

# Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <http://orca.cf.ac.uk/94627/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Brannan, Matthew J., Fleetwood, Steve, O'Mahoney, Joseph and Vincent, Steve 2017. Meta-analysis: A critical realist critique and alternative. *Human Relations* 70 (1) , pp. 11-39. 10.1177/0018726716674063 file

Publishers page: <http://dx.doi.org/10.1177/0018726716674063>  
<<http://dx.doi.org/10.1177/0018726716674063>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



# Meta-Analysis: A Critical Realist Critique and Alternative

## Abstract

Meta-analysis has proved increasingly popular in management and organisation studies as a way of combining existing empirical quantitative research to generate a statistical estimate of how strongly variables are associated. Whilst a number of studies identify technical, procedural and practical limitations of meta-analyses, none have yet tackled the *meta-theoretical* flaws in this approach. We deploy critical realist meta-theory to argue that the individual quantitative studies, upon which meta-analysis relies, lack explanatory power because they are rooted in quasi-empiricist meta-theory. This problem, we argue, is carried over in meta-analyses. We then propose a ‘Critical Realist Synthesis’ as a potential alternative to the use of meta-analysis in organisation studies and social science more widely.

## Keywords

Meta-analysis, meta-theory, critical realism, ontology, epistemology, aetiology, open and closed systems, tendencies

## 1. Introduction

According Bornmann & Mutz (2015) the quantity of published research doubles every nine years. This increases the appeal of methods that facilitate the integration and synthesis of existing research. Recently, social scientists, especially those working in management and organisation studies (MOS), have developed three basic methods to synthesise existing research: *systematic review*, *meta-interpretation* and *meta-analysis* (MA). This paper adds to a significant body of literature dedicated to critically evaluating MA. To date, critical evaluation has, primarily, engaged with the technical, procedural and practical problems of MA, and has, implicitly, presumed that resolving these problems is both necessary and sufficient to make MA more effective. Whilst these debates are welcome, they do not address the meta-theoretical underpinnings of MA, which is the focus of this paper. Our argument is that MA is of limited use in *explaining* the kind of social or organizational phenomena of

interest to readers of *Human Relations* – and cognate journals. This is not due to technical, procedural and practical problems in the application of MA, but due to the flawed meta-theory underpinning MA. Instead, we propose an alternative that we refer to as ‘*critical realist synthesis*’ (CRS), rooted in an entirely different meta-theory – a term we use to include methodology, ontology, epistemology, aetiology, and concepts of explanation, prediction and theory.

To make this argument, we start with an overview of MA, and provide a brief synopsis of some extant criticisms or what we term its ‘known problems’. We then provide a critical realist<sup>1</sup> critique of the meta-theoretical underpinnings of MA by drawing on two highly cited recent pieces of MA to illustrate our argument. By way of contrast, we then outline CRS and argue, whilst it ostensibly provides less ‘certainty’ than MA, CRS generates greater explanatory power, and is based on more realistic ontological premises.

## 2.0 What is MA?

MA first appeared in the field of medicine in 1904 as a method of aggregating data from experimental research. After World War Two it expanded into the fields of psychology, education and social science research. The primary aim of contemporary MA is to compute a weighted mean of effect size between phenomena; the secondary aim is to identify moderating (and mediating) variables. Let us take a closer look:

Meta-analysis, literally the statistical analysis of statistical analyses, describes a set of procedures for systematically reviewing the research examining a particular effect, and combining the results of independent studies to estimate the size of the effect in the population ... The outcome of a meta-analysis is a weighted mean effect size which reflects the population effect size more accurately than any of the individual estimates (Ellis 2010: 94-5).

[M]eta-analysis is ... a method that estimates an overall ‘effect-size’ of a range of studies from the individual effect sizes of each individual study, thus giving

---

<sup>1</sup> We recognise that CR is a broad church, encompassing original critical realism, dialectical critical realism, the ‘spiritual turn’, and other variants. The theorising we propose here is based on the original tenets of critical realism (Bhaskar 1975), but is equally compatible with (though perhaps less interesting to) any of the subsequent versions.

greater 'power' to the overall statistic. It does this by calculating a mean of means of means: in the original study, a mean is taken of the effects of a particular variable for all points in a study, then variables are averaged to provide an overall effect size (mean) for that study, and then the effect sizes of a number of studies are averaged in the MA procedure (Weed 80-81).

An effect can be the result of a treatment revealed in a comparison between groups (e.g., [medically] treated and untreated groups) or it can describe the degree of association between two related variables (e.g., treatment dosage and health). An effect size refers to the magnitude of the result as it occurs, or would be found, in the population. (Ellis 2010: 4, 6-7)

*'Effect size'* is a measure of the association or relationship between two variables across a range of carefully selected studies. Such analysis presumes that values of independent variables will be related to, or associated with, values of dependent variables if they are observed to regularly occur together with sufficient frequency, with statistical techniques being deployed to identify this association and its properties. On the presumption that the association is causal, independent variables are thought to have a (causal) effect on dependent variables. The term *'effect size'*, then, refers to the magnitude of the association between independent and dependent variables. This then forms the basis for testing meta-hypotheses.

If the MA fails to explain an (arbitrary) 75% of the variance, or we know in advance that there are significant differences in effect sizes across studies, then a moderator analysis can be conducted:

Moderation represents the idea that the magnitude of the effect of an antecedent (e.g., organizational structure or strategy) on firm outcomes depends on contingency factors, such as the uncertainty and instability of the environment ... [M]oderation refers to the conditions under which an effect varies in size, whereas mediation refers to underlying mechanisms and processes that connect antecedents and outcomes (Aguinis et al., 2013: 1-2)

MA assumes that the effective aggregation of information creates greater statistical power than that derived from any single study, and that results from individual studies are

generalisable to a larger population: to ‘translate statistical relations into successful recipes for individual organizations’ (Hodgkinson and Rousseau 2009). The ability to determine causes and effects is ostensibly enhanced as the population grows (as more studies are added to the MA) and inconsistencies between results are quantified and assessed. Moderators and mediators can also be included in an attempt to ‘explain’ variation between results and the presence of forms of bias.

