

26. Risikoscheue Frauen und wettbewerbsorientierte Männer – wie relevant sind Geschlechterdifferenzen in Neigungsmaßen? Antworten aus der Meta-Forschung

Miriam Beblo und Eva Markowsky

Prof. Dr. Miriam Beblo ist Professorin der Volkswirtschaftslehre am Fachbereich Sozialökonomie der Universität Hamburg. Als angewandte Mikroökonomin forscht sie zu den Themen Arbeit, Familie, Gender und Migration. In der Lehre versucht sie, den praktischen Nutzen theoretischer und empirischer Erkenntnisse aus der VWL für die Politik in den Mittelpunkt zu rücken. Sie selbst hat dabei an vielen Stellen ihrer beruflichen Laufbahn Erfahrung mit wissenschaftsbasierter Politikberatung gesammelt und wirkt in zahlreichen politischen Beratungsgremien mit. Sie engagiert sich in verschiedenen Gremien der Nachwuchs- und Forschungsförderung. Zudem ist sie Mitglied im Vorstand der Deutschen Vereinigung für sozialwissenschaftliche Arbeitsmarktforschung (SAMF e.V.)

Prof. Dr. Eva Markowsky ist Juniorprofessorin für Quantitative Ökonomik an der Universität Potsdam. Sie forscht zu Arbeitsmarkt-, Migrations- und Bildungsökonomik mit Schwerpunkt auf Geschlechtergleichheit und Diversity. Zudem engagiert sie sich in Meta-Science und Methodenforschung. Sie studierte und promovierte an der Universität Hamburg. Eine weitere Leidenschaft von ihr betrifft die Kommunikation von wirtschafts- und sozialwissenschaftlicher Forschung an die interessierte Öffentlichkeit. Sie ist u. a. Mitglied der Research Group REAL sowie Stipendiatin der Joachim-Herz-Stiftung im Add-on Fellowship Interdisziplinäre Ökonomik.

Zusammenfassung:

- Individuelle Präferenzen bezüglich Risiko, Wettbewerb, Selbstvertrauen und Altruismus beeinflussen ökonomische Entscheidungen, was zu Geschlechterlücken in Erfolgen führen kann.
- Meta-Analysen helfen, die wachsende Zahl empirischer Einzelstudien systematisch zu erfassen und deren Schlussfolgerungen kritisch zu prüfen. Sie sind damit ein wichtiges Instrument zur Validierung empirischer Forschung.

- Eine Meta-Analyse zu Geschlechterdifferenzen im Wettbewerbsverhalten zeigt deutlich geringere Unterschiede als bisher angenommen.
- Auch Meta-Studien zu Risikoneigung, Selbstbewusstsein und Altruismus belegen: Geschlechterunterschiede sind meist klein und stark kontextabhängig.

26.1 Einleitung

Der Arbeitsmarkt ist geprägt von Geschlechterdifferenzen in vielerlei Dimensionen. In manchen Dimensionen, wie der Erwerbsquote, haben sich die Unterschiede zwischen Frauen und Männern deutlich verringert, weil im Verlauf der letzten Jahrzehnte immer mehr Frauen auf dem Arbeitsmarkt tätig geworden sind. In anderen Dimensionen dagegen, wie dem Umfang der Erwerbsarbeit, dem Arbeitsvolumen (Gesamtzahl der Erwerbsstunden), den hierarchischen Positionen und der Bezahlung, gab es zwar leichte Annäherungen zwischen weiblichen und männlichen Beschäftigten, dennoch blieben Lücken erhalten.

Die geschlechterspezifische Einkommenslücke kann dabei als aggregiertes Maß auch der anderen Dimensionen von Ungleichheit auf dem Arbeitsmarkt bezeichnet werden, da sie die kumulierten Unterschiede in der bezahlten Tätigkeit, Arbeitszeit und Position in einer einzigen Größe zusammenbringt. Ein Teil des geschlechterspezifischen Entgeltunterschieds lässt sich auf den Umfang der Arbeitszeit, das Tätigkeitsprofil, die Branche oder die Position im Betrieb zurückführen, die Frauen im Vergleich zu Männern typischerweise einnehmen. Ein anderer Teil ist durch keinen dieser Faktoren erklärbar, denn auch bei ähnlichen Tätigkeiten und ähnlichen Qualifikationen verbleibt eine Entgeltlücke. In der Berichterstattung von Destatis zum *gender pay gap* wird diese regelmäßig belegt und als »bereinigte«, an anderer Stelle auch als »angepasste« Entgeltlücke, bezeichnet.

Um zum einen die Ursachen für die verbleibende bereinigte Entgeltlücke näher zu beleuchten sowie zum anderen die Ursachen für unterschiedliche Karriereverläufe von Frauen und Männern jenseits von strukturellen Benachteiligungen (wie der Zuweisung von Sorgeverantwortung, Mitspracherechten, Diskriminierung und sozialen Normen) tiefer zu ergründen, hat die ökonomische Geschlechterforschung in den vergangenen zwanzig Jahren zunehmend die Rolle von Präferenzen und Neigungen in den Blick genommen. Insbesondere die Risikoneigung, die Wettbewerbsneigung und das Selbstvertrauen sind individuelle Merkmale (im Englischen häufig als *traits* bezeichnet), von denen gut dokumentiert ist, dass sie großen Einfluss auf ökonomisch relevantes individuelles Verhalten haben. Sie beeinflussen beispielsweise die Wahl von Studiengängen und Berufen, und können damit auch auf den ökonomischen Erfolg, gemessen am Bildungs-

niveau und Einkommen, einwirken (Buser u.a. 2014; Reuben u.a. 2017). Es gibt natürlich weitere ökonomisch bedeutsame Neigungen, wie zum Beispiel Altruismus, für welche Geschlechterdifferenzen dokumentiert sind. Beim Altruismus ist der Zusammenhang mit dem individuellen Arbeitsmarkterfolg aber nicht so eindeutig. Einerseits könnten Unterschiede im altruistischen Verhalten dergestalt, dass Frauen pro-sozialer eingestellt wären als Männer, zu geringeren Geschlechterlücken auf dem Arbeitsmarkt beitragen, wenn sie dadurch mehr Erfolg in Beruf und Karriere hätten (Falk/Hermle 2018; Kosse/Tincani 2020). Andererseits könnte altruistisches Verhalten auch ökonomisch nachteilig sein, insbesondere in Wettbewerbssituationen und Gehaltsverhandlungen.

