2006). Peer Review: A Flawed Process at the Heart of Science and Journals. *Journal of the Royal Society of Medicine*, 99, 178–182. <https://doi.org/10.1258/jrsm.99.4.178>
- Squazzoni, F., & Gandelli, C. (2012). Saint Matthew Strikes Again: An Agent-based Model of Peer Review and the Scientific Community Structure. *Journal of Informetrics*, 6(2), 265–275. <https://doi.org/10.1016/j.joi.2011.12.005>
- Steiner, P. M., & Wong, V. C. (2018). Assessing Correspondence Between Experimental and Non-experimental Estimates in Within-Study Comparisons. *Evaluation Review*, 42(2), 214–247. <https://doi.org/10.1177/0193841X18773807>
- Stone, P. (2011). *The Luck of the Draw – The Role of Lotteries in Decision Making*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199756100.003.0001>
- Stone, P. (2013). Access to Higher Education by the Luck of the Draw. *Comparative Education Review*, 57(3), 577–599. <https://doi.org/10.1086/670663>
- Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-Discontinuity Analysis: An Alternative to Ex Post Facto Experiment. *Journal of Educational Psychology*, 51(6), 309–317. <https://doi.org/10.1037/h0044319>
- Treischl, E., & Wolbring, T. (2020). *Wirkungsevaluation*. Beltz.
- van der Linden, W., & Wise, L. (2014). Erosion of Peer Review? *Educational Measurement: Issue and Practice*, 33(4), 3. <https://doi.org/10.1111/emip.12044>
- VolkswagenStiftung (2019a). *Experiment! – Auf der Suche nach gewagten Forschungsideen*. <https://www.volkswagenstiftung.de/unsere-foerderung/unsere-foerderung-im-ueberblick/experiment>
- VolkswagenStiftung (2019b). *Information zur Antragstellung 100*. https://www.volkswagenstiftung.de/sites/default/files/downloads/MB_100_d.pdf
- VolkswagenStiftung (2023). *Wieviel Risiko braucht die Förderung von Wissenschaft?* <https://www.volkswagenstiftung.de/de/veranstaltungen/wieviel-risiko-braucht-die-foerderung-von-wissenschaft>
- Weber, C. (2022, 2. September). Forschungspolitik – Gehen Sie über Los. *Süddeutsche Zeitung*. <https://www.sueddeutsche.de/wissen/forschung-forschungspolitik-entscheidung-foerderung-zufall-1.5649973>
- Weiner, B. (1985). An Attributional Theory of Achievement Motivation and Emotion. *Psychological Review*, 92(4), 548–573. <https://doi.org/10.1037/0033-295X.92.4.548>
- Wolbring, T. (2014). Kausalanalyse und Wirkungsevaluation. Potential Outcomes, Graphenmethodologie und ihre Anwendung am Beispiel der Bologna-Reform. *Zeitschrift für Evaluation*, 13(2), 243–270.
- Wooding, S. (2020). Heuristics, not Plumage: A Response to Osterloh and Frey’s Discussion Paper on ‘Borrowed plumes’. *Research Policy*, 49(1), a103871. <https://doi.org/10.1016/j.respol.2019.103871>

- Yaqub, O. (2020). JIFs, Giraffes, and a Diffusion of Culpability: A Response to Osterloh and Frey's Discussion Paper on 'Borrowed Plumes.' *Research Policy*, 49(1), a103868. <https://doi.org/10.1016/j.respol.2019.103868>
- Yokoyama, S. (1996). Japanese Grant System is a Lottery. *Nature*, 384, 402–403. <https://doi.org/10.1038/384402a0>

Dr. Rüdiger Mutz | Universität Zürich | Center of Higher Education and Science Studies (CHESS) | Plattenstrasse 54 | CH-8032 Zürich |
E-Mail: ruediger.mutz@uzh.ch

Prof. Dr. Tobias Wolbring | Friedrich-Alexander-Universität Erlangen-Nürnberg |
Fachbereich Wirtschafts- und Sozialwissenschaften | Lehrstuhl für Empirische
Wirtschaftssoziologie | Findelgasse 7/9 | D-90402 Nürnberg |
E-Mail: tobias.wolbring@fau.de

Prof. Dr. Hans-Dieter Daniel | Universität Zürich | Psychologisches Institut |
Professur für empirische Hochschulforschung | Binzmühlestrasse 14 | Box 28 |
CH-8050 Zürich | E-Mail: hans-dieter.daniel@uzh.ch