

De waarschijnlijkheid van observaties

Citation for published version (APA):

Lakens, D. (2014). De waarschijnlijkheid van observaties. *Algemeen Nederlands Tijdschrift voor Wijsbegeerte*, 106, 49-53. <https://doi.org/10.5117/ANTW2014.1.LAKE>

DOI:

[10.5117/ANTW2014.1.LAKE](https://doi.org/10.5117/ANTW2014.1.LAKE)

Document status and date:

Gepubliceerd: 01/01/2014

Document Version:

Uitgevers PDF, ook bekend als Version of Record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

De waarschijnlijkheid van observaties

Daniël Lakens

ANTW 106 (1): 49–53

DOI: 10.5117/ANTW2014.1.LAKE

Radder gaat met zijn discussie over reproduceerbaarheid in op een belangrijk en actueel thema binnen empirische wetenschappen in het algemeen, en de psychologie in het bijzonder. Hoewel zijn analyse van de verschillende types van reproduceerbaarheid nuttig is, is de conclusie die hij trekt over wat wetenschappers moeten doen niet alleen ongefundeerd, maar ook schadelijk voor een cumulatieve wetenschap. In deze reactie licht ik kort toe waarom empirische wetenschappers de aanbeveling van Radder in het verleden al hebben genegeerd, en waarom ze dat in de toekomst ook moeten blijven doen.

Toen het onderzoek uit CERN naar de Higgs boson in juli 2012 aan de wereld gepresenteerd werd, werden er conclusies getrokken op basis van de gegevens van twee onderzoeksteams, die beide hetzelfde gevonden hadden. Radder beargumenteert dat het niet nuttig is om zulke reproducties uit te voeren, en geeft de voorkeur aan (conceptuele) replicaties. Het lijkt hem beter “primair te streven naar replicaties en alleen reproducties van het hele proces (door anderen) uit te voeren als daar aanwijsbare theoretische of praktische redenen voor zijn.” Waarom wordt het advies van Radder in een van de belangrijkste empirische ontdekkingen van de afgelopen decennia volledig genegeerd, ondanks de enorm hoge kosten die de reproductie van een experiment met zich meebrengt in deze context? De reden is dat empirische wetenschap meer is dan begrijpen. Minstens zo belangrijk is de taak van empirisch onderzoek om uitspraken te doen over de waarschijnlijkheid van observaties, om zo te bepalen welke effecten het begrijpen waard zijn, en welke beter aan toeval kunnen worden toegeschreven. Deze kant van het verhaal wordt door Radder volledig genegeerd, maar een bespreking van het nut van reproducties en replicaties zonder het meenemen van de waarschijnlijkheid dat empirische observaties iets zeggen over de werkelijkheid kan nooit het hele verhaal vertellen.

Radder schrijft over waarnemen, maar bedoelt daarmee veelal ‘statistische inferentie’. Het bestaan van de Higgs boson werd geïnferreed op basis van de waarschijnlijkheid waarmee een bepaald patroon in de data werd gevonden. De grenswaarde van de onwaarschijnlijkheid waarmee deze patronen toegeschreven kunnen worden aan het toeval is in de natuurkunde 5 sigma, in de psychologie $p < .05$, maar is in alle disciplines gebaseerd op vergelijkbare statistiek. Metingen hebben variantie, en hoe meer data je hebt, hoe kleiner de variantie, en hoe nauwkeuriger de statistische inferenties die je maakt zijn. Radder praat over studies alsof de waarnemingen alleen betwist kunnen worden op basis van hun conceptuele interpretatie. Als dat zo was, en elke uitkomst van een studie in dezelfde mate een vermindering in de waarschijnlijkheid dat een theoretisch idee waar is zou betekenen, dan was de empirische wetenschap tot mislukking gedoemd. Er zou geen sprake zijn van progressieve en degeneratieve onderzoekslijnen (Lakatos: 1970), en alle empirische vraagstukken zouden verzanden in filosofische discussies.