These benefits have been asserted in some sections of the MOS field, wherein the value of MA has even expanded beyond the realm of synthesis, and towards claiming the generation of new knowledge:

Beyond overcoming difficulties associated with individual studies such as sampling error, measurement error and restriction of range, MA enable an analyst to synthesise the findings of primary studies to test hypotheses that were not testable in those studies (Eden, 2002: 841)

Having outlined the basic premise of MA, we now briefly outline the known technical, technical, procedural and practical issues with the practice of MA, before moving on to our realist critique.

## **2.1 Known problems**

The many technical, procedural and practical (i.e. collection of source data) challenges involved in conducting MA have been detailed by a number of authors. First, there is a lack of agreement on the basic methods to assess effect size, which in turn produce significantly differing results. Whilst calculating effect size requires the subtraction of the mean of the control group from the mean of experimental group and dividing the difference by the standard deviation, there is no agreement on how this standard deviation is calculated (see Glass, 2000; Hough and Hall 1994).

The practical task of constructing a sample also provides a number of issues for MA. The MA literature seldom discusses inclusion criteria for data (Rousseau et al., 2008: 491), despite the fact that these cannot be generalised across MA. Inclusion criteria are thus ultimately judgement calls that vary by the research topic and researcher preferences, yet they impact

clearly upon the calculation of effect sizes as they define the source material that constitute the analysis. This leads to a problem with publication bias, as published results tend to be those that show strong statistical outcomes (Rousseau, et al., 2008). Thus, MA tends to over-represent positive results whilst dramatically underreporting those that are null. This has led some to argue that results reported as statistically significant may have inbuilt exaggeration bias (Rossi, 1987).

Relatedly, the validity of effect sizes is a function of the homogeneity of included studies (Miller, 1987). This poses a paradox as studies with large sample sizes are privileged, which mitigates against the possibility of pooling sufficient homogeneity in terms of research foci, especially in social science research. Whilst on the face of it, greater inclusion seem to follow the internal logic of MA by increasing the scope and sample size of the analysis, Coyne et al., (2011:224) show that including very small scale research in MA is likely to lead to ‘overestimate effects’ which statistical techniques cannot correct.

The extent to which source studies can be combined is also dependent upon the degree of similarity (in terms of definitions, interpretations of key variables, and the deployment of data capture techniques) between studies (Linden and Priestley, 2009). However, the codification of the process through which this is achieved, is often significantly truncated or even omitted in publication. Similarly, the nuanced way in which theories and concepts inform the design and operationalisation of the original studies is crucial. Data from original studies require manipulation and tabulation to perform MA and, given that these were generated for alternative purposes, it is problematic to match the theoretical perspective of the meta-analyst and the original research, if the original data is even accessible at all (Cowton, 1998). The consequence of this is that effect size analysis may therefore amalgamate statistical findings based on differing interpretations of the theoretical hypothesis as well as differentially operationalised constructs of study.

A further challenge for MA relates to the quality of source data: any given range of source data is likely to display variability in terms of the extent to which they possess internal (elimination of bias) and external validity (Franke, 2001). The MA analyst takes for granted that what the original analyst did to code the data into concepts is reliable, so one’s measures could be very different across studies. This means that the extent to which the results can be generalised to their target population is at best questionable. Moreover:

method variance is pervasive, ubiquitous, almost invariably in social and behavioural science, each array of measurements ... contains variance associated with the method. Any obtained relationship between two such units can be due to method variance (Fiske 1982: 82).

Generalisation on the basis of studies with reliability issues will therefore accentuate rather than reduce, or correct for, error, and may reflect manipulations of non-comparable independent variables and their effects on non-comparable dependent measures.

### **3.0 A critical realist critique of MA**

Whilst the technical, procedural and practical issues with meta-analyses are notable, our critique is not based upon these. Indeed even if these problems were resolved, our critique, which is *meta-theoretical*, would remain. To the best of our knowledge, no meta-theoretical critique of MA has been undertaken (although, see Pawson 2004). Let us start with establishing some basic terms and ideas that will inform the rest of the paper.

First, we use the term 'causal mechanism' generically, to refer to things like 'social structures', 'cultural structures', 'institutions', 'conventions', 'norms', 'rules', and so on. An HRM practice, or a discourse could, for example be causal mechanisms. The term 'mechanism' carries no connotations of simple additive effects or determinism. It simply refers to a thing that has causal powers or, in layperson's language, the ability to affect things. A causal mechanism is *causal* in virtue of the powers it possesses as derived from its properties. The causal powers of any mechanism only become enabled when enacted by human agents. When, therefore, we refer to a mechanism causing this or that, we always have in mind an *agentially enacted mechanism*.

Second, we use the term '*quantitative empirical studies*' to refer to those studies employing *quantitative data and statistical research techniques*, typically regression analysis. They should not be confused with *qualitative empirical studies* such as ethnographies, case studies, in-depth interviews, participant observation and such like<sup>2</sup>.

---

<sup>2</sup>Some research appears to be qualitative because, for example, it uses interview techniques. But, if the

Third, for CRs, the objective of social science is not to predict but to *explain*. This is achieved by identifying, and theorising: an appropriate, (*qua* relevant) agent (A); an appropriate causal mechanism<sup>3</sup> (M); how agent (A) interprets, and enacts mechanism (M), generating tendencies (T) towards outcome (O); other mechanisms, often referred to as ‘the context’, (M<sup>c</sup>) that dispose agent A to interact with M and not (say) N. Any putative explanation can then be empirically substantiated - i.e. successfully tested, which does not involve simply testing quantitative hypotheses. We refer to this as generating *theoretically informed and empirically substantiated explanations*.

Fourth, quantitative empirical studies, that provide the source material for MA, are rooted in a meta-theory we call ‘*quasi-empiricism*’<sup>4</sup> and comes with a ‘*chain of meta-theoretical concepts*’ (Fleetwood 2014), especially *ontology, epistemology, methodology, aetiology*, and concepts relating to *open and closed systems, theory, prediction and explanation*<sup>5</sup>. Let us look a little closer at this chain of meta-theoretical concepts.

