

KNO-020 · Review, Consensus and Disagreement

KNO · Ways of Knowing

Juan Segurola

v0.6 · EN

All Rights Reserved

Article Abstract

Science uses several social procedures to expose claims to criticism, but those procedures do not all do the same job. A manuscript may be reviewed before publication, a published result may face replication or correction, and a community may later describe a position as a consensus while substantial questions remain open. Disagreement can arise from evidence, methods, models, assumptions, uncertainty or values, and those sources need not have the same remedy. This article asks how review, consensus and disagreement should be interpreted as evidence about scientific evaluation. It separates first-order evidence from judgments about that evidence, tracks what publication changes and what it does not, and asks what makes agreement or disagreement informative rather than merely visible across scientific communities and changing evidence.

In 60 Seconds

Peer review, publication, consensus and disagreement are stages or signals within scientific evaluation, not substitutes for the underlying evidence. Reviewers can expose weaknesses before publication, but different reviewers may reasonably emphasize different risks. Publication widens scrutiny rather than ending it. Consensus becomes more informative when it records convergence across evidence, methods and critical examination, while remaining open to revision. A minority view is not disqualified by being a minority, but neither does minority status grant equal evidential support. Productive disagreement identifies where interpretations diverge and what observation, analysis or replication could discriminate between them. Corrections and retractions can reveal weaknesses, but they can also show that the scholarly record remains open to repair. The practical task for a reader is therefore to ask what was reviewed, what evidence supports the claim, where disagreement lies, how independent the supporting routes are, and what could change the evidential balance in scientific practice.

Why This Question Matters

Public discussion often compresses scientific evaluation into slogans: “peer reviewed” is treated as a seal of truth, “scientific consensus” as a vote, or “scientists disagree” as proof that knowledge has collapsed. Those shortcuts obscure the distinction between evidence and the social procedures used to scrutinize it. The distinction matters when papers conflict, expert committees issue assessments, minority interpretations challenge prevailing views or a correction changes a published record. Knowing what review and consensus can and cannot establish helps readers judge controversy without either worshipping authority or mistaking disagreement for epistemic paralysis.

1. Peer review is a filter, not a truth machine

A journal review asks qualified readers to examine a manuscript before an editorial decision. The process can identify unclear reasoning, missing controls, unsupported claims, methodological problems or relevant literature that authors have overlooked. COPE describes reviewers as playing an important role in the integrity of the scholarly record and requires reviews to be objective, constructive and appropriately matched to expertise. [1]

That role is substantial but limited. A reviewer usually sees a manuscript, its reported methods and the evidence supplied by the authors. Reviewers do not rerun every experiment, independently regenerate every dataset or guarantee that unreported assumptions are harmless. Empirical research on peer review has also documented variability in reviewer judgments and concerns about reliability, bias and predictive validity. [2][3]

Claim C01 - Declared synthesis. Peer review is best understood as a structured opportunity for expert criticism before or around publication, not as a procedure that establishes truth. The first part is directly supported by [1][2]; the “not truth” contrast is an interpretive synthesis from the documented scope and limitations of the process.

Where That Leaves Us

Peer review can expose weaknesses before publication, but its reach is bounded by what reviewers can inspect and by the information authors provide. The epistemic status is therefore provisional: review adds structured criticism without certifying truth. Once that limit is clear, the next question is what to make of disagreement when competent reviewers reach different judgments.

2. Disagreement among reviewers is not a malfunction by itself

Two competent reviewers can read the same paper differently. They may specialize in different methods, notice different failure modes or judge the importance of limitations differently. That disagreement can be useful because it exposes places where a manuscript depends on assumptions that were not obvious to all readers.

Editorial practice does not reduce this to vote counting. Nature Genetics explicitly tells reviewers that other reviewers may have different expertise or views, and states that editorial decisions are not made by counting votes or numerical ranks. Editors weigh the arguments raised by reviewers and authors. [4]

Claim C02. Reviewer disagreement can reveal differences in expertise, assumptions and evidential weighting rather than simply indicating that one reviewer is incompetent. [2][4]

Where That Leaves Us

Reviewer disagreement can be informative because different expertise and assumptions expose different vulnerabilities, and editorial decisions need not reduce those judgments to a vote. The evidence supports disagreement as a diagnostic signal, not as proof that a manuscript is sound or unsound. The next question is what changes when a claim passes from review into publication.

