RESEARCH · RESEARCH DESIGN

Double-Blind Methodologies and the Elimination of Bias in Knowledge Production

Why knowing who received which treatment can bias research, the history of blinding from Clever Hans to the placebo-controlled trial, what single, double and triple blinding mean, evidence that unblinded studies exaggerate effects, and how blinding is applied in peer review, hiring, data analysis and commercial research.

ModQu research article · 2,128 words · 21 references (APA 7th edition)

Researchers, participants and assessors are human. When they know which treatment a person received, or which result would support a favoured hypothesis, that knowledge can shape what they do, what they see and what they record, usually without any intention to deceive. **Blinding**, also called masking, is the practice of keeping such knowledge from the people whose behaviour or judgement could be affected by it. A **double-blind** study is one in which neither the participants nor the researchers who interact with them know who is in which group. Blinding is most closely associated with clinical trials, but the principle applies wherever human judgement enters the production of evidence: in peer review, hiring, forensic science, data analysis and market research. This article explains why blinding matters, how it developed, what the different levels of blinding mean, what evidence shows about its effects and how it can be applied beyond the laboratory.

The problem blinding solves

Bias from knowledge of treatment can enter research at several points.

Participants who know they are receiving a new treatment may expect to improve and report feeling better, or may behave differently, for example by following advice more carefully. Those who know they are in a control group may be disappointed, drop out or seek treatment elsewhere. Beecher (1955) drew attention to the size of **placebo effects**, reporting that a substantial proportion of patients improved after receiving inactive treatments; later analyses argued that some of this apparent improvement reflected natural recovery and other factors, but the broader point that expectations can affect reported outcomes is well established. Orne (1962) showed that participants in experiments often try to work out what the researcher expects and act accordingly.

Those delivering a treatment may, if they know who is receiving it, give those participants more attention, encouragement or care, so that the comparison is no longer between treatments alone.

Outcome assessors may see what they expect to see, especially when measurement involves judgement. Rosenthal and Fode (1963) told student experimenters that some laboratory rats had been bred to be good at learning mazes and others to be poor at it. In fact, the rats had been assigned at random, yet the rats labelled as bright performed better, apparently

because of subtle differences in how the students handled and recorded them. Rosenthal and Jacobson (1968) reported a related effect in classrooms, where teachers' expectations about certain pupils appeared to influence those pupils' progress, although the size and replicability of the classroom effect have been debated.

Data analysts who know which group is which, and what result is hoped for, may make analytic choices, such as which observations to exclude or which model to use, that favour the expected result. Nickerson (1998) reviewed extensive evidence of **confirmation bias**, the tendency to seek, interpret and remember information in ways that support existing beliefs.

A short history

An early demonstration of why blinding is needed came from a horse. In the early 1900s, a horse in Berlin known as Clever Hans appeared to answer arithmetic questions by tapping his hoof. Pfungst (1911) showed that Hans could answer correctly only when the questioner knew the answer and was visible to the horse. Hans was responding to tiny, unintentional changes in the questioner's posture and expression as he approached the correct number of taps. When the questioner did not know the answer, the horse's ability disappeared.

Kaptchuk (1998) traces the use of blind assessment further back, to investigations in the eighteenth century, including the French royal commission of 1784 that tested Franz Mesmer's claims for "animal magnetism" by blindfolding subjects, and to the use of blind testing in nineteenth-century debates over homeopathy. The modern placebo-controlled, double-blind clinical trial became established in the mid-twentieth century and is now the standard for testing many medical treatments.

Levels of blinding

The terms "single-blind", "double-blind" and "triple-blind" are widely used but inconsistently defined. Devereaux et al. (2001) asked physicians how they interpreted these terms and examined medical textbooks, finding many different definitions for each. For this reason, reporting guidelines such as CONSORT ask researchers to state exactly who was blinded, rather than relying on labels (Schulz et al., 2010).

