Context Based Inferences in Research Methodology: The Role of Culture in Justifying Knowledge Claims

Colin W. Evers and Mark Mason

Introduction

The aim of this paper is to explore some ways in which cultural contexts can influence research methodology, particularly the kinds of inferences that methodology can sanction. Since the concept of culture is quite wide and includes, *inter alia*, both beliefs and practices, we focus here on beliefs, using the generic term 'theory' to refer to more systematically interrelated bodies of beliefs. At issue then is how substantive theories, or networks of beliefs, typically embedded in cultures, can influence methodology. For the purposes of this discussion, methodology is construed broadly as an applied epistemology, that is, as an account of knowledge and the nature of justification. The argument that follows traces the interaction between substantive theoretical assumptions and methodology in terms of the most familiar types of inference, notably induction, hypothetico-deduction and abduction. Because of its scope and generality, the argument that follows is largely philosophical, drawing on ideas in theory of knowledge and in philosophy of science. (For a useful overview, see Godfrey-Smith, 2003a.) Even the oft-drawn distinction between qualitative and quantitative research is subsumed by the main categories of inference discussed here. Some examples and illustrations are provided to demonstrate how these ideas can be applied to the operation of methodologies in comparative contexts. However, these are kept to a minimum as the papers that follow in this special issue provide, in their various ways, much more detail.

Inductive Inference

Induction is the practice of making inferences from the particular, or even a small subset of examples, to a much larger set. In understanding how it operates, it is useful to draw an old distinction, namely that between enumerative induction and analytical induction. (Znaniecki, 1934; Godfrey-Smith, 2003b.) Enumerative induction involves the usual processes of making statistical inferences based on a subset of examples any of which has an equal chance, often ensured by randomness, of being selected from a larger set of cases. For this type of inference, the number of cases matters. Analytical induction, on the other hand, seeks to find what is common among objects or phenomena and to use these commonalities as the basis for making inferences about the entire class of similar objects or phenomena. As Znaniecki (1934, p. 251) describes the method: it "abstracts from the given concrete case characters that are essential to it and generalizes them, presuming that in so far as essential, they must be similar in many cases". Here, it is in principle possible to make inferences from just one case. While each of these types of induction has had its defenders, for example, Reichenbach, who saw all non-deductive inference as statistical estimation, and Dewey, who saw it as a quest for a representative kind with the right features (Godfrey-Smith, 2003, p. 585), their methods of warranting inferences from subset to wider population are

different. We consider, first, enumerative induction, whose procedures are well understood and standard fare in statistical inference.

Enumerative Induction

Since, let us suppose, we are trying to estimate the number of F's that are G's in a population based on a sample size of N, we can use the fact that the statistical properties of all random samples of size N are binomially distributed. We can then use this to infer, within some confidence interval (and given a sufficiently large N), the chance of the sample's statistics estimating those of the population.

In order to make this model of inference work, we need sufficient background knowledge to prevent two sources of possible error. The first requires prevention of sampling bias which undermines randomness. Suppose our F's are people and our G's are speakers of Cantonese and we are comparing the proportion of Cantonese speakers in two jurisdictions. To ensure the equal likelihood of population members in a jurisdiction's sample, we need to correct for any unevenness in the distribution of F's who are G's. A sample drawn from southern China, for example, will be statistically different from one drawn from northern China. Randomization procedures in experimental setups are often much easier to achieve than in *ex post facto* studies, the latter being a type of research that is much more common in comparative education. The difficulty here is that the knowledge required for delivering an appropriate stratification of samples can outrun the knowledge being sought for the aggregate. That is, knowledge of a coarse-grained aggregate requires knowledge of a fine-grained breakdown. The process of justification here need not be flatly circular if we have proxy measures for the distribution of Cantonese speakers. However, the point is still the same: knowledge of randomness is not epistemologically cheap.

There are other ways in which sampling bias can occur. A well-known example is the requirement for participants to sign an ethics consent form. Systematic differences can result in samples depending on whether or not jurisdictional cultures are conducive to the practice of signing consent forms. Perceptions of the researcher as an insider or an outsider can also affect samples. (See Katyal and King, this volume.)