3. Publication is a stage in scrutiny, not its endpoint

Once a paper is published, scrutiny broadens. Other researchers may try to reproduce calculations, repeat experiments, test different populations, challenge measurements or compare alternative explanations. The National Academies distinguishes reproducibility using the same data and computational steps from replicability across studies with new data, and emphasizes that failed replication can sometimes expose limitations or open new scientific questions. [5]

Publication therefore changes a claim's social status without freezing its epistemic status. A paper can later be corrected, qualified, contradicted or retracted. Ioannidis' influential 2005 essay formalized how bias, low power, flexibility and prior odds can make published claims unreliable in some settings; the paper itself later received a correction to one table, a useful reminder that publication and subsequent correction belong to the same record. [6][7]

Claim C03. Publication expands the community able to test a claim; it does not end the process by which that claim is evaluated. [5][6]

Where That Leaves Us

Publication makes a claim more widely inspectable and gives it a durable scholarly status, but it does not freeze its epistemic status. Replication, reanalysis, correction and contradiction can still change the balance of evidence. That leaves a further question: if publication is not finality, what exactly should scientific consensus mean when communities later report broad agreement?

4. Consensus is not the same thing as unanimity

Scientific consensus can mean several things, so the term should be used carefully. In formal consensus-study processes, the National Academies describes consensus reports as evidence-based positions produced by expert committees from gathered information and deliberation, followed by independent peer review. [8]

That is different from asking a room to raise hands. The epistemic value comes from the evidence and the procedures used to challenge interpretations, not from unanimity as such. A mature consensus may still contain uncertainty, unresolved subquestions and legitimate minority interpretations.

Claim C04 - Declared synthesis. In scientific assessment, consensus is most informative when it records convergent expert judgment grounded in an explicit body of evidence and critical deliberation. [8] directly supports the institutional model; the generalization beyond that model is a synthesis.

Where That Leaves Us

Consensus is most informative when it records evidence-based expert convergence under procedures that expose reasoning to challenge; it is not equivalent to unanimity or a headcount. That conclusion is well supported institutionally but remains generalizable only with care. The next question is why agreement reached through partly independent evidential routes should carry more weight than repeated agreement built on shared dependencies.

5. Agreement becomes more informative when routes to it are partly independent

Ten papers that reuse the same dataset, assumption or calibration are not equivalent to ten independent lines of inquiry. Convergence is more informative when different measurements, designs, models or populations fail in different ways. This is the same reason replication and methodological diversity matter: shared weaknesses can create shared conclusions.

The National Academies notes that consistency across independent studies can strengthen confidence while also warning that non-replication has multiple possible causes and must be interpreted in context. [5]

Claim C05. The evidential weight of agreement depends partly on how independent the contributing evidence and methods are. This is a declared synthesis grounded in the logic of replication described in [5].

Where That Leaves Us

Agreement gains evidential force when contributing studies, methods or measurements do not all share the same failure modes. Independence is rarely absolute, so the claim is comparative rather than binary. This leaves a practical difficulty: even a strong majority can be wrong, so how should a minority position be treated without either dismissing it or granting it automatic parity?

6. A minority view is not automatically wrong

History contains cases in which minority positions became mainstream after new evidence, better instruments or improved theory. That possibility is important because peer review and expert judgment can be conservative. Benda and Engels found evidence that peer review can perform reasonably on average while also identifying tension between review and innovative ideas, with bias and preconceptions posing risks. [3]

But the symmetrical mistake is just as serious: the existence of a minority view does not show that the majority is poorly supported. A disagreement has evidential force only to the extent that the competing position supplies methods, observations or arguments capable of changing the balance of evidence.

Claim C06 - Declared synthesis. Minority status should neither disqualify a claim nor grant it evidential parity. [3] supports the risk of conservatism; the parity rule is an explicit methodological synthesis.

Where That Leaves Us

Minority status is not an evidential defect by itself, and review systems can be conservative; yet dissent also does not create equal support where evidence is unequal. The warranted position is therefore conditional on methods, observations and arguments rather than social rank. The next question is how disagreement can be localized so that competing interpretations become testable instead of merely opposed.

7. Disagreement is most useful when it can be localized

“Scientists disagree” is often too coarse to be informative. The disagreement may concern the raw observations, how variables were measured, which cases were excluded, which model is appropriate, how uncertainty was propagated, whether a causal interpretation is identified, or how much weight to assign to competing evidence.

Localizing disagreement matters because different disagreements require different tests. A calibration dispute may be settled by better measurement. A model dispute may require a discriminating prediction. A causal dispute may require a different design. A disagreement about values or acceptable risk cannot be settled by pretending it is merely a measurement problem.