Gelukkig is dat niet zo. In plaats van eindeloze argumenten over de conceptuele interpretaties van onderzoeksuitkomsten, waarbij elke onderzoeksuitkomst dezelfde waarde krijgt toegeschreven, staat een combinatie van statistische inferentie en theoretische voorspellingen ons toe om te differentiëren in de mate waarin empirische resultaten ons inzicht geven in de waarschijnlijkheid dat uitspraken waar zijn. De empirische cyclus is hiermee onder meer in staat geweest om twijfel zoals deze door Radder geuit wordt over of “experimenten in de psychologie betrouwbare en zinvolle kennis kunnen opleveren” weg te nemen. Waar deze discussie 30 jaar geleden nog erg populair was (Gergen: 1985), heeft de empirische werkelijkheid theoretische vingeroefeningen zoals het in twijfel trekken van de psychologie als empirische wetenschap al lang ingehaald (al is het psychologen te verwijten dat ze zich te weinig aantrekken van deze aanhoudende kritiek binnen andere disciplines om de moeite te nemen deze te weerleggen).

Laat ik een voorbeeld geven. Het sociaalpsychologische ‘voet-in-de-deur’ effect beschrijft hoe mensen meer geneigd zijn om met een groot verzoek in te stemmen na een klein verzoek, dan wanneer het grote verzoek direct gedaan wordt. Externe validiteit is hier geen enkel probleem – al in de originele studie werden gewone mensen thuis opgebeld door een onbekend onderzoeksbureau met de vraag of ze kort wilden aangeven welke schoonmaakproducten ze gebruikten (het kleine verzoek) waarna het grote verzoek volgde: Of een team van het onderzoeksbureau enkele uren ongestoord en in volledige vrijheid door het huis van de persoon die opgebeld werd mocht lopen om te kijken hoe en waarmee het huis schoon

werd gehouden. Als mensen alleen het grote verzoek kregen, ging minder dan 25% van de mensen hiermee akkoord. Na een klein verzoek, was dit meer dan 50%.

Als allebei deze percentages de lezer wat hoog lijken, dan is het misschien nuttig om te vermelden dat dit onderzoek in 1966 gedaan werd (Freedman & Fraser: 1966). We leven niet langer in een tijd waarin we vreemden na telefonisch contact de toestemming geven enkele uren in ons huis rond te lopen. Het onderzoek kan dus nooit meer gereproduceerd worden. Dat is geen enkel probleem voor de psychologie, omdat we goed kunnen voorspellen wanneer het voet-in-de-deur effect werkt (Burger: 2009). Zo'n goed begrip hebben we niet van alle psychologische effecten. Radder merkt terecht op dat *social priming* een gebied is waar de waarschijnlijkheid dat effecten optreden nog veel minder zeker is, laat staan dat we de effecten goed genoeg begrijpen zodat we ze kunnen voorspellen.

Juist onder deze omstandigheden, wanneer de waarschijnlijkheid van een effect nog relatief onzeker is, is de reproduceerbaarheid van onderzoek van essentieel belang voor een empirische wetenschap. Het advies van Radder om te streven naar conceptuele replicaties is precies datgene wat in *social priming* onderzoek de afgelopen 20 jaar vooral gedaan is, en wat heeft geleid tot een degeneratieve onderzoekslijn. Als je niet kunt voorspellen wanneer een effect optreedt, is het belangrijk dat je een effect in ieder geval kunt reproduceren. Dit is ook de reden waarom Koole en Lakens (2012) stellen dat onderzoek reproduceren een “sure and simple way to improve psychological science” is. Juist als we onderzoekers de gelegenheid geven om te vluchten in ad-hoc reddenaties over mogelijke verklaringen voor het uitblijven van een effect, wat bij conceptuele replicaties altijd mogelijk is, zijn we verzekerd van een degeneratieve onderzoekslijn (LeBel & Peters: 2011).

Radder bekritiseert deze aanpak omdat de zwakke theoretische basis in sommige subdomeinen van de psychologie (zoals bij *social priming* onderzoek) te inadequaat is om de materiele realisering van het experiment volledig te controleren. In een voetnoot verwijst hij naar de mogelijkheid dat triviale details zoals de dag van de week en de kleur van de wanden effecten kunnen beïnvloeden. Dit is vanzelfsprekend een contradictio in terminis. Als deze details werkelijk triviaal zijn, dan beïnvloedden ze het effect niet, en als ze het effect beïnvloeden, zijn ze niet triviaal.