Knowledge of theory as an antecedent to enumerative induction takes a more basic form. In the case of researching persons who are Cantonese speakers, theoretical judgments of course need to be made about such matters as: What counts as Cantonese in relation to regional variations? How much does a speaker need to know? How fluent should be their usage? What is the boundary for learners to be so classified? But a general version of this issue applies to all F's and G's regardless of what they are. For, as Goodman (1955, 1972) has argued, there is no such thing as a class of similar objects, where objects are individuated by properties. That is, contrary to intuition, two swans have as many properties in common as a swan and a duck. In order to classify objects or episodes into kinds, or similarity classes, we need to appeal to some prior weighting of properties, picking some as more important than others. Quine makes this point in relation to how a child learns, by induction from examples, the word 'red':

> Logically, as long as *a, b,* and *c* are three and not one, there is exactly as much difference between *a* and *b* as between *a* and *c*; just as many classes, anyway, divide *a* from *b* (i.e., contain one and not the other) as *a* from *c*. For the child, on the other hand, some differences must count for more than others if the described process of learning 'red' is to go forward at all. (Quine, 1957, p. 231)

Popper (1959, pp. 420-422; 1963, pp. 58-64) goes much further, using the result to argue that the concept of induction is logically incoherent. The logically prior requirement for weighting properties means that similarity is always similarity-for-us. "Generally, similarity and with it repetition, always presupposes the adoption of a point of view" (Popper, 1959, p. 421). For Popper, this means that experience functions not as a source of inductive inference, but rather as material to confirm or disconfirm prior theories, that is, those points of view, initially innate, that come before particular experiences. Theory development is thus driven by disconfirmation, the mechanism that tells us our theories are inadequate. (A formal proof of Goodman's thesis can be found in Watanabe, 1969, pp. 276-377. He calls the proof the Theorem of the Ugly Duckling. For a recent critical review of these ideas see Decock and Douven, 2011.) We won't pursue the consequences for this result here (although it will figure in our discussion of analytic induction), except to note that the main general consequence is that theories that underwrite the identification of similarities in one context need to be scrutinized for their appropriateness in other contexts. The importance of culture emerges from the fact that key similarities, such as ways of defining 'otherness', can be culture-laden.

So far we have been examining both the necessity for, and the role of, prior theory in sampling for enumerative inductive inference. In comparative studies using this methodology, sampling bias can occur where culture-laden theories that drive random selection in one jurisdiction will operate less randomly if used in another jurisdiction. Researchers who conduct large-scale international comparative studies such as TIMSS or PISA are expected to allow for this.

A Hawthorne effect, which is different from a sampling bias, occurs when the act of inquiry itself functions to influence the data obtained. Here, inquiry is behaving as a kind of confounding variable. There are many ways that this can occur. Here are two. In the original series of studies that gave the effect its name, participants at the Hawthorne Works factory of Western Electric were found to improve their productivity primarily in an effort to please the researchers who were experimenting with the kinds of work conditions that might make for productivity gains. (For a discussion, see Hoy and Miskel, 2001, pp. 14-16.) Analogously, research participants may, when asked their views on some topics, give answers that are designed to please the researcher rather than to say what they really think.

A more common version of the Hawthorne effect is where fear of sanction prevents participants from expressing what they think. Something like this might have been occurring in the large-scale study of Hong Kong Chinese attitudes towards transgender people and their civil rights conducted by King (2008). Despite evidence of prejudice from a pilot study, an examination of the way Hong Kong law treated transgender people, and a study of language usage (the most common Cantonese term for a transgender person translates as

"monster"), some 50% of randomly selected respondents comprising a large sample were neutral in their answers to Lykert-scale questions used to measure prejudice. The reason that Hawthorne effects can be particularly troublesome in comparative research is because perceptions of what will please another or what will attract sanctioning are culture-laden and can vary among jurisdictions. What prompts a perception of sanction in Hong Kong can be quite different in Thailand.

Analytic Induction

Analytic induction bases its inferences from the particular to the general on the claim that the particular is an instance, or a representative, of a kind. The argument for representativeness is not statistical. Rather, it is a claim made from theoretical (or analytical) considerations. Some examples will help clarify the nature of this type of inference.

Consider, first, evidence for the claim that sugar dissolves in water. In enumerative induction the procedure would be to observe an appropriately large number of dissolvings. That is, numbers matter for sustaining this type of inference. However, with analytical induction, it is the chemical properties of sugar that loom large. Once it is established, perhaps by analysis of its molecular features, that sugar is the kind of substance that dissolves in water, numbers of instances are not important. It is the identification of sugar as an example of a kind that dissolves in water that matters. The argument extends beyond the natural kinds that science trades in to social kinds as well. A bachelor is an instance of a social kind, defined constitutively by a jurisdiction's distinction between married and unmarried adult males. The inferential machinery for determining the proportion of bachelors whose educational attainments exceed a certain level is entirely different from that for determining the proportion who are not married. In the second inference we make use of how the term "bachelor" is understood as a way of identifying bachelors as belonging to a particular social kind. Note that social kinds can be created by a jurisdiction in ways that depart from determination by ordinary linguistic usage. Thus, while the dictionary might define spouse as "one's husband or wife", governments sometimes define the term differently to meet wider social agendas. Hence, in the interests of promoting greater equality, Australian legislation, in 2008, expanded the definition of "spouse" for tax and superannuation purposes to include same-sex couples, thus bypassing the dictionary's reference to a husband or a wife.