Für alle genannten Neigungsmaße werden in der Verhaltensökonomie i.d.R. unterschiedliche Ausprägungen zwischen Frauen und Männern vermutet. Entsprechend wurden zu allen Neigungsmaßen bereits zahlreiche experimentelle Studien und Überblicksartikel verfasst. Nach gängiger Lesart lassen die Ergebnisse der Einzelstudien auf statistisch wie ökonomisch signifikante Unterschiede zwischen den Geschlechtern schließen. Auch Mukesh Eswaran (2014) zieht diese Schlussfolgerung im zweiten Kapitel *Do Women and Men Behave Differently in Economic Situations?* seines in vielen Seminaren zu Gender Economics eingesetzten Lehrbuchs *Why Gender Matters in Economics*. Bei genauerer Betrachtung und systematischer Sichtung der experimentellen Befunde, zum Beispiel im Rahmen von Meta-Analysen, zeigt sich allerdings, dass die vermeintlich gesicherten Geschlechterdifferenzen viel weniger robust sind als gemeinhin angenommen. Stattdessen zeigen sich mitunter sogar große Ähnlichkeiten in den Verhaltensmaßen von Frauen und Männern sowie eine starke Abhängigkeit des Verhaltens von konkreten Entscheidungssituationen und Kontextbedingungen.

Unserer Auffassung nach ist die systematische Sichtung und Neubewertung von empirischer Evidenz eine immerwährende Herausforderung und gleichzeitig Anspruch, um die Integrität und Validität von empirischen Ergebnissen und daraus gezogenen Schlussfolgerungen zu gewährleisten. Im Folgenden stellen wir dar, warum die Methode der Meta-Analyse geeignet ist, um der Ergebnisbandbreite von Einzelstudien und dem Phänomen des so genannten *empirical shift* innerhalb der Wirtschaftswissenschaften sinnvoll zu begegnen. Wir reflektieren die Vor- und Nachteile meta-analytischer Methoden und liefern eine kurze Einführung in ihre konkrete Anwendung am Beispiel unserer eigenen Meta-Analyse zur geschlechterspezifischen Wettbewerbsneigung (Markowsky/Beblo 2022). Hiermit wollen wir eine Reflektion über die vermeintlich gesicherte empirische Evidenz zu Geschlechterunterschieden in ökonomischen Neigungsmaßen – und darüber hinaus – anregen.

26.2 Meta-Analyse – was ist das?

Ursprünglich in der Psychologie und medizinischen Forschung entwickelt, hat die Meta-Analyse sich auch in den Wirtschaftswissenschaften als ein wertvolles Instrument etabliert, um umfangreiche und heterogene Bestände an empirischen Studien systematisch zusammenzufassen. Die Verlagerung von theoretischer hin zu empirischer Forschung in den Wirtschaftswissenschaften in den vergangenen gut 40 Jahren ist unter dem Schlagwort *empirical shift* dokumentiert (Angrist u.a. 2017). Diese Entwicklung bietet einerseits enormes Potenzial, da sie das verfügbare Wissen über ökonomische Zusammenhänge stetig erweitert und Handlungsempfehlungen auf eine breitere empirische Basis stellt. Andererseits ist die stetig und schnell wachsende Zahl empirischer Ergebnisse immer schwieriger zu verarbeiten und droht zu einer Fragmentierung des Wissens zu führen. Die Methode der Meta-Analyse ermöglicht es nun Forschenden, die Ergebnisse zahlreicher Studien zu bündeln, Muster zu erkennen und daraus robustere und verallgemeinerbare Schlussfolgerungen zu ziehen.

In der Ökonomik zielten die frühen Anwendungen der Meta-Analyse vor allem darauf ab, Spezifikationsfehler in den Einzelstudien zu identifizieren und zu korrigieren, weshalb sich hier schnell die Methode der Meta-Regressionsanalyse etablierte. Diese nutzt Regressionstechniken, um zu analysieren, wie unterschiedliche Studiendesigns und Spezifikationsmerkmale – etwa die statistischen Methoden oder Auswahl erklärender Variablen – als sogenannte Moderatoren die Variabilität der Ergebnisse in ökonometrischen Studien beeinflussen. Dabei werden die Ergebnisse der einzelnen Studien als abhängige Variable auf die Moderatoren regressiert. So können studienübergreifend nicht nur relevante Kontextvariablen identifiziert werden, sondern es ergeben sich oft auch wertvolle Hinweise für weitergehende Analysen (Stanley/Doucouliagos 2012, S. 7). Manchmal bieten Meta-Analysen sogar die Möglichkeit, Fragen zu beantworten, die sich in Einzelstudien nicht adäquat untersuchen lassen. So hinterfragt beispielsweise die Meta-Analyse von Weichselbaumer/Winter-Ebmer (2007) das theoretische Argument, dass Diskriminierung in einem Wettbewerbsmarkt nicht bestehen kann und der freie Wettbewerb somit zu stärkerer Lohngleichheit führt. Dafür analysieren sie, wie sich unterschiedliche Wirtschaftsbedingungen und der Grad der Marktoffenheit einer Volkswirtschaft, auf die in Einzelstudien ermittelten nationalen Lohnunterschiede auswirken. Eine andere Autorengruppe um Card (2018) untersucht aktive Arbeitsmarktpolitiken in einer Meta-Studie und nutzt die Variation im Studiendesign und -timing innerhalb eines umfangreichen Literaturkorpus, um aufzuzeigen, wie verschiedene Arbeitsmarktprogramme unterschiedliche Bevölkerungsgruppen erreichen und wie deren Wirkung vom Konjunkturzyklus abhängt. In beiden Fällen lassen sich durch Meta-Analysen wichtige Einblicke gewinnen,

indem die Heterogenität im Studiendesign und in den Daten einer großen Zahl von Studien ausgenutzt wird. So werden Aspekte beleuchtet, die bei einer Analyse der Primärdaten allein entweder gar nicht oder nur sehr aufwändig und daher kostenintensiv hätten berücksichtigt werden können.

