**Title: Replication is an epistemic principle in material theory of induction**

**Abstract**

In John Norton’s Material Theory of Induction, background facts provide the warrant for inductive inference and determine evidential relevance. Replication, however, is excluded as a principle of inductive logic. While Norton argues replication lacks the precision and methodological clarity to serve as a material principle of inference, I argue that replication nonetheless functions as an epistemic principle of induction. I examine how replication contributes to epistemic justification within both externalist and internalist frameworks and show that its role extends beyond procedural repetition. Replication acts as a reliable belief-forming process for identifying stable facts and inferences. This reframes MTI as a theory shaped not only by local facts but by how scientists determine which facts can function as background warrant.

**1. Introduction**

John Norton’s Material Theory of Induction [MTI] proposes that scientific induction is licensed by background facts [BFs] that provide the epistemic warrant for inductive logic. Norton’s theory compellingly challenges the notion of universal inductive schemas such as inductive generalization, hypothetical induction, and probabilistic induction (Norton, 2003). As part of this account, Norton dismisses replication as a principle of inductive logic, arguing that it lacks clear application boundaries, defined procedural thresholds, and consistent evidential import (Norton, 2015). Because replication does not consistently indicate when an outcome should count as evidence, Norton concludes it cannot ground inductive logic. But this overlooks how replication is already used in practice to determine which inferences and BFs are taken as epistemically reliable. Replication, in practice, justifies scientists’ beliefs about which inferences and BFs are stable and reliable. While replication does not meet Norton’s criteria for a principle of logical consequence, it serves an important epistemic function.

In Section 2A, I provide an overview of the MTI, and in Section 2B, I discuss Norton’s claim that replication does not play a formal role in inductive logic, even if it may have a practical or epistemic function. In Sections 3A–3C, I address Norton’s hasty dismissal of replication, as presented in Chapter 3 of *The Material Theory of Induction* and the 2015 paper *Replicability of Experiment*. Specifically, I argue that Norton’s classification of BFs, as it relates to replication, relies on a metaphysical distinction, whereas the relevant issue is epistemic, whether BFs are known. Further, in response to Norton’s claim that replication is irrelevant without prior knowledge of enabling BFs, I present cases from scientific and medical practice where facts have been widely adopted based solely on consistent replicability. Finally, I refute Norton’s view that science does not treat replication as foundational to induction by showing how multi-center research explicitly relies on replication to establish reliable BFs for experimental protocols.

In Section 4, I differentiate between two meanings of induction, as identified by de Grefte (2020): the logical and the epistemic. While replication is not a formal principle of *inductive implication*, understood as the logical relationship of support between propositions, I argue that it functions as an epistemic principle in *inductive inference*, understood as the method by which beliefs are formed from prior beliefs (de Grefte, 2020). In Section 5, building on de Grefte’s argument that MTI is committed to epistemic externalism, I explore how replication can be understood differently depending on one’s epistemic commitments, whether externalist or internalist.

This paper argues that replication is an essential epistemic principle, one that enables the operational reliability of inference and BFs over time.

**2. The Material Theory of Induction and the Replicability of Experiment**

**2A. Material approaches to induction**

To illustrate MTI’s view on the role of BFs in justifying induction, Norton (2021) shows how domain-specific BFs can support an inference that would otherwise be, to use Norton’s term, “fragile”. A key function of BFs is to narrow the inductive risk inherent in applying any inductive schema. By relying on local knowledge, an inference can be warranted by the specific BFs relevant to the domain, even when the inductive schema employed is weak.

Norton provides the example of Marie Curie’s work with radium chloride [RaCl₂]. Curie had only a single sample of isolated RaCl₂, yet she inferred that *all* RaCl₂ would be isomorphic to barium chloride [BaCl₂], based on their shared structural properties. At face value, this seems like a fragile inductive inference: generalizing from a single instance to all, an inference unlikely to hold. However, Norton argues that Curie’s induction was warranted by BFs in mineralogy. Curie’s induction drew on Huay’s principle, which states that each crystalline substance generally has a single characteristic crystallographic form. The concept of isomorphous groups, where crystalline substances with analogous chemical compositions tend to have closely similar crystal forms, further supported the inference.

Because RaCl₂ and BaCl₂ both form monoclinic chlorides, Norton argues that it was deductively valid, given Huay’s principle, to infer that RaCl₂ would exhibit the same properties as BaCl₂. Although Curie’s inference might be categorized as inductive generalization, inference to the best explanation, or Bayesian reasoning, Norton argues that these forms do not provide the warrant. Instead, it is the local BFs, such as Huay’s principle and the structural analogy to BaCl₂, that justify the inference. These mineralogical BFs functioned as warranting BF that, in Norton’s account, justified Curie’s induction (Norton, 2021; Shenker, 2020).

**2B. Norton’s view on replication in induction**

1. Top of Form

Norton identifies three issues preventing replication from serving as a formal principle of inductive logic: (a) the difficulty in distinguishing between experiments, (b) vagueness in the concept of a "good indicator," and (c) replication’s failure to provide (or success in providing) clear guidance on the inductive significance of experimental outcomes (Norton 2015). I will return to each of these issues in section 4, where I examine each in turn to show why replication does not function as a general inductive principle but can still contribute to justification within MTI.

For now, let me turn briefly to issue (c): Norton contends that the evidential significance of successful replication depends on prevailing BFs. Norton argues that replication alone does not dictate whether an experimental result should be accepted or rejected; rather, it is the BFs that determine the inductive import of evidence. Norton illustrates this by presenting scenarios where (i) replication *was not* successful and yet was largely ignored by the scientific community and where (ii) replication *was* successful, yet the evidence was not imported by the scientific community (Norton 2015).

