TECHNIUM
SOCIAL SCIENCES JOURNAL

Vol. 28, 2022

# A new decade for social changes

www.techniumscience.com

ISSN 2668-7798

Technium Social Sciences Journal
Vol. 28, 170-202, February, 2022
ISSN: 2668-7798
www.techniumscience.com

# The difficulty of making claims to knowledge in social science

**Stephen Gorard[1], Yiyi Tan[2]**

[1 2]Durham University Evidence Centre for Education

s.a.c.gorard@durham.ac.uk

**Abstract**. This paper considers three common types of claim to research knowledge, and the relative difficulty of making each type of claim in an empirically and logically justified manner. Before this, the paper looks at some more general issues often raised when discussing knowledge in social science, such as the nature of truth and justified belief, the existence of "isms" and paradigms treated like fashion accessories that one can adopt or not at will, and the intrinsic limitations of how we get to know about the "stuff" we might want to make research claims about. The idea of this early section is to remove some potential obstacles, before arguing that none of these issues is relevant to the rest of the paper about the nature of claims in their most generic form, independent of things like specific methods of data collection. The first type of claim we identify is a fully descriptive one that only summarises the data observed. This is the easiest and safest kind of claim, but even these might suffer from non-random errors and inaccuracies. However, their biggest limitation is their lack of any wider utility. The second kind of claim is a generally descriptive one that makes statements about unobserved data on the basis of a fully descriptive claim. Here we meet Hume's problem of induction. These claims have two parts, and the inductive part cannot seemingly be justified by logic, inferential statistics (whether Fisher or Neyman-Pearson style), Carnap's inductive probabilities, or even necessarily by Popper's falsification process. The third type is a causal claim, which we argue must also be a general claim. We develop a model, based on the work of Mill and Bradford-Hill, of what a plausible causal claim entails. But it still has all of the problems emerging from the first two types of claim, and adds a further problem created by our inability to assess causes directly. The paper concludes by suggesting how social science can proceed most safely in practice.

**Keywords**. Causal claims, Descriptive claims, Induction, Philosophy of science

## Introduction

This paper is about the empirical and logical foundations of claims to knowledge in social science. The paper starts by trying to clear some of the clutter involved in discussing claims to knowledge through a preliminary consideration of the role of "truth" (continued in later sections). It then introduces a typology of three kinds of claims to knowledge and the logical difficulties each of these face, before drawing out some proposed implications for the conduct and understanding of social science.

## What is truth?

Traditionally, it has been assumed that knowledge of the kind represented by social science claims has to be true, in order to be knowledge (Nagel 2014). If a claim is not true then it cannot be knowledge, by definition. For example, the tracking theory of knowledge, proposed by Nozick (1981) and others, suggests that we only know something if that something is factually true, we correctly believe it to be true, but that if it were not true then we would not believe it. However, if we remove the idea of factual truth, the best we can do is to imagine knowledge to be something for which we have a justifiable belief - as long as we would change that belief if new evidence or argument emerged to make it no longer justified.

Of course, there is such a thing as truth (and untruth) derived from logic (or maths), but this deductive kind of truth contains no new information. Information here is linked to uncertainty, and is measured in terms of how much uncertainty the information reduces. Outside maths and formal logic, truth that contains actual information is an ideal and we emerge only with best bets or justified beliefs.

For some commentators there is no such thing as truth, even as an ideal. For example, “realities are discursive; that is, there is no direct access to a reality 'outside' discourse” (Maclure 2003, p.180). Research here is merely the deconstruction of meaning rather than a search for the truth (or preference) or practicality (what works). However, it would be a category mistake to say that some social science research descriptions are not meant to be imagined as “true”, else why should we be concerned with them? Research offers no certainties (Chalmers 1999), but denying the possibility that there is any means of judging knowledge claims to be more or less true would make research a completely pointless activity (Gomm 2004).

Truth, according to Howe (1988), is a normative concept, like “good”. It is what works in practice, and for the present, because that is how we recognise its truth (see the later section on causation). Where research has been testable, and has practical consequences, a kind of evolutionary natural selection has led, over time, to this universality of approach. Research findings, and the models based on them, represent a simplified description of a real-world system that assist us with calculations and predictions. They do not represent complete truth, and are good and useful only in so far as they enable us to make good decisions or improve performance (West and Harrison 1997).

**Theory as fashion choice?**

At a recent research seminar, the speaker described asking a group of US and UK based researchers to help rate the quality of some pieces of research. The result was wide disagreement, split mostly along national lines. The speaker described the US readers as being mostly “positivist” and the UK readers as mostly “interpretivist”. For the speaker, this difference in avowed philosophy explained the divergence in ratings. The two groups were using different criteria to judge the security of research claims.

If this distinction were a valid one then research ratings based on the two sets of criteria would be incommensurable, and no overall review of the evidence on any topic could attempt to synthesise what is known. Some commentators have said exactly that. For example, Sale et al. (2002) claim that “The quantitative paradigm is based on positivism. Science is characterized by empirical research” (p.44). Whereas, “In contrast, the qualitative paradigm is based on… multiple realities. [There is] no external referent by which to compare claims of truth” (p.45). The idea of such paradigms is discussed further below.

For the present, it is clear that recognising the existence of genuine multiple perspectives for any evidence does not mean the end of truth as an ideal. We could, for example, view one research site in terms of its efficiency, economy, heating, and lighting etc. Each account so generated may be accurate (true), but they are also, quite properly, orthogonal to all of the others. We cannot, because of this, seriously assert that anything must be true.

The key point here is that weighting each study in terms of its quality or trustworthiness is a key part of assessing the overall picture of the research literature on any topic. We cannot simply ignore research because it is from a different perspective. That would lead to bias. But we cannot assemble the full body of evidence with appropriate weighting, if that weighting is substantially determined not by what the research was but which of the two avowed philosophies (above) was adopted by the reader. We could not compare purportedly different kinds of research directly.

The problem is actually much worse than this because the number of such avowed philosophies in research is much greater than these two. The list includes realist, relativist, and constructivist, among many others, plus variants created by adding prefixes such as "neo-" or "post". Thus, we could have neo-realists or post-positivists and many other examples.

If each of these many perspectives uses different criteria to judge the quality of any piece of research then we will not be able to have wide inter-rater reliability or agreement in judgements. This means we cannot synthesise all evidence properly, to benefit the public who pay for it and whose lives may be affected by its use.

Are all of these perspectives needed, and how should researchers select which one is "best"? And how much difference do they really make to how research is conducted and judged?

*The many "isms"*

Positivist accounts might draw a useful distinction between logical or mathematical reasoning, which uses already defined elements in a deductive process to generate new insights, and new knowledge generated by experience. These are what Hume (1962) termed the relations of ideas, which are actually tautological containing no new information and merely and restating assuming the premises on which they are based, and matters of fact, which are necessarily more tentative and error-prone, with generalisations from them based on induction. Whereas deduction works from the general to the more particular, induction works from particular observations or data points to a more general conclusion. There are, of course, things that are neither matters of facts or of reason, such as religions, theology, or perhaps introspection, and personal judgements of aesthetics, and so on. But positivism would traditionally not be concerned with them, if they were not open to empirical inspection. It is an Aristotelean rather than Platonic approach to focussing on what is deemed the safest kinds of knowledge.

As this paper explains in more detail later, there are problems both with the process of reasoning itself, and with making safe claims based on observations. Because such claims are linguistic (or at least symbolic), knowledge is intertwined with the meaning of any language or language term used to describe that knowledge (Quine 1960). To some extent at least, any reported observation (in its most general sense as a data point) can appear differently to different observers, with different language terms or prior experiences. We have no idea whether such observations really are different to each observer, just as we have no way of knowing whether two people experience the same colour when looking at something described as "blue" in colour.

Also, all research data collection is theory-laden rather than truly objective (Phillips 1999). When setting out to test a practical question, imagining the eventual argument structure on which a knowledge claim will be based helps to ascertain the form of the data it would require, and so helps the planning of research (Toulmin 1958). Even such an apparently basic operation as the measurement of a length involves acceptance of a series of theories about the nature of length and the isomorphic behaviour of numbers (Berka 1984). As with number and length, so also with many of our basic concepts and classifications as used in social science – such as sex, time, place, family, class or ethnicity. These concepts can be tremendously powerful and useful, but they remain theoretical and so should be susceptible to change. Theory, in this sense, is part of our ordering of all experience (Dubin 1978).

These widely accepted provisos have led some commentators to adopt a slightly modified position termed post-positivism. This is still based on logic and empiricism, but clarifies what should have been clear from the start – there is no one objective truth about any research claim. Truth is at least partly about meaning, and any observation depends partly on the perspective of the observer (Phillips 1992). There is no simple correspondence between research evidence and an external reality. This leads to the useful concept of relativity, of examining phenomena from different viewpoints in an attempt to provide a way of expressing any research findings so that it includes all of these viewpoints, and would therefore be understandable from each (Turner 2002). Perhaps most famously, Einstein (1920) produced theories of special and general relativity in physics which can demonstrate both the importance of observer standpoints, <u>and</u> how the phenomenon under investigation can be understood/resolved at a meta-level for all standpoints. There can be many true descriptions of a finite set of events - as long as they are consistent with each other (Boghossian 2007).

However, interpretivism has also appeared as an alternative idea or explanation of knowledge claims, antithetical to the more scientific approach of post-positivism. Like post-positivism it also does not suggest that objective truth is possible, because the language and prior experiences of researchers can affect their observations. But it implies that post-positivism assumes a correspondence between observation and reality, and/or that it does not permit subjectivity. Neither is true.

If interpretivism takes the subjectivity of experiences to a further extreme, and spurns relativity, then observers cannot see beyond their own perspectives, and so cannot validate or perhaps even communicate their version of knowledge to others (Scauso 2020). Similarly, constructivists and relativists who believe that research findings do not stem from some commonly glimpsed externality, but are created <u>solely</u> by the research process, cannot triangulate findings between methods or individual observers. For them, relativity is not possible. Each observation would be considered unique and transitory.

This approach has the same kind of problems as a discredited extreme form of neo-positivism based on the idea that entities only exist if and when they are measurable (see Cook and Payne 2002). The problem with this "worst kind" of relativism is that it is inherently contradictory, being itself based upon a universal claim about the truth of each perspective (Ramakers 2002, p.631). It is standard practice for relativists therefore to claim to know something about the nature of all truth claims which is not possible if their theory is correct (Winch and Gingell 1999). In its extreme form, something like relativism <u>cannot</u> be true. If the claims of relativism are objectively true then its claims are false (not everything is relative), and *vice versa* (Nagel 1997). Note that this problem does not arise in traditional ideas of knowledge/justified belief. Here, some things can be relativist and socially constructed, like morals or aesthetic judgements. They could even be considered post-modern, as in architecture. But other things, like the existence of mountains, are not socially constructed even though we may learn from others how to identify and name them. Otherwise, we are led to the absurd position of Latour (1998), who said of the idea that Ramses II had died from tuberculosis that it was impossible because tuberculosis had not yet been identified in the lifetime of Ramses. Latour claimed that it is equivalent to saying that Ramses had been killed by a machine gun that had not been invented yet. This extreme view means that mountains and stars could not have existed before humans, and so a constructivist is led to claim that the world has not existed very long, and that there were no such thing as dinosaurs, for example.

Relativism is therefore absurd, and false by its own account. There must be some facts (even if some are mistaken, see above). If there were no facts, but only statements of opinion according to a theory, then is that itself a fact. If it is a fact, and so there are facts, then relativism is false - and it would be hard to believe there are facts about opinions/beliefs but not about mountains. If the idea that statements are only believable in relation to a theory is not a fact, then there must be a second theory about why we accept the first theory, and so on *ad infinitum* (Boghossian 2007).