Ein häufig vorgebrachtes Argument gegen die Validität meta-analytischer Forschung ist das sogenannte *Garbage in, garbage out*-Problem. Der zentrale Punkt hierbei ist, dass Meta-Analysen nur dann verlässliche Ergebnisse liefern, wenn sie auf methodisch sauberen und qualitativ hochwertigen Einzelstudien basieren. Werden minderwertige oder fehlerhafte Studien einbezogen, führt dies zwangsläufig zu einer Verzerrung der Meta-Ergebnisse. Eine hochwertige Meta-Analyse erfordert daher strenge Qualitätskontrollen, was im Falle intransparenter und mitunter fragwürdiger Publikationspraktiken empirischer Forschung eine Herausforderung darstellen kann. Zudem ist die Bewertung der Qualität einer Studie wiederum Gegenstand subjektiver Einschätzung der Meta-Forschenden (Anderson/Kichkha 2017, S. 58). Ein weiteres ernsthaftes Problem für Meta-Analysen ist die möglicherweise verzerrte Veröffentlichungswahrscheinlichkeit, der so genannte Publikations-Bias. Der Trend, dass vor allem statistisch signifikante oder theoriekonforme Ergebnisse veröffentlicht werden, wird sowohl durch Editor:innen und Gutachter:innen im Publikationsprozess als auch durch die Autor:innen selbst begünstigt. Das sogenannte *File-drawer*-Problem beschreibt die Tatsache, dass nicht-signifikante Studien oft gar nicht zur Veröffentlichung eingereicht werden und somit für Meta-Analysen nicht zugänglich sind. Diese Verzerrung wird durch fragwürdige, aber verbreitete Praktiken wie *p-hacking* und *specification searching* verstärkt, bei denen Daten oder Analysen so lange manipuliert werden, bis sich signifikante Ergebnisse ergeben. Die selektive Veröffentlichung von »positiven«, signifikanten, theoriekonformen, oder die gängige Auffassung untermauernden Ergebnissen kann dazu führen, dass Meta-Analysen ein verzerrtes Bild der tatsächlichen Zusammenhänge liefern. Diese systematische Überschätzung von Effekten durch Publikations-Bias stellt eine ernsthafte Bedrohung für die Validität meta-analytischer Forschung dar. All diesen Herausforderungen muss methodisch begegnet werden, um mit Hilfe der Meta-Analyse zu aussagekräftigen und verlässlichen Ergebnissen zu gelangen.

26.3 Meta-Analyse praktisch: Geschlechterdifferenzen in der Wettbewerbsneigung

Der Forschungsstand zu Geschlechterdifferenzen in der Wettbewerbsneigung stellte sich im Jahr 2020 folgendermaßen dar: Zahlreiche Einzelstudien wiesen auf erhebliche Unterschiede zwischen weiblichen und männlichen Teilnehmenden in Wettbewerbsexperimenten hin, angefangen mit der grundlegenden Experimentalstudie von Niederle/ Vesterlund (2007). Für dieses Originalexperiment wurden Anfang der 2000er Jahre studentische Teilnehmende rekrutiert, die im Experimentallabor einer US-amerikanischen Universität Rechenaufgaben lösen sollten und dafür einmal mit einem Stücklohn und einmal mit einem Wettbewerbslohn (beste Leistung unter Vieren) kompensiert wurden. Anschließend wurden die Teilnehmenden für weitere Rechenaufgaben vor die Wahl zwischen Stück- und Wettbewerbslohn gestellt. Hierbei wählten die männlichen Teilnehmenden deutlich häufiger den Wettbewerbslohn als die weiblichen. In der Folge wurden zahlreiche Wettbewerbsexperimente von unterschiedlichen Autor:innengruppen weltweit durchgeführt, die verschiedene Aspekte des Originalexperimentes beibehielten oder diese bewusst variierten. Narrative Überblicksartikel von Niederle (2017) und Shurchkov/Eckel (2018) interpretierten die Ergebnisse der Einzelstudien als klaren Beleg für einen Geschlechterunterschied in der Wettbewerbsneigung. So schrieb Niederle (2017, S. 490) in ihrem Handbook-Artikel »Gender« beispielsweise: »the existence of a gender gap in tournament entry has stood the test of replication«. Eine Meta-Analyse zur Verallgemeinerbarkeit des Befundes lag zu dem Zeitpunkt noch nicht vor. Wir beschlossen, diese Lücke zu schließen und entwickelten die Meta-Studie »When do we observe a gender gap in competition entry? A meta-analysis of the experimental literature« (Markowsky/Beblo 2022).