Norton does not treat replication as a universal inductive principle but still recognizes that it can play a practical evidential role when guided by appropriate BFs. In analyzing the role of replication, Norton categorizes BFs into two kinds. Class A BFs, which “specify conditions under which the effect or process of interest will manifest in a veridical experimental outcome,” and Class B BFs, which “specify the conditions conducive to spurious experimental outcomes. These conditions simulate a veridical experimental outcome, when the sought effect or process is not present; or they may interfere sufficiently to produce an unsuccessful outcome, when the effect or process is present” (Norton 2015). Armed with this distinction, Norton argues that the function of replication is to test for and isolate Class B BFs, that is, confounding conditions that could simulate a veridical result (Norton 2015). Norton states:

“With any experiment, we can be uncertain whether appropriate facts in classes A and B prevail. Successful replication does not test all of them. Rather it tests whether certain unfavorable confounding conditions of class B are present” (Norton 2015).

Furthermore, without already established Class A BFs, an experimental paradigm is not justified as a serious scientific endeavor. Norton argues you need both Classes A and B to be hospitable:

“The inductive inference to this conclusion is warranted by appropriate facts in class[es] A and B” (Norton 2015).

“For successful replication requires the facts in classes A and B above to be hospitable” (Norton 2015)

For scenario (i), where replication *was not* successful and yet was largely ignored by the scientific community, Norton examines a series of ether wind experiments conducted in the early 20th century, focusing on how BFs shape the inductive import of experimental replication. The Michelson-Morley experiment (1887) was designed to detect the presence of an ether wind, a hypothesized medium through which light propagated. However, the experiment produced a null result, suggesting that either the ether did not exist or that its effects were undetectable. Decades later, Miller (1925) conducted a similar experiment, reporting positive evidence for the ether wind. Miller’s findings appeared to contradict the Michelson-Morley result and reignited debate over the existence of the ether (Norton 2015).

Given the implications of these results for the theory of special relativity, Einstein was forced to respond to Miller’s claims. If Miller’s findings were correct, they would directly challenge the principle of the constancy of the speed of light, a cornerstone of relativity. Einstein dismissed Miller’s results, suggesting that an undetected temperature effect, a Class B BF, had compromised the experiment. Later, in 1955, Shankland et al. re-examined Miller’s data and replicated Miller’s experimental setup, demonstrating that temperature variations could indeed reproduce the effect Miller observed. This finding ultimately discredited Miller’s results and reinforced the conclusions drawn from the Michelson-Morley experiment. Norton presents this case to show how a failed replication, when influenced by known or unknown BFs, can be dismissed by scientists, who then retain previously accepted conclusions. In this instance, Miller’s positive ether-drift result was rejected due to a suspected confounding factor (a Class B BF). Norton uses this to emphasize that the inductive weight of any replicated evidence depends on the prevailing BFs. Replication, in Norton’s view, does not speak for itself; it is the BF that determines how replication is interpreted and whether it counts as meaningful evidence (Norton 2015).

For scenario (ii), where replication *was* successful, yet the evidence was not imported by the scientific community, Norton examines the history of intercessory prayer experiments, which aimed to test whether prayer could influence health outcomes. Figures such as Richard Dawkins were particularly vocal in their rejection of these experiments, arguing that they lacked the necessary grounding in established scientific facts. In *The God Delusion*, Dawkins writes that “the very idea of doing such experiments is open to a generous measure of ridicule” (Norton 2015). Without Class A BFs that could support the inference, even replicated results were dismissed as epistemically insignificant. Norton cites a case where the authors of such a study, comprised of a Catholic, a Protestant, and a Jew, “became convinced that the very idea of testing distant prayer scientifically was fundamentally unsound” (Norton 2015), given they had no good reason to expect the effect of the process of interest (supernatural intervention) to manifest.

Norton uses these cases to argue that replication is not itself a principle of induction; rather, its epistemic significance depends on whether appropriate BFs are present. This example supports Norton’s broader claim that induction is determined by localized material facts rather than a universal schema (Norton 2015). It also aligns with the material approach to reproducibility, which holds that replication alone is insufficient to establish inductive warrant unless supported by appropriate BFs.

**3. Norton’s Hasty Dismissal of Replication’s Epistemic Role**

**3A. Conjecture – Einstein’s mere hypothesis**

Norton’s conclusions in the series of ether wind experiments suffer from three key problems. Norton overlooks (i) the speculative nature of Einstein’s initial inference, (ii) misinterprets how replication contributed to discovering previously unconceived BFs, and (iii). wrongly assumes a rigid distinction between Class A and Class B BFs. Here, I will examine these issues in detail, showing how Norton’s treatment of replication fails to align with science as practiced.

Bottom of Form

(i) Einstein’s untested inference to the best explanation

Einstein, at the time of inference, likely relied on a *general* BF, that temperature can confound experiments, but did not yet have a *particular* BF tailored to the ether-drift experiments. What Einstein had was merely a hypothesis, a conjecture, not a warranted BF. This distinction matters: a general BF can motivate an idea, but without specificity, it remains speculative. Replication is what transforms this speculative link into a stabilized BF, guiding the confirmation, and eventual acceptance of the temperature explanation as a warranted BF. Norton fails to fully acknowledge the benefit of hindsight Norton has in evaluating Einstein’s inference: what began as a conjecture only gained epistemic legitimacy after successful replication by Shankland et al. confirmed temperature as the relevant BF. Furthermore, our ability to recognize stable inferences relies on the fact that they have not been overturned by subsequent experiments. Had these findings not been replicated, the entire chain of dependent scientific claims would have to be reconsidered.

‘Stability’ here refers to the ability of an inference or BF to remain justified across varied contexts of replication. Rather than guaranteeing universal repeatability, replication contributes to inductive warrant by helping identify which inference and BFs are stable enough to support further inference. A stable BF endures empirical testing and can function as a reliable grounding in material inferences. Section 4 considers how different forms of replication, both successful and failed, interact with BF stability in specific historical cases.