Anyway, there are many examples of social scientists who claim to be relativists, for example, while behaving with respect to the ideas of others as nothing of the sort (Postan 1971). If truth is actually unique to each observer's perspective then perspectives (like interview quotations) cannot be aggregated, and more general statements cannot be justified. Yet comparison and aggregation of views is exactly what purported relativists and interpretivists do in practice. Relativism is not the same as the more useful and simplifying idea of relativity, even though some commentators seem to conflate the two.

*The notion of "paradigms"*

These various "isms" are often defended by advocates as being paradigms – paradigms that researchers can apparently select at will, as though they were fashion items.

Kuhn's (1970) theory of paradigms suggests that groups of investigators tend to settle within a norm-referenced framework to try and solve closely defined 'puzzles'. A paradigm is a set of accepted rules and norms of reviewing within any field, for solving a scientific question that it is deemed possible to find a solution to in the near future. This distinguishes a paradigm from the many important and interesting questions that do not have an answer at any particular stage of progress or knowledge (Davis 1994), or for which the idea of an "answer" does not even make sense.

The normal science conducted within such a paradigm could also be simply passive and uncritical rather than genuinely cumulative in nature. It could be based on practices that differ from those stated (i.e. where there is deceit, either of the self or the audience). As Lakatos (1978) pointed out, maintenance of the *status quo* in any scientific endeavour 'is [often] achieved by censorship' (p.44). Normal science may therefore give an appearance of harmony, and of fitting together, because its practitioners conceal their actual methodological divergence in practice (Gephart 1988).

Such normal science frameworks are periodically disrupted to such an extent that there is a paradigm shift which eventually settles down to a new puzzle(s). The shift may have a variety of determinants, but common ones would be new evidence based on a new way of looking at the puzzle, a genuinely new idea, or just a change in general acceptance of existing critique. Often, a new paradigm emerges because a procedure or set of rules has been created for converting a more general query into a puzzle. The shifts from Newtonian physics to relativity to quantum physics are often cited as examples. Progress can be made both by working within a paradigm (the human genome project, for example), or by a paradigm shift (perhaps represented by Schrödinger's 1994, "What is life"?). Both can be invaluable.

However, this term "paradigm" has been misappropriated in the context of "isms", and is now commonly used to mean something very different in social science. It no longer refers to a communal effort in a field of research to solve a closely defined problem, which might undergo a radical shift on the basis of evidence. Instead it is used to defend a conservative approach to research, entailing a very limited set of specific research methods, not intended to be affected by contradictory evidence or any ideas of a different nature (Perlesz and Lindsay 2003). Worse, because of this link between purported paradigms and methods of data collection/analysis, new researchers are taught that using a specific research method means adopting an entire "paradigm". Most commonly this occurs with "paradigms" termed "qualitative" (interpretivist, concerned with meaning, often using interviews and observation) and "quantitative" (positivist, concerned mostly with pattern, based on measurements). Even worse, work conducted in these two traditions is then thought to be so different that it cannot be compared or mixed or used in complementary ways. For example, "Because the two paradigms do not study the same phenomena, quantitative and qualitative methods *cannot* be combined for cross-validation or triangulation purposes" (Sale et al. 2002, p.43).

This existence of these two supposed paradigms and their incommensurability is held predominantly and vocally by those espousing the "qualitative" one. The "quantitative" one exists largely as a counterpoint created to help explain the existence of the "qualitative" one. In fact though, numbers and narratives are routinely combined in real-life research, and meaning/experience is used in social science both to identify problems such as patterns of inequality, and to help understand them. Researchers who are not part of the "qualitative" silo are generally inclusive of types of evidence in the way that all of us actually are in real-life. When we act pragmatically in our non-research lives we do not usually invoke a paradigm as our starting point. In preparing a large formal party, for example, we might use documents, conversations, and numeric accounts together in an unproblematic way. We would not reject the advice of the caterer simply because it was expressed verbally, nor would we refuse to calculate the amount of food or drink required simply because that would involve numbers. To do either would be pointlessly inefficient whether we were planning a party or conducting research. When "we think about investigations carried out in the normal course of our daily lives, how often do measuring and counting turn out to be essential to our purposes' (Crotty 1998, p.15).

In fact, the terms qualitative and quantitative do not denote paradigms, they do not help to understand the research process, and their supposedly bespoke approaches are not even in tension (Gorard with Taylor 2004). They represent what Holmwood and Stewart (1991) would call a "non-productive" idea. Non-productive ideas start from a premise that social experience is confounding, contradictory and paradoxical, because that is how it often appears to new or naïve researchers. This tradition is therefore able to preserve its prior supposed theories for much longer, because even when its ideas are tested (which they rarely are) a theory is not deemed to fail when it is contradicted by experience (Sullivan 2001). Contradiction is simply assumed to be further confirmation for the initial idea(s) that the world is contradictory (the notion of "falsification" is discussed in a later section of the paper).

The idea of paradigms and the role of grand theory in the supposed qualitative paradigm have therefore become a kind of self-supporting religious faith, antithetical to the logic of research. At heart, research should be sceptical rather than dogmatic, and the research results based on that stance in other fields have been astounding (Kroto 2003). Such productive research has offered powerful resistance to authoritarian epistemologies, especially those of clericalism which promote the importance of doctrine over experience. Good research also has a long tradition of simplicity, in reporting, in theory and attempted explanations. This has helped the wider critique of evidence and ideas, to keep research grounded and of practical use - thereby avoiding the chance of a Sokal hoax, for example (Bricmont and Sokal 2001).

In practical fields such as education, housing, health, crime and so on, we are primarily concerned with substantial arguments and should therefore, according to Toulmin (1958), ground our claims in the practical context of each situation, rather than in the abstract principles that earlier philosophers and religious leaders wished to impose on us. However, it might be no exaggeration to say, in the twenty-first century, that the growth of research is still being retarded as it was in previous centuries by a kind of reactionary clericalism (Steele 2002).

*Resolving the schism?*

With so many "isms" apparently available to a new researcher, the situation can be portrayed as meaning that researchers can examine them all like a rack of clothes in a shop, and pick which they find most attractive. This would be theory as fashion choice. However, this is a false sense of choice.

It is perfectly possible for a philosopher to have radical views of the nature of knowledge that are logically consistent with our everyday observations - such as solipsism, or even a belief that everything that happens is random. But ethically, it makes less sense to hold any such views while conducting applied and publicly-funded research. If a "paradigm" were to privilege particular forms of evidence then this would lead to knowledge relativism in which the different parties cannot even argue coherently with each other, since each party can legitimately reject the very nature of the argument used by the other party (note that this is a very different situation to querying the quality of their argument or evidence). If all research were to lead only to the finding that the results depend on the prior perspective of the researcher, then it would cease to be funded and cease to be listened to by outsiders with any respect at all.

Some researchers invalidly but "evidently believe that the choice of a research method represents commitment to a certain kind of truth and the concomitant rejection of other kinds of truth" (Snow 2001, p.3). World views do not logically entail the use of specific methods (Guba 1990), but may only be thought to be so due to a common confusion between the logic of designing a study and the method of collecting data (de Vaus 2001, Geurts and Roosendaal 2001, Gorard 2013). "Research should be judged by the quality and soundness of its conception, implementation and description, not by the genre within which it is conducted" (Paul and Marfo 2001, pp. 543-545), and methods should be used "as a tool serving the questions pursued, rather than allowing them to constrict the range of inquiry" (Beyer 1992, p.62).

Analysis of the conduct of real-life research suggests that it is sometimes done better (i.e. more sceptically, more robustly, and more open to inspection) and sometimes not. But there are no systematic differences in the practical approach actually used that stem from any prior "paradigm" (just as with the formal party planning). The actual philosophy adopted by researchers makes no obvious difference to how they proceed (Rorty 1999), because "avowed philosophies" are not actually acted on in practice (Postan 1971, p.ix). In fact, the revealed difference between purported "philosophies" is often to do more with the topic choices of those advocating, for example, interpretivism (an interest solely or largely in human meaning and personal experiences), and their methods (favouring ethnography), than with any more substantial philosophical underpinning.

Hacking (1999, p.67) suggests that researchers ideas' like purported paradigms that make no discernible difference and are never tested are not part of the research process at all. They are irrelevant, and merely a voice for their users' own "rage against reason". There are many shared assumptions whatever methods are used (Denscombe 2002). A consideration of how social science research is actually done, rather than how commentators often claim it should be done, suggests that nearly all studies proceed in the same way – contextualised, value-attuned and largely consistent with post-positivism (Eisenhart and Towne 2003).

**A simple typology of research claims**

Based on the above points, and until it may become necessary, we therefore consider research claims in this paper at a meta-level, rather than specific to paradigms, isms or research methods. We start by proposing a simple typology of three distinct kinds of claims to knowledge, as used in social science (and elsewhere). These are:

- *Fully descriptive* claims based on a knowingly limited set of observations or data points, where the claims concern only those data points. A simple example might be reporting interview data from a number of participants, not intended to represent a wider population of possible participants.
- *Generally descriptive* claims are based on a limited set of observations, just like fully descriptive claims. But here the observations are intended to be used to make a more general claim to knowledge. A common example might be the attempt to generalise the findings from a sample of participants to a wider population of cases not participating in the research.
- *Causal claims* must, we think for reasons discussed later, also be general claims, based again on a limited set of observations. They add a further conceptual element to general claims, by claiming that some observations were created (or modified, influenced or impacted) by other observations. An example might be the proposition that gaining a particular educational qualification tends to lead to a higher-paid job.

We are concerned here primarily with claims based on empirical evidence. Purely deductive claims are also possible by themselves, but insofar as they are truly deductive they are tautological and so contain no new information (Shannon and Weaver 1949, and see below).

Each type of claim requires more than the one above it, not in terms of evidence but mostly in terms of the assumptions needed to justify them. Each type of claim is therefore increasingly hard to justify logically. However, they all have several aspects in common. All start with some observations (or data points of any kind). These observations are the “facts” which underpin each claim. But in social science these “facts” would not actually be factual, because these observations could be biased, mistaken, misunderstood, misrecorded or misreported. They are an attempt to portray something valid about the “stuff” makes up the world we are trying to research.

The typology is further illustrated in terms of two dimensions in Table 1. The rows represent whether claims are merely descriptive of existing data, or whether they go beyond the data available. The columns represent whether the claims are causal or not. The three relevant cells have a few examples of claims of the kinds that will be discussed in the paper.

Table 1 – A simple typology of evidence-based claims, with examples

| | Descriptive claims | Causal claims |
|---|---|---|
| Claims based solely on data | **Fully descriptive claims**<br>This is a raven<br>At least one raven is black<br>58% of respondents were employed<br>Most of the employed had degrees | **Internal causal claims** |

| | | |
|---|---|---|
| Claims going beyond the data | **Generally descriptive claims**<br>All/most ravens are black<br>No ravens are orange<br>On average, poorer children get lower school qualifications | **General causal claims**<br>Lack of education reduces the likelihood of employment<br>Using this programme increases average attainment at school |

The paper now discusses each of these claims further, and the problems that can be faced when making them. In doing so, it looks at the correspondence (or not) between research claims and an external reality (the "stuff").

**Fully descriptive claims**

Fully descriptive claims, such as those shown in Table 1, can be useful in social science. They are usually (even if disguised) in a format comparing the number of observations with a certain characteristic (nC) to the total number of observations made (n). They can be about unique events (e.g. this thing observed is a raven), about the relative number of cases or proportions (e.g. 58% of respondents in this survey reported being employed), and compared across groups, (e.g. in this survey more of the employed, than the unemployed, had degrees), or over space or time (e.g. this problem is getting worse).

For fully descriptive claims, the n matters. Claims are more substantive the larger n is. For example, "15% of 100 ravens are not black" is more substantive than "this thing is a raven", but less substantial than "15% of 1,000 ravens are not black". And like all empirically-founded claims, they are more trustworthy when the data collection is clear, independent, replicable and so on (Gorard 2021). All research is like a warranted argument. If the reported observations were not true how else can we explain their appearance? If we cannot find a better explanation then we might accept the descriptive claim for the time being – done properly this could provide a justified belief in the claim.