Who can be blinded, and why

Each group's knowledge can bias a different part of the research

Who is blinded	Bias prevented	Example
Participants	Expectations change behaviour or reported outcomes	Identical placebo tablets
Those delivering treatment	Extra attention or care for one group	Staff given coded supplies
Outcome assessors	Seeing what they expect to see	Assessor not told the patient's group
Data analysts	Analytic choices that favour a hoped-for result	Group labels scrambled until plan is fixed
Reviewers and evaluators	Judging the person, not the work	Names removed from papers or applications

Based on Schulz & Grimes (2002); Schulz et al. (2010) ModQu.com

Figure 1. Who can be blinded, and what blinding each group prevents. Reports should state exactly who was blinded rather than relying on labels such as double-blind. Based on Schulz and Grimes (2002) and Schulz et al. (2010); table by ModQu.

As commonly used:

- **Single-blind** usually means that participants do not know which treatment they receive.
- **Double-blind** usually means that neither participants nor the researchers who deliver treatment or assess outcomes know.
- **Triple-blind** usually adds the people who analyse the data, or a committee monitoring the trial.



Figure 2. Levels of blinding as the terms are commonly used. Because definitions vary, reporting guidelines ask authors to name each group that was blinded. Based on Devereaux et al. (2001) and Schulz et al. (2010); diagram by ModQu.

Schulz and Grimes (2002) emphasise that blinding is distinct from **allocation concealment**, which prevents those enrolling participants from knowing or predicting the next assignment before it is made. Allocation concealment protects the randomisation; blinding protects what happens after it. Both are needed.

Evidence that blinding matters

Studies that compare trials with and without blinding provide direct evidence of its effects. Wood et al. (2008) analysed data from meta-analyses of clinical trials and found that trials without adequate blinding reported larger treatment effects than blinded trials, but only when outcomes were subjective, such as patient-reported symptoms or clinician assessments. For objective outcomes, such as death, the difference was negligible. This pattern is what would be expected if the exaggeration comes from judgement and expectation.

Hróbjartsson et al. (2012) examined trials in which the same outcome had been assessed by both blinded and unblinded assessors. Unblinded assessors produced substantially more favourable estimates of treatment effects, exaggerating odds ratios by about a third on average. Holman et al. (2015) searched the life sciences literature for experiments recorded blind and not blind and found that non-blind studies tended to report larger effects and a higher proportion of significant results.

Bias from knowledge of treatment is therefore not a hypothetical concern. It is large enough to change conclusions, and it is largest exactly where judgement is involved.

How blinding is achieved

In drug trials, blinding usually relies on identical-looking placebos, prepared and coded by a pharmacy or third party, so that neither patients nor staff can tell them apart. Boutron et al. (2006) reviewed reports of trials of drug treatments and catalogued the methods used, such as matching the appearance, taste and smell of active and placebo treatments, and using a "double dummy" design when two treatments differ in form.

Many interventions cannot be blinded in this way. Participants know whether they had surgery or physiotherapy, attended a training course or not, or saw a new website design. In such cases, researchers blind whoever they can: outcome assessors who do not know the participant's group, analysts who work with coded group labels, or committees that judge outcomes from records with identifying information removed.

Blinding can also fail. Participants may guess their treatment from side effects, and staff may guess from patterns of response. Bang et al. (2004) proposed an index for assessing whether blinding was successful by asking participants and staff to guess the assignment, though interpreting such guesses is not straightforward, because a treatment that works may itself reveal which group a person is in.

Beyond clinical trials

The principle of blinding applies wherever judgement could be influenced by knowledge that should not matter.

Peer review

In **single-blind** peer review, reviewers know the authors' identities but authors do not know the reviewers'. In **double-blind** review, both are hidden. Snodgrass (2006) reviewed the literature on the two approaches in computer science and related fields. Tomkins et al. (2017) conducted an experiment at a major computer science conference in which submissions were randomly assigned to reviewers who either could or could not see author information. Reviewers who could see author information were significantly more likely to recommend acceptance of papers from famous authors and highly ranked institutions. Okike et al. (2016) sent the same manuscript, with and without a prominent author's name, to reviewers for an orthopaedics journal; the version with the prominent author was much more likely to be recommended for acceptance.