(ii) The role of replication in discovery

The ether wind experiments illustrate how replication played a crucial role in identifying previously unconceived BFs, as Einstein hypothesized. Norton acknowledges that replication tests for the presence of confounding Class B BFs, those that might simulate a veridical result, by systematically varying operators, materials, or conditions (Norton 2015, 240). However, Norton frames replication as a method for ruling out potential sources of error, not as one that identifies or confirms BFs that positively support an inference (Class A), but only as a tool for excluding confounding influences (Class B). Yet Shankland et al.’s replication did more than eliminate a possibility. By deliberately manipulating temperature and reproducing Miller’s results, they identified temperature as a specific BF that shaped the observed effect. Once confirmed, this BF no longer functioned as a suspected confounder (Class B) but contributed to reliable inference. In this way, replication not only helped uncover a previously unrecognized BF, but also helped establish that this BF, initially treated as a confounder, could in fact support inference and thus belonged among those warranting epistemic significance.

(iii) The problematic Class A vs. Class B distinction

Norton’s classification of BFs is *metaphysical* in nature, determined according to whether they enable or frustrate experiments concerning the effect of interest. However, the scientifically relevant distinction is epistemic, determined by whether the pertinent BFs are known. Norton’s classification of BFs into Class A and Class B is not a logical distinction.

The temperature effect, initially treated as a Class B BF in Miller’s experiment, was reclassified as a Class A BF in Shankland et al.’s study after replication empirically confirmed temperature as the confounding factor responsible for Miller’s misleading veridical outcome. According to Norton’s definitions, a Class B BF is a factor that interferes with the observation of the target effect, whereas a Class A BF is one required for producing it. Norton presents this as a metaphysical distinction that applies regardless of whether the factor is recognized. However, as seen in Miller’s case, temperature fluctuation was interfering but unrecognized, so it was a Class B BF. In Shankland’s study, replication revealed its role and it was deliberately incorporated into the reasoning, making it a Class A BF. This confirmation transformed it from a merely suspected Class B BF into a known inductively relevant Class A BF, enabling reliable inferences about the failure of Miller’s experiment and supporting the broader rejection of the ether theory. This shift, from being treated as a background condition to being recognized as a causally relevant factor, illustrates how the epistemic role of a BF depends on its experimental validation, demonstrating that these categories are not logical in nature. Their classification depends on contextual interpretation within experimental practice, not on formal or logical principled criteria. A better classification would distinguish between *unknown BFs*, those unconceived and undiscovered and therefore not controlled for, and *known BFs*, which are both conceived and discovered. Once a BF is found to be epistemically significant, it shifts from Class B to Class A, showing that the distinction is an artifact of experimental design and interpretation rather than an inherent logical distinction. According to Norton, replication identifies previously unrecognized Class B BFs. However, these BFs can transition into Class A once their relevance is confirmed.

Seen this way, replication should be framed epistemically in terms of stability: *a stabilizing replicate* confirms prior results, while a *destabilizing replicate* challenges reliability and may lead to the discovery of new BFs. In this way, replication serves as a tool for discovering both kinds of BFs. Norton’s reliance on the Class A/B distinction downplays replication’s epistemic significance by treating it as merely uncovering the existence of Class B BFs. Once recognized as relevant, a BF can guide further induction, not because it is metaphysically fixed, but because it has proven itself stable through repeated engagement.

**3B. Class A background fact is not required**

Norton argues that a Class A BF is necessary for an experiment to be considered evidentially significant, and without one, replication becomes irrelevant. Norton’s primary example is the prayer experiment, where the supposed absence of a Class A BF—supernatural intervention, which in scientific terms would be considered a mechanism of action (MOA)—is used to justify dismissing experimental results. Norton claims that without an identified Class A BF, experimental findings lack evidential significance. However, this argument relies on a carefully selected example. Numerous cases in science demonstrate that replication retains evidential weight even when a Class A BF, or what is commonly referred to as MOA, is unknown. Consider the following examples:

(i) Metformin: A drug validated through replication, despite an unknown MOA

Metformin, a widely used first-line treatment for type 2 diabetes (Diabetes Care 2018), exemplifies how an intervention can be deemed reliable based on reproducibility alone, even when its MOA is not fully understood. Originally introduced as a diabetes treatment in the UK and other European countries in 1958 (though first discovered in 1922), metformin’s efficacy was established through decades of consistent clinical results. Scientists observed its effectiveness long before they understood its molecular mechanisms (Bailey 2017).

While some aspects of metformin’s MOA, such as its activation of AMP-activated protein kinase (AMPK) in the 1980s[[1]](#footnote-1), have since been elucidated, its precise molecular target remains unidentified to this day. Nevertheless, metformin has been continuously used and recommended based on its reproducibility in controlling blood glucose levels[[2]](#footnote-2). The absence of a fully defined Class A BF (i.e., a known MOA) has not prevented its widespread adoption, demonstrating that replication alone can be a sufficient basis for scientific and medical acceptance (Bailey 2017).

(ii) Fetal Bovine Serum (FBS): A standard reagent without a fully characterized MOA

Fetal Bovine Serum (FBS) has been a standard supplement for promoting cell growth in cell culture since its introduction by Theodore Puck in the late 1950s. Despite its widespread use, the precise active components responsible for its effects remain incompletely understood (van der Valk et al. 2018). FBS has become a standard reagent in cell culture research, biotechnology, and pharmaceutical production due to its capacity to support robust and reproducible experimental results. However, its composition is inherently variable, fluctuating between batches and across different regions and seasons, making it difficult to fully characterize as a consistent material. Nevertheless, despite these uncertainties, FBS remains widely used because its biological effects have been repeatedly demonstrated across various experimental conditions and research settings, reinforcing confidence in its reliability (van der Valk et al. 2018).