In his work on speech acts, Searle (1969, pp. 33-37) distinguishes regulative rules (rules used to regulate pre-existing practices) from constitutive rules (rules used to create practices). A rule such as "drive on the left hand side of the road" regulates the practice of driving. Rules, such as the rules of chess, create the game. That is, the game is brought into existence by the rules that define the various moves that pieces can make and what counts as winning or losing. Searle (1995) develops these notions into a systematic theory of social reality, especially in relation to the role of language as a condition for creating social facts. However, we shall not pursue these ideas here. The point we want to make is that this conceptual apparatus can be pressed into the service of making analytic inductions. By making modest assumptions about the desirability of avoiding vehicle collisions and the efficacy of standardising features of road usage, the importance of amassing large numbers of

observed instances of people driving on the left (or right) diminishes. Indeed, one instance can suffice to settle the question of which convention a jurisdiction has adopted, since there are only two options. In this case, analytic induction works by having meagre empirical evidence cast in the role of eliminating one explanation over its only other plausible rival. It is functioning as a form of inference to the best explanation.

Where different languages embody, or express, different cultural assumptions, analytic induction needs to accommodate these to ensure that inferences can be used for comparative purposes. Take the case of understanding the link between talented performance and giftedness. There is a venerable tradition, deriving from the essentialism of both Plato and Aristotle, for seeing talented performance as arising out of a person's possession of the attribute of giftedness. The use of IQ testing to allocate types of educational provision falls within this broad tradition. But the Confucian cultural tradition takes a different turn. The Chinese term 'cheng cai', used to denote talented performance, embeds within it the notion that such performance is the result of development. (See Evers and Wu, 2006, pp. 521-522, and Wu, 2005, for more detail.) Analytical inferences about promoting talented performance made from within an essentialist cultural tradition will therefore not carry over into a culture embodying non-essentialist assumptions. A related point can be made about the role of student effort, or 'hard work', in achieving learning outcomes as measured by international comparative studies. Essentialist and non-essentialist views of performance will support differing analytical inferences.

If we take seriously Searle's (1995, 1998) view of the nature of institutional reality as sustained in part by constitutive rules, then very large parts of social life will be amenable to inferences by means of analytic induction. For example, the classification of persons as teachers within educational systems is not decided by, say, biological considerations, but by the constitutive rules that define what a teacher is: such as, meeting selection criteria leading to the issuance and acceptance of a role-defining employment contract. Knowledge of these rules, and the many others that define the institution of education within a jurisdiction, will constrain or extend the scope for inference from empirical data in ways both distinctive to the jurisdiction and relatively independent from the number of cases. Moreover, the theories of social life that cultures contain will be an important ingredient in the formulation of the constitutive rules that define institutional reality. In this way, cultural knowledge becomes a requirement for understanding the social realities to be investigated and the inferences one can analytically make from those investigations.

Hypothetico-Deductive Inference

The methodology of hypothetico-deductive inference is a pattern of epistemic justification whose development and adoption was motivated largely by perceived weaknesses in inductive techniques, particularly enumerative induction. It comes out of the philosophical tradition of logical empiricism. Instead of trying to justify theories by inference from instances, the process of justification was reversed. Theories came first and their justification depended on whether observed instances confirmed them. That is, empirical theories permitted, hypothetically, the deduction of empirical consequences. These theories were then tested in various ways to see if the hypothesized consequences could actually be

observed under the conditions posited by the theory. In common with enumerative induction, numbers mattered, at least when it came to confirmation. The more confirmations piled up, the more justified the theory was supposed to be. But there was another aspect of testing that was also important: disconfirmation. And here, merely one observed counter-example to a predicted outcome could prove troublesome, arguably falsifying a theory being tested.