Aside from any inevitable errors in data collection, bias caused by missing data, and mistakes in analysis, such descriptions can be deemed "factual". They simply report what was observed (or believed to have been observed). They should do so fully, and transparently, so that readers can check the accuracy of every claim.

Such simple descriptions can help define an issue or problem, they can set the context for a more general study, and sometimes they can be powerful in their own right. Of course, a larger number of such studies could be conducted, and then combined to create a larger dataset. This does not always help, but it can do. For example, the term statistics derives from knowledge about the state. Political arithmetic is the simple descriptive portrayal of numeric "facts" about the state, such as levels of poverty, ill health, and infant mortality. This approach can lay bare a problem or the level of inequality, in a way that is hard for politicians and others to ignore. It can be invaluable, and has been so in the past. Nevertheless, it is just a start. Even with political arithmetic, readers are quickly moved to ask whether these figures are equally true everywhere, why they arise, and what should be done to ameliorate them. These more complex but interesting questions cannot be answered by mere description.

Therefore, perhaps the major problem with most fully descriptive claims is why anyone would want to make them. Simply describing the characteristics or experiences of a limited number of cases is not always or often useful. Readers would immediately want to know if these findings are special or permanent or true more widely. They would want to know the usefulness of the findings. Research is more than story-telling.

Technium Social Sciences Journal
Vol. 28, 170-202, February, 2022
ISSN: 2668-7798
www.techniumscience.com

But often, fully descriptive claims are more like journalism or novel-writing, reporting what happened, or who said what, only partially and with little rigour. This might occur because of the purported incommensurability of observer perspectives (see above). Such claims are of little greater interest for social science.

Locke (1979) thought that observation of evidence was intrinsically inferior, in its truth claims, to logic, and that testimony, or second hand observation, was worse again. One can try to minimise the problems with any observed data, and there is a wide literature on how to do so. It might help if the observations were automated, replicated, made by people who were unaware of the purpose of the research (blinded), made by people with no vested interest in the results, checked for the "reliability" of several observers, collected about the same phenomena in different ways, and so on. Failure to consider such assistance suggests that the observer is really only interested in what they think is going on, and not concerned that others are persuaded by the "truth" of their account (to have justified belief).

Nevertheless, whatever is tried, in any reasonably large set of observations there will be errors. A key point to note here is that people have no reason to assume that these errors will be "random" in nature (randomness is discussed further below). Bias, by definition, is not a chance occurrence. Research has long suggested that misrecording or misreporting data is not random, and tends to favour the prior beliefs of the researcher (Adair 1973). This means that researchers have no easy, or even technically complex way, of estimating the scale and impact of such errors, let alone of correcting them. Care and judgement are needed, but something like inferential statistics cannot help (see next section). There is no randomisation or probability to assess (Gorard 2021).

Even fully descriptive claims usually involve more than the observations themselves. There will be some kind of analysis as well. To continue the example above, maybe the researcher will report how many interview participants responded in a particular way, or whether participants with a specific characteristic responded more frequently in a particular way. The count of participants itself may be in error. Any form of analysis can be conducted wrongly in practice, or misapplied to the context. And the more complex it is the more likely it is to be in error, and the more any initial errors in the data will propagate. And just like errors in the original observations, any subsequent analytical errors will not be random in nature. A mistake in counting or in classification cannot be addressed or even identified by any process predicated on randomisation. These problems, and many more, will arise for any empirical claim. Care, simplicity (parsimony, see below), and transparent judgement are the main ways we can think of to deal with them.

**Generally descriptive claims**

Generally descriptive claims go beyond the data that they are based on, to make statements at least partly about data or cases that have not been observed. They might suggest patterns or rules about cases from which no evidence has been collected. A hypothetical example might be the claim that "all ravens are black" (or more realistically for social science perhaps, most ravens are black). In order to prove the statement that all ravens are black through observation we would need to see all ravens (and to know that we had seen all ravens). The numerator (number of black ravens) and the denominator (number of all ravens) must be exactly the same. Barring errors, this would be a fully descriptive (population) claim rather than a generalisation. This is not very realistic for many research purposes, where we would hardly ever know how many cases there were in a population nor whether we have really observed them all. As Hume (1962) and others have noted, without seeing all ravens we cannot prove that all ravens are black. This is the problem of induction.

By definition therefore, a generally descriptive claim must be based on observing fewer than all ravens (or whatever). This gives the claim two components - an empirical fully descriptive basis, and an inductive part. The empirical part is how many ravens (n) have been observed and how many of these had the characteristic C (or nC), of being black in this example. For the universal positive statement to be true, n must equal nC, otherwise it has been falsified.

The inductive part is the extension of the empirical report about some number of ravens all being black to the claim of "all" possible ravens being black. This part is not empirical, and cannot be empirical for any n less than the total number of ravens (or whatever). How can the inductive part of a general claim be justified, given that it cannot be empirical (by definition)?

*Inferential statistics*

One widespread approach used in attempting to justify the inductive part of a claim involves the use of inferential statistics (significance tests, standard errors and related constructs). However, using traditional inferential statistics to make more generalisable claims involves making several unrealistic assumptions (Gorard 2021). We can only use these techniques when the cases involved have been fully randomised – either through random selection from a known wider population, or by random allocation of a population to two or more groups. Neither situation applies to our examples, or to any of the other claims in Table 1. If we knew the wider population of ravens, or non-black things, for example, in order to be able to select a random sample, then we would often only need to count the population. No generalisation would usually be needed. If we do not know the wider population then we cannot randomise cases from it.

The kinds of probabilities involved in inferential statistics are anyway only those that might apply to an ideal game of chance. If we know that a six-sided die is unbiased then we can state that the probability of rolling a 2 in one trial is 1/6. We can say that the probability of getting two 2s in a row is 1/36 etc. Put another way, if we know everything relevant about a situation like this then we can easily calculate the probability of any specific set of occurrences. But this is never the case in real-life research. And the reverse is not possible. We cannot use a specific set of occurrences that we observe to tell us about everything else relevant (Hume 1962).

Anyway the whole approach makes no sense in attempting to justify the inductive part of claim. Imagine we were trying to assess the likelihood that all ravens were black. Observing just one non-black raven makes any statistical analysis redundant. So we must assume that all observations so far have been of black ravens. We already know that the claim "no ravens are black" is false, because we have observed at least some black ravens. Therefore we also know that some ravens are black. Again, no inferential statistics are needed. If we "test the hypothesis" that all ravens are black, then the p-value for however many black ravens have been observed will always be one (100%). Here the inferential statistics approach is useless.

If instead we want to test the idea that "most ravens are black" we would need to specify a precise figure for what "most" means in order to compute a p-value. The p-value we get will depend on the figure we choose to represent the notion of most. We know that if we "test" the claim that all but one of very many ravens were black, then the p-value of obtaining all black observations of ravens in a limited sample will be very high – almost as high as if we assumed all ravens were black. If all but one raven is black then any raven you spot is very likely to be black. And conversely the probability of observing a non-black raven is very low. Both

Technium Social Sciences Journal
Vol. 28, 170-202, February, 2022
ISSN: 2668-7798
www.techniumscience.com

probabilities depend on knowing exactly how many ravens there are from the outset. The p-value tells us nothing that our, otherwise arbitrary, assumption does not do already. We are no closer to knowing if indeed all but one raven is black, or more or less. The same kind of conclusion would be so, whatever the precise figure used for the initial assumption. The calculation is entirely tautological and yields no more information. It is just another, more technical and less accessible, way of restating the initial assumption.

An analyst could use something like the principle of indifference to decide on the key initial but arbitrary assumption for inferential statistics, about the likelihood of the initial hypothesis being true before collecting new evidence. For example, in the absence of any knowledge to the contrary they might assume that one of two possible outcomes was equally likely (50% or equipoised). The initial assumption might be about whether a specified hypothesis were true or not. However, this 50% likelihood will not be true in a real-life situation, even with no prior knowledge, and in most real-life situations there will be at least some prior evidence. Traditional inferential statistical analysis simply ignores such prior evidence and eschews any context in an anti-scientific way.

Carnap (1955) has anyway demonstrated that the principle of indifference is difficult to apply, and can be challenged. If we have no knowledge, the principle states, we assume that all outcomes are equally probable. But what does this really mean? For example, imagine a large bag of marbles of three colours (blue, red and yellow). We do not know how many marbles there are of each colour. If our assumption is that the first ball will be blue, then by the principle of indifference the hypotheses B that the ball will be blue and its inverse B’ that the ball will not blue are evens. There is then a 50% chance of a blue ball and a 50% chance of either red or yellow. On the other hand, a hypothesis that the first ball will either be blue or red means that this outcome is 50% likely and so there is a corresponding 50% chance of yellow. But no marbles have been selected yet. And making up a hypothesis cannot affect the number of marbles of each colour. Using the principle of indifference leads us to a contradiction.

Perhaps more importantly, the principle assumes that we know from the outset how many possible outcomes there are. We may not know this, even if we think we do. If there are more options in reality than are catered for in setting up initial equipoise probabilities then those probabilities will be wrong. So the principle of indifference does not work as intended.

Added to this is the problem that even if social science researchers knew the full population, they would rarely, if ever, have a set of fully randomised cases in real-life due to missing data – non-response, attrition and so on. And randomness is a mathematical <u>necessity</u> for undertaking the computations involved in inferential statistics. But the real killer blow is that even if a social scientist actually had an ideal set of randomised cases and knew everything necessary about the population, they could then only compute the probability of their specific set of observations occurring, given the initial assumptions about their general claim. Therefore, they still cannot use inferential statistics to <u>test</u> their general claim. They cannot use the statistical result to assess whether their claim is true, how likely it is to be true, or how likely it is to have arisen by chance (Gorard 2021).

Inferential statistics do not work to help establish general claims. Frequency statistics, comparisons, and modelling (not inferential statistics) can be used, of course, but only with fully descriptive claims – whether these fully descriptive claims are stand-alone, or form the basis for a more general or even a causal claim. However, none of these approaches can help to justify the inductive part of a general claim.

*Inductive probability*

Commentators have long concluded that probabilities of the kind represented by inferential statistics are not needed anyway (Jeffreys 1948). For them, everything needed for social science claims can be expressed in terms of what Carnap (1955) called inductive probabilities. This is perhaps what most early statisticians intended statistics to be, until Fisher and others devised what they envisaged as a more precise statistical probability approach.

Carnap (1955) and others have tried to create a more coherent way of computing the likelihood of an inductive statement being true, based on additional evidence emerging. This is different to wanting a larger n for a fully descriptive in order for the claim to be more interesting or taken more seriously – which is always a judgement, not a calculation. Instead, an inductive probability computation is based on the assumption that every new observation changes the likelihood of an inductive claim by a specific and calculable amount.

Imagine a general claim H that all observations of X have the characteristic Y, with pXY as the prior or unconditional probability of H being true. Imagine also that H is “resistant” to the complete generalisation that all X have characteristic Y, by a constant value lamda. Here lamda could entail a range of factors, but it is usually and mostly taken to represent the size of the population, and therefore how big n would have to be to have made a general claim fully descriptive in theory. However, it could also be represented as the amount of information that would be needed to reduce uncertainty about the claim (Shannon and Weaver 1949). This resistance to generalisation is key, because otherwise the posterior probability is too sensitive to new evidence, based on traditional Bayesian analyses (Gorard 2002a). According to Carnap’s inductive method, the posterior probability of X having characteristic Y in light of new evidence E (or pXY|E) is equal to:

$$\frac{\text{(the new number of observations with characteristic Y + lamda.pXY)}}{\text{(the total number of observations+lamda)}}$$

For example, if the prior probability of H were 0.5 (based on the principle of indifference, above) then the formula would be:

$$\frac{(nY+lamda/2)}{(n+lamda)}$$

In this example, if lamda were 0 then the inductive probability of H would be the number of observations of X with characteristic Y over the number of observations of X (or nY/n). This would be the simple proportion of black ravens, for example. Using a lamda of 0 (assuming no resistance to generalisation) would be what is termed the frequentist approach in statistics – to ignore the resistance of the claim to generalisation, and always to assume a prior probability of exactly 0.5 (above). New evidence thus completely over-rides any prior knowledge in a clearly unjustified way.