(iii) Another example of Norton’s Class A based dismissal

Another example Norton uses to argue that replication cannot confer evidential weight in the absence of a Class A BF is homeopathy. Norton states:

“Critics of homeopathy (such as me) will refuse to accept that a controlled trial of a homeopathic remedy can demonstrate the remedy’s efficacy, for the remedy contains no active ingredients.” (Norton 2015)

I, too, am a critic of homeopathy. However, Norton’s argument conflates two separate issues: (1) the ability of replication to establish empirical effects and (2) the requirement of a molecular mechanism for scientific credibility regarding the MOA. Contra Norton, replication can indeed establish that a treatment produces an effect, even if the MOA is unknown. But Norton’s point gains plausibility from the fact that no purported explanation of the MOA could be taken seriously if it lacked an active ingredient, this is a key distinction. If rigorous, well-controlled studies repeatedly demonstrated an effect beyond placebo, where rigor includes a clinically meaningful outcome with valid measurement, then even without an identified MOA or active ingredient, the evidence would warrant serious consideration. However, in the case of homeopathy, replication has failed to confirm its efficacy, and there is no evidence that homeopathic remedies produce any therapeutic effect beyond placebo, making it scientifically untenable (Cukaci et al. 2020). Norton dismisses homeopathy for the wrong reasons. As demonstrated by the cases of metformin and FBS, scientists routinely accept empirical results based on reproducibility alone, even when the MOA remains unknown.

**3C. Role of replication as practiced**

Norton, in *Replicability of Experiment, 2015*, states:

“While the schema favored by philosophers fail to be universal, this paper (Replicability of Experiment 2015) asks whether such a principle can be found in the explicit lore of practicing scientists.”

Scientific practice clearly shows that replication is a guiding principle. This aligns with Norton’s own methodological proposal: that the norms governing inductive inference should be uncovered by examining science as practiced. Scientists in multi-research center collaborations aim to standardize protocols, controlling for technical effects to ensure experimental results are reproducible across different settings. To achieve this, they actively identify and eliminate sources of variability between laboratories, compare discrepancies in reproducibility within the literature, and assess inconsistencies across published studies to determine whether methodological differences account for variations in results. Additionally, research centers collaborate directly, systematically testing protocols across laboratories to isolate sources of irreproducibility and optimize experimental conditions.

(i) Examples of standardization and reproducibility efforts

One example of this approach is the optimization of induced pluripotent stem cell (iPSC) differentiation protocols. Research centers collaborate to refine protocols for differentiating iPSCs into specific cell types. Given the sensitivity of iPSC differentiation to slight variations in conditions, standardization is necessary for reproducibility across labs. Scientists identify and eliminate inconsistencies in methodology to ensure robust and reliable differentiation outcomes (Volpato et al. 2018).

Another example is the study of C9orf72 BAC transgenic mouse phenotypes, where researchers working with these mice have encountered discrepancies in observed phenotypes across different laboratories. Multi-center collaborations have been essential in determining why certain phenotypic traits fail to replicate. By comparing experimental conditions, researchers can pinpoint variables that contribute to these inconsistencies, ultimately leading to more reliable scientific conclusions (Mordes et al. 2020).

(ii) Replication as an embedded scientific practice

Replication is also embedded in experimental design, particularly in biological research, where it serves as a foundation upon which new discoveries are built. One of the explicit manifestations of this practice is the convention of using prior findings as the basis for the first figure in new studies. In many biological studies, the first figure (often Figure 1A), reproduces key findings from previous research. This practice ensures that subsequent experiments build on a foundation of reliable, validated results (Kenyon et al 1993; Libina et al 2003). Similarly, experiments frequently include positive controls, which are often replications of previously published findings. This practice ensures that the experimental platform is functioning as expected and that any new results can be meaningfully interpreted.

Contrary to Norton’s claim that replication plays no fundamental role in evidential import, the explicit practices of working scientists suggest otherwise. Furthermore, replication can be an essential part of experimental design, with many studies explicitly reproducing previous findings to build upon prior research and serve as a foundation for new discoveries. Scientific practice treats replication as a fundamental method for determining which claims are stable enough to serve as BFs.

**4. Replication as an Epistemic, Not Logical, Principle**

**4A. Why Norton Misclassifies Replication**

de Grefte (2020) distinguishes between two meanings of induction. The first, *inductive implication*, refers to the logical relationship of support between two or more propositions. The second, *inductive inference*, describes the process of forming new beliefs based on prior ones. This distinction, between logical structure and epistemic practice, is crucial when evaluating the role of replication in MTI. Norton describes MTI as a theory of inductive logic, explaining how different propositions can inductively support one another (Norton, 2020, 2003a, 2003b, 2010, 2014; Shenker 2020; de Grefte 2020). However, as de Grefte points out, Norton does not clearly distinguish between *inductive implication* and *inductive inference* in MTI. I agree with de Grefte’s claim, as shown in Section 3, Norton assesses replication’s logical status through its epistemic role without clearly distinguishing the two. This leads Norton to assess replication’s epistemic role using criteria appropriate only for *inductive implication*. Replication isn’t a logical principle because it’s doing epistemic work. This misclassification explains why Norton concludes that replication fails as a principle of inductive logic. In Section 4, I reclassify replication not as a failed logical principle, but as a successful epistemic one. I take up each of these objections below to show how they illuminate the epistemic work replication performs in scientific practice.

(a) The distinction between two experiments cannot be clearly stated

Norton suggests that if replication is to be treated as a precise inductive principle, then the distinction between a single experiment and two distinct experiments must be clearly defined, but this, Norton argues, is problematic. However, scientists already address this ambiguity in practice, and the vagueness does not diminish replication’s epistemic role, as will be discussed further below. The distinction is handled locally by expert scientists who determine what counts as a valid replication. Strictly speaking, no experiment can ever be exactly the same, a problem highlighted by experimenter’s regress[[3]](#footnote-3) (Collins 1985). Importantly, replication’s flexibility allows it to function effectively across a range of methodological contexts, making it a useful tool for evaluating scientific inferences.

To clarify this flexibility, I draw on the classification of replication proposed by Ioannidis and colleagues (Goodman et al. 2016). They distinguish three forms of replication:

* **Method reproducibility (MR):** repeating the exact experimental protocol—for example, using the same knockout mouse line, tissue, RNA isolation method, and qPCR settings to measure gene expression.
* **Result reproducibility (RR):** obtaining the same empirical result by repeating the exact method or using sufficiently similar procedures—for instance, another lab uses a similar or the same knockout model and qPCR approach and observes the same downregulation.
* **Inference reproducibility (IR):** reaching the same scientific conclusion—that the knocked-out gene downregulates a particular pathway—even if one lab uses RNA expression and another uses protein expression data.