Although these ideas were developed with scientific theories in mind, particularly theories from physics, they also came to enjoy wider influence, owing, in part, to the work of Herbert Feigl (1951, 1953) in which he attempted to show how the ideas could be applied to social science. The weaknesses in this pattern of knowledge justification, where they emerged in discussions of scientific theories (e.g., Quine, 1960; Hanson, 1968, 1972; Kuhn, 1970; Feyerabend, 1975) are even more pronounced in social science. One theme that is common to objections to hypothetico-deductive methods of justification is that the methodology must implicitly assume parts of the very theories concerning whose merits it is attempting to adjudicate.

To see this in relation to one objection, consider the following. For testing to work, it supposes that there is a clear distinction between the theory under test and the observations used for confirmation or disconfirmation. Otherwise, observations would lack the epistemic warrant to justify theories. But to figure in the logic of testing, observations must be described in some language (a point that is at one with our earlier account of the role of prior theory in making similarity judgments). Because the linguistic terms employed will be embedded in some theory or another, the challenge was to describe observations in such a way that their role in testing does not beg the question in favour of the theory under test. The presumption in favour of the language of behaviour, and hence behavioural science, was that it would provide the required level of epistemic objectivity. Describing human behaviours as “bar pressings”, or as “rotating an arm through an arc of 90 degrees”, seemed to invest observations with this objectivity.

Unfortunately, the distinction between a mere happening (raising one’s arm) and an action (bidding at an auction, acknowledging a friend, saluting, brushing away a mosquito) does real predictive and explanatory work in social science. The causal antecedents of a bid and a salute are entirely different, as are their causal consequences. Collapsing them together under the label “arm movement” disqualifies them from figuring in the adjudication of different accounts of what people are doing in social contexts. This is not evidence of the effects of a neutral observation language. Rather, it is a consequence of adopting the theoretical presuppositions embedded in behaviour theory-laden accounts of observations.

The effects of such theory-ladenness on hypothetico-deductive patterns of justification in research methodology are, naturally enough, sensitive to culture. For example, in comparing the administration of schools across culturally different jurisdictions, it’s not just the evaluation of what can count as good leadership that can vary. If we suppose that, in general terms, leadership involves the exercise of influence to achieve organizational goals, then the very classification of actions as leadership will be culture-laden. This is because the nature of influence, and the extent and pattern of its propagation through a social network,

are sensitive to culture-laden understandings of how influence should be legitimized, exercised and bounded. Thus, in contrasting Chinese Confucian views of leadership with some Western models, Wong (2001) notes that the ethical humanism behind the virtue of promoting harmony, common in Confucian thought, constrains the notion of leadership in ways that are different from significant Western notions of transactional or transformational leadership. Acts that look like leadership in one context may fail to be so classified in another context.

There is a deeper problem lurking behind the matter of theory-ladenness that needs to be mentioned as it is quite general. The logic of hypothetico-deductive confirmation and disconfirmation was developed within the framework of an extensional logic. This notion is technical but it can be easily unpacked. A sentence in an extensional system will remain true under different substitutions of referring expressions provided those expressions refer to the same objects. Here's an example. Take the sentence: "The school on the corner of Third and Elm streets has the best academic performance in the state". Now if the school on the corner of Third and Elm is also, as a matter of fact, the school with the largest student population, then in an extensional system, it will also be true that "the school with the largest student population has the best academic performance in the state". But now consider a non-extensional context, one that is intensional (with an 's'), or referentially opaque. Suppose we have the sentence "Tom believes that the school on the corner of Third and Elm streets has the best academic performance in the state". If Tom does not know that the school on the corner of Third and Elm is the school with the largest student population, then despite the expressions referring to the same object, it does not follow that "Tom believes that the school with the largest student population has the best academic performance in the state" is true. The upshot is that empirical evidence confirming or disconfirming intensional statements needs to be formulated in the right way. That is, Tom will go on believing what he does about the school on the corner of Third and Elm, even if he is persuaded that the school with the largest student population is not the best academically performing school in the state.

The example scales up. Theories of social phenomena that omit use of what are called propositional attitudes – terms such as "believes", "knows", "understands", "thinks", "fears", and the like – risk missing significant causal and predictive features of the social landscape. But theories that include them must accept that their role is infused with interpretations that need to take into account cultural understandings. So for evidence to figure in hypothetico-deductive inferences under these conditions, cultural considerations cannot be avoided.