As lambda rises, the inductive probability of H tends towards pXY (the prior or unconditional probability of all X also being Y). With high lamda the impact of each new observation is less, but never zero. For example, if pXY is 0.5 and lamda is 1,000, then the probability of H after one new observation of XY is 501/1,001, or around 0.5005. This is only just a bit bigger than the unconditional probability, without new evidence, of 0.5.

Technium Social Sciences Journal
Vol. 28, 170-202, February, 2022
ISSN: 2668-7798
www.techniumscience.com

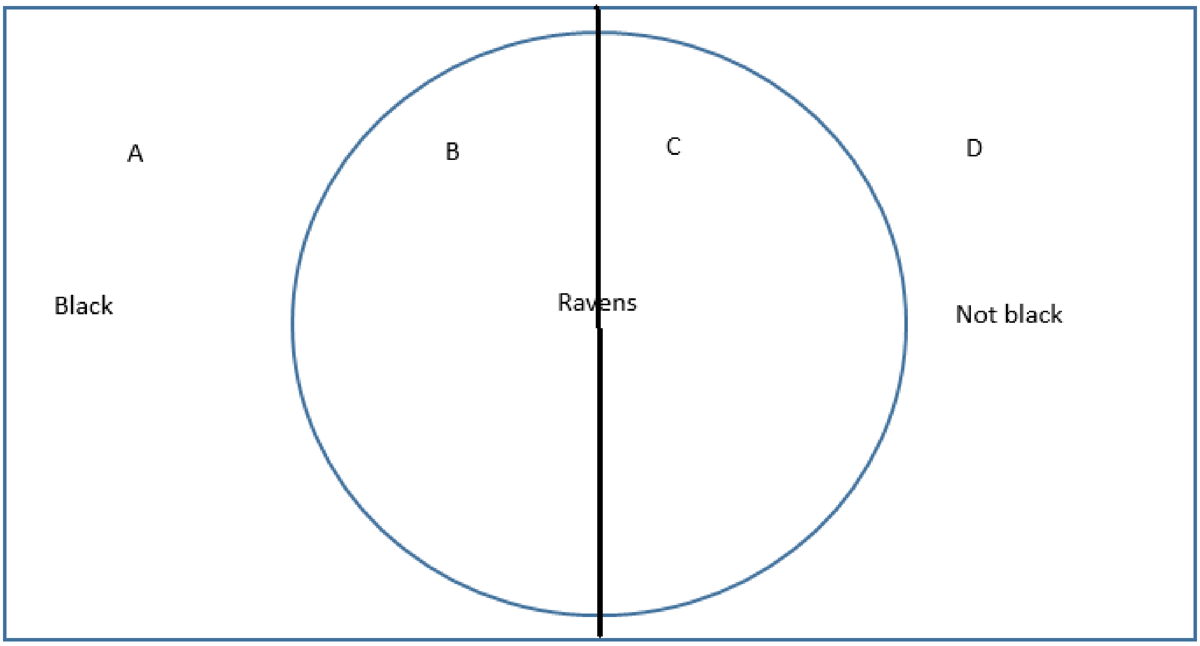
If pXY is 0.1 at the outset, and lamda is 10, then one observation of X with characteristic Y changes the probability of H after this new evidence to around 0.182 (2/11). A further observation of the same kind would change pXY|E to 0.256 (2.82/11). After a sequence of 10 observations of X with characteristic Y, then the probability of H after this new evidence would be 0.55 (11/20).

*Hempel's paradox*

Before considering the value and validity of such inductive probabilities any further, we discuss what is required of new evidence for a universal general claim by looking at the ramifications and suggested resolutions of what has been termed Hempel's paradox. The key issue is to determine how large n and nC would be in Carnap''s formula above. We have so far assumed that nC is simply the number of black ravens, for example. But the equivalence criterion (below) suggests that something like nC could be much larger in reality, and this would make a substantial difference to the calculation of any inductive probability.

Figure 1 represents the example of a generally descriptive claim that "all ravens are black". It is a universal set with four marked subsets. A is everything that is black but not a raven (b.(not r)). B is everything that is black and a raven (b.r). C is everything that is a raven and not black ((not b).r). And D is everything that is not black and not a raven ((not b).(not r)). Then A, B, C and D can be used to refer to these subsets (for any claim).

Figure 1 – Venn diagram illustrating a general claim




If true, the statement "all ravens are black" would mean that C must be empty (there are no non-black ravens). The statement has no immediately obvious implications for A (other things can also be black) or for D (other things can still be not black, even if all ravens are black). It is more ambiguous whether the statement has any necessary implications for B. If there are ravens, then they must be black and so included in subset B. But on one interpretation, there could be no ravens, and B could be empty. Here the statement is interpreted more precisely as being “if there are such things as ravens then they would all be black”.

It seems that observing (or gathering evidence of) a case in each of these four subsets would have different implications for the statement “all ravens are black”. Observing a black non-raven (A) appears to make no difference to any other subset. It seemingly neither confirms nor denies the statement. Observing a black raven (B) would seem to provide some evidence or confirmation for the general claim, but makes no difference otherwise. Observing a non-black raven (C) would falsify the general statement (all ravens are black). What difference does observing a non-black non-raven (D) make? At first sight, to take a standard example, observing a green apple appears to be irrelevant to the claim that “all ravens are black”.

However, if we consider the statement that “all ravens are black” (or if something is a raven it is black) more closely, this is logically equivalent to the statement “if something is not black it is not a raven”. Whenever the first statement is true then the second is also true, and *vice versa*. If the second statement is false then the first is also false, and *vice versa*. We consider observation of a black raven to be evidence for the first statement, and therefore it must also be evidence of the second statement. But it then follows that observation of a green apple (for example) which is evidence for the second statement is equally also evidence for the first statement. Observing a green apple (or whatever) provides at least <u>some</u> evidence that all ravens are indeed black.

This is what has been termed Hempel’s Raven Paradox (Hempel 1945a, 1945b). In logic, apparently, observing anything that is not a raven and not black provides the same kind of evidence for the statement that “all ravens are black” as the observation of a black raven. This sound incorrect, because we would not normally agree that seeing a green apple told us anything about the colour of ravens. How can we resolve this?

Maher (1999) phrases this problem in terms of three plausible sounding principles. The first has been termed Nicod’s condition (Nicod 1930). Given no other context, observing an object X that has characteristic Y is treated as confirmatory evidence that all X are also Y. The second principle, termed the equivalence condition, is that where evidence confirms a claim, then the same evidence also confirms any proposition that is logically the same as that claim. The final plausible principle (Nicod’s criterion) is that, given no other context, observing something that is neither X nor Y provides no evidence on whether all X are also Y. Green apples are, at first sight, deemed irrelevant to whether all ravens are black.

It is clear that these principles lead to a contradiction, and cannot all be true. The first two together imply that green apples are some kind of evidence for the claim that all ravens are black. The last one denies this. Which, if any, of them are right or wrong?

*A “Bayesian” solution?*

In real-life, as researched by social scientists, we might assume that subset D is bigger than subset A. There are, in effect, more non-black things than black things in the world. Both A and D will be bigger than both B and C. Only a small fraction of the things in the world are ravens. And we might also assume that B is bigger than C. Even if not all ravens are black, many patently are. So in terms of scale: D>A>B>C. Theis relative frequency of cases in each part of Figure 1 could help to resolve the paradox.

For the moment, we are only concerned with subsets B and D. An observation in A apparently makes no difference to whether C is empty (but see below), and an observation in C settles that C is not empty. Only an observation in B or D can alter our belief that all of subset B.C is in B, but without settling the matter. In traditional philosophy of knowledge only the observation of ravens would be allowed to either confirm or falsify the statement that "all ravens are black". The observation of other things of any colour are deemed to make little or no difference (Nicod's criterion, above).

We might assume that the number of ravens in the world, while perhaps large, remains finite. So, it could be argued that observing one more black raven takes us one step closer to proving that *all* ravens are black (where nC/n is closer to 1). However, the number of things that are not black and not ravens is much larger than the number of ravens. In this case, observing a green apple might indeed have an implication for the truth of the statement about ravens but the increase in evidence would be correspondingly much smaller than that provided by observing a black raven (Good 1960). This coincides with common sense, and seems to resolve the apparent paradox. A thing that is not a raven provides some small confirmation by not being a raven of the kind/colour that would falsify the claim that all ravens are black. Or more simply, it reduces the number of things that, before observation, could have been ravens, and so the number of ravens that could have been not black. It therefore slightly increases our confidence that all ravens are black, but not by as much as observing an actual black raven. This argument could be a called a "Bayesian solution", based on the relative estimated frequency of items in each subset.

The argument relies on relative frequencies, probabilities or similar numeric values. But Hintikka (1970) suggested that the solution does not lie in relative frequencies. They considered the general claim that "all men are tall", which they said was logically the same as "all short people are women" (or more strictly all short people are not men). Allowing an individual's sex to be binary in this way for the purposes of Hintikka's illustration, the equivalence condition means that observing a short non-man provides confirmation that all men are tall. This example, Hintikka claims, is not like "all ravens are black" where we assume that there are more black things than ravens, and this relative frequency affects our understanding of the paradox. In the absence of clear knowledge that there are many fewer men than short people, we still find it hard to accept that observing a short non-man is evidence that all men are tall. The solution to the raven paradox is <u>not</u> therefore about the relative frequency of observing each term in the claim.

However, the number of "things" that are not ravens and not black might reasonably be so large as to be infinite, so that counting them would be a "super task". If so, then observing one more of these would make no (or little difference) to the statement that all ravens are black (*Swinburne 1971)*. If so, <u>Nicod's criterion stands, the paradox remains, and Bayesian solutions do not work.</u> Observing a non-black non-raven cannot help confirm than all ravens are black. The equivalence criterion appears false. How can that be?

*Resolving the "paradox" otherwise*

A lot of debate about the paradox centres on the relevance of having additional background knowledge to the claim or not. Imagine we knew that we could be in one of two possible universes. In one there would be a million birds overall, including 100 ravens all of which are black. In the other universe there would be a million birds and 1,000 ravens of which 999 were black. Observing a black raven is much more likely in the second universe. Therefore observing a black raven would, oddly, be evidence of living in a universe where not all ravens were black. According to this example by Good (1967), maybe it is Nicod's condition that is wrong. The paradox only seems solid in the absence of further knowledge. Perhaps evidence does not affect our inductive probability as we imagine, or maybe not at all. See also Maher (1999), who rejects the whole idea of a Bayesian resolution based on relative frequencies.

Other examples have been suggested to make a similar point about Nicod's condition. For example, if we make the claim that "all humans are less than 3 metres tall" because we have never observed anyone much more than 2 metres tall, then observing someone 2.9 metres tall may make us more likely to reject our initial claim. Here, observing someone 2.9 metres might make the claim that "all humans are less than 3 metres tall" weaker even though the new observation ostensibly supports it (i.e. 2.9 is less than 3).

Another way of resolving the apparent paradox might be to imagine that there was no such thing as ravens – that subsets B and C are both empty. It is then not clear that the second step (the equivalence condition) makes sense. It would then say, anything that is not black is not also something that does not exist. This is surely the same as the tautology that anything that exists is not something that does not exist. However, this attempt at resolution is not useful for two reasons. We do have evidence that ravens exist. And anyway ravens are merely meant to be a scientifically neutral example of any object, event, or thing that we are interested in. As Gorard and Tan (2021) argued, social scientists are ethically committed to investigating things that actually exist.