This framework can support the idea that replication is not merely about identical experiments (MR) but rather about establishing the robustness of findings across varying conditions. In this context, an important epistemic distinction in replication must be considered: *direct* vs. *indirect* *(or conceptual)* replication and the type of reproduction (Derksen & Morawski, 2022). Direct replication involves using the exact methods to reproduce results (M1 → R1), whereas indirect (or conceptual) replication relies on differing methods to achieve similar outcomes (M2 → R1). In theory, indirect replication allows more opportunities for the discovery of BFs because it diversifies its method, enabling the exploration of a larger range of possibilities. Where we choose to place our confidence in replication depends on these distinctions.

Consider the example of reproducing gene inhibition to observe phenotypic consequences. This goal does not, on its own, dictate a single method of replication, as multiple experimental routes can be used to achieve it. In biology, it is common practice to inhibit a gene or its product at the RNA or protein level to observe a phenotype. This can be achieved through DNA mutation, interfering with RNA expression (e.g., RNA interference, RNAi), or direct inhibition of protein function, respectively. Successful replication across all three methods increases confidence in the observed effects, while failed replication can generate new hypotheses.

For example, consider a case where a DNA mutation fails to replicate the phenotypic effects observed with RNAi (Van Gilst et al. 2005; Brock et al. 2006). One well-supported BF is that mutation of an essential gene can trigger compensatory mechanisms during development (Lambie and Kimble, 1991). In contrast, RNAi knockdowns, which are often incomplete or tissue-specific, may avoid triggering such compensation and can also produce off-target effects (Kennedy et al. 2004; Dillin, 2003). A mismatch between mutant and RNAi phenotypes may therefore prompt new hypotheses — for instance, that the gene plays a role during development or that its function is localized. In some cases, such mismatches have been attributed to developmental compensation, where compensation was proposed and developmental relevance was later supported by additional genetic evidence (Brock et al. 2006; Ashrafi, 2007). Simply replicating gene inhibition using a single method would not have revealed these functional distinctions or led to these insights.Top of Form

Taken together, these examples show that the ambiguity Norton identifies, specifically the inability to sharply distinguish between experiments, is not a flaw but a key feature of replication’s epistemic value. This flexibility allows replication to operate across diverse scientific contexts. Its lack of formal strictness is exactly what enables replication to guide inference and belief revision in practice.

(b) The concept of a ‘good indicator’ is too vague

Norton critiques replication by arguing that it relies on vague inductive terms, particularly the term ‘good indicator,’ which he claims is too imprecise to function as a logical principle. For replication to serve as a valid principle of inductive logic, it would need to meet formal standards of precision and context-independence. However, this critique only applies if replication is assumed to function as a formal principle of inductive logic.

In science as practice, the vagueness Norton highlights is routinely navigated through methodological norms. The heuristic of “more replication is better” allows for an evolving understanding of scientific claims, rather than adhering to a singular, formal standard. The more a result is replicated, the stronger its epistemic justification becomes. This aligns with detachment in induction (Norton, 2003), where repeated instances of Result Reproducibility (RR) or Inference Reproducibility (IR) progressively increase confidence in the underlying claim until it is treated as sufficiently warranted to be accepted without proviso. Instead of requiring formal precision, replication increases reliability through repeated testing across varying conditions, which is how scientists handle uncertainty.

While a single replication may lend some support, only a robust pattern of replications across independent studies significantly strengthens the epistemic weight of a finding. Each successful replication adds to the cumulative justification for scientific inferences, helping to distinguish reliable results from weakly supported ones. This progression of increasing confidence is not about guaranteeing truth but about building justification over time. The vagueness Norton identifies poses no problem for my account, since I do not claim that replication is a principle of *inductive implication*. Its epistemic function does not depend on meeting the *inductive implication* criteria that Norton requires for inclusion in a formal logic of induction.

(c) Replication provides no guidance on whether an experimental outcome is epistemically inert

This concern is addressed in detail in Section 3A&B. The key argument is that replication plays a crucial role in establishing justification. Norton’s objection, that replication offers no guidance on whether an experimental outcome is epistemically inert, misses the fact that replication provides a practical process for distinguishing reliable from unreliable inferences.

While replication may not serve as a universal principle of *inductive implications*, its epistemic function is clear: it guides scientists in determining which results are, in Norton’s terms, epistemically significant. That is, which replication outcomes meaningfully contribute to the justification of a claim, and which are epistemically inert—where the outcome fails to support or undermine the claim's justification.

**4B.** **Replication as Epistemic Principle**

Norton critiques replication for lacking the precision and methodological clarity required of a logical principle, features central to his view of inductive implication (Norton 2015, 229). Norton’s assessment sets aside broader accounts of logic, such as probabilistic or context-sensitive systems, which Norton does not consider in his critique. Nonetheless, this reflects a category mistake: replication is not a logical principle, but an epistemic one. Norton overlooks replication’s crucial epistemic role in justifying beliefs, particularly in relation to BFs. Replication provides evidential support for the stability of BFs, thereby justifying the inferences drawn from them and the claims those inferences support.

Replication is about how scientists come to believe which BFs, evidence, or inferences are stable and reliable. Scientific practice does not require certainty; it requires stable, justified, workable belief. That is what replication provides. While unsuited for inclusion as a principle of induction based on implication, replication remains foundational to scientific knowledge. Justification in practice is shaped by local judgment, evolving standards, and iterative refinement. By focusing narrowly on replication’s inductive implication inadequacy, Norton overlooks its central epistemic role.

Replication plays its epistemic role not by meeting the formal structure of inductive implication but through how it operates in practice: it is contextual, revisable, and pervasive. Without accounting for replication and its epistemic role, I argue MTI’s model of justification remains incomplete. In Section 5, I examine how treating replication as a justificatory principle influences MTI’s account of epistemic justification.