A second objection concerning the purported relationship between evidence and theory in hypothetico-deductive reasoning is about the difference between the amount of evidence available and the explanatory and predictive scope of theories. Theories, because of their generality, always outrun the empirical evidence offered in their support. A useful analogy is the way a line may be fitted to a finite set of data-points. The line is like a theory and the data-points are the available empirical evidence. In this set-up, there is any number of lines (of varying complexity) that can be drawn through the same set of points. We say of this situation that the empirical evidence always underdetermines a theory. This raises the

question of how to understand the role of confirmation in theory justification, since an arbitrary number of different theories are being confirmed by the same evidence. One understanding, proposed by Kuhn, is that when it comes to justification among these theories, or paradigms "the proponents of competing paradigms practice their trades in different worlds" (Kuhn, 1970, p. 150). People who inhabit these worlds are looking at the 'same' evidence from different perspectives. (The scare quotes signify that even the evidence will come to be of a different kind when viewed from a different theoretical perspective.) The sun rising comes to be reconceptualised as the earth rotating. Kuhn (1970, pp. 111-135).

The methodological point is that if theory choice is being determined solely by confirmation, then it is insufficient for choosing among all the theories that fit the empirical evidence. And in this case, other factors come into play. Kuhn accords a large place to sociological factors in explaining large-scale changes in scientific theories, or the adoption of new paradigms. Here, again, is where culture can play a determinate role in this aspect of the workings of hypothetico-deductive justification. For on this view, culture shapes both the understanding of evidence and the choices made among competing theories or accounts of that evidence. And here, again, attempts to build up theories or accounts of phenomena in different jurisdictions for purposes of comparison, in using this methodology, will need to be especially vigilant in controlling for similarities among purported categories of comparison.

Problems with confirmation as a method of theory adjudication prompted a focus on disconfirmation: for, while confirmation was insufficient for the task, different theories that were supported by the same body of known evidence could be tested against their differing predictions concerning the next observation. That is, observation could justify a theory by falsifying its rival theories. Karl Popper (1959, 1963) is the best known advocate of this approach. For Popper, all scientific knowledge is provisional. The best theories are those that have withstood the most rigorous tests and have survived, that is, have not been falsified, or disconfirmed, by empirical evidence. But this does not guarantee their warrant into the future, as even the most well-established theories may come to grief on some future test. The aim of this pattern of justification is therefore modest: to establish theories that can be tested rigorously and that survive these tests. As the full range of Popper's ideas has been systematically presented and discussed in comparative education, owing to the work of Brian Holmes (1972, 1981a, 1981b) and his critics (e.g. Singh, 1985) we shall focus on just one issue – the problem of holism – since this provides an important entry for considerations of culture in falsificationist methodology.

When a theory implies an observation statement that is up for testing, logically, it is always some conjunction of statements within the theory that support the implication. So if the observation statement is falsified by what is observed, then the logical implication is that at least one of the statements comprising the conjunction is false. But which one (or more)? In responding to this challenge, Quine (1951, p. 43) famously remarked that "Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system". That is, strictly speaking, there is nothing in the logic of disconfirmation that constrains our choice. Lakatos (1970) attempted to deal with the problem by suggesting that a methodological decision be made to protect some part of a theory – its "hard core"

(Lakatos, 1970, p. 133) – by use of a "protective belt" of auxiliary hypotheses. Falsification applies to claims within the protective belt, thus leaving the central claims of the theory immune to revision, at least until the series of theories with the same hard core (called a "research program") begins to exhibit signs that it is degenerating. (For an account of such signs, see Lakatos, 1970, pp. 116-122.)

Lakatos's recommendation captures an important feature of research in science, since particular scientific theories can often be located within larger sets of shared assumptions forming research programs. Moreover, as Kuhn observed, the process whereby researchers become inducted into these research programs is a form of enculturation. Now the point to be made here is that culture can play an important role in shaping what counts as the hard core of theories in social science. Thus, to use our earlier example, theories of leadership that are shaped by the procedurally unrevisable notion that harmony is to be preserved will impose a different set of choices on the logic of disconfirmation, and hence what the empirical evidence supports, than theories that insulate from revision the kind of individualism behind, for example, Western notions of transformational leadership. That is, the actual methodology of justifying a theory of leadership can be affected by the role of culture in demarcating what is to count as a theory's hard core. So, comparative researchers building up theories, based on empirical evidence, concerning the nature of leadership as construed in culturally different jurisdictions, may need to allow for the logic of disconfirmation to work in different ways: in particular, that the class of claims that can be falsified, the protective belt, may be quite different in each case.