In ihrem Vorgehen folgen die meisten Meta-Studien folgendem fünfteiligen Schema: I) Studienrecherche, II) Studiensichtung und -filterung, III) Kodierung, IV) Datenaufbereitung und -bereinigung, V) statistische Analyse. In unserer eigenen Meta-Analyse recherchierten und sichteten wir deshalb in den Arbeitsschritten I) und II) zunächst die vorhandene internationale experimentelle Evidenz zu geschlechterspezifischen Unterschieden in Wettbewerbspräferenzen. Wir fanden 110 Einzelstudien aus den Jahren 2007 bis 2021 mit über 400 Effektstärken aus 30 Ländern und sechs Kontinenten. Der analytische Ausgangspunkt jeder Meta-Analyse ist die Identifizierung der interessierenden Effektstärke, also des zentralen Zusammenhangs, der in den zu Grunde liegenden Primärstudien analysiert wurde. In ökonomischen Kontexten sind das oft Regressionskoeffizienten, geschätzte Elastizitäten oder auch einfache Differenzen in Ausprägungen. In unserem Anwendungsbeispiel der Geschlechterdifferenzen

in der Wettbewerbsneigung betrachteten wir als Effektstärke die Differenz in der Wettbewerbseintrittsrates zwischen Frauen und Männern in ökonomischen Experimenten. Dies ist typisch für Anwendungsbeispiele aus der experimentellen Ökonomik, wo Effektstärken häufig Unterschiede in den Mittelwerten der abhängigen Variable zwischen zwei Gruppen darstellen.

In den anschließenden Arbeitsschritten III) Kodierung und IV) Datenaufbereitung und -bereinigung differenzierten wir zwischen »rohen« Lücken, also den Unterschieden in den beobachteten Anteilen an Frauen und Männern, die freiwillig in den Wettbewerb treten, und »residualen« Lücken – den Unterschieden, die verbleiben, wenn die Studienautor:innen weitere beobachtbare Merkmale der Studienteilnehmer:innen (zum Beispiel ihre Leistungsfähigkeit, ihre Risikopräferenzen oder ihr Selbstvertrauen) mithilfe eines Regressionsmodell berücksichtigen. Darüber hinaus kodierten (also erfassten) wir alle Merkmale von Heterogenität zwischen den Studien. Das betraf die Gruppe der Teilnehmenden (Studierende, Heranwachsende, Kinder, nicht-studierende Erwachsene), die Art des Experiments (Feld-, Labor-, Online-) und weitere konkrete Bedingungen des Wettbewerbs.

Im Folgenden wollen wir uns auf den entscheidenden Arbeitsschritt V), die statistische Analyse, konzentrieren: Der erste Teil einer Meta-Analyse besteht oft in einer Meta-Zusammenfassung, also der Berechnung eines Gesamteffekts über alle Studien hinweg. In der Praxis ist das ein gewichteter Mittelwert der einzelnen Effektstärken, wobei das gewählte Gewicht vom zu Grunde liegenden meta-analytischen Modell bestimmt wird. Das einfachste Modell ist laut Stanley/Doucouliagos (2012) das der sogenannten gemeinsamen Effekte (*common effects*), in welchem angenommen wird, dass alle gesammelten Effektstärken der gleichen Grundgesamtheit mit einem gemeinsamen Mittelwert entstammen und dann mit normalverteiltem Sampling-Fehler geschätzt werden. Im Gegensatz dazu geht das Modell der zufälligen Effekte (*random effects*) davon aus, dass die Primärergebnisse mehreren unterschiedlichen Grundgesamtheiten entstammen. Beispielsweise erscheint es sinnvoll anzunehmen, dass die Preiselastizitäten des Angebots bestimmter Güter von den Merkmalen räumlich oder zeitlich begrenzter Märkte abhängen. In einer Meta-Analyse von Schätzungen solcher Elastizitäten wäre es demnach unrealistisch anzunehmen, dass in Wahrheit eine gemeinsame Elastizität existiert und Abweichungen der Schätzung rein statistischer Natur sind. Vielmehr würde man die einzelnen Märkte als getrennte Grundgesamtheiten betrachten, die jeweils eine eigene zu Grunde liegende Elastizität aufweisen. Im *random effects* Modell weichen die einzelnen

Effektstärken deshalb zufällig vom wahren Effekt ES_0 ab.¹ Der Gesamteffekt wird immer als gewichteter Mittelwert der Primärergebnisse (aus den Studien) ermittelt, wobei Ergebnisse von Studien mit präziser gemessenen Werten mehr Gewicht erhalten.²

In unserer Meta-Zusammenfassung folgten wir dem Modell der zufälligen Effekte. Sie ergab einen Gesamteffekt, d.h. eine gewichtete mittlere Effektstärke ES , von 0,13. Diese Zahl ist so zu interpretieren, dass Männer über alle Studien hinweg 13 Prozentpunkte häufiger eine kompetitive Entlohnung im Experiment wählen als Frauen. Ein geschlechtsspezifischer Unterschied ist somit zwar statistisch eindeutig nachweisbar, aber nur in der Größenordnung eines Drittels der Differenz von 38 Prozentpunkten, welche in der ursprünglichen Arbeit von Niederle/Vesterlund (2007) gemessen wurde.

Interessanterweise berichten auch diejenigen Studien, die am strengsten dem Niederle-Vesterlund-Design folgten, zwar überdurchschnittlich große Effektstärken – also größere Geschlechterunterschiede im Wettbewerbsverhalten als in der Gesamtheit der Studien. Dennoch zeigen auch diese Studien geringere Effektstärken als die Originalstudie. Die Bandbreite reicht in dieser Teilstichprobe von 0 bis 53 Prozentpunkten – bei einer mittleren Effektstärke von 24 Prozentpunkten (siehe Illustration im Forest Plot der Effektstärken in Beblo/Markowsky 2022, Abbildung 1). Über alle Experimentalstudien hinweg lag der unterste Messwert sogar bei -42 Prozentpunkten; hier wählten die teilnehmenden Frauen den Wettbewerb also sogar deutlich häufiger als die teilnehmenden Männer. Wir unternahmen daraufhin einen Test auf Heterogenität der Effektstärken. Dieser bestätigte, dass sich die gemessenen Effektstärken so stark zwischen den Studien unterscheiden, dass eine Meta-Regression zur Identifikation von Moderatoren des Effektes angeraten erschien.