Claim C07 - Declared synthesis. The most productive scientific disagreements specify the level at which competing interpretations diverge and what observation or analysis could discriminate between them. This is a methodological synthesis of the review, replication and model-testing logic developed across KNO-013, KNO-016 and the present article.

Where That Leaves Us

A disagreement becomes scientifically useful when we can specify whether it concerns observations, measurement, model choice, causality, uncertainty or values, because different sources of disagreement require different tests. This is a methodological synthesis rather than a universal taxonomy. Once disagreement is localized, the next question is whether a stable consensus should be treated as settled or merely well supported for now.

8. Consensus can be stable without being final

A consensus may remain stable for decades because successive tests continue to support it. It can also change quickly when a new result removes a central assumption or when a new instrument opens a previously inaccessible regime. Stability therefore tells us something about the history of successful challenge, but it does not create logical immunity from revision.

The National Academies' treatment of reproducibility and replicability explicitly allows non-replication to contribute to discovery rather than treating inconsistency as automatically pathological. [5]

Claim C08. A scientific consensus can be well supported and still remain revisable in principle. [5] [8]

Where That Leaves Us

A stable consensus can reflect repeated survival under criticism without becoming logically immune to revision. Its stability is evidence about the history of successful challenge, not a guarantee concerning every future observation. That distinction leads naturally to the scholarly record itself: if claims remain revisable after publication, what should we learn when an explicit correction appears?

9. Correction is evidence that scrutiny continued

Corrections and retractions are sometimes presented as evidence that science “failed.” They can instead show that the record remained open to repair. This does not mean every correction is trivial or that institutional self-correction is always fast. It means the relevant question is whether errors can be detected, documented and propagated through later work.

The 2022 correction to Ioannidis' 2005 paper is a simple example: the journal preserved the original article, linked the correction and specified the exact table error. [7]

Claim C09. A visible correction can be evidence of continuing quality control, although the existence of a correction does not by itself measure the reliability of the rest of the record. [7]

Where That Leaves Us

A visible correction can show that scrutiny continued and that a specific error was documented in the public record, although the existence of a correction does not certify the reliability of everything else. The epistemic gain is auditability, not absolution. The next question is therefore what information a claim of review or consensus must expose if readers are to audit it rather than trust a label.

10. Consensus claims need an audit trail

A statement such as “experts agree” is weaker than a statement that identifies who was assessed, what question they considered, what evidence base they reviewed, how disagreements were handled and when the assessment was made. Formal consensus-study processes are valuable partly because they make at least some of that path inspectable. [8]

The same principle applies to review. COPE emphasizes declared conflicts, relevant expertise, confidentiality and constructive reasoning because a review gains legitimacy from how it was performed, not merely from the label “peer reviewed.” [1]

Claim C10. Claims about review or consensus are more auditable when the process, participants, scope and evidence base are explicit. [1][8]

Where That Leaves Us

Claims about review or consensus are more auditable when they identify participants, scope, procedure and evidential basis. That does not guarantee a good outcome, but it makes the route to the judgment inspectable. Once this audit trail is visible, a deeper limit appears: can agreement or disagreement itself substitute for the first-order evidence under discussion?

11. What disagreement cannot do

Disagreement alone cannot make two claims equally probable. It cannot substitute for missing evidence. Nor can consensus alone turn weak measurements into strong ones. Both agreement and disagreement are second-order information: they tell us something about how a community has processed evidence, but they do not replace the first-order evidence itself.

Claim C11 - Declared synthesis. Expert agreement and disagreement are evidence about the state of evaluation, not substitutes for the observations, measurements and arguments being evaluated.

Where That Leaves Us

Expert agreement and disagreement are second-order information about evaluation, not replacements for observations, measurements and arguments. This is a declared synthesis and should not be mistaken for a theorem derived from one source. The next question is practical: given that social signals are informative but limited, how should a reader combine them with first-order evidence when encountering a contested claim?

12. A practical reading rule

When you encounter a contested claim, ask four questions:

1. **What is the first-order evidence?** What was measured, observed or derived?
2. **What was the review process?** Who examined the claim, with what expertise and what opportunity to challenge it?
3. **What exactly is agreed or disputed?** Data, method, model, causal interpretation, uncertainty, or value judgment?
4. **What would change the balance?** A replication, a new measurement, a discriminating experiment, a corrected analysis, or a new theoretical constraint?

That procedure does not guarantee the right answer. It prevents the words *peer reviewed*, *consensus* and *controversy* from doing more epistemic work than they can support.

Claim C12 - Declared synthesis. The strength of a review or consensus claim should be evaluated together with the underlying evidence, process and possible discriminating tests.