Abductive Inference

This is, perhaps, the most common yet least understood model of research methodology in social science. One reason for its ubiquity is that it functions as a commonsense default option for situations where information is relatively scarce and where decisions have to be made among alternatives. And a reason for its being not well understood is that its criteria for theory choice tend to be perceived as vague, or not as precise as other methods purport to be. Originating in the work of C. S. Peirce (1931 – 1958), the main idea of abductive inference is to adjudicate hypotheses, or theories, on the basis of their explanatory powers. Peirce (1931-1958, Volume 5, p. 189) claims that abductive arguments have the following schematic form:

> The surprising fact, C, is observed.
> But if A were true, C would be a matter of course.
> Hence, there is reason to suspect that A is true.

Here is how we might use this schema to give a plausible explanation of our earlier example of road rules.

> The surprising fact (to a visitor from another jurisdiction) is that C is observed, namely, someone is driving their car on the right-hand side of the road.
> But if it were the regulative rule in this jurisdiction that A, namely, everyone drives on the right, then C would be a matter of course.

> Hence, there is reason to suspect that A is true.

As it is given, the schema says nothing about the merits of alternative explanations. To remedy this omission, we consider Josephson and Josephson's (1994, p. 5) version:

> D is a collection of data
> Hypothesis H explains D.
> No other hypothesis can explain D as well as H does.
> Therefore, H is probably true.

This version lacks the clause about surprise that gives rise to motivation. It also omits reference to theories, trading instead in single hypotheses. These can be captured in the version proposed by Haig (2009, p. 221):

> P1, P2, ... are surprising empirical phenomena.
> Theory T explains P1, P2, ...
> No other theory can explain P1, P2, ... as well as T does.
> Therefore, T is accepted as the best explanation.

With this schema in mind, attention now shifts to what it is that makes one theory better at explaining the relevant phenomena than another theory. Work on providing an answer to this question is often called the study of inference to the best explanation (IBE), and there are many candidate accounts of IBE. (For an overview, see Haig, 2009.)

One of the lessons to be learned from the difficulties that were encountered by hypothetico-deductive accounts of justification, and which prompted the development of ideas about IBE, is the limits entailed by relying on empirical adequacy as the sole criterion for a theory to be warranted. But in fact, as Paul Churchland has observed, in addition to empirical virtues a theory may also enjoy "superempirical" epistemic virtues. Here is his argument, quoted at some length:

> Since there is no way of conceiving or representing 'the empirical facts' that is completely independent of speculative assumptions, and since we will occasionally confront alternatives on a scale so comprehensive that we must also choose between competing modes of conceiving what the empirical facts before us *are*, then the epistemic choice between these global alternatives cannot be made by comparing the extent to which they are adequate to some common touchstone, 'the empirical facts'. In such a case, the choice must be made on the comparative global virtues of the two global alternatives, the T1-plus-the observable-evidence-therein-construed, *versus* T2-plus-the-observable-evidence-therein-(differently)-construed. That is, it must be made on *superempirical* grounds such as relative coherence, simplicity, and explanatory unity. (Churchland, 1985, pp. 41-42.)

These epistemic virtues of theory, which include consistency (contains no contradictions), coherence (the various parts of the theory fit together), simplicity (an absence of *ad hoc* claims), comprehensiveness (explains more rather than less) and explanatory unity (uses the

same claims to explain a wide variety of phenomena), together with empirical adequacy, are known collectively as the virtues of coherence justification. On this view, theory T offers the best explanation of phenomena P1, P2, and so on, if T is more coherent than its rivals.

Unfortunately, without some mode of simplification, the application of any model of coherence justification becomes computationally intractable for large bodies of belief. (Millgram, 2000.) However, in explaining our beliefs as located within a global system of belief, two features of that system are problematic. For, as Fodor (1983, pp. 104-119) has observed, this total body of belief is isotropic and Quinean. It is isotropic because an explanation for a belief can come from any part of the total system. And it is Quinean because, as we have seen, the merits of a piece of empirical evidence for a theory are determined by the global epistemic properties of the theory. In the light of this, the most useful simplifying strategy adopted in practice is to build explanations that cohere with prior well-established bodies of theory. Thus, developments in physics would need to cohere with established bodies of mathematics, or developments in biology would be constrained by the demand to cohere with prior theories in physics or chemistry. In this way, prior accepted theory scaffolds the task of subsequent coherence justification of developing theories. And in social science theories of culture can act as both constraints that scaffold coherence testing of explanations of social phenomena, as well as providing sources of further explanatory material.