Strawson (1952) suggests instead that any proposition of the format "all X are also Y", assumes from the outset that there is such a thing as X (and presumably Y). This would make the resolution above impossible. However, it may suggest a different solution. If "all X are also Y" presupposes that X exist then it cannot be logically and completely the same as "all non-Y things are not X" which does not explicitly assume that X exists. If the two statements really have this different implication, then again the equivalence criterion does not apply. Testing the two would presumably involve different actions in research terms.

In addition, perhaps new evidence affects our knowledge (or estimates) of the number of things in sub-sets B and D differently. If the evidence suggests that "all ravens are black" this tells us about the colour of all ravens. It therefore adds to our estimate of the number of black things in the world, but does not necessarily mean that there are any more ravens than we envisaged before accepting that all ravens are black.

There are several variations of this idea that the two propositions deemed equivalent in the equivalence condition are actually subtly different. Observing a black raven would falsify the claim that "no ravens are black" (a universal negative claim), in the same way that observing a non-black raven would falsify the universal affirmative claim that "all ravens are black". However, observing a non-black non-raven does not falsify the claim that "no ravens are black". Therefore observing a black raven might provide crucially different information to observing a non-black non-raven (Scheffler and Goodman 1972). This, in turn, would make the equivalence condition false.

The equivalence condition can also lead to a seeming contradiction. Consider the claims that "all ravens are white" and "all ravens are black". They are contradictory. Yet the equivalence condition states that observing a non-black non-raven like a green apple provides support for the claim that "all ravens are black". By the same equivalence a green apple also provides support for the contradictory claim that "all ravens are white" because a green apple is a non-white non-raven. The equivalence condition takes us into strange territory.

*Returning to inductive logic*

As noted so far, there are both philosophical and practical problems to face in using the inductive probability approach outlined previously. It is not clear exactly what the resistance constant lamda is in any context, and indeed whether it is anything more than the size of the relevant population. It is also not clear what the prior probability should be (the same issue arises for all so-called "Bayesian" models). Carnap's formula never allows for all cases to be observed, because it takes no account of the population size, meaning that if all ravens have been observed and all were black then the probability of all ravens being black would still remain less than 1. It also seems that n must always equal nC. It is impossible for nC to be greater than n (there cannot be more black ravens than there are ravens). But it is also impossible for nC to be less than n, because the observation of even one non-black raven immediately negates or falsifies the claim that all ravens are black. The issue then ceases to be one about probability or likelihood (if it ever was).

Inductive probability anyway usually concerns not exact frequencies but judgements about the strength of support for a general claim in light of existing and new evidence. Or about the relative strength of two claims on the same issue. It could be part of a psychological process of helping to decide on a "best bet" on the basis of the existing evidence (Gorard 2002b), rather than a precise calculation about the true underlying state of the world.

However, the biggest obstacle to the kind of inductive probability proposed by Carnap and others is how it fails to overcome longstanding objections to the "logic" of induction (Hume 1962). We are dealing with a general claim that has two elements – the empirical part (nC/n), and an inductive part that goes beyond the evidence to state that all future observations will be the same as all past observations. Each new piece of evidence E which is relevant to our claim H, must alter the probability of H given E so that it is greater than the unconditional probability of H. It means that $p(H|E)>p(H)$. But the empirical part of H will be directly confirmed by E, while it is not clear that the inductive part is changed at all, or if so how it is changed. Popper (1992) argued that the part of the claim that is empirically-founded is supported by any new evidence purely <u>deductively</u>. It is clear that observing a new black raven increases the number of black ravens that have been observed. This is a tautology because it is deductive. Support for or confirmation of the claim is not inductive; nor is it probabilistic. It relates only to that sub-part of a general claim that is actually fully descriptive. Meanwhile, according to Popper, the part of H that is not deductively supported by new evidence remains unchanged.

Popper and Miller (1983) explain the problem slightly differently. The hypothesis H is the same as (H OR E) AND (H OR NOT E), by expansion. The first part is the bit that follows deductively from the evidence E. The second part is the bit of H that goes beyond E (the inductive part). It is the bit that new evidence can never confirm. Evidence that provides deductive support for a hypothesis cannot support the inductive part of a hypothesis. And statistical probability cannot handle induction, and is irrelevant to it. In fact then, there is no role for probability of any kind. The first part of the claim is handled deductively/descriptively, and the second part remains a problem for the logic of all social scientific claims.

*Claims like - most ravens are black*

Another relatively simple attempted resolution to Hempel's paradox might be to consider that the claim "all ravens are black" is unrealistic. A weaker alternative claim would be "most ravens are black" or, as Gaifman (1979) puts it, "nearly all ravens are black". Now the paradox is not as clear. "Nearly all ravens are black" is not the strict equivalent of "nearly all non-black things are not ravens". Not in the same way that all ravens are black is the same as all non-black things are not ravens. If we substitute "nearly all" for "all" in the phrasing of the paradox, the paradox disappears, or at least it weakens enough to allow a loophole.

However, in almost all other respects a move from universal claims like “all ravens are black” to considering more particular claims like “most ravens are black” does not help with induction. We would still need either to observe all ravens and note that more than half were black, or we would need to know exactly how many ravens there were, and to observe as many as needed (more than half) to know that most were black. Both of these would lead to fully descriptive claims. To make a general claim about “most” is no easier in practice or principle than to make a claim about “all”.

Note that this is very different from fully descriptive claims such as “some ravens are black” or more accurately “at least one raven is black”. These might require only one observation in order to be fully proven. This is an important distinction, even though Aristotle and others might classify both claims involving “most” and “some” as being particular in nature. Rather than say that claims based on “most” are really universal in nature, even though they are very like them, we prefer to express them as generally descriptive claims rather than fully descriptive ones.

In real-life research, we are in a situation analogous to having a very large bag of marbles, of which we do not know who many there are, what colours they might be, or how many there would be of each colour. If we shake the bag and blindly remove one marble at random and it is blue, what does this tell us about the remaining marbles in the bag? Not a great deal it seems. It does not help to decide whether any, most, or all of the other marbles are blue (or not). Therefore, it cannot help us estimate either the frequentist or inductive probability of the remaining marbles being blue (or not). Nor does it really help to have withdrawn several other marbles previously. For any large number of marbles, removing one more offers the same lack of knowledge about the remaining ones as when we withdrew the first one (Gorard 2021).

We can build up a collection of observed marbles, by taking more out of the bag. This is like the empirical part of any claim. The more observations we have the stronger any pattern (or perhaps lack of pattern) will be in the observed marbles. But these fully descriptive observations are necessarily silent on the colour of the marbles remaining in the very large bag. Speculation about the remaining contents of the bag is like the inductive part of a general claim. It is not based on evidence – marbles outside the bag do <u>not</u> tell us about the contents of the bag. Nor is speculation deductive in any way, so we cannot use a *modus tollens* form of argument to help, for example.

If the probability of the remaining marbles being blue (or whatever) depended on how many have already been observed, compared to how many there are altogether, as it is made to in inductive probability calculations, then we would need to know more about the population than we generally do know in social science. If the equivalence condition is valid, then the denominator in any real-life situation is infinite (or so large that it does not matter if it is infinite). If the nominator is the number of black ravens observed so far, then the denominator is far greater than the number of possible ravens. It would include also the number of all non-black things. If the denominator is infinite then each new observation cannot affect the proportion of objects seen.

*The problem of induction in generally descriptive claims*

If one wants to make a more general claim to knowledge – taking a Popperian example (see below) – that all swans are white, then observing one white swan is not enough to sustain the claim. Even 100 or one million observations are not enough. Replicating observations in this way does not seem to help establish the general claim (as with the bag of marbles). All one can truly do, even after observing one million white swans, is to make the fully descriptive claim that one million swans are white. And, as above, even this claim is not absolutely clear, because of the possibility of misclassifying, misrecording and so on. Any more general claim would have to be tentative unless or until one has seen and judged the colour of all swans. And how would one know, in practice, that they had seen all swans?

In the same way as in Popper's example, all of our claims based on research data are limited (even where they have been replicated and peer-reviewed). They must be seen as tentative. With induction however, the replication of our results is not actually that important. One can see many white swans without the claim being true, and a research "finding" can be replicated many times and still be wrong. Hume (1962) introduced this "skeleton in the cupboard of philosophy" - that the process of inductive reasoning has no logical foundation. Yet induction has often been used as the chief criterion of demarcation between what is considered "science" and what is not.

Popper (2002) suggested a way around this, by highlighting the notion of falsification. This kind of testability, he said, was the true difference between science and all else. One cannot, for example, conclude with logical certainty that all swans are white merely from repeated observation of white swans (induction). But one can falsify the claim that all swans are white by just one observation of a non-white swan. Thus, progress comes chiefly from falsifying theories not from further confirmation of them. This is an attractive idea. But is it true that Popper's falsification evades the use of induction?

In formal logic, the statements "A entails B" and "Here is an A which is not B" form a contradiction. Neither can be said to falsify the other because one would not know which, if any, of the statements was true. One only knows that both cannot be true. There is no logical justification for saying that the example of "A which is not B" means that "A entails B" is false. Since A and B are ideal terms we do not attempt to tinker with them and overcome the contradiction. Contradiction is not the same as falsification.

The idea of falsification arises from the fact that these nouns and adjectives are not logical entities. They are names for real-world things, and in that real world there is bias, misclassifying and so on. In the real world, where A and B become swan and white, one can at least consider the possibility that only one of the propositions is falsified by the contradiction. This is what Popper does without making this step explicit. He then states that it is clear which proposition is wrong - so clear that the alternative is usually dismissed as merely playing with words (Thouless 1974). But perhaps this supposed clarity is, like induction, actually only a habit of mind.

In the example, Popper proposes that we change the definition of swan to include the possibility that some swans are not white, and does not even bother to argue against the alternative. Nevertheless, the other way out of the contradiction is equally logical. We could change the definition of black to exclude the possibility of being applied to swans. Thus, the thing that looks like a black swan is actually not a swan because it is black. The choice is between changing our definition of swan or of black. In this example, people prefer changing the definition of the least familiar term, and black is a much more familiar term than swan. If the same is true in every example of falsification then what seems like a logical argument for falsification is actually an appeal to the same non-logical phenomenon of familiarity that underlies induction. When observation leads us to question a belief because it brings two beliefs into contradiction people tend to stick with the most familiar of the two concepts. This suggests that Popper's notion of falsification does not actually eliminate inductive logic at all (see also Goodman 1973). Familiarity breeds certainty in a way that appears logically unjustified.

*Technium Social Sciences Journal*
*Vol. 28, 170-202, February, 2022*
*ISSN: 2668-7798*
*www.techniumscience.com*

Take another example of a claim – that all doors are rectangular in shape. Many doors are rectangular, partly because people can control their shape. But some are not – perhaps in igloos, or the International Space Station. There will also be cases of different shapes that we are not sure are doors. The problem, as with the "bias" in falsification, lies in the use of words. Words are not like the categorical, algebraic or logical variables A, B etc. Our name categories, like door or swan, are imposed on things that could really be continuously variable in nature. One might, in theory, line up everything in the world, in order of "doorness", and somewhere there would be examples that are genuinely hard to classify (maybe some kinds of windows). We all know this. As Russell (1903) and others have shown, putting things into clearly delineated sets may not work. So, the induction problem is not one of knowledge *per se* but stems mostly from language and the use of categories.