Hiring and evaluation

Goldin and Rouse (2000) studied the adoption of "blind" auditions by major orchestras in the United States, in which candidates played behind a screen so that the jury could not see them. Using audition records, they estimated that the screen increased the likelihood that women would advance in auditions, and that the switch to blind auditions could explain part of the increase in the proportion of women in orchestras. Later commentators have questioned how strong the statistical evidence in that study is, but blind evaluation has nonetheless been widely adopted in hiring and grading, for example by removing names from applications or examination scripts.

Data analysis

Blind analysis hides the results of interest from analysts until all analytic decisions have been made. It has been used in particle physics for decades. Klein and Roodman (2005) describe techniques such as adding a hidden offset to the measured value, or hiding the data in the region where a signal is expected, so that analysts cannot tune their methods to produce a desired result. MacCoun and Perlmutter (2015) argued that blind analysis should be adopted more widely in the social and biological sciences, for example by having a colleague scramble group labels or add noise to the outcome until the analysis plan is final.

Blinding in commercial research

Many commercial research methods already use forms of blinding, and others could benefit from it.

- **Blind product tests.** Taste tests and product trials often remove branding so that respondents judge the product rather than the brand. Comparing blind and branded tests shows how much the brand itself contributes.
- **Concealing the sponsor.** Revealing who commissioned a survey can bias responses, as respondents may say what they think the sponsor wants to hear. Many surveys therefore describe the sponsor only in general terms until the end, and then disclose it, consistent with ethical guidelines.
- **Blind coding.** When open-ended responses or interviews are coded into categories, coders who do not know which group a respondent belongs to, such as test or control, or which hypothesis is favoured, are less likely to code in a biased way.
- **Blind analysis of tests.** In A/B tests and other experiments, analysts can work with coded labels until the analysis plan is fixed, reducing the temptation to choose the metric or segment that makes a favoured version look best.

A practical checklist

When planning any study in which judgement or expectation could affect the result, it is worth asking:

- Who could know which group each participant is in, and could that knowledge affect what they do or record?
- Can participants be kept unaware of their assignment, for example with an identical-looking alternative?
- Can outcomes be assessed by someone who does not know the assignment, or measured objectively?
- Can the analysis plan be fixed before the analyst sees results by group?
- If blinding is not possible for some groups, how will the resulting risk of bias be acknowledged in the report?
- Will the report state exactly who was blinded and how, rather than using a label alone?

Limits and costs

Blinding is not always possible, and it is not free. Some treatments cannot be disguised, and attempts to blind them may be unconvincing or unethical. Placebo procedures in surgery, for example, raise serious ethical questions. Blinding can make research more complex and expensive, and in some settings, such as action research or participatory studies, openness with participants is part of the method.

Where full blinding is impossible, partial blinding is still valuable. Blinding outcome assessors and analysts is often feasible even when participants and practitioners know the treatment, and the evidence suggests that blinding assessors is where much of the benefit lies when outcomes involve judgement (Hróbjartsson et al., 2012).

Conclusion

Blinding protects research from a simple fact: people see and do what they expect. From Clever Hans to modern clinical trials, the evidence shows that knowledge of which treatment someone received, or which result is hoped for, can shape behaviour, perception and analysis enough to change conclusions. Hiding that knowledge from participants, practitioners, assessors and analysts, wherever possible, removes one of the most persistent sources of bias in knowledge production. The principle extends well beyond medicine, to peer review, hiring, data analysis and commercial research, and it is often one of the cheapest ways to make evidence more trustworthy.