**5. Epistemic Consequences of Treating Replication as a Principle in MTI**

de Grefte argues that MTI carries a commitment to epistemic externalism, since the justification of inductive beliefs depends on BFs that exist independently of an individual's beliefs. Changes in these BFs alter the epistemic status of beliefs, making justification contingent on external conditions. Because material postulates serve as the warrant for inductive inferences, a belief is justified only if the relevant BFs actually obtain. If the material fact underpinning a belief is false, absent, or inapplicable, then the belief was never justified, even if it appeared so to the agent. On this view, justification is externally grounded: it depends on how the world is, not on what the agent knows or can access. In this section, I will briefly explore the epistemic consequences of adopting replication as an epistemic principle in MTI and show how replication can function within both externalist and internalist accounts of epistemic justification[[4]](#footnote-4).

(i) Externalist reading

The role of replication in externalist justification can be interpreted in at least three ways. First, it can be viewed as an epistemic tool for filtering BFs. Second, replication itself may count as a belief-forming process that satisfies reliabilist criteria. Third, replication can serve as a meta-methodological tool for confirming that a method (or process) is reliable. Each view offers a distinct but compatible reading of replication’s epistemic significance within MTI.

The first reading builds on de Grefte’s externalist framework. Replication contributes to epistemic justification by exerting a selective pressure on BFs that justify induction, such that only those that are stable and reproducible tend to persist in scientific practice. This epistemic filtering is not driven by agent judgment, but by the empirical structure of replication itself: BFs that support reliable inference (Class A) continue to be used, while unstable BFs or those that confound observation (Class B) fall out of circulation. Although replication is not itself the warranting mechanism and does not guarantee the truth of any particular BF, it remains epistemically indispensable. By revealing which BFs consistently persist across repeated investigation, replication helps determine which BFs can be reliably appealed to as sufficiently stable and reliable in the world to support inductive reasoning.

On another level, the externalist account of knowledge can be understood through a reliabilist lens (Goldman, 1979), where replication itself constitutes a belief-forming process that tends to produce true beliefs. According to this view, a belief is knowledge if it is true and is caused by a process that reliably produces true beliefs. Crucially, justification depends not on the agent’s access to reasons or evidence, but on the actual reliability of the belief-forming process. That process must be truth-conducive in the real world, whether or not the agent can verify its reliability internally. Thus, reliabilism holds that justification is determined externally, by the belief’s causal history, rather than internally, by the agent’s reflective access.

If replication as a method consistently yields a result under varying conditions, and that process causes the agent to believe R, the belief is justified because it arises from a truth-tracking mechanism, even if the agent lacks reflective access to why it’s justified. In this framework, replication functions as a reliable belief-forming process, a method that typically produces true beliefs, thereby justifying the resulting belief within an externalist framework. While it is not the warranting condition in MTI, replication contributes to epistemic justification by acting as a process that reliably generates true beliefs, making it fully compatible with externalist commitments. Replication supports externalist justification because stable and true facts, those that hold across varying conditions, tend to be replicable, even when their underlying structure remains unknown to the agent.

A third interpretation treats replication not as the belief-forming process itself, but as a meta-methodological tool that evaluates the reliability of those processes. In this view, replication does not directly form beliefs or serve as a warranting condition but instead confirms whether a given method consistently yields stable, truth-tracking outcomes. For example, consider how indirect strategies like computational protein structure prediction (e.g., AlphaFold) gain credibility over time. When their predictions are later confirmed by direct experimental methods like X-ray crystallography or cryo-EM across many cases, replication does not just confirm a single prediction, it retroactively validates computational modeling as a trustworthy method for inferring structure. In this sense, the resulting belief, that a given protein has a particular structure, can be treated as a scientific fact, not because it was directly observed, but because it was produced by a method shown to be reliable through repeated replication. Replication here does not justify each individual claim directly; it justifies the method, and by extension, makes belief in its outputs epistemically valid under a reliabilist framework. This shows how replication can support the reliability of an indirect strategy even though the method (replication) is not itself belief-forming in the traditional (reliabilist) sense. Thus, while replication does not directly produce first-order justified beliefs, it enables second-order justification by generating warranted beliefs about the reliability of the methods that do.

(ii) Internalist reading

On the internalist account of knowledge, replication supports justification by providing accessible epistemic reasons that can be evaluated by agents. When an experimental result has been replicated multiple times, it gives a scientist internal grounds to believe it is a true proposition, and crucially, to know that they know it, in line with the internalist’s emphasis on reflective justification. Importantly, replication may be performed by the agent or by others; both contribute to internal justification, though in different ways. Direct replication by the agent offers firsthand epistemic access, while replicated findings by others, especially when published or widely accepted, can also be internally accessed through trust in other’s reproducibility of the result. Scientists value both modes, treating replication as a shared epistemic resource that enables internal justification through either personal verification or reliable uptake of others’ findings. This distinction between firsthand and shared replication becomes especially clear in historical cases where justification shifted over time as more replications became available.

For example, Huay’s principle, which holds that each crystalline substance generally has a single characteristic form, may have been justified either by Curie’s own prior observations of crystalline substances or through Curie’s access to others’ replications (i.e., an internalist account via firsthand or shared replication). However, at the time of Curie’s inference about radium chloride (RaCl₂), Curie lacked direct evidence that the principle applied to this specific substance. From the perspective of Norton’s MTI, Curie’s inference was nonetheless justified externally because Curie knowingly or unknowingly relied on what would ultimately prove to be the correct BF: that RaCl₂ shared isomorphic properties with BaCl₂. On the contrary, under an internalist standard, Curie did not have sufficient justification at the time, as Curie lacked replicated evidence confirming that RaCl₂ was monoclinic. This confirmation came only later, when additional samples of RaCl₂ were examined and found to conform to Huay’s principle. Thus, the replication of Huay’s principle across RaCl₂ samples eventually stabilized the inference, transforming it from a provisional claim into a well-justified scientific fact.