Where That Leaves Us

A disciplined reader asks what was measured, how the claim was reviewed, where disagreement lies and what evidence could discriminate between alternatives. The rule does not guarantee the right answer; it prevents institutional labels from doing more epistemic work than they can support. The final question is one of scope: which neighboring KNO articles own the evidence-building and reproducibility issues that this article deliberately leaves outside?

13. Boundary and inherited cases

KNO-020 owns the **social evaluation layer**: how review, consensus and disagreement organize criticism after evidence has been produced. It does not replace KNO-016, which asks how evidence itself is assembled, or KNO-021, which will treat openness, reproducibility infrastructure and practices that make checking possible.

Five earlier KNO articles delegated specific disagreement cases here. Their disposition is deliberately limited. KNO-010's disputes over species and rank remain examples of classification disagreement; KNO-011's permissive and restrictive definitions of measurement remain examples of definitional disagreement; KNO-012's "experimenter returns" problem remains a case of disagreement about controls and intervention; KNO-013's realism versus empirical adequacy remains a model-interpretation dispute; and KNO-016's statement that peer review and consensus belong here is fulfilled by the present article's treatment of social scrutiny. None of those cases is re-litigated here. They remain illustrative cases in their owner articles, while KNO-020 supplies the general framework for interpreting the disagreement.

Where That Leaves Us

KNO-020 now closes the delegated social-evaluation cases without duplicating their scientific content: earlier articles retain their specific disputes, while this article supplies the general framework for interpreting review, consensus and disagreement. KNO-016 owns evidence construction and KNO-021 owns open-science and reproducibility infrastructure. With those boundaries fixed, the article can now state what is known, uncertain and revisable about social evaluation.

What We Know, and How

- Structured peer review can expose problems before publication, but reviewer judgments vary and are not certificates of truth. [1][2][3]
- Editorial decisions can incorporate conflicting reviewer advice rather than simply count votes. [4]
- Reproducibility and replicability extend scrutiny beyond publication and can strengthen, qualify or challenge claims. [5]
- Formal scientific consensus processes can combine evidence review, expert deliberation and independent review. [8]
- Corrections can preserve an auditable record of errors found after publication. [7]

What We Still Don't Know

There is no universal numerical conversion from “degree of expert agreement” to probability that a claim is true. Different fields also use different review models, evidential standards and mechanisms for building consensus. The reliability of those processes depends on the question, the evidence, the incentives and the independence of the people and methods involved.

What Could Change Our Minds

Evidence that a review model systematically fails to detect important errors, that a consensus process suppresses relevant counterevidence, or that a minority model repeatedly makes better predictions should reduce confidence in the prevailing assessment. Conversely, repeated independent tests that eliminate plausible alternatives should increase it.

If You Remember Only One Thing

Review, consensus and disagreement organize criticism; their epistemic value depends on the evidence, procedures and discriminating tests they make visible.

Related Articles

- KNO-012 · Experiments and Variable Control
- KNO-013 · Models, Maps and Reality
- KNO-014 · Causality and Correlation
- KNO-015 · Uncertainty and Probability
- KNO-016 · How Evidence Is Built: Putting the Tools Together

- KNO-021 · Open Science and Reproducibility

References

1. COPE Council. *COPE Ethical Guidelines for Peer Reviewers*. Version 2, September 2017. DOI: 10.24318/cope.2019.1.9.
2. Bornmann, Lutz. "Scientific peer review." *Annual Review of Information Science and Technology* 45(1): 197-245 (2011). DOI: 10.1002/aris.2011.1440450112.
3. Benda, Wim G.G.; Engels, Tim C.E. "The predictive validity of peer review: A selective review of the judgmental forecasting qualities of peers, and implications for innovation in science." *International Journal of Forecasting* 27(1): 166-182 (2011). DOI: 10.1016/j.ijforecast.2010.03.003.
4. *Nature Genetics*. "Peer Review." Editorial policies, current web policy. No DOI; live publisher policy page.
5. National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press, 2019. DOI: 10.17226/25303.
6. Ioannidis, John P. A. "Why Most Published Research Findings Are False." *PLOS Medicine* 2(8): e124 (2005). DOI: 10.1371/journal.pmed.0020124.
7. Ioannidis, John P. A. "Correction: Why Most Published Research Findings Are False." *PLOS Medicine* 19(8): e1004085 (2022). DOI: 10.1371/journal.pmed.1004085.
8. National Academies of Sciences, Engineering, and Medicine. "Consensus Study Reports" institutional description, National Academies Press. Live institutional page; no DOI for the description itself.