A good example of this can be seen in attempts to explain the factors that make for student learning in schools. In a very large scale study that involved collating results from some 800 meta-analyses, that summarized over 50,000 studies that in turn reported on the learning of some 200+ million students, Hattie (2009) developed a ranking, based on a calculation of effect size (ES), of the various factors that influence student learning. Here are some of his findings:

Piagetian programs, ES = 1.28
Formative evaluation, ES = 0.90
Average of all interventions, ES = 0.40
Principals/school leaders, ES = 0.36
Television, ES = -0.18
Mobility, ES = -0.34

While it is tempting to use this kind of ranking as a policy guide for what sorts of interventions and programs should have the highest priority for development and implementation, adding explanatory hypotheses drawn from social and cultural theory can alter the nature of conclusions made by abductive inferences. There are two points that we wish to note. The first is that there is a mixture of kinds of studies that have been aggregated. For example, to mention two, some are experimental studies conducted under conditions that allow for the manipulation of variables, while others are *ex post facto* studies that make do with what is occurring in the environment. Now the social relations of inquiry under these two regimes need not be the same. In particular, they can differ systematically in relation to the amount of variability that the set-ups permit. Take the effect size of leadership on student learning. Overwhelmingly, the most common kinds of

studies are *ex post facto*. But in the jurisdictions where many of these studies were conducted, for example, the USA, Canada and the UK, the constitutive and regulative rules governing the appointment of school leaders seem designed precisely to reduce variability in levels of competence in job performance. So a low effect size might be an artefact of the dominant research design applied within the social arrangements of a particular jurisdiction. This means that research design itself can function as an influence on the amount of effect size measured. Now if the proportion of different research designs is different for groups of factors being investigated for their effect on student learning, then a ranking of effect sizes that does not take this into account will be compromised. Precisely how much it is compromised is a matter for further investigation. However, in this way, considerations of culture and society can figure in abductive inferences for an alternative ranking of effect sizes, a ranking that may also differ in different jurisdictions depending on the differential effect of the social and cultural relations of inquiry.

A second, more general point, can be made about the role of culture in abductive inference. The social and cultural arrangements under which research is conducted may not obtain in places where it is being considered for application. Since, for *ex post facto* studies, these conditions are approximately the same as those that provide the causal context in which the phenomena under investigation occur, the use of research from one jurisdiction to support arguments for the adoption of policies and practices in another amounts to supposing that research findings can be sustained under counterfactual circumstances. However, this can be justified in abductive inference only if the causal impact of different social and cultural contexts is taken into account: that is, if these contexts, when theorized, also sustain the same inference to the best explanation. To determine justification, similar studies would need to be undertaken in those different contexts where information about the contexts figured in abductively derived research conclusions.

Conclusion

The argument throughout has been that for a number of major types of patterns of inference employed in different research methodologies used in social science, the role of knowledge of contexts, and in particular cultural contexts, figures explicitly in the different logics of justification. This matters a great deal in comparative research because it means that knowledge of culture is more than just another feature of a society that needs to be researched. Rather, it shows that this knowledge plays a decisive role in the kinds of inferences that can be made. To sum up, for enumerative induction, cultural considerations can influence prior judgments of similarity needed to classify data in advance of inductive inference. For analytical induction, prior cultural understandings are a source of theoretical meaning, as well as knowledge of the regulative and constitutive rules that shape or define the phenomena about which inferences are to be made. For hypothetico-deductive patterns of justification, cultural understandings help shape the boundaries of what empirical evidence is able to refute or not, as well as providing a source of theoretical vocabulary for that empirical evidence. And for abductive reasoning, inference to the best explanation draws on cultural knowledge in order to fashion explanations that are relevant to the contexts that frame the phenomena under investigation. Because these can vary across regions of comparative study, the methodologies that figure in these comparative findings,

that is, their very processes of inference, are themselves subject to a comparative framework and not relatively neutral adjudicators of what claims are justified.

REFERENCES

CHURCHLAND, P.M. (1985) The ontological status of observables: in praise of the superempirical virtues, in P.M. CHURCHLAND & C.A. HOOKER (Eds*) Images of Science* (Chicago, University of Chicago Press.)

DECOCK, L. & DOUVEN, I. (2011) Similarity after Goodman, *Review of Philosophy and Psychology* 2(1), pp. 61-75.

EVERS, C.W. & WU, E.H. (2006) On generalising from single case studies: Epistemological reflections, *Journal of Philosophy of Education,* 40(4), pp. 511-526.

FEIGL, H. (1951) Principles and problems in theory construction in psychology, in W. DENNIS (Ed.) *Current Trends in Psychological Theory* (Pittsburgh, University of Pittsburgh Press).

FEIGL, H. (1953) The scientific outlook: Naturalism and humanism, in H. FEIGL & M. BRODBECK (Eds) *Readings in the Philosophy of Science* (New York, Appleton-Century-Crofts).