A similar dynamic can be seen in the ether drift experiments. When Miller’s experiment produced positive results for the ether wind, Einstein hypothesized that temperature fluctuations had confounded the results. At the time, this was merely speculation, an unverified hypothesis based on general knowledge that temperature settings could affect experimental outcomes. From an internalist perspective, Einstein lacked sufficient justification for this inference, as Einstein had no direct empirical evidence that temperature fluctuations were responsible for Miller’s findings. From an externalist perspective, however, Einstein’s inference is considered justified under MTI because it relied on a BF that ultimately proved correct. Norton’s externalist view treats Einstein’s claim as justified at the time, because it relied on a BF that turned out to be true. The later replication by Shankland et al. confirms that temperature effects were indeed responsible, making the earlier justification retrospectively apparent, though not retrospectively created. It was only through this subsequent replication that Einstein’s inference was validated, confirming that temperature was a relevant BF and securing the reliability of the Michelson-Morley null result. In this way, the example illustrates how replication supports justification differently under internalist and externalist frameworks: Einstein’s claim was initially justified externally, but later became justified internally once replication revealed the truth of the BFs.

Any scientific claim, no matter how well epistemically supported by BFs, remains a single instance until replicated. Replication serves as the means by which BFs are tested, refined, and solidified into reliable epistemic tools. Without replication, we lack a systematic method for distinguishing stable inferences from those that remain speculative. While de Grefte argues that MTI implies a commitment to epistemic externalism, I suggest that once replication is taken seriously as an epistemic principle, MTI can be reinterpreted to accommodate internalist justification as well. This reinterpretation of MTI, centered on the epistemic role of replication, highlights the overlap between externalist and internalist accounts: replication tracks truth-conducive conditions in the world while also providing justification accessible to agents through transparent and repeatable methods.

(iii) Pragmatist approach

The role of replication in MTI can also be understood through a pragmatist framework of truth, one that builds on Norton’s own appeal to science-as-practiced. If replication is the central process by which BFs are stabilized, then the justification within MTI should be reinterpreted as also emerging from the sustained success of inquiry or practice itself, as exemplified by the case of metformin discussed above. This shift depends on how truth is conceived, whether as correspondence to antecedent facts or as something that emerges from the success of inquiry. Norton emphasizes that BFs must be true in order to justify induction, much like a deductive schema must be valid to license a conclusion:

“Facts have to be true for it to do the warranting. The best we can do is to assure ourselves with a great degree of confidence that it is so (true)... The warranting character of the facts is similar to the deductive inference schemas. In order for a deductive inference to be valid, you need a good schema. In order for an inductive inference to be cogent then the background facts has to be a truth…” (John Norton, Bar-Hillel Lecture, 17 Dec 2020, *Material Induction*, YouTube, 1:23:25)

For the pragmatist, "truth" emerges from, perhaps, what inquiry would converge on overtime (C.S. Peirce), or what proves itself through usefulness in experience (W. James) (Peirce 1878/1992; James 1907). On this view, *true* BFs are provisional and revisable, gaining epistemic weight through their sustained success in supporting inquiry across time and context. Replication, in this interpretation, is a core process through which facts earn their reliability and are treated as true for the purposes of guiding inference in practice. Accordingly, the externalist commitment to BFs as metaphysically prior in MTI becomes untenable. Furthermore, MTI no longer functions as a logic of induction in Norton’s sense, since it lacks the fixed warranting structure that, in Norton’s account, licenses inference independently of context. Instead, MTI becomes a descriptive account of how scientists focus on local, domain-specific inference and stabilize scientific facts without appeal to universal schemas.

This perspective can be further clarified by considering how a pragmatist might reinterpret justification within MTI. The pragmatist perspective reconceives justification in induction as something emerging from the ongoing success of inquiry. Here, Norton’s BFs can be seen as provisional, domain-stabilized forms of *background knowledge* whose justification depends on their sustained utility in *guiding* inquiry. On this view, BFs are historically contingent features of reasoning in practice, made durable through replication and continued use. As in MTI, schemas do not justify induction but are useful tools for inquiry. Curie’s inference on RaCl₂ relied on the prior success of Haüy’s principle within crystallographic practice. While the principle or schema did not justify the inference, it guided her reasoning, and justification accrued as the inference held up under repeated testing and continued use. For the pragmatist, such *background knowledge* supports inference by shaping what is taken to be a promising direction or reasonable hypothesis. Curie’s inference would not have been intelligible had Haüy’s principle not already proven useful in past inquiries. They constrain the directions inference (or inquiry) can take by drawing on what has previously worked, allowing scientists to generate plausible claims even without direct confirmation.

**6. Conclusions**

This paper demonstrates that replication is an essential epistemic principle that underpins the justification of scientific claims. Norton, in MTI, contends that induction relies on local BFs rather than a universal schema and argues that replication does not warrant induction. However, Norton’s perspective conflates the logical and epistemic dimensions of induction. Replication not only determines which BFs become stable but also governs how scientific inferences gain reliability over time. This undermines Norton’s claim that BFs alone justify inference independently of methodological features like replication. Instead, replication must be recognized as a process by which BFs, or background knowledge, earn their epistemic status, making it indispensable to scientific reasoning. Once replication is acknowledged as an epistemic principle, it reshapes how BFs justify inference, not just by tracking reality, but by showing stability accessible to scientific agents. Replication may ultimately serve not only to support justification but to act as one of its central foundations in scientific inference.

**Author statement**

The author is solely responsible for the conception, development, and writing of this article. All views expressed and all content presented are the author’s own.

**Acknowledgements**

I thank Nathan Hawkins for his kind and wonderful mentorship and assistance throughout the development of this paper. I am also grateful to Dilara Halim and Alejandro V. Rifo for their thoughtful input from the perspective of practicing scientists.

**References**

Ashrafi, K. (2007). Obesity and the regulation of fat metabolism. *WormBook*, 1–20.

Bailey, C. J. 2017. “Metformin: Historical Overview.” *Diabetologia* 60 (8): 1566–1576.