References

Bang, H., Ni, L., & Davis, C. E. (2004). Assessment of blinding in clinical trials. *Controlled Clinical Trials*, *25*(2), 143–156. https://doi.org/10.1016/j.cct.2003.10.016

Beecher, H. K. (1955). The powerful placebo. *JAMA*, *159*(17), 1602–1606. https://doi.org/10.1001/jama.1955.02960340022006

Boutron, I., Estellat, C., Guittet, L., Dechartres, A., Sackett, D. L., Hróbjartsson, A., & Ravaud, P. (2006). Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: A systematic review. *PLoS Medicine*, *3*(10), Article e425. https://doi.org/10.1371/journal.pmed.0030425

Devereaux, P. J., Manns, B. J., Ghali, W. A., Quan, H., Lacchetti, C., Montori, V. M., Bhandari, M., & Guyatt, G. H. (2001). Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA*, *285*(15), 2000–2003. https://doi.org/10.1001/jama.285.15.2000

Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review, 90*(4), 715–741. https://doi.org/10.1257/aer.90.4.715

Holman, L., Head, M. L., Lanfear, R., & Jennions, M. D. (2015). Evidence of experimental bias in the life sciences: Why we need blind data recording. *PLOS Biology*, *13*(7), Article e1002190. https://doi.org/10.1371/journal.pbio.1002190

Hróbjartsson, A., Thomsen, A. S. S., Emanuelsson, F., Tendal, B., Hilden, J., Boutron, I., Ravaud, P., & Brorson, S. (2012). Observer bias in randomised clinical trials with binary outcomes: Systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ*, *344*, Article e1119. https://doi.org/10.1136/bmj.e1119

Kaptchuk, T. J. (1998). Intentional ignorance: A history of blind assessment and placebo controls in medicine. *Bulletin of the History of Medicine, 72*(3), 389–433. https://doi.org/10.1353/bhm.1998.0159

Klein, J. R., & Roodman, A. (2005). Blind analysis in nuclear and particle physics. *Annual Review of Nuclear and Particle Science*, *55*, 141–163. https://doi.org/10.1146/annurev.nucl.55.090704.151521

MacCoun, R., & Perlmutter, S. (2015). Blind analysis: Hide results to seek the truth. *Nature*, *526*(7572), 187–189. https://doi.org/10.1038/526187a

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Okike, K., Hug, K. T., Kocher, M. S., & Leopold, S. S. (2016). Single-blind vs double-blind peer review in the setting of author prestige. *JAMA*, *316*(12), 1315–1316. https://doi.org/10.1001/jama.2016.11014

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776–783. https://doi.org/10.1037/h0043424

Pfungst, O. (1911). *Clever Hans (the horse of Mr. von Osten): A contribution to experimental animal and human psychology* (C. L. Rahn, Trans.). Henry Holt.

Rosenthal, R., & Fode, K. L. (1963). The effect of experimenter bias on the performance of the albino rat. *Behavioral Science*, *8*(3), 183–189. https://doi.org/10.1002/bs.3830080302

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. Holt, Rinehart and Winston.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*, *340*, Article c332. https://doi.org/10.1136/bmj.c332

Schulz, K. F., & Grimes, D. A. (2002). Blinding in randomised trials: Hiding who got what. *The Lancet*, *359*(9307), 696–700. https://doi.org/10.1016/S0140-6736(02)07816-9

Snodgrass, R. (2006). Single- versus double-blind reviewing: An analysis of the literature. *ACM SIGMOD Record*, *35*(3), 8–21. https://doi.org/10.1145/1168092.1168094

Tomkins, A., Zhang, M., & Heavlin, W. D. (2017). Reviewer bias in single- versus double-blind peer review. *Proceedings of the National Academy of Sciences*, *114*(48), 12708–12713. https://doi.org/10.1073/pnas.1707323114

Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Jüni, P., Altman, D. G., Gluud, C., Martin, R. M., Wood, A. J. G., & Sterne, J. A. C. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: Meta-epidemiological study. *BMJ*, *336*(7644), 601–605. https://doi.org/10.1136/bmj.39465.451748.AD