1 Der Gesamteffekt wird mit folgender Gleichung modelliert: $ES_{ij} = ES_0 + \theta_{ij} + \varepsilon_{ij}$, mit $\theta_{ij} \sim N(0, \tau^2)$ und $\varepsilon_{ij} \sim N(0, \sigma^2)$, wobei die Ergebnis-spezifische »wahre« Effektgröße aus zwei Komponenten besteht: ES_0 und dem »Zufallseffekt« θ_{ij} . τ^2 ist ein Maß für die Heterogenität zwischen den Primärergebnissen, über die Varianz der Stichprobenfehler hinaus, unabhängig sowohl von ES_0 als auch ε_{ij} , und ist ebenfalls zu schätzen.

2 Im Fall der common effects entsprechen die Gewichte der inversen Varianz $1/SE_i^2$, während das random effects Modell zusätzlich die geschätzte Studienheterogenität (τ^2) integriert, so dass sich Gewichte von $1/(SE_i^2 + \tau^2)$ ergeben (Stanley/Doucouliagos, 2012, S. 46). Stanley/Doucouliagos (2015) schlagen noch eine dritte Modellvariante vor, welche sie als unrestricted weighted least squares (unrestricted WLS) bezeichnen. Hier werden die Gewichte als $1/\phi SE_i^2$ definiert, wobei die multiplikative Komponente ϕ zur Varianz des Schätzers aus dem common effects Modell ergänzt wird. Die Autoren zeigen, dass die unrestricted WLS-Methode zu identischen Berechnungen des studienübergreifenden Gesamteffekts führt, jedoch zu realistischeren, nämlich breiteren, Konfidenzintervallen, wenn es Hinweise auf Heterogenität zwischen den Einzelergebnissen gibt. Darüber hinaus liefert das Modell zuverlässigere, unverzerrte Ergebnisse bei Vorliegen von Publikations-Bias.

Eine Meta-Regression kann helfen, den Einfluss von heterogenen Daten, Spezifikationen oder Methoden auf die Studienergebnisse zu ergründen. Die Meta-Regression erlaubt auch, mögliche Verzerrungen durch unbeobachtete Variablen und ähnliche methodische Herausforderungen zu analysieren. Im nächsten Schritt regressierten wir deshalb die einzelnen Effektstärken auf alle relevanten Dimensionen von Heterogenität zwischen den Studien.³

Heterogenität kann dabei auf unterschiedlichen Ebenen entstehen. Beispielsweise werden in vielen Studien mehrere Spezifikationen zur Schätzung eines Zusammenhangs berichtet, welche man in der Meta-Analyse gerne alle berücksichtigen möchte. Werden Effektstärken auf der Ebene der Spezifikation verglichen, gibt es Studienmerkmale, die auf alle Spezifikationen innerhalb derselben Studie zutreffen, wie beispielsweise die Autor:innen selbst, sowie andere Merkmale, die auf Ebene der Spezifikation variieren, wie beispielsweise die Einschließung oder Auslassung einer bestimmten Kontrollvariablen oder das verwendete Regressionsmodell.⁴

In unserer Meta-Analyse identifizierten wir als relevante Moderatoren für den gemessenen geschlechterspezifischen Unterschied im Wettbewerbseintritt vor allem die Design-Merkmale des Experiments: Wer nimmt daran teil? Welche Art von Aufgabe ist zu erledigen? Wie genau sind die Wettbewerbsbedingungen? Gegen wen wird gespielt? sowie einige andere Verschiedenheiten. Unsere Meta-Regression bestätigte zunächst die Vermutung aus der Meta-Zusammenfassung, dass die größten geschlechterspezifischen Unterschiede im Wettbewerbsverhalten in denjenigen Studien gefunden werden, die das Niederle-Vesterlund-Design am strengsten einhalten. Die Unterschiede zwischen Frauen und Männern waren also am stärksten ausgeprägt in Laborexperimenten mit Studierenden und bei Rechenaufgaben. Dagegen waren sie bei anderen Altersgruppen (Kindern oder älteren Erwachsenen), bei verbalen Aufgaben und in anderen Umgebungen als dem Labor fast vernachlässigbar gering. Wer was und wo entscheidet, hat also gravierende Auswirkungen auf die gemessene Lücke in der Wettbewerbsneigung.

Die Meta-Regressionsanalyse deckte außerdem auf, unter welchen Wettbewerbsbedingungen das Verhalten von Frauen und Männern kaum voneinander abweicht: Beim Wettbewerb innerhalb einer gleichgeschlechtlichen Gruppe

3 Die Schätzgleichung lautet: $ES_{ij} = \beta_0 + \gamma X_{ij} + \varepsilon_{ij}$, wobei X_{ij} einen Vektor beobachteter Merkmale der Studie darstellt und γ den Vektor der zugehörigen Koeffizienten.

4 Diese Mehrebenenstruktur bringt auch eine weitere Form von stochastischer Abhängigkeit mit sich: Spezifikationen, die aus derselben Studie stammen, können nicht als voneinander unabhängig betrachtet werden. Der methodische Umgang damit entspricht in der Meta-Analyse weitgehend dem in Primäranalysen, zum Beispiel durch die Verwendung von Mehrebenenmodellen oder durch cluster-robuste Schätzung der Varianz.