Brock, T. J., Browse, J., & Watts, J. L. 2006. “Genetic Regulation of Unsaturated Fatty Acid Composition in *C. elegans*.” *PLoS Genetics* 2 (7): e108. https://doi.org/10.1371/journal.pgen.0020108

Collins, H. M. 1985. *Changing Order: Replication and Induction in Scientific Practice*. London: Sage Publications.

de Grefte, Job. 2020. “Epistemic Benefits of the Material Theory of Induction.” *Studies in History and Philosophy of Science Part A* 84: 99–105.

Derksen, Maarten, and Jill G. Morawski. 2022. “Kinds of Replication: Examining the Meanings of ‘Conceptual Replication’ and ‘Direct Replication.’” Perspectives on Psychological Science 17 (5): 1490–1505.

Diabetes Care. 2018. *Diabetes Care* 41 (Suppl. 1).

Dillin, A. (2003). The specifics of small interfering RNA specificity. *Proceedings of the National Academy of Sciences of the United States of America*, 100(11), 6289–6291.

Gilst, M. R. V., Hadjivassiliou, H., Jolly, A., & Yamamoto, K. R. 2005. “Nuclear Hormone Receptor NHR‑49 Controls Fat Consumption and Fatty Acid Composition in *C. elegans*.” *PLoS Biology* 3 (2): e53. https://doi.org/10.1371/journal.pbio.0030053

Goodman, S. N., Fanelli, D., and Ioannidis, J. P. A. 2016. “What Does Research Reproducibility Mean?” *Science Translational Medicine* 8 (341): 341ps12.

James, William. 1907. *Pragmatism: A New Name for Some Old Ways of Thinking*. New York: Longmans, Green, and Co.

Kennedy, S., Wang, D., and Ruvkun, G. 2004. “A Conserved siRNA-Degrading RNase Negatively Regulates RNA Interference in *C. elegans*.” *Nature* 427 (6975): 645–649.

Kenyon, C., Chang, J., Gensch, E., Rudner, A., and Tabtiang, R. 1993. “A *C. elegans* Mutant That Lives Twice as Long as Wild Type.” *Nature* 366 (6454): 461–464.

Lambie, E. J., & Kimble, J. 1991. “Two Homologous Regulatory Genes, *lin-12* and *glp-1*, Have Overlapping Functions.” *Development* 112 (1): 231–240.

Libina, N., Berman, J. R., and Kenyon, C. 2003. “Tissue-Specific Activities of *C. elegans* DAF-16 in the Regulation of Lifespan.” *Cell* 115 (4): 489–502.

Mordes, D. A., Morrison, B. M., Ament, X. H., Cantrell, C., Mok, J., Eggan, P., et al. 2020. “Absence of Survival and Motor Deficits in 500 Repeat *C9ORF72* BAC Mice.” *Neuron* 108 (4): 775–783.e4.

Norton, John D. 2003a. “A Little Survey of Induction.” In *Scientific Evidence: Philosophical Theories and Applications*, edited by Peter Achinstein, 9–34. Baltimore: Johns Hopkins University Press.

Norton, John D. 2003b. “A Material Theory of Induction.” *Philosophy of Science* 70 (4): 647–670.

Norton, John D. 2010. “There Are No Universal Rules for Induction.” *Philosophy of Science* 77 (5): 765–777.

Norton, John D. 2014. “A Material Dissolution of the Problem of Induction.” *Synthese* 191 (4): 671–690.

Norton, John D. 2015. “Replicability of Experiment.” *Theoria* 30 (2): 229–248.

Norton, John D. 2021. *The Material Theory of Induction*. Calgary: University of Calgary Press.

Peirce, C. S. 1992. “How to Make Our Ideas Clear.” In *The Essential Peirce: Selected Philosophical Writings, Volume 1 (1867–1893)*, edited by Nathan Houser and Christian Kloesel, 124–141. Bloomington: Indiana University Press. Originally published 1878 in *Popular Science Monthly*, 12: 286–302.

Shenker, Orly. 2020. “John Norton Bar-Hillel Lecture – Material Induction.” YouTube video, December 17.

van der Valk, J., Bieback, K., Buta, C., Cochrane, B., Dirks, W. G., Fu, J., et al. 2018. “Fetal Bovine Serum (FBS): Past – Present – Future.” *ALTEX* 35 (1): 99–118.

Volpato, V., Smith, J., Sandor, C., Ried, J. S., Baud, A., Handel, A., et al. 2018. “Reproducibility of Molecular Phenotypes after Long-Term Differentiation to Human iPSC-Derived Neurons: A Multi-Site Omics Study.” *Stem Cell Reports* 11 (4): 897–911.

Zhang, C. S., Li, M., Ma, T., Zong, Y., Cui, J., Feng, J. W., Wu, Y. Q., Lin, S. Y., Han, J., & Lin, S. C. 2022. “Low-dose metformin targets the lysosomal AMPK pathway through PEN2.” *Nature* 603 (7903): 159–165.

1. Since then, metformin was shown to activate AMPK via inhibition of mitochondrial complex I, which decreases ATP production and raises the AMP/ATP ratio indirectly activating AMPK. However, this effect requires supra-physiological concentrations. More recently, Zhang et al. (2022) reported a lysosome-dependent mechanism, where metformin binds PEN2, recruiting the v-ATPase subunit ATP6AP1 to activate AMPK directly at clinically relevant doses. [↑](#footnote-ref-1)
2. For example, metformin’s use was justified by decades of replicable clinical results before its molecular target was identified. See Bailey (2017). [↑](#footnote-ref-2)
3. The “experimenter’s regress” (Collins 1985) exposes a circular epistemic problem: the validity of an experimental result is judged by whether it replicates prior results, but whether two experiments are considered the same depends on whether their results match. [↑](#footnote-ref-3)
4. I do not attempt to resolve the problem of induction here (as was discussed in de Grefte 2020), nor do I take replication to introduce a novel justificatory logic. [↑](#footnote-ref-4)