Omdat Radder geen aandacht besteedt aan de statistische waarschijnlijkheid waarmee experimenten ons iets leren over de werkelijkheid, is hij niet in staat om de vraag te stellen die er werkelijk toe doet, voordat hij zijn conclusies trekt over het belang van reproducties en replicaties.

De vraag die een empirische wetenschap zich in de eerste plaats stelt is: Hoe waarschijnlijk is het dat empirische observaties toeval waren, en hoe waarschijnlijk is het dat deze observaties gemodereerd worden door onbekende factoren zoals de dag van de week, de kleur op de muren, etc.? Hoe moeilijker het is om een effect te reproduceren, des te waarschijnlijker het is dat een effect toeval was. Zoals met alle probabilistische oordelen geeft een reeks van gefaalde reproducties geen zekerheid, maar het maakt het wel steeds waarschijnlijker dat een eerder gerapporteerd effect toeval is. Op een gegeven moment mogen we aannemen dat de bewijslast komt te liggen bij de onderzoeker die zegt dat het effect bestaat, in plaats van bij de persoon die niet in staat was om het te reproduceren.

De conceptuele replicaties die Radder voorstelt werken uitstekend als het initiële effect met een redelijke mate van betrouwbaarheid is waargenomen. Maar de neiging om onderzoeksresultaten die misschien puur toeval zijn te interpreteren is juist de valkuil waar de sociale psychologie na 20 jaar uit probeert te klimmen. Deze inzichten moeten meegenomen worden in een wetenschapsfilosofie die in enige mate relevant probeert te zijn voor andere vakgebieden. Er zijn zonder twijfel grote problemen binnen enkel de psychologie. Maar de analyse van die problemen, en mogelijke oplossingen van die problemen, wordt het beste gedaan door psychologen zelf (bijv., Meehl: 1990; Nosek, Spies, & Motyl: 2012). Het zou mooi zijn als wetenschapsfilosofen hieraan bij konden dragen, maar dan is een beter begrip van het belang van statistische inferenties in een empirische wetenschap wel essentieel.

Literatuur

- Burger, J.M. (1999) The Foot-in-The-Door Compliance Procedure: A Multiple-Process Analysis and Review, *Personality and Social Psychology Review* 3, pp. 303-325.
- Freedman, J.L., & Fraser, S.C. (1966) Compliance without Pressure: The Foot-in-The-Door Technique, *Journal of Personality and Social Psychology* 4, pp. 195-202.
- Gergen, K.J. (1985) The Social Constructionist Movement in Modern Psychology, *American psychologist* 40, pp. 266-275.
- Koole, S.L., & Lakens, D. (2012) Rewarding Replications: A Sure and Simple Way to Improve Psychological Science, *Perspectives on Psychological Science* 7, pp. 608-614.
- Lakatos, I. (1970) Falsification and the methodology of scientific research programmes, in: I. Lakatos en Musgrave (red.) *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press, pp. 91-196.
- LeBel, E.P., & Peters, K.R. (2011) Fearing The future of Empirical Psychology: Bem's (2011) Evidence of Psi as A Case Study of Deficiencies in Modal Research Practice, *Review of General Psychology* 15, pp. 371-379.

- Meehl, P.E. (1990) Appraising and Amending Theories: The Strategy of Lakatosian Defense and Two Principles That Warrant It, *Psychological Inquiry* 1, pp. 108-141.
- Nosek, B.A., Spies, J.R., & Motyl, M. (2012) Scientific Utopia: II. Restructuring Incentives and Practices to Promote truth over Publishability, *Perspectives on Psychological Science* 7, pp. 615-631.

Over de auteur

Daniël Lakens is universitair docent Mens-Techniek Interactie aan de Technische Universiteit Eindhoven. Hij is experimenteel psycholoog. Een aantal publicaties: Lakens, D. & Evers, E. (2014), 'Sailing from The Seas of Chaos into The Corridor of Stability: Practical Recommendations to Increase The Informational Value of Studies.' *Perspectives on Psychological Science*. Nosek, B.A., & Lakens, D. (2014), Registered Reports: A method to Increase the Credibility of Published Results. *Social Psychology*. Koole, S.L., & Lakens, D. (2012), Rewarding Replications: A Sure and Simple Way to Improve Psychological Science. *Perspectives on Psychological Science*.