(statt einer gemischtgeschlechtlichen), beim Wettbewerb gegen sich selbst (indem versucht wird, die eigene Leistung in einer früheren Runde zu übertreffen, statt die einer anderen Person), wenn vergangene Erfolge in den Wettbewerb eingehen (statt noch zu leistende), wenn es mehrere Gewinner:innen gibt (statt nur einer:s einzigen) und bei verschiedenen Interventionen während des Experimentes. Hierzu zählen zum Beispiel Feedback zur relativen Leistung oder positive Diskriminierung, wie die Einführung einer Geschlechterquote unter den Gewinner:innen oder eines Bonussystems. All diese »Design-Elemente« erwiesen sich in unserer Meta-Regression als wirkungsvolle Mittel zur Verringerung oder sogar Beseitigung der Geschlechterlücke im Wettbewerb. Dabei verringert beispielsweise positive Diskriminierung von Frauen die geschlechterspezifische Lücke in der Wettbewerbsneigung um durchschnittlich 26 Prozentpunkte, *ceteris paribus*. Ausgehend von einer mittleren Effektstärke von 13 Prozentpunkten bedeutet das eine vollständige Umkehr der Lücke, so dass Frauen in solchen Umgebungen sogar häufiger den Wettbewerb wählen als Männer. Auch wenn die anderen Design-Elemente nicht annähernd so stark wirken wie die positive Diskriminierung, reduzieren sie den Geschlechterunterschied doch jeweils merklich.

Eine Meta-Regression der residualen Geschlechterlücke in den Studien, d.h. solche, bei denen weitere beobachtbare Merkmale der Studienteilnehmer:innen, insbesondere andere Neigungen, in den Regressionsanalysen der Studien berücksichtigt worden sind, zeigte zudem, dass die Wettbewerbslücke völlig unabhängig von den Leistungsmessungen der teilnehmenden Frauen und Männer ist. Der Unterschied ist also nicht dadurch getrieben, dass Frauen in den auszuführenden Aufgaben schlechtere Leistungen erbringen als Männer, und aus diesem Grund seltener in den Wettbewerb treten. Darüber hinaus scheint das in den Studien gemessene Wettbewerbsverhalten zwar mit den individuellen Risikopräferenzen und dem Selbstvertrauen der Teilnehmenden zusammenzuhängen, insbesondere, wenn diese einfach abgefragt und nicht extra gemessen werden. Trotzdem können diese anderen Neigungsmaße die Geschlechterlücke im Wettbewerbsverhalten nicht vollständig erklären. Obwohl einige Autor:innen argumentieren, dass die individuelle Wettbewerbsneigung bereits durch Risikoneigung und Selbstvertrauen abgebildet sei und es folglich kein eigenständiges *competitiveness trait* gäbe (Gillen u.a. 2019), verbleibt nach unseren Ergebnissen ein eigenes ökonomisch wie statistisch relevantes Maß für die Wettbewerbsneigung – oder zumindest die Geschlechterlücke darin.

Soweit zum Vorgehen bei einer Meta-Analyse und unsere eigene Anwendung auf die Frage geschlechterspezifischer Unterschiede bei der Wettbewerbsneigung: Im Gesamtergebnis sind die Unterschiede klein und stark kontextabhängig.

26.4 Jenseits von Wettbewerb: Meta-Ergebnisse zu Geschlechterdifferenzen bei anderen Neigungsmaßen

Ähnlich wie bei der Wettbewerbsneigung lassen sich auch die Meta-Befunde zu den anderen Neigungsmaßen zusammenfassen, zu denen bereits entsprechende Analysen vorlagen.

Risikoneigung: Die systematischen Sichtungen der experimentellen Ergebnisse zur Risikoneigung legen nahe, dass es keine robusten Geschlechterdifferenzen in der Risikoaversion gibt. Stattdessen betont Nelson (2015) in ihrer Meta-Zusammenfassung die große Ähnlichkeit im Verhalten von Frauen und Männern und die große Varianz der Risikobereitschaft innerhalb jeder Geschlechtergruppe. Die Hälfte der von ihr gesichteten Studien zeigt keinen signifikanten Unterschied zwischen Männern und Frauen. In einer zweiten Meta-Studie analysieren Filippin/Crosetto (2016) über 60 Studien, welche den Holt-Laury-Test zur Messung der Risikobereitschaft verwenden. Der Test stellt die Teilnehmenden vor folgende Wahl zwischen zwei Lotterien: Sie können sich entweder für eine sicherere Lotterioption mit geringerer Auszahlungserwartung oder eine riskantere Variante mit höherer Auszahlungserwartung entscheiden. Durch schrittweise Erhöhung der Gewinnwahrscheinlichkeit bei der riskanteren Lotterie lässt sich der Punkt bestimmen, an dem Menschen von der sichereren zur unsichereren Option wechseln – und damit ihre Risikoneigung messen. Die überwiegende Zahl von Einzelstudien findet demnach zwar, dass Frauen risikoaverser sind als Männer. Allerdings ist der Unterschied nur in knapp 13 Prozent der Studien statistisch signifikant und selbst dann bewerten die Autor:innen ihn als nicht substantiell, also nicht ökonomisch relevant. Interessant ist auch hier wieder die Heterogenität der Effektstärken. Ein Geschlechterunterschied entsteht nämlich vor allem, wenn es statt der Wahl zwischen zwei unsicheren Lotterien eine so genannte *safe option*, eine sichere Auszahlungsoption, gibt. Diese erweist sich insbesondere für weibliche Teilnehmende als relativ attraktiver. Ebenso spielt abermals der Kontext des Experiments eine Rolle dafür, ob ein Geschlechterunterschied in der Risikoneigung sichtbar oder gemessen wird.

Selbstvertrauen: Beim Selbstvertrauen deuten Meta-Analysen zunächst auf leichte Unterschiede nach dem Geschlecht hin. Allerdings finden Bandiera u.a. (2022) keine starke statistische Evidenz dafür, dass Männer generell selbstbewusster wären als Frauen. Stattdessen trauten sich sowohl Frauen als auch Männer eher zu viel zu in Anbetracht ihrer Leistungen (*overconfidence*). Zudem gibt es eine große Variation der Geschlechterdifferenz im Selbstvertrauen über die Studien hinweg und je nach Untersuchungsbereich, ob es zum Beispiel um das eigene Körperbild, um kognitive Fähigkeiten (Gentile u.a. 2009) oder um technologische Fähigkeiten geht (Christensen 2023). Bezüglich letzterer hat sich

der geschlechterspezifische Unterschied über die vergangenen drei Jahrzehnte verkleinert. Am ausgeprägtesten ist er zudem in westlichen Ländern des globalen Nordens, während in Teilen Afrikas das Selbstvertrauen der untersuchten Frauen sogar höher liegt als das der Männer (Christensen 2023).