Note that the potential problems with falsification as a process and as a criterion demarcating science and non-science, are not the same as the deliberate vagueness and incoherence adopted by some "researchers". Referring to the theories of Freud or Marx, Hollis (1994, p.72) comments that "these theories were awash with confirming evidence but for the unsatisfactory reason that their adherents could square them with whatever happened". Rather than specifying in advance the conditions under which a theory would be deemed to be false, adherents of big theories often defend their position in advance by arguing against logic itself. For example, Maclure (2003, p.134) treats poststucturalism as just such a big theory. She says, "by 'theory' I have in mind that loose collection of continentally influenced approaches with literary and/or psychoanalytic leanings that often go under the names of poststructuralism, postmodernism, deconstruction and discourse analysis". She defines it as follows - 'Perhaps the nearest one could get to a common characteristic of poststructuralism would be a radical suspicion of reason, order and certainty' (p.180). Therefore, this is a theory that can be defended against contrary evidence because it rejects the very notion of logic on which contradiction is based by conflating reasonable doubts about certainty in social science with doubts about reason itself.

*Summary of descriptive claims*

"Generally descriptive" claims have all of the same problems as fully descriptive claims, but they are also problematic in making statements about observations that have not been made (and that might never be made). They require something like induction. Although describing a sample in fully descriptive work has some technical problems, it is a relatively simple process in terms of logic. However, going beyond the sample to make assumptions about cases not in the sample has a much shakier logical foundation.

Again, inferential statistics cannot help. If one already knew how many marbles there were of each colour in the bag, then one could easily compute the probability of obtaining any combination of colours when sampling from the bag. However, without already knowing what is in the bag, revealing a subset of marbles does not permit one even to say what the probability was of those marbles being revealed – let alone compute the probability of the next marble from the bag being revealed as blue (or any other colour). The sample of marbles does not reveal what colour the other marbles in the bag are. It says nothing about the remaining contents of the bag.

To apply the analogy to research, if we already know what is in the bag then we do not need to do the research to find out. And if we do not know what is in the bag then we can only find out by taking out every single marble. It may seem therefore, that generally descriptive claims can be made about populations, where the entire population is observed (or otherwise has data collected about it). In a sense, this is true. But we prefer to envisage this as being an extreme example of fully descriptive work. There is no generalisation beyond the cases actually observed. In any case, social science population data will nearly always be incomplete due to non-response, dropout, or simply missing values about some existing cases. And the problems of miscounting etc. apply to population data as much as they do to anything else.

**The nature and habit of causal claims**

The final type, causal claims, are even more problematic than mere descriptions. Causes and effects are ideas used to describe a firm impression that people have about the way the world works. Events and processes have a regularity and time sequence that offer both an explanation for why things occur, and even a way of controlling them. Social science, perhaps more than other fields, is pervaded by what Abbot (1998, p.149) called 'an unthinking causalism", which appears to be worsening over time (Robinson et al. 2007). Correlations, patterns and even just perceptions and opinions are routinely presented by researchers in very definitive causal terms. We all need to be clearer about what it means to make the strong claim that something causes something else.

Hume (1962) described cause and effect as an immutable habit of mind – people are pre-disposed to see regularities in their environment and ascribe something like causation to them. This may have been a valuable evolutionary heuristic when time was short and a quick decision was needed. But it can lead to mistakes and superstition in the longer term. Across his different writings Hume seemed to be somewhat ambivalent about causation (Coventry 2008). On the one hand, as a "matter of fact", all that one has to support the existence of causation is the observed regularities of nature. One cannot use Hume's "relation of ideas" to deduce causation logically from any such available facts or regularities. But Hume also suggested that causal claims are, and must be, testable propositions about knowledge.

Causes cannot be deduced just from observing effects (Blalock 1964). Seeing a light bulb going off does not, by itself, allow the observer to deduce whether it has been switched off, or there is a power failure, or the bulb is broken, for example (Salmon 1998). Similarly, effects can rarely, if ever, be deduced simply by observing their possible causes. Who would have thought before experiencing it that striking a flint could create a fire, for example? How could we tell what an unknown switch might turn on?

Potentially, causal models are also very complex. Any event could be the effect of a large number of contributing causes. All of these causes might be needed to create the effect, but be insufficient in isolation. All causes might work only within a given context, or only in combination (Emmet 1984). A fire needs oxygen, flammable material, and ignition (a flame). One can say that the flame causes the fire, but it does not do so alone, and a variety of causes could be sufficient to create the effect, with none of them strictly necessary. One might start a fire with a lighter, a match, a flint, or a magnifying glass. Also, any cause or combination of causes could have more than one effect. Starting a fire causes combustion of the flammable material, but it also causes heat and light, among other things.

These issues are all problematic for Hume's idea of cause and effect as having a constant conjunction. If C is caused by both A and B in combination, then the correlation between A and C in isolation may be zero. The same thing arises for B and C in isolation. We may therefore be unable to predict exactly what the effects of a set of causes might be, because of the complexity of their interaction. Instead, one might predict their effects in probabilistic terms, or after controlling for everything else. An example of a *ceteris paribus* causal model could be the erosion of a river bank caused by a meandering river (Corbi and Prades 2000). There is no doubt that the river bank will erode over time even though it is not possible to be precise about the exact pattern it will form. This is reasonable, but makes it hard to test any causal model in practice.

There have been many attempts, since Hume, to describe the elements needed to establish a strong causal claim. For Mill (1882) a cause has three key elements. It should be clearly related to the effect (correlated through observation), it must precede the effect, and there must be no plausible alternative explanations for the effect other than the cause.

The first of these elements, the association of the putative cause and its effect, is certainly a *sine qua non*. Commentators might say that a correlation is not the same as causation, but not having a correlation between two things surely means that neither is the cause of the other. So one can test a causal claim by falsification, to the extent that one can assess a correlation as part of a fully descriptive claim (see above).

Of course, assessing such a correlation may not be easy in practice. In some of the natural sciences one might clone cells, or find identical particles. Hume considered billiard balls, which are also similar to each other, and may be envisaged as interchangeable. In social science, however, one cannot usually expose the same people or organisations both to a research process and not. This means that the results of causal research in social science is not generally clear-cut. People might use statistical approaches to express the nature of causal models, and this may lead others to imagine them as being probabilistic (Goldthorpe 2001). But actually, they reflect the limitation of our understanding, and not necessarily the reality of the world (Shafer 1996).

Viewing causation as a stable association between two phenomena, as Mill and Hume do, creates several problems. It is clearly wrong to suggest that a singular event cannot have a cause or causes, but there can be no repeated association between singular events – such as the onset of the Second World War. In a sense, all events are singular in terms of time, place, context, and the actors involved. Mill's criteria are best understood as describing how one can identify causes, and are not necessarily characteristics of all cause:effect sequences. Where one can observe or repeat very similar situations, such as striking a billiard ball in Hume's account, it is much easier to test a proposed causal model than when faced with a complex causal question about a one-off process, such as what caused the outbreak of the Second World War.

In both classical and operant conditioning, it has been shown that the association of two things leads the conditioned subject to behave in the presence of one thing as though it implied the presence of the other (Skinner 1971). Skinner's pigeons "learnt" to pull a lever which had always accompanied the release of a pellet of food in the past. The conditioned subjects do this whether the lever is mechanically releasing the pellet or not – for a time at least. To an observer, the pigeons seem to behave as though the lever is a cause. In intermittent reinforcement schedules, where the pellet appears on only some occasions, this behaviour is even stronger – it will take more examples of no pellet after pulling the lever to un-condition the subject than it would if the pellet had previously always appeared.

Further, Skinner's accidental reinforcement schedule is a powerful reminder of the dangers of allowing causal models to be based only on association. In accidental reinforcement schedules, providing pellets at random tends to reinforce whatever behaviour the subject was involved in at the time. That behaviour is then more likely to be repeated by the subject, and so more likely to coincide with the next random arrival of a pellet. It is a kind of confirmation bias. The more the pigeons perform the ritual the more likely it is that food will randomly follow one of the performances. This continuously reinforces the ritual. Eventually, the subject repeats an endless superstitious ritual of one behaviour, only intermittently reinforced by the arrival of a

Technium Social Sciences Journal
Vol. 28, 170-202, February, 2022
ISSN: 2668-7798
www.techniumscience.com

pellet, so making the apparent association resistant to un-conditioning. The response becomes self-fulfilling.

These findings suggest that the kind of imagined probabilistic causation, commonly reported in social science, will paradoxically be an even stronger habit of mind than Hume's constant conjunction idea. And this is so, even though it is actually more likely to be an erroneous association than a constant conjunction would be, partly because of the complexity of deciding whether a purported cause that is only sometimes "effective" is actually a cause at all. And partly because it may be accidental (a superstition). Our task as researchers is to identify ad avoid such superstitions as far as possible.

Mill's second element is also problematic. It is not necessarily true that a cause must precede an effect. The two can be contemporaneous. Some observations which are seemingly in a temporal sequence may actually be reciprocal (Hagenaars 1990). One can accept causes simultaneous with their effects, such as where a ball rests on a cushion, and the cushion is causing the ball not to drop further (Mackie 1974). If we drop two balls into a bowl, we can model the final resting places of both balls mathematically, but we cannot use this to decide which ball is "causing" the other to be displaced from the centre of the bowl. The events are surely mutually determined (Garrison 1993). Mathematical statements or systems of equations can describe such systems but they cannot express either intention or causality. They can be used to show that systems are, or are not, in equilibrium, and to predict the actual change in the value of one variable if another variable is changed. However, it is important to recall that this prediction works both ways. If $y=f(x)$ then there will be a complementary function such that $x=f'(y)$. Which variable is the dependent one (on the left-hand, predicted side) is purely arbitrary. Nothing in mathematics, logic, or statistical analysis can overcome this limitation.

In fact, all one can say, with some conviction, is that our present models do not permit a reverse sequence of causation. The effect cannot come before its ultimate cause. Student attainment at age 16 cannot cause their attainment at age 11, in any real sense (but see Gorard 2013).

Mill's third element is the need for an explanation. It is correct that such an explanation must be the simplest and most plausible. A causal explanation describes a process that shows how the cause could generate the effect. A good explanation must be easy to test, and must make the fewest assumptions necessary to provide a mechanism linking cause and effect. The proposed effect must be capable of change, and it must be capable of being changed by the proposed cause (de Vaus 2001).

A good example is the clear relationship between smoking and lung cancer. The statistical conjunctions and the observations from laboratory trials with animals were explained by the isolation of carcinogens in the smoke, and the pathological evidence from diseased lungs. These combined to create an explanatory theory.

However, it is not clear that an explanation is essential to a causal claim. It is possible to switch a light on and off without understanding how it works. The fact that it does work is part of what shows that the switch is the cause of the light going on and off. This suggests that the explanatory mechanism is the least important part of any causal model. If it is clear that altering the presence or strength of a potential cause works in the sense of changing an effect, then it matters less if the mechanism is fully understood or not. And, of course, even the most convincing explanation possible is of little consequence if the potential cause has no discernible effect in practice.

Bradford-Hill (1966), and others working on the links between smoking and lung cancer, proposed a somewhat tougher set of scientific conditions than Mill, for the identification of a causal link. Some of these are clarifications of Mill's elements, establishing rules for how and when Mill's elements will have been established. For example, the first element, correlation, must be found in different studies, led by different researchers, using different methods and differing cases. This additional specification is good practice, but is not a philosophical component of a causal model. In addition, Bradford Hill tried to address the lack of constant association in some contexts, by saying that the frequency of association between the cause and effect must be substantial compared to the frequency of either X or Y in isolation. They no longer have to be constantly conjoined. This is an appropriate rule, but it does make the identification of causes harder (as with intermittent reinforcement).

Mill's sequence element is divided into two parts by Bradford-Hill. So, a cause must be able to predict the effect (as discussed above). And the cause must come before the effect (but see limitations of this idea for contemporaneous events, above).