### ***Ontology***

Observed (empirical) *events or states of affairs*, are the ultimate phenomena about which quasi-empiricists collect data - e.g. size of organisations; presence of teamwork; being female; employee performance. If these events are observed (or proxied) in terms of quantity or degree, they become variables – i.e. *quantified* events. The ontology consists, therefore, of *observed* events or states or affairs that are unique, unconnected or *atomistic*.

### ***Epistemology***

Whilst quantitative empirical researchers are probably aware that the variables they measure represent causal mechanisms, broadly conceived, their focus is always on the events these

---

analysis ends up quantifying the data from interviews, then it is more accurate to define this as quantitative research.

<sup>3</sup>We use the singular here (agent and mechanism) for ease of exposition. Most of the time, we will have to use the plural (agents and mechanisms).

<sup>4</sup>Some use the term, ‘positivism’. We think ‘scientism’ is a more accurate description, but stick with the term ‘quasi-empiricism’ to avoid getting embroiled in philosophical definitions that are tangential. For elaboration of positivism/empiricism in MOS see Donaldson (1996, 2003, 2005); Johnson & Duberley (2000); and Fleetwood & Hesketh (2010).

<sup>5</sup>For a CR interpretation of quasi-empiricism’s meta-theoretical underpinnings see Danermark (2002); Sayer (2010); Fleetwood (2014) and Fleetwood & Hesketh (2010).

mechanisms generate<sup>6</sup>. If, as presumed, particular knowledge is gained through observing events, more *general* or ‘scientific’ knowledge is gained only if these events manifest themselves as *regularities in the flux of events or states of affairs*<sup>7</sup>. This is usually styled ‘whenever event  $x_1 \dots x_n$  then event  $y$ ’ or  $y = f(x_1 \dots x_n)$ .

Together, this ontological and this epistemological position imply a ‘flat’ ontology – the assumption that all that exists are events (or actions) and people’s perceptions of these events (Table 1).

<b>Domain</b>	<b>Entity</b>
Empirical	Experiences & perceptions
Actual	Events & actions

*Table 1. A ‘flat’ ontology*

### ***Methodology***

The method of quasi-empiricism seeks to generate predictions, typically in the form of hypotheses to be refuted or supported via the collection of quantitative data. The only phenomena that feature in quantitative empirical research are those capable of being transposed into *variables* - i.e. the quantified expression of events. What cannot be quantified adequately is omitted.

### ***Aetiology***

---

<sup>6</sup>For example, Roth et al, (2010: 275) refer to the ‘weakening of stereotypes or other related mechanisms’; Shirom et al, (2008: 1376) refer to ‘coping mechanisms’; and Subramony (2010: 747, 759) refers to ‘feedback mechanisms’ and ‘goal setting and reinforcing mechanisms’.

<sup>7</sup>For ease of exposition, we drop the term, ‘states of affairs’ and refer, simply, to ‘events’.

The notion of causation pre-supposed by quasi-empiricism is referred to as the *regularity view of causation*. As its ontology is of observed atomistic events, its concept of causality cannot be conceived of in terms of anything other than events and their regularity. As the epistemology of quasi-empiricism is reliant upon identifying event regularities, its conceptualisation of causation requires knowledge of event regularities. To know the cause of increased organisational performance is, for example, to know that it is regularly preceded by the introduction a bundle of HRM practices. More generally, to know the cause of event  $y$ , requires us to know (no more than) that event  $x$ , or events  $x_1, x_2 \dots x_n$ , is/are regularly conjoined to event  $y$ .

It is worth adding that conclusions are often, usually implicitly, given a universal and general 'twist', along with a spurious precision. For example, in their analysis of performance related pay (PRP) (Gielen et al. 2010: 291) write that: 'PRP increases productivity at the firm level by 9%'. It is not clear if this is understood as a 'one off', or whether this is supposedly generalizable to all firms. If the latter, the '9%' looks to be an example of spurious precision.

### ***Open and closed systems***

The quasi-empiricist commitment to causality as regularities in the flux of events requires that social or organisational systems are theorised or modelled as if they are *closed systems*, defined thus: *Parts of the socio-economic world characterised by regularities in the flux of events (or states of affairs) of the form 'whenever event  $x$  then event  $y$ ', or  $y = f(x)$  are closed systems, and parts of this world not so characterised are open systems* (See Bhaskar 1978, Lawson 1995 and Fleetwood 2016). Crucially, statistical techniques like regression analysis not only presuppose, but only work in, closed systems. Methodologically speaking, quantitative empirical researchers of organisation studies must 'engineer' closed systems (only in theory, because a real open system such as an organisation cannot be closed) so they can write things like:

Hypothesis 1: Empowerment-enhancing bundles [of HRM practices] will be positively correlated with business outcomes (Subramony 2009: 748).

This translates to 'whenever empowerment-enhancing bundles (EEB), then business outcome', or 'outcome =  $f(\text{EEB})$ ' and, by definition, this describes a closed system.

Variations in regularity are generally specified probabilistically or stochastically, as random processes occurring in the *ontic* domain. Probability is a *measure* of the likelihood of an event occurring. The re-conceptualisation of stochastic event regularities using the concepts of probability, might be styled ‘whenever event  $x$ , then on *average* event  $y$ ’, or  $y = f(x + \varepsilon)$  or more accurately, ‘whenever the realised value of the (independent) variable measuring event or state of affairs  $x$ , then the conditional mean<sup>8</sup> of the (dependent) variable measuring event or state of affairs  $y$ ’. The error term ( $\varepsilon$ ) presents random influences on the dependent variable  $y$  and consequently converts the mathematical model linking  $y$  to the  $x$  into a stochastic or statistical model representing the population of interest (Downward 2015: 210). If an empirical researcher managed to identify a stochastic event regularity (perhaps over a restricted space/time) then s/he will have identified a *stochastically closed system*. Henceforth, we use the phrase ‘*event regularities, probabilistically specified*’, to refer to the kind of associations identified by statistical techniques such as regression analysis and MA. (Fleetwood 2016).