Altruismus: Eswaran (2014) beschreibt in dem schon zitierten zweiten Kapitel seines Buches eine Reihe von experimentellen Tests (Ultimatum-Spiel, Diktator-Spiel, Gefangenendilemma), mithilfe derer auf altruistisches Verhalten geschlossen werden kann. Die von ihm zitierten Studien zeigen in der Regel ein größeres Ausmaß an Altruismus bei den weiblichen als bei den männlichen Teilnehmenden. Auch hier jedoch zeichnen Meta-Analysen ein differenzierteres Bild. Zwar verhalten sich weibliche Teilnehmende in Diktatorspielen, in denen eine Person allein über die Aufteilung z.B. einer Geldsumme bestimmen kann, im Durchschnitt altruistischer: Sie geben ihren Mitspielern im *giving game* mehr ab (Bilén u.a. 2021) und nehmen ihnen im *taking game* weniger weg (Flage 2024). Allerdings variiert der Geschlechterunterschied erheblich und hängt stark von den Spielregeln ab – insbesondere von den Kosten des altruistischen Verhaltens und der leistungsabhängigen Verteilung des zugewiesenen Geldes. Doñate-Buendía u.a. (2022) zeigen in ihrer Meta-Analyse eine signifikante Interaktion zwischen dem Geschlecht und sozialer Distanz auf: Demnach geben Frauen mehr an Fremde und Männer mehr an Familie/Freunde.

26.5 Fazit

Zusammengefasst liefert die Meta-Betrachtung von Geschlechterunterschieden in ökonomisch relevanten Präferenzmaßen ein differenziertes Bild. Bei der Wettbewerbsneigung sind die Unterschiede zwischen den Geschlechtern deutlich geringer als in den am häufigsten zitierten und oft verallgemeinerten Einzelstudien. Sie sind außerdem stark abhängig vom jeweils getesteten Kontext. Unter bestimmten Bedingungen, beispielsweise wenn es nicht nur eine, sondern mehrere Gewinner:innen des Wettbewerbs geben kann, verschwinden sie fast vollständig. Auch durch gezielte Maßnahmen, beispielsweise das zur Verfügung stellen von Informationen zur relativen Leistung, lassen sich Geschlechterdifferenzen im Wettbewerbseintritt stark verringern. Für die Risikoneigung finden Meta-Analysen ebenfalls keine robusten Geschlechterdifferenzen. Die Unterschiede sind meist nicht statistisch signifikant und wenn doch, dann nicht ökonomisch relevant. Relevant erscheint hingegen die Erkenntnis, dass Frauen häufiger als Männer eine sichere Option wählen, wenn diese zur Verfügung steht, während die Wahl zwischen zwei unsicheren Auszahlungen kaum unterschiedlich verläuft. Beim Selbstvertrauen zeigen sich zwar leichte

Unterschiede zwischen Frauen und Männern, diese variieren aber stark nach Domäne und geografischer Region. Auch Altruismus ist eine Neigung, bei der gemeinhin Geschlechterunterschiede vermutet werden. Von allen Neigungs- und Präferenzmaßen scheinen sie hier auch bei meta-analytischer Betrachtung am ehesten vorzuliegen. Allerdings ist der Zusammenhang mit dem Arbeitserfolg, ob sich letztendlich mehr oder weniger Altruismus ökonomisch auszahlt, nicht abschließend geklärt. Insgesamt verdeutlichen die hier betrachteten Meta-Studien, dass Geschlechterunterschiede in den Neigungen zu Wettbewerb und Risiko sowie Selbstvertrauen weitaus weniger robust sind als häufig postuliert. Stattdessen zeigen sich große Ähnlichkeiten zwischen den Geschlechtern und eine ausgeprägte Abhängigkeit von spezifischen Kontexten und Bedingungen.

Diese Erkenntnis ist besonders relevant für die Ableitung politischer und institutioneller Handlungsempfehlungen: Die starke Kontextabhängigkeit der Geschlechterunterschiede deutet darauf hin, dass durch geeignete Rahmenbedingungen durchaus ein Umfeld geschaffen werden kann, indem Frauen und Männer ähnliche Entscheidungen treffen können und wollen. Da Feldstudien belegen, dass Wettbewerbsneigung, Risikoverhalten und Selbstvertrauen mit realen Karriereentscheidungen korrelieren (zum Beispiel Buser u. a. 2014; Reuben u. a. 2017), können selbst kleine Anfangsunterschiede in diesen Neigungen ökonomische Ergebnisse beeinflussen, die sich im Lebensverlauf zu großen Geschlechterdifferenzen kumulieren. Unter Umständen sind aggregierte Maße für Geschlechterregalität, wie der eingangs erwähnte Gender-Pay-Gap, auch deshalb von so großer Persistenz gekennzeichnet, weil große Geschlechterunterschiede vor allem in gut bezahlten und mit großer Entscheidungsmacht ausgestatteten sogenannten *high-profile jobs* zu finden sind. Dies sind genau die Tätigkeiten, die in einem anspruchsvollen, stark kompetitiven und risikobehafteten Arbeitsumfeld unter Druck ausgeübt werden, so dass hier ein großes Einflusspotenzial für die diskutierten Neigungen vermutet werden kann (Sandberg u. a. 2013; Bertrand 2011). Wenn es durch geeignete Maßnahmen gelingen sollte, die geschlechterspezifischen Unterschiede bei wichtigen biografischen Entscheidungen zu reduzieren, sollten sich auch die Unterschiede in den Bildungs- und Arbeitsergebnissen verringern lassen.