The third element is again the requirement for a coherent, plausible, workable mechanism explaining how the cause can influence the effect. But it should also be widely "agreed" and "consistent with prior knowledge". Again, these additional requirements sound sensible for practice (or in a legal case), but again they do not form part of the actual logic of causes. Something could be correctly identified as a cause by only one person, or incorrectly identified by many. And that identification might create a scientific revolution that is not consistent with what was assumed to be prior "knowledge". Also missing from the Bradford-Hill account is the "elimination of sensible alternative explanations". This elimination can be through robust testing of all possible explanations, or it can be based on an argument such as that the best explanation is the most parsimonious one. Plausibility alone is not enough for a theory.

Bradford-Hill adds a fourth element in two parts. There must be a reduction in the effect after removing or reducing the cause. And there must be an increase in the cause after the introduction of, or increase in, the cause. This is useful. It adds a requirement that deliberate variation in the appearance or strength of the cause must yield a change in the effect. Put another way, one must not consider a causal model established unless it has been robustly tested (though an experimental design or similar) on several occasions. However, it still assumes the idea of a constant conjunction, and that the effect has only one cause. Neither is necessarily true.

A key point is not whether one can explain why a cause has an effect, but whether it can be demonstrated to have an effect at all. Causation can best be viewed via the impact of an intervention. Does the proposed causal model work in practice, under controlled and rigorously evaluated conditions? Since causes are not susceptible to direct observation, but what they cause is effects, then at least those effects must be observable (like a light coming on, when the switch is pressed). We need evidence that controlled interventions have altered the level or presence of the potential cause, and so produced changes in the purported effect that cannot be explained in any other way.

Gorard (2002, 2013) reformulated all of these elements into a simplified model of causal evidence for social science, consisting of four main criteria. These criteria do not require *constant* conjunction. They allow cause and effect to appear at the same time (but not with the cause after the effect). They include the need for intervention studies. And they insist that the explanation must be warranted by the full body of evidence available. Note that causes will exist even if they are not known about. These criteria concern what we would need to know in order to state that a causal model exists.

For X (a possible cause) and Y (a possible effect) to be in a causal relationship:

1. X and Y must be repeatedly associated (correlated). This association must be strong and clearly observable. It must be replicable, and it must be specific to X and Y.

2. X and Y must proceed in a suitable sequence. X must always precede or appear with Y (where both appear), and the appearance of Y must be safely predictable from the appearance of X.

3. It must have been repeatedly demonstrated that an intervention to change the strength or appearance of X clearly changes the strength or appearance of Y. Of course X may not be the true proximal cause of Y, and may only have an impact on Y via an intermediary process or event.

And possibly:

4. It helps to have a coherent mechanism to explain the causal link. This mechanism must be the simplest available without which the evidence cannot be explained. Put another way, if the proposed mechanism were not true then there must be no simpler or equally simple way of explaining the evidence for it. However, this criterion is unlike the other three. The former are relatively objective or descriptive "facts" about the world, while the latter is more about our understanding of those facts.

In this formulation, a causal model cannot be for a specific event because of the need for replication of the impact of an intervention, and the repeated observations needed to establish any correlation. A causal claim cannot be only descriptive of existing data because it must be replicable. This means that Table 1 really only has three cells that are suitable for further consideration in terms of social scientific claims (the fourth is shaded out). Because of the emphasis on replication, and the need for repeated association and interventions, a causal claim can be taken more seriously when both the number of such repetitions and the number of total attempts are large. The numbers of attempts and successes matter, and this is independent of the types of data being collected as evidence.

*Summary of causal claims*

Causation is not something that can be observed directly. It is not even, like the unobserved cases in a general claim, something that could ever be observed. Instead, it is concept that we use to try and explain regularity in findings. Again, it should not really need to be stated that one cannot demonstrate or prove the existence of a causal model underlying observations through any technical or probabilistic means, such as inferential statistics.

Causation is always only ever a hypothetical construct, but it is one that underpins all social science research. It is of course possible to envisage a world without causation, but that would be a world in which research was pointless. It is also possible to envisage a world in which there was sometimes causation and sometimes not, but that would be unparsimonious.

There can be no direct evidence that observations are either caused, or somehow just random events that might seem patterned, in the same way as there are sequences in a table of random numbers (Arjas 2001). Either explanation fits the facts. So using either as an explanation for observed phenomena involves making an assumption not contained in any data. To use both to explain observations involves making *two* assumptions, and is therefore un-parsimonious. It is hard enough to establish whether causes exist or not. To allow them to exist alongside unrelated phenomena would make most social scientific propositions completely un-testable (Gorard 2013).

We therefore assume that social science is interested in causation, and that causation will be built on something like the four principles above (especially repeated interventions), which have emerged from centuries of philosophical, legal and practical considerations.

**Conclusions**

In this paper, we have been considering claims to knowledge. The paper considers three different claims to knowledge, namely, "fully descriptive", "generally descriptive" and causal claims. These are all common in social science, and each type of claim requires more assumptions than the previous one. After discussing their methodological and logical foundations, this paper describes some of the limitations in the nature of these three claims.

Fully descriptive claims can suffer from non-random errors and inaccuracies in observations, and analysis. Otherwise, fully descriptive claims are based solely on evidence, and so are the easiest to justify. They can have powerful impacts, but in practice are often more trivial in nature.

Generalised claims stem from fully descriptive ones, and so have the same issues. In addition, generally descriptive claims can also be questioned because of the long-standing problem of induction. Even the notion of falsification might not be able to help with this.

Finally, causal claims are the most problematic of the three. While widely assumed, causation cannot be observed directly. Making more general claims, including causal ones, involves going beyond the available evidence and so such claims are harder to justify. We argue that causal claims must also be general ones in social science, and so the paper focuses most on universal general claims. These claims have two components, an empirical part and an inductive part.

Neither part can be addressed through inferential statistics. None of the problems identified in this paper relating to each kind of knowledge claim is probabilistic in nature. From the early discussion about errors in collecting and assembling data to the lack of a logical foundation for causal claims, the kind of uncertainty introduced in all knowledge claims cannot be addressed by significance tests (or confidence intervals, and related paraphernalia).

Nor can we use the seemingly more promising approach of Carnap's inductive probabilities, once the paper has discussed possible resolutions of Hempel's paradox about general claims. There is no obvious Bayesian solution based on relative frequencies. The problem is not really about probabilities at all, and increasing the available evidence any further can only confirm the first component of a general or causal claim.

According to Russell (1996 pp. 649-700), Hume "proved that pure empiricism is not a sufficient basis for science" nor, by implication, for social science. But Russell also conceded that if we just overlook this problem enough to allow the principle of induction to hold, then scientific research makes sense. We can still make (tentative) general and causal claims while also suggesting that our claims are based on evidence. However, this is a big "if". Russell admits that it is hard to formulate a valid argument why this one concession to non-empirical non-deductive faith should be made, but all others rejected.

Nevertheless, it is consistent for anyone conducting social science research to assume that causation exists, else why would they bother to do the research? It is the same argument that could be made about the implications of solipsism. Researchers are logically and ethically required to accept that causation is a real possibility. If we genuinely reject the idea of causation then research, and trying to improve social conditions, become pointless. Research can only make sense if researchers can make a difference (i.e. have an effect). However, our preference for causation (invisible and without form) is still no more than a kind of religious belief.

Popper (2002) famously suggested that the problem of induction is solved by a focus on falsification. All claims are tentative. The best tentative claims are those most easily exposed to being falsified. They should be unambiguously expressed, parsimonious, and yield testable propositions. This testability, and the possibility of being surprised by a new finding, are at the heart of good social scientific claims. We can make progress in social science, as in many other fields, by testing claims to "destruction" to see if they survive. The lack of logical or empirical foundation for induction behoves all researchers to be sceptical of all claims – their own and others.

Theory helps us to decide what and how we research. It helps us to measure and explain. It can be crucial in the transfer of research findings to new settings, and an important end-product of research. Above all, theories are cheap. They allow us to consider alternative positions simultaneously, to make progress in parallel, and to accept failure without great cost. A theory is a tentative explanation, used for as long as it usefully explains or predicts real-world events, not an end in itself. As soon as theory itself becomes an obstacle to what or how research is conducted then it becomes worse than useless. Above all, theories must be subject to testing and then discarded whenever they consistently do not match up to empirical observation. Theories will always be under-determined by the evidence on which they rest, since any finite number of observations can be explained by a potentially infinite number of theories.

It is clear that theory or explanation is the least important element of a causal claim (or indeed any claim). If two items are unrelated to each other then neither can be the cause of the other. If the apparent effect appears before the apparent cause then the causal claim is considered wrong. If varying the cause never produces a change in the apparent effect then the causal claim is wrong. But the explanation is not about what is true. It is about people's understanding of what is true. Just as a cause can exist without anyone noticing it, so a cause can exist even if no one understands how it works. The fact that it works (or at least appears to work) is enough. Theory is often playing catch up, in trying to explain new findings, as well as sometimes generating new ideas to test.

Rather than worrying about rather petty distinctions between things like constructivism and social constructs, given that no one is suggesting that we have direct experience of an objective reality, we should be more concerned with finding better ways of describing what we do experience (Rorty 1999). This would help avoid a waste of research time and effort that could be devoted to useful reform and improving popular justice in the real world inhabited by the people who actually fund the research (Howe 1995).

The simplest explanation of any observation(s) is the best, not because it is proven or more likely, but because it is easier to test than any more complex ones. Adding needless elements to the explanation makes it confusing. Explanations must be trimmed down to the minimum needed, and this also makes them easier to test. Maybe this is what is happening with the preference for changing the definition of swan rather than black, in the verbal form of a syllogism. Maybe it seems simpler to test the colour of swans than to test what things colours should or should not be attached to. This needs for simplicity leads to a warranting principle (Gorard 2002). Before drawing a conclusion, we need see whether the observations (data/evidence) can be explained at least as well by any simpler conclusion.

*Technium Social Sciences Journal*
*Vol. 28, 170-202, February, 2022*
*ISSN: 2668-7798*
*www.techniumscience.com*

*Care in making/establishing claims*

Given the rather unsatisfactory nature of all knowledge claims, we suggest that the best way to help avoid being misled lies in care and good judgement. All claims are clearly contingent. A claim that is unfalsifiable is useless and may be damaging. Transparency of data and analysis help. It can help others understand how the data was assembled and its flaws, and to judge whether the research is the best bet for them to act on for the present.

Even logical claims are suspect when the elements in an argument refer to real-life things, and are not solely symbolic. Ignoring the implications of Gödel's incompleteness demonstrations (Smullyan 1992), logical deductions are generally considered to be necessarily true (and 2+2 must equal 4, by definition). If we assume that all swans are black then it must be true that none are white. This would be logical. And must be necessarily true within the context of the assumption. But these are tautologies, and they contain no information of the kind that would reduce uncertainty as defined in Information Theory by Shannon and Weaver (1949). However, in real-life research we may miscount to five, or misclassify a swan etc. Since our data is always tentative, we can never be sure of any claim or finding, even though the analysis might be perfect.

There is no point at all in trying to establish whether an invalid or untrue claim is then more generally true. It has not even been established as a valid fully descriptive claim. Generalisation is something that is only relevant once a secure claim has been established. As discussed, there is no way of knowing for certain whether a secure finding is also true of other cases not involved in the research (whether, indeed, <u>all</u> swans are white). Attempted falsification, in imaginative ways, might help but as we have illustrated there are problems even in that. And, as above, statistical analysis must be silent on this.

Moving from merely descriptive claims to causal ones introduces more barriers to security. This means that the move in research from one to the other should start only with the securest claims and ensure that a similar regard to the safety of findings is applied to every stage.

In addition to adhering to elements of rigour such as the design or scale of any research, there will always remain an important place for judgement and clarity in interpreting any set of results.

**Acknowledgements**

The authors would like to thank Nadia Siddiqui and Jonathan Gorard for acting as sounding boards for parts of this paper.