### ***Theory***

A theory is often said to have a *predictive* dimension containing statements delivering predictions such as ‘ $y$  will follow  $x$ ’; and an *explanatory* dimension containing statements delivering ‘explanation’ which, amounts to the same thing. ‘Theory’, then, becomes reduced to a set of statements designed to enable predictions, usually, in the form of hypotheses. We describe this as ‘theory’- i.e. with scare quotes - because, in our example, a ‘theory’ that explains an increase in organisation performance, is reduced to a statement to the effect that ‘a bundle of HRM practices were introduced’. Whilst other information, perhaps identifying the relevant causal agentially enacted mechanisms, is sometimes included, it is, strictly speaking, not necessary. This is sometimes referred to as ‘ultra-empiricism’ or ‘measurement without theory’.

### ***Prediction and (lack of) explanation***

---

<sup>8</sup>Or conditional expectation, conditional expected value or conditional distribution. The approach is consequently often referred to as the ‘average economic regression’ approach (Downward 2015: 2011).

Prediction for quasi-empiricism is based upon induction from past regularities in the flux of events. This conflates prediction and explanation. This illicit conflation is commonly referred to as the ‘symmetry thesis’, whereby the only difference between explanation and prediction relates to the direction of time (i.e. if x predicts y, then x is said to ‘explain’ y). For example, if the introduction of teamworking was found to predict an increase in profitability, then the former, would be said to ‘explain’ the latter. This conflation manifests itself in the way independent variables are commonly referred to as ‘explanatory variables’, and/or ‘predictors’ of the magnitude of dependent variables. This is, however, a misconception. Imagine that a regression analysis identifies an association between the introduction of teamworking and an increase in profitability, or put another way, imagine that the introduction of teamworking *predicts* the increase in profitability. Is anything explained by this? The answer is no. A prediction, even a successful one, *explains nothing*. A regression analysis, even one that successfully identifies an association between independent and dependent variables, does not reveal *why* the association comes about and, therefore, lacks *explanatory power*.

### **Summary**

The lack of explanatory power in individual quantitative empirical studies, rooted as they are in quasi-empiricist meta-theory, is the result of their commitment to the particular *chain of meta-theoretical concepts* – i.e. an ontology of events or states of affairs; causality as event regularity; epistemology based upon identifying event regularities probabilistically specified; a method of building theoretically closed systems to engineer the event regularities that generate predictions to be tested qua hypotheses; and theory as sets of statements that ‘set up’ the event regularities as predictions, which are then conflated with ‘explanations’. As these studies cannot generate explanations, they cannot generate *theoretically informed and empirically substantiated explanations either*. Unfortunately, this problem is not restricted to individual quantitative empirical studies but, as we will see below, carries over into MA more generally.

We envisage two potential responses from advocates of MA. First, they might find, demand, or carry out individual studies, including quantitative studies, which do have explanatory power, or insist on them being used as the appropriate basis for MA. Second, they might counter-argue that MA does not lack explanatory power. All MA has sections referred to as

'theory', 'literature review', 'hypothesis building', or some such, and it is in these sections that theoretically informed and empirically substantiated explanations can be found. Unfortunately, these responses will not work. Apropos the first response, whilst qualitative empirical research is essential in the search for theoretically informed and empirically substantiated explanations (Ackroyd and Karlsson 2014; see also Edwards et al., 2014), it is precisely this material that is excluded from MA:

'weed out all those papers that do not report data ... as well as those studies that are based on the analysis of qualitative data (e.g., ethnographies...and case studies). Getting rid of these types of papers is straightforward (Ellis 2005: 98).

Furthermore, in order to find or carry out quantitative studies that *do* have explanatory power, they would have to be rooted in an alternative meta-theory, one not committed to the chain of meta-theoretical concepts noted above. Yet, quantitative empirical researchers cannot just abandon their commitment to this or that meta-theoretical concept, because these concepts only 'work' as a complete package. The alternative, which we propose later, is that we should abandon this entire chain of meta-theoretical concepts, and replace it with an alternative.

### ***3.1. Illustrating the meta-theoretical problems with MA***

To illustrate our critique, we have selected two recent, highly-cited examples of meta-analyses, published in top ranked journals in the authors' areas of interest. In the first paper, Reichl et al., (2014) explore the relation between work-nonwork conflict and burnout by conducting a meta-analysis of 86 relevant studies, which allows for an analysis of 220 coefficients. In the second, Subramony (2009) explores the relationship between bundles of HR practices and specifically defined organizational outcomes. This is achieved through a MA of 65 relevant studies, which allows for an analysis of 239 separately reported effect sizes. Both are examples of 'best practice' MA and the criticisms we raise apply to all the examples of MA we are familiar with.

*Reichl et al.*

First, in a (half-page) section entitled 'theoretical framework', Reichl et al. mention 'several theoretical reasons to expect relations between work–nonwork conflict and burnout' (p.982-3). After a very short theoretical discussion they refer the reader to six sources where, presumably, the theoretically informed and empirically substantiated explanations informing their MA might be found. Further inspection, however, reveals this not to be the case. One study is just another MA; two are 'standard' quantitative studies seeking empirical regularities; three offer theoretical insight, but are *not* qualitative empirical studies, and two are extremely dated. Their 'theoretical framework' section, then, offers little or nothing in the way of theoretically informed and empirically substantiated explanation.

Second, Reichl et al.'s MA tells us that work–nonwork conflict was correlated with emotional exhaustion and cynicism, but these relations were moderated by gender, age, family characteristics and cultural norms. They are aware of "important gaps in our knowledge about underlying processes [i.e. causal mechanisms] and moderating variables" (pg 980), and their remedy is to obtain "theoretically derived moderators" - i.e. to theoretically derive the moderating causal mechanisms. Whilst this looks like a potential source of theoretically informed and empirically substantiated explanation informing their MA, further inspection reveals this not to be the case. Apropos the moderator variable gender: one study is a theoretically informed quantitative analysis; four are 'standard' quantitative studies, despite one having 'multi-method' in the title; and two are dated. Concerning the moderator variable family characteristics: two are 'standard' quantitative studies and two are meta-analyses. For the 'age' variable, there is only a 'standard' quantitative study. Apropos the moderator cultural norms: four studies are 'standard' quantitative studies; one is another MA; and two are overviews/reviews. None of these references offer the kind of theoretically informed and empirically substantiated *explanation* that would be needed to derive the moderating causal mechanisms. This point is developed in more detail later.