Die zentrale Herausforderung besteht nun darin, die in Experimentalstudien identifizierten förderlichen Kontextbedingungen auf reale Arbeitsmarkt- und Bildungssituationen zu übertragen. Wie lassen sich die im Labor- oder Online-Experiment erprobten Bedingungen, die zu ähnlichem Verhalten von Frauen und Männern führen, in das echte Leben transferieren? Meta-Studien ergänzen und systematisieren nicht nur die hierfür relevanten Grundlagen, sie generieren auch wichtige Hinweise auf Handlungsoptionen für Politik, Unternehmen und andere Organisationen. Damit liefern sie gleichstellungspolitische Ansatzpunkte für

die Gestaltung einer geschlechtergerecht(er)en Arbeitswelt mit Wettbewerbssituationen, die selbstbewusste und risikobehaftete Entscheidungen verlangen.

Literatur

- Anderson, Richard/Kichkha, Areerat (2017), Replication, meta-analysis, and research synthesis in economics, in: *American Economic Review*, Jg. 107, H. 5, S. 56–59.
- Angrist, Joshua/Azoulay, Pierre/Ellison, Glenn/Hill, Ryan/Lu, Susan Feng (2017), Economic research evolves: Fields and styles, in: *American Economic Review*, Jg. 107, H. 5, S. 293–297.
- Bandiera, Oriana/Parekh, Nidhi/Petrongolo, Barbara/Rao, Michelle (2022), Men are from Mars, and Women Too: A Bayesian Meta-analysis of Overconfidence Experiments, in: *Economica*, Jg. 89, Ausg. S1, S. 38–70.
- Bertrand, Marianne (2011), New perspectives on gender, in: Orley Ashenfelter/David Card (Hg.), *Handbook of Labor Economics*, Bd. 4b, Amsterdam, S. 1543–1590.
- Bilén, David/Dreber, Anna/Johannesson, Magnus (2021), Are women more generous than men? A meta-analysis, in: *Journal of the Economic Science Association*, Jg. 7, S. 1–18.
- Buser, Thomas/Niederle, Muriel/Oosterbeek, Hessel (2014), Gender, Competitiveness, and Career Choices, in: *The Quarterly Journal of Economics*, Jg. 129, H. 3, S. 1409–1447.
- Card, David/Kluve, Jochen/Weber, Andrea (2018), What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations, in: *Journal of the European Economic Association*, Jg. 16, H. 3, S. 894–931.
- Christensen, MacKenzie (2023), Tracing the Gender Confidence Gap in Computing: A Cross-National Meta-Analysis of Gender Differences in Self-Assessed Technological Ability, in: *Social Science Research*, Jg. 111.
- Doñate-Buendía, Anabel/García-Gallego, Aurora/Petrović, Marko (2022), Gender and other moderators of giving in the dictator game: A meta-analysis, in: *Journal of Economic Behavior and Organization*, Jg. 198, S. 280–301.
- Eswaran, Mukesh (2014), *Why gender matters in economics*, Princeton.
- Falk, Armin/Hermle, Johannes (2018), Relationship of gender differences in preferences to economic development and gender equality, in: *Science*, Jg. 362, H. 6412.
- Filippin, Antonio/Crosetto, Paolo (2016), A Reconsideration of Gender Differences in Risk Attitudes, in: *Management Science*, Jg. 62, H. 11, S. 3138–3160.
- Flage, Alexandre (2024), Taking games: a meta-analysis, in: *Journal of the Economic Science Association*, Jg. 10, S. 255–278.
- Gentile, Brittany/Grabe, Shelly/Dolan-Pascoe, Brenda/Twenge, Jean/Wells, Brooke/Maitino, Alissa (2009), Gender Differences in Domain-Specific Self-Esteem: A Meta-Analysis, in: *Review of General Psychology*, Jg. 13, H. 1, S. 35–45.
- Gillen, Ben/Snowberg, Erik/Yariv, Leeat (2019), Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study, in: *Journal of Political Economy*, Jg. 127, H. 4, S. 1826–1863.
- Kosse, Fabian/Tincani, Michaela (2020), Prosociality predicts labor market success around the world, in: *Nature Communications*, Jg. 11, Art. 5298.

- Markowsky, Eva/Beblo, Miriam (2022), When do we observe a gender gap in competition entry? A meta-analysis of the experimental literature, in: *Journal of Economic Behavior & Organization*, Jg. 198, S. 139–163.
- Nelson, Julie (2015), Are Women Really More Risk-Averse than Men? A Re-Analysis of the Literature Using Expanded Methods, in: *Journal of Economic Surveys*, Jg. 29, H. 3, S. 566–585.
- Niederle, Muriel (2017), Gender, in: Kagel, John H./Roth, Alvin E. (Hg.), *The Handbook of Experimental Economics*, Bd. 2, Princeton, S. 481–562.
- Niederle, Muriel/Vesterlund, Lise (2007), Do women shy away from competition? Do men compete too much?, in: *Quarterly Journal of Economics*, Jg. 122, H. 3, S. 1067–1101.
- Reuben, Ernesto/Wiswall, Matthew/Zafar, Basit (2017), Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender, in: *The Economic Journal*, Jg. 127, H. 604, S. 2153–2186.
- Sandberg, Sheryl (2013), *Lean in. Women, Work, and the Will to Lead*, New York.
- Stanley, Tom/Doucouliagos, Hristos (2012), *Meta-regression analysis in economics and business*, New York.
- Shurchkov, Olga/Eckel, Catherine (2018), Gender Differences in Behavioral Traits and Labor Market Outcomes, in: Averett, Susan/Argys, Laura/Hoffmann, Saul (Hg.), *The Oxford Handbook of Women and the Economy*, Oxford, S. 481–512.
- Weichselbaumer, Doris/Winter-Ebmer, Rudolf (2007), The effects of competition and equal treatment laws on gender wage differentials, in: *Economic Policy*, Jg. 22, H. 50, S. 236–287.