FEYERABEND, P.K. (1975) *Against Method* (London, Verso).

FODOR, J.A. (1983) *The Modularity of Mind* (Cambridge, MA, MIT Press).

GODFREY-SMITH, P. (2003a) *Theory and Reality: An Introduction to the Philosophy of Science* (Chicago, University of Chicago Press).

GODFREY-SMITH, P. (2003b) Goodman's problem and scientific methodology, *Journal of Philosophy*, 100(11), pp. 573-590.

GOODMAN, N. (1955) *Fact, Fiction and Forecast* (Sussex, Harvester Press).

GOODMAN, N. (1972) *Problems and Projects* (New York, Bobbs-Merrill).

HAIG, B.D. (2009) Inference to the best explanation: a neglected approach to theory appraisal in psychology, *American Journal of Psychology* 122(2), pp. 219-234.

HANSON, N.R. (1968) *Patterns of Discovery* (New York, Cambridge University Press).

HANSON, N.R. (1972) *Observation and Explanation: A Guide to the Philosophy of Science* (London, George Allen & Unwin).

HATTIE, J.A. (2009) *Visible Learning* (London, Routledge).

HOLMES, B. (1972) Comparative education as a scientific study, *British Journal of Educational Studies* 20(2), pp. 205-219.

HOLMES, B. (1981a) *Comparative Education: Some Considerations of Method* (London, Allen & Unwin).

HOLMES, B. (1981B) Sociology and comparative education: A comparative educationist's perspective, *International Review of Education* 27(4), pp. 397-410.

HOY, W.K. & MISKEL, C.G. (2001) *Educational Administration: Theory, Research, and Practice* (New York, McGraw-Hill).

JOSEPHSON, J. & JOSEPHSON, S. (1994) *Abductive Inference: Computation, Philosophy, Technology* (Cambridge, Cambridge University Press).

KING, M. (2008) *Hong Kong Chinese Attitudes towards Transgenderism and Transgender Civil Rights* (Unpublished PhD thesis, The University of Hong Kong).

KUHN, T. (1970) *The Structure of Scientific Revolutions* (2nd edition, enlarged, Chicago, University of Chicago Press).

LAKATOS, I. (1970) Falsification and the methodology of scientific research programs, in I. LAKATOS & A. MUSGRAVE (Eds) *Criticism and the Growth of Knowledge* (London, Cambridge University Press).

MILLGRAM, E. (2000) Coherence: The price of the ticket, *Journal of Philosophy*, 97(2), pp. 82-93.

PEIRCE, C.S. (1931-1958) *Collected Papers* (Cambridge, MA, Harvard University Press).

POPPER, K.R. (1959) *The Logic of Scientific Discovery* (London, Hutchinson).

POPPER, K.R. (1963) *Conjectures and Refutations* (London, Routledge).

QUINE, W.V.O. (1951) Two dogmas of empiricism, *Philosophical Review* 60, pp 20-43. Cited as reprinted in W.V.O. QUINE, (1961) *From a Logical Point of View* (New York, Harper & Row).

QUINE, W.V.O. (1957) The scope and language of science, *British Journal for the Philosophy of Science*, 8, pp. 1-17. Cited as reprinted in W.V.O. QUINE (1976) *The Ways of Paradox and Other Essays* (Cambridge, MA, Harvard University Press), pp.228-245.

QUINE, W.V.O. (1960) *Word and Object* (Cambridge, MA, MIT Press).

SEARLE, J.R. (1969) *Speech Acts* (London, Cambridge University Press).

SEARLE, J.R. (1995) *The Construction of Social Reality* (New York, The Free Press).

SEARLE, J.R. (1998) *Mind, Language, and Society* (New York, Basic Books).

SINGH, B.R. (1985) A major shortcoming of Professor Holmes' comparative methodology: The question of falsifiability, *Quality and Quantity*, 19, pp. 155-165.

WATANABE, S. (1969) *Knowing and Guessing* (New York, John Wiley & Sons).

WONG, K.C. (2001) Culture and educational leadership, in K.C. WONG & C.W. EVERS (Eds) *Leadership for Quality Schooling* (London, Routledge/Falmer).

WU, E.H. (2005) Factors that contribute to talented performance: A theoretical model from a Chinese perspective, *Gifted Child Quarterly*, 49(3), pp. 231-246.

ZNANIECKI, F. (1934) *The Method of Sociology* (New York, Rinehart & Company, Inc.).