*Subramony (2009)*

Let us turn our attention now to the other example of MA: Subramony's paper on HRM bundles and firm performance.

[HRM] bundles consisting of multiple complementary practices are typically considered superior to individual best practices in influencing firm performance.

This study investigates the relationship between three such bundles (empowerment, motivation, and skill-enhancing) and business outcomes.... Although it makes conceptual sense to categorize individual HRM practices into these bundles, there is *insufficient empirical evidence regarding both their proposed synergistic properties and the magnitude of bundle effects on firm performance*. I propose to bridge this gap in the strategic HRM literature by investigating the relationship between empowerment, motivation, and skill bundles and various business outcomes; *clarifying the synergistic properties of these bundles* by comparing their effects to those of individual HRM practices; and demonstrating the usefulness of these bundles in relation to high-performance work systems (HPWSs) (Subramony 2009: 745-6, emphasis added)

To say that there is insufficient empirical evidence regarding ( $\pi$ ), the proposed synergistic properties of bundles of HRM practices, is entirely correct. Subramony's observation that there is insufficient empirical evidence regarding ( $\Omega$ ), the magnitude of bundle effects on firm performance has valid and invalid elements. It is *invalid* in the sense that there are actually many quantitative empirical studies seeking to identify the magnitude of bundle effects on firm performance. It is, however, *valid* in the sense that what evidence there is does not support the existence of the statistical association he believes exists. Subramony proposes to 'bridge this gap' by (a) investigating the relationship between these bundles and business outcomes; (b) clarifying the synergistic properties of these bundles by comparing their effects to those of individual HRM practices; and (c) demonstrating the usefulness of these bundles in relation to HPWS. Notice, however, that there are two 'gaps' - i.e. ( $\pi$ ) and ( $\Omega$ ). At best MA can deal with ( $\Omega$ ) by engaging in (a) and (c). What MA cannot do, however, is deal with ( $\pi$ ) via (b). It cannot bridge the gap of insufficient empirical evidence regarding the proposed synergistic properties of bundles of HRM because to do this would require theoretically informed and empirically substantiated explanations of *why* empowerment, motivation, and skill-enhancing HRM practices cause increased performance. MA cannot get anywhere near delivering explanations of this kind.

This said, as with Reich et al., Subramony's MA is not entirely devoid of theoretically informed explanations - although few of them are empirically substantiated. He writes:

The combination of multiple empowerment-enhancing practices...is likely to be synergistic because of the potential complementarities among these practices. For instance, allowing autonomous work teams to manage the production of a component or provision of a specific service can enhance employees' sense of responsibility and autonomy within the constraints of their work role. Additionally, the provision of voice and upward feedback mechanisms can help employees view themselves as part of a larger organizational system, leading them to engage in discretionary behaviors, including suggesting improvements to the products, services, or work processes; assuming increased responsibilities; and volunteering (e.g., serving on joint management-worker task forces). Also, the presence of multiple empowerment-related practices is likely to signal a coherent organization wide commitment to employee empowerment, which is likely to result in reciprocation in the form of in-role and extra-role behaviors (Subramony 2009: 748).

Subramony's brief explanations for the existence of synergies are not unreasonable, but any competent researcher in this field could come up with reasonable explanations about dis-synergies. The fact is, we do not really know which is the case because there are insufficient theoretically informed and empirically substantiated explanations of the proposed synergies. Moreover, MA brings us no closer to obtaining these explanations because it focuses our attention on identifying statistical associations, such as that underlying 'Hypothesis 1: Empowerment-enhancing bundles [of HRM practices] will be positively correlated with business outcomes' (Subramony 2009: 748).

Let us consider an example of how statistical techniques used in MA can often lead us further into obscurity.

By calculating the composites of relevant effect sizes within each study, I created the empowerment, motivation, and skill bundles. For instance, if a given study provided correlations between training and productivity and selection and productivity, a single composite score was created to reflect the combined effect of both the skill-enhancing practices of training and selection on productivity (Subramony 2009: 753).

Subramony takes past research showing correlations between training and productivity, and selection and productivity, and combines them into a single composite score reflecting the combined effect of training and selection on productivity. Whatever the advantages of doing this are, they have to be weighed against the dis-advantages. And the main disadvantage is this; to know that there are correlations between training and productivity, and selection and productivity is not to explain anything – i.e. we remain in the dark as to why these correlations come about. But then to combine them into a single score reflecting their *combined* effect on productivity leaves us with an even more complex statistical association about which we actually understand even less. *We are moving further away from generating theoretically informed and empirically substantiated explanations, not getting closer to them.*

Note that this has nothing to do with missing moderating or mediating variables. Indeed, if it turned out that additional moderating or mediating variables were needed, the problems would get even worse: we would end up with yet more variables, and yet more associations between them, and be no closer to deriving theoretically informed and empirically substantiated explanations.

Moreover, what can be done, practically, with Subramony's argument that:

‘firms can benefit from the adoption of high-performance HRM practices... as long as these practices also are complementary. Thus, instead of simply increasing the number of HRM practices...firms could derive positive returns by enhancing synergy among these practices (Subramony 2006: 759).

The only way this finding could be of substantive, or practical use, would be if it enabled an HR manager to successfully predict (solely on the basis of past event regularities) that the implementation of a bundle of complementary high-performance HRM practices would be followed by increased organisational performance in some future period. Even if an HR manager was prepared to implement the bundle, s/he would need to know a great deal more about how exactly to ‘enhance synergy among these practices’ than can be provided in such research. For the practitioner, therefore, these exhortations require a peculiar leap of faith that diminishes their own insight, experience and expertise. The implication is that the HR manager should sacrifice any experienced insight as to *why* certain HR practices may, or may

not, work in their own context, and instead follow the numbers. One is left feeling that the HR professional might be usefully replaced with an algorithm.

## Summary

Section 3 established that the lack of explanatory power that characterises individual quantitative studies, rooted in quasi-empiricist meta-theory, is the result of its commitment to a particular chain of meta-theoretical concepts. *Unfortunately, this problem carries over into MA.* Despite MA having dedicated 'theoretical' sections, the latter carry little in the way of explanation and, therefore, can offer little prospect of theoretically informed and empirically substantiated explanations. The explanations contained in MA are as lacking in explanatory power as the individual quantitative studies upon which they are based. Allow us to make the point more forcefully: if one individual quantitative study lacks explanatory power, then synthesizing scores of them does not increase the explanatory power.