**References**

[1] Abbot, A. (1998) The causal devolution, *Sociological Methods and Research*, 27, 2, 148-181
[2] Arjas, E. (2001) Causal analysis and statistics: a social sciences perspective, *European Sociological Review*, 17, 1, 59-64
[3] Berka, K. (1983) *Measurement: its concepts, theories and problems*, London: Reidel
[4] Beyer, J. (1992) Researchers are not cats, 65-72 in Frost, P. and Stablein, R. (Eds.) *Doing exemplary research*, London: Sage
[5] Blalock, H. (1964) *Causal inferences in nonexperimental research*, Chapel Hill: University of North Carolina Press
[6] Boghossian, P. (2007) *Fear of knowledge*, Oxford: Oxford University Press

[7] Bradford-Hill, A. (1966) The environment and disease: Association or causation?, *Proceedings of the Royal Society of Medicine,* 58, 285
[8] Bricmont, J. and Sokal, A. (2001) Science and sociology of science: beyond war and peace, in Labinger, J. and Collins, H. (Eds.) *The one culture?*, Chicago: University of Chicago Press
[9] Carnap, R. (1955) *Statistical and inductive probability*, carnapstat.pdf (cmu.edu)
[10] Chalmers, A. (1999) *What is this thing called science?*, Milton Keynes: Open University Press
[11] Cole, S. (1994) Why doesn't sociology make progress like the natural sciences?, *Sociological Forum,* 9, 2, 133-154
[12] Cook, T. and Payne, M. (2002) Objecting to the objections to using random assignment in educational research, in Mosteller, F. and Boruch, R. (Eds.) *Evidence matters: randomized trials in education research*, Washington: Brookings Institution
[13] Corbi, J. and Prades, J. (2000) *Minds, causes, and mechanisms*, Oxford: Blackwell
[14] Coventry, A. (2008) *Hume's theory of causation*, London: Continuum
[15] Crotty, M. (1998) *The foundations of social research*, Thousand Oaks: Sage
[16] Davis, J. (1994) What's wrong with sociology?, *Sociological Forum*, 9, 2, 179-197
[17] de Vaus, D. (2001) *Research design in social research*, London: Sage
[18] Denscombe, M. (2002) *Ground rules for good research*, Buckingham: Open University Press
[19] Dubin, R. (1978) *Theory building*, New York: Macmillan Press
*[20] Einstein, A. (1920) Relativity: The special and general theory, Henry Holt and Company.*
[21] Eisenhart, M. and Towne. L. (2003) Contestation and change in national policy on "scientifically based" education research, *Educational Researcher*, 32, 7, 31-38
[22] Emmet, D. (1984) *The effectiveness of causes*, London:Macmillan Press
[23] Gaifman, H. (1979) Subjective probability, natural predicates and Hempel's Ravens, *Erkenntnis*, 14 (2): 105–147. doi:10.1007/BF00196729
[24] Garrison, R. (1993) Mises and his methods, pp.102-117 in Herbener, J. (Ed.) *The meaning of Ludwig von Mises: contributions in economics, sociology, epistemology, and political philosophy*, Boston: Kluwer Academic Publishers
[25] Gephart, R. (1988) *Ethnostatistics: Qualitative foundations for quantitative research*, London: Sage
[26] Geurts, P. and Roosendaal, H. (2001) Estimating the direction of innovative change based on theory and mixed methods, *Quality and Quantity*, 35, 407-427
[27] Goldthorpe, J. (2001) Causation, statistics, and sociology, *European Sociological Review*, 17, 1, 1-20
[28] Gomm, R. (2004) *Social research methodology: a critical introduction*, Basingstoke: Palgrave Macmillan
*[29] Good, I. (1960) The paradox of confirmation, The British Journal for the Philosophy of Science, 11 ,42, 145–149*
[30] Good, I. (1967) The white shoe is a red herring, *British Journal for the Philosophy of Science*, 17, 4, 322, doi:10.1093/bjps/17.4.322
[31] Goodman, N. (1973) *Fact, fiction and forecast*, New York: Bobs-Merrill
[32] Gorard, S. (2002a) Can we overcome the methodological schism?: four models for combining qualitative and quantitative evidence, *Research Papers in Education*, 17, 4, 345-361

[33] Gorard, S. (2002b) The role of causal models in education as a social science, *Evaluation and Research in Education*, 16, 1, 51-65
[34] Gorard, S. (2002c) Fostering scepticism: the importance of warranting claims, *Evaluation and Research in Education*, 16, 3, 136-149
[35] Gorard, S. (2013) *Research Design: Robust approaches for the social sciences*, London: SAGE
[36] Gorard, S. (2021) *How to make sense of statistics: Everything you need to know about using numbers in social science*, London: SAGE
[37] Gorard, S., with Taylor, C. (2004) *Combining methods in educational and social research*, London: Open University Press
[38] *Gorard, S. and Tan. Y. (2021) The logic of social scientific claims, Social Sciences Journal, 24,1, 113-124*
[39] Guba, E. (1990) The alternative paradigm dialog, pp.17-27 in Guba, E. (Ed.) *The paradigm dialog*, London: Sage
[40] Hacking, I. (1999) *The social construction of what?*, London: Harvard University Press
[41] Hagenaars, J. (1990) *Categorical Longitudinal Data: Log-linear, panel, trend and cohort analysis*, London: Sage
*[42] Hempel, C. (1945a) Studies in the logic of confirmation I, Mind, 54 (13), 1–26, doi:10.1093/mind/LIV.213.1*
[43] *Hempel, C. (1945b) Studies in the logic of confirmation II, Mind, 54 (214), 97–121, doi:10.1093/mind/LIV.214.97*
[44] Hintikka, J. (1970) Inductive independence and the paradoxes of confirmation, in Rescher, N. (ed.) *Essays in honor of Carl G. Hempel: a tribute on the occasion of his sixty-fifth birthday*, Synthese Library, Dordrecht
[45] Hollis, M. (1994) *The philosophy of social science*, Cambridge: Cambridge University Press
[46] Holmwood, J. and Stewart, A. (1991) *Explanation and social theory*, London: Macmillan
[47] Howe, K. (1988) Against the quantitative-qualitative incompatibility thesis, *Educational Researcher*, 17, 10-16
[48] Howe, K. (1995) Two dogmas of educational research, *Educational Researcher*, 14, 10-18
[49] Hume, D. (1962) *On human nature and the understanding*, MacMillan
[50] Jeffreys, H. (1948) *Theory of Probability*, 2nd ed.,The Clarendon Press, Oxford
[51] Kroto, H. (2003) Chemists should remove scientific inventions, *Times Higher Educational Supplement*, 18/4/03, p.13
[52] Kuhn, T. (1970) *The structure of scientific revolutions*, Chicago: University of Chicago Press
[53] Lakatos, I. (1978) *The methodology of scientific research programmes*, Cambridge: Cambridge University Press
[54] Latour, B. (1998) Ramses II est-il mort de la tuberculose?, *La Recherche*, 307, 84-85
[55] Locke, J. (1979) *An Essay Concerning Human Understanding*, New York: Oxford University Press
[56] Mackie, J. (1974) *The cement of the universe*, Oxford: Clarendon Press
[57] MacLure, M. (2003) *Discourse in Educational and Social Research*, Buckingham: Open University Press

[58] Maher, P. (1999) Inductive Logic and the Ravens Paradox, *Philosophy of Science*, 66, 1, 50–70, doi:10.1086/392676
[59] Mill, J. (1882) *A System Of Logic, Ratiocinative And Inductive, being a connected view of the principles of evidence, and the methods of scientific investigation* (8th Edition), New York: Harper & Brothers
[60] Nagel, T. (1997) *The last word*, Oxford: Oxford University Press
[61] Nagel, J. (2014) *Knowledge*, Oxford Press
[62] Nicod, J. (1930) *Foundations of geometry and induction*, Routledge
[63] Nozick, R. (1981) *Philosophical Explanations*, Cambridge, MA: Harvard University Press
[64] Paul, J. and Marfo, K. (2001) Preparation of educational researchers in philosophical foundations of inquiry, *Review of Educational Research*, 71, 4, 525-547
[65] Perlesz, A. and Lindsay, J. (2003) Methodological triangulation in researching families: making sense of dissonant data, *International Journal of Social Research Methodology*, 6, 1, 25-40
[66] Phillips, D. (1992) *The social scientist's bestiary*, Oxford: Pergamon Press
[67] Phillips, D. (1999) How to play the game: a Popperian approach to the conduct of research, in Zecha, G. (Ed.) *Critical rationalism and educational discourse*, Amsterdam: Rodopi
[68] Popper, K. (2002) *The logic of scientific discovery*, Routledge
[69] Popper, K. and Miller, D. (1983) A Proof of the impossibility of inductive probability, *Nature*, 302 (5910): 687
[70] Popper, K. (1992) *Realism and the Aim of Science*, Routledge
[71] Postan, M. (1971) *Fact and relevance: essays on historical method*, Cambridge: Cambridge University Press
[72] Quine, W. (1960) *Word and Object*, MIT Press
[73] Ramakers, S. (2002) Postmodernism: a 'sceptical' challenge in educational theory, *Journal of Philosophy of Education*, 36, 4, 629-651
[74] Robinson, D., Levin, J., Thomas, G., Pituch, K. and Vaughn, S. (2007) The incidence of 'causal' statements in teaching-and-learning research journals, *American Educational Research Journal*, 44, 2, 400-413
[75] Rorty, R. (1999) Phony science wars, review of Hacking, I. (1999) The social construction of what?, London: Harvard University Press, in *The Atlantic Monthly online*, November 1999
[76] Russell, B. (1903) The principles of mathematics, New York: W. W. Norton & Company
[77] Russell, B. (1996) *History of Western philosophy*, Routledge
[78] Sale, J., Lohfeld, L. and Brazil, K. (2002) Revisiting the quantitative-qualitative debate: implications for mixed-methods research, *Quality and Quantity*, 36, 43-53
[79] Salmon, W. (1998) *Causality and explanation*, New York: Oxford University Press
[80] Scauso, M. (2020) Interpretivism: Definitions, trends, and emerging paths, *International Studies*, https://doi.org/10.1093/acrefore/9780190846626.013.522
[81] Scheffler, I. and Goodman, N. (1972) Selective Confirmation and the Ravens, *Journal of Philosophy*, 69, 3, 78–83
[82] Schrödinger, E. (1994) *What is life?*, Cambridge University Press
[83] Shafer, G. (1996) *The art of causal conjecture*, London: MIT Press

[84] Shannon, C. and Weaver, W. (1949) *The mathematical theory of communication*, University of Illinois Press
[85] Skinner, B. (1971) *Beyond freedom and dignity*, Hackett Classics
[86] Smullyan, R, (1992) Gödel's Incompleteness Theorems, Oxford: Oxford University Press
[87] Snow, C. (2001) Knowing what we know: children, teachers, researchers, *Educational Researcher*, 30, 7, 3-9
[88] Steele, T. (2002) The role of scientific positivism in European popular educational movements: the case of France, *International Journal of Lifelong Education*, 21, 5, 399-413
[89] Strawson, P. (1952) *Introduction to Logical Theory*, Methuan & Co: London
[90] Sullivan, A. (2001) Cultural Capital and Educational Attainment, *Sociology*, 35, 4, 893-912
[91] *Swinburne, R. (1971) The paradoxes of confirmation – A survey, American Philosophical Quarterly, 8, 318–330*
[92] Thouless, R. (1974) *Straight and crooked thinking*, London: Pan
[93] Toulmin, S. (1958) *The uses of argument*, Cambridge: Cambridge University Press
[94] Turner, D. (2002) *The class struggle: the place of theory in education?*, inaugural lecture, School of Humanities and Social Sciences, Glamorgan University
[95] West, M. and Harrison, J. (1997) *Bayesian forecasting and dynamic models*, New York: Springer
[96] Winch, C. and Gingell, J. (1999) *Key concepts in the philosophy of education*, London: Routledge