Does this mean that all attempts to synthesise existing research are doomed to failure? We think not, but only if we turn to an alternative approach, that we call critical realist synthesis (CRS), that is rooted in an entirely different meta-theory. It is to this that we now turn.

## 4.0 A critical realist alternative<sup>9</sup>

In order to see exactly where CRS differs from MA, we present CR's chain of meta-theoretical concepts, in the same format as we did for quasi-empiricism in part 3.

### *Ontology*

As well as the actual and the empirical (Table 1) CRs recognise the existence of the 'deep' (Table 2). This stratified ontology is also emergent, meaning that entities existing at one 'level' are rooted in, but irreducible to, entities existing at another 'level'. For example, the social is rooted in, but irreducible to the biological, which is rooted in, but irreducible to the chemical, which is rooted in, but irreducible to the atomic, and so on. (Elder-Vass 2010). *Social* reality is also *transformational*; Agents reproduce or transform a set of pre-existing mechanisms. Society

---

<sup>9</sup>For an introduction of how we might go about gaining these insights, see Dirpal (2015).

continues to exist only because agents reproduce or transform the mechanisms that causally condition their social actions.

<b>Domain</b>	<b>Entity</b>
Empirical	Experiences & perceptions
Actual	Events & actions
'Deep'	Structures, mechanisms, tendencies, powers

*Table 2. A stratified or laminated ontology*

In a social world, characterised by stratification, emergence, transformation, and, typically, configurations of interacting causal mechanisms, it is unsurprising to find partial, approximate, rough-and-ready regularities or patterns in the flux of events. Following Lawson (1997, 2003: 81-83 and 105-7), we refer to these as '*demi-regs*', which can be styled as 'whenever event x, *then sometimes, but not always* event y'; for example, 'women *sometimes, but not always*, look after children more than men'. A system wherein *demi-regs* predominate, is an open system. Thus, whilst any explanations CR generate should 'fit' with the statistical record, the statistical record explains nothing in itself (see also Porpora, 2015).

### ***Epistemology***

With the recognition that events do not often manifest as regularities and that *something* must govern an irregular flux of events, the emphasis of CR investigation switches from the domains of the empirical and actual to the deep: the causal mechanisms that govern the flux of events. For example, we noted the claim that the introduction of PRP increases productivity at the firm level by 9% (op. cit). CRs might re-focus attention towards the mechanisms by which the relationship might occur: the motivation of some, but not all, individuals to increase their effort towards those metrics which are being measured, or the impact of labour relations on such motivation. In weighing up explanations, CR's accept the possibility of judging between competing claims because they reject the claim that to accept epistemic relativism is to accept *judgmental* relativism. That said, there is no gainsaying the difficulty involved with this,

especially when such judgement requires far more than simply carrying out statistically-based hypothesis testing to see which competing theories have greater explanatory power<sup>10</sup>.

### ***Aetiology***

The parts of the social world not characterised by event regularities (i.e. open systems) are still governed by something. This something cannot be a law as this would produce constant regularities. So instead, CRs use the term tendency to depict *the (transfactual) way of acting of a thing (or things) with properties* (Fleetwood 2009). A tendency is not an empirically observable pattern as a tendency can be in play and yet not manifest itself empirically, as it be counteracted by other mechanisms (Fleetwood 2012: 13).

To illustrate causation, CRs seek what Hesketh & Fleetwood (2010) refer to as 'thick explanation' - i.e. the kind of explanation that requires *hermeneutic* information - i.e. information relating to a range of human cognitive activities such as understanding, intention, purpose, meaning, interpretation, reason and so on. We do not, however, know what the cause of the action is, one does not understand it, until we know the intention that underlies it, that is, until we know *why* the agent did what s/he did. If, to explain an action is to give a causal account of it, then to explain an action is to give an account of *why* the actor did what s/he did. Whilst exploring motivations is always difficult, these can be explored using interviews. Sims-Schouten & Riley (2014) and Smith & Elger (2014) show, for instance, how interview-based research facilitates the probing of agent's own understandings of causal relations in organizational contexts.

### ***Methodology***

As the social world is an open system, mechanisms *cannot* be induced or deduced, but must instead be *retroduced* and *retrodicted*. Retroduction 'consists of a movement [...] from the conception of some phenomenon of interest to a conception of some totality or thing, mechanism, structure or condition that is responsible for that given phenomena' (Lawson, 2003: 145). It usually involves asking a specific kind of question: 'What thing, if it existed,

---

<sup>10</sup> That said, it would be remiss of us not to point out that CRs have no single criterion for identifying what constitutes explanatory power, and the concept is in need of further elaboration (Lawson 2003: chapter 4; Fleetwood & Hesketh 2010: chapter 6).

might account for the existence of P'? and might end up identifying Q as the thing in question. Retrodiction is used when we are relatively ignorant about the mechanisms in operation that are causing the phenomena under investigation. When there is little or no existing theory to act as a guide, we must take a voyage of discovery, make hypothetical conjectures, requiring the 'scientific imagination' (See Lewis 1999). We use what we do know to explain what we do not know.

### *Open and closed systems*

In *open* systems, theoretically informed claims must be framed in transfactual terms. Transfactual claims cannot, however, be empirically substantiated by testing quantitative hypotheses. Consider two hypotheses: the first is typical of quasi-empiricism and the second is transfactual.

*H<sub>1</sub> Workers assembled into a team increase profit*

*H<sub>2</sub> Workers assembled into a team tend to increase profit*

The intuition underlying H<sub>1</sub> is something like 'workers assembled into a team raise the probability that profit will increase'. This presumes the existence of a (ontic) stochastic regularity, which can be re-conceptualised probabilistically, between assembling workers into a team, and the resulting increase in profit. H<sub>1</sub> can be tested using 'normal' statistical techniques. In complete contrast, the intuition underlying H<sub>2</sub> is something like 'workers assembled into a team have the causal power to raise profit, but sometimes this power is actualised and sometimes it is not'. This gives rise to a demi-regularity, rather than a stochastic regularity, and thus cannot be re-conceptualised probabilistically. H<sub>2</sub> cannot be tested using the 'normal' statistical methods rendering quasi-empirical methods such as MA unsuitable for open-systems theorising.

### *Prediction and explanation*

CRs hold that, in open systems 'thick' explanation is probably our only guide to the future. If, for example, one can uncover, and explain, the causal mechanisms (e.g. HR practices) that, when drawn upon by workers and managers, increase organisational performance, then one has an explanation of the increase in performance. Such an explanation would allow one to

understand the tendencies generated when workers and managers engage with HR practices. If one understands these tendencies one can make *tendential predictions*.

Importantly, and in contrast to the empiricist tradition, which focuses only on what actually happens, powers or tendencies for critical realist are *transfactual*, and therefore point to the *potential* of entities. Thus, given the appropriate context (i.e. products, production regimes, labour relations systems) one mechanism may have more potential to increase performance than another, even if this potential is continually negated by countervailing tendencies. This is important because, unlike MA, it points to theorising the possibilities of future social events, caused by agentially enacted mechanisms, even if these events have *not* occurred in the past.

### ***Theory***

For CRs, theory consists of statements that deliver causal explanations. We can illustrate this by returning to our previous example: if we want to explain the tendency for team-working to increase productivity, we might look to existing theory about the relations within teams, seeking to develop new insights about (i) exactly how teams (as bundles of causal mechanisms) raise productivity; (ii) how agents are engaged with them; and (iii) the complex set of interactions between the bundles themselves and between the agents.

### ***Summary***

In sum, a CR chain of meta-theoretical concepts can be contrasted to that of quasi-empiricism in which MA is rooted (Table 3).

	<b>Quasi-empiricism</b>	<b>Critical realism</b>
<b>Ontology</b>	Atomistic, observable, events No agency-structure approach, only rational agents as individuals.	Social world is stratified or laminated, emergent, transformational, systemically open, complex, becoming, processual & relational.

		Agents & structures (i.e. mechanisms) are distinct but related.
<b>Epistemology</b>	Knowledge derives from (a) observing (b) event <i>regularities</i> .	Knowledge derives from uncovering causal mechanisms. Epistemically but not judgementally relativist.
<b>Aetiology</b>	Humean: causality as event regularity. Laws, law-like relations & functional relations.	Separates Humean causality from causality as powers & tendencies. Powers & tendencies replace laws, law-like relations & functional relations.
<b>Methodology &amp; research technique</b>	Some version of the covering law method Engineering event regularities via theoretically closed systems. Exclusively quantitative data & techniques such as regression, analysis of variance, factor analysis & meta-analysis	Causal-explanatory. Explanation comes via uncovering & understanding causal mechanisms. Ethnography, participant observation, in-depth interviewing critical discourse analysis, action research, archaeology, deconstruction & genealogy accepted. Mainly uses qualitative techniques, but the role of (some) quantitative techniques is debated.
<b>Objective</b>	Prediction. To construct & test predictions & hypotheses to establish whether claims are true or false.	Explanation. Claims must be empirically substantiated, but quantitative hypothesis testing is inappropriate.
<b>Explanation</b>	Explanation confused with prediction.	Explanation is 'thick' - an account of the operation of causal mechanisms. Explanation is confused with prediction.
<b>Prediction</b>	Prediction confused with explanation. Prediction based on inductive generalisations. Spurious precision.	Tendential prediction based on knowledge of causal mechanisms. Tendential prediction is not precise, but not spurious either.
<b>Theory</b>	Vehicle for delivering predictions	Vehicle for delivering causal-explanatory accounts.
<b>Mode of inference</b>	Deduction & induction	Retroduction & retrodiction

*Table 3 comparative aspects of Quasi-Empiricism and Critical Realism*

With the meta-theoretical framing completed, we can now take the first tentative steps to show how the above CR meta-theory might be used to guide CRS.

## 5.0 Critical realist synthesis (CRS)

In order to help generate theoretically informed and empirically substantiated explanations, we have amended Pawson’s (2004) realist review process, designed for policy interventions, to deal with synthesis more generally (Table 4). As one’s approach to CRS will vary considerably depending on its purpose, the steps in this table are not meant to be sequential, compulsory or exhaustive, but instead provide a broad steer that is intended to guide CRS.

Define the scope of the synthesis	Identify the focus / question	<ul style="list-style-type: none"> <li>• Who are the agents involved?</li> <li>• What is the intervention / mechanism being studied?</li> <li>• What are the contexts for its use?</li> <li>• What are the intentions of those who use it?</li> <li>• What is its intended impact?</li> </ul>
	What mechanisms are assumed?	<ul style="list-style-type: none"> <li>• Search for assumed mechanisms / theories in the literature / policies</li> <li>• Group, categorise or synthesise theories</li> <li>• Design set of theories / mechanisms to be explored / tested</li> </ul>
Search for and appraise the evidence	Search for the evidence	<ul style="list-style-type: none"> <li>• Decide and define review strategy</li> <li>• Define search sources, terms and methods to be used</li> <li>• Set the thresholds for stopping searching at saturation</li> </ul>
	Appraise the evidence	<ul style="list-style-type: none"> <li>• Does the research address the key mechanisms being studied?</li> <li>• Does the research support the conclusions it makes?</li> </ul>
Extract and synthesise findings	Extract the results	<ul style="list-style-type: none"> <li>• Identify important mechanisms, contexts, entities, stratification etc.</li> <li>• Capture these in tables / templates</li> <li>• Undertake retroduction to identify further mechanisms.</li> </ul>
	Synthesise findings	<ul style="list-style-type: none"> <li>• What are the common mechanisms, entities, properties and outcomes?</li> <li>• Produce ‘thick’ descriptions of the operation of the mechanism(s)</li> <li>• How and why do these vary in different contexts?</li> <li>• How do these compare to the ‘assumed’ mechanisms?</li> </ul>
	Draw conclusions	<ul style="list-style-type: none"> <li>• What are the tendencies of the key mechanisms?</li> <li>• What are the key contextual factors which have an impact on the operation of the mechanism(s)</li> </ul>

*Table 4 The process of CRS*

We now explore each step in more detail.

### *Define the scope of the synthesis*

The scope of CRS will often be in the form of a question like: ‘how does mechanism M, when enacted by agent A, tend to alter outcome O?’. This approach will also work for clarification questions, such as ‘what are the properties of mechanism M?’ or ‘why does outcome O often occur in context P’. As CRs accept systemic openness, CRS is not only restricted to providing a ‘thick’ explanation of the agentially enacted mechanism, but also of the different contexts in which the mechanism might generate a tendency to O - or qualitatively different versions of O - as well as the unintended consequences of mechanism M’s tendency. This also permits consideration of the ways in which the outcome might react back, in a later time period, on mechanism M and agent A.

In clarifying the purpose of the review, it is also useful to know the mechanisms which are claimed or assumed in the relevant literature or policy. For example, in exploring the question *do bundles of HRM practices improve performance?*, it might be useful to identify the assumptions which are made, or theories that are drawn upon, when this is presumed to be the case, as these provide a focus for the review which can be explored systematically later. This step is not always necessary.

### *Search for, and appraise the evidence*

As CRS focuses on identifying agents and mechanisms, it need not restrict itself to statistical studies or indeed studies from any specific discipline, including CR. For example, O’Mahoney (2011) reviews the social constructionist identity literature, much of which explicitly rejects realism, to retroduce the entities, powers, and mechanisms involved with identity construction. Indeed, the ecumenical nature of CR’s review of the literature allows it to use this breadth to identify similar causal mechanisms working in a variety of contexts. Moreover, as Ackroyd and Karlsson suggest, the CR researcher is marked by their ‘eclecticism’ when it comes to matching innovative methods to collect new data indicative of the existence and character of causal mechanisms (2014:22). For example, Pawson (2005) – which is expanded upon below - seeks to understand the impact of ‘public disclosure’ on recalcitrant behaviour (i.e. does ‘naming and shaming’ work?). To investigate this, he drew on a wide range of public disclosure policies, from Megan’s Law and school league tables, to

hospital star ratings and naming prostitute's clients. This allows exploration of similar mechanisms but in very different contexts, permitting the identification of the particular contexts which were more likely to generate a tendency for disclosure to affect behaviour.

In collecting studies, quantitative work (e.g. those using regression analysis to identify statistical associations) should be treated with caution. Instead of dismissing them, however, we would check to see if, *in addition to* the (non-explanatory) statistical data, there is something that might help us to create theoretically informed and empirically substantiated explanations. Instead, we are far more favourably disposed to past *qualitative* empirical research. In both cases (and recalling section 3.0) we would be asking ourselves: Does this past research help us deepen our understanding of the appropriate agents and mechanisms, how agents and mechanisms interact, and the other mechanisms (i.e. 'the context') that dispose this agent to interact with one mechanism and not another.

CR accepts that different disciplines may use different terms to describe similar mechanisms – though where these terms differ they may be more or less accurate. For example, 'enculturation', 'socialisation', 'institutionalisation', 'indoctrination', 'learning', and 'disciplining', might be used in different traditions to describe the ways in which societies inform and (re)create the individuals which inhabit them. Such terminological diversity should always be critically appraised, as such terms are not apolitical, and for example, can range from strong managerialism ('workers can learn to be more efficient') to critical ('workers are indoctrinated through induction programmes'). However, such diversity should also be embraced as providing potentially useful alternative perspectives on how the agentially enacted causal mechanisms operate and relate. More specifically, terms captured in the review may operate at different (sometimes emergent) levels - for example, socialisation and learning may be different (and related) forms of indoctrination.

In short, then, a CR review of a subject should cast its net wide, searching not only for key words (such as HRM, bundles and performance) but also examining historical texts, and different disciplines for similar mechanisms that may have operated in different contexts. Thus, for the literature review, the search terms and sources would usually be wider than we might expect in a standard structured literature review (Tranfield et al. 2003). Once the relevant literature has been collected, and this would usually be an ongoing process, it needs

to be appraised, both in terms of ensuring the research actually addresses the mechanism(s) under study, and its internal validity - that the data actually supports the conclusions it makes.

### *Extract and synthesise findings*

The purpose of a CRS analysis is to identify the agentially enacted causal mechanisms by reviewing extant literature. In CR-oriented studies, these mechanisms will often (though not always) be explicit. However, when reviewing non CR-oriented literature the analysis can often proceed in two steps. The first is to identify the agentially enacted mechanisms stated within the literature, and the second is to retroduce and retrodict further conditions of possibility for these. It is important to note that CRS does not require a rejection of any research that is not CR in orientation. Let us consider two examples of this latter point.

First, O'Mahoney's (2011) review, mentioned above, takes the statements of ostensibly anti-realist authors concerned with identity construction to identify the key agentially enacted mechanisms involved. For example, he draws upon an article by Thomas and Davies (2005) that details how Kate, a personnel manager in the police service, draws on discourses of femininity and parenthood to resist performative employment discourses. O'Mahoney's first step is to identify the context stated explicitly by the authors, including Kate herself, her job, the Police, and the various discourses which, for CRs, are causal mechanisms. The second step is to retroduce implied mechanisms and powers, such as the power of the Police service to employ and discipline workers, and terminate contracts, and the (agential) power of Kate to learn skills and reproduce them. We also learn much about discourses – e.g. that they can be resisted, that some discourses (such as femininity) exist in tension with others (such as masculinity), and that individuals exercise some form of free-will in choosing to engage with or resist them. This allows O'Mahoney to argue that, contrary to the anti-realist protestations of social constructivists, their research can contribute towards the kind of theoretical informed and empirically substantiated explanations sought in CRS.

Second, in seeking to understand the impact of 'public disclosure' on recalcitrant behaviour Pawson (2005) examines the policy literature to identify the common mechanisms which are specified (Figure 1).

Insert figure 1 here.

He then reviews the history and operation of the various disclosure policies to identify when the mechanisms lead to positive and negative outcomes (Figure 2). For example, he notes that,

‘Megan’s Law swept onto the statutes following the enormous public outcry at the brutal death [of a child]. The courts responded to the wave of sentiment that ‘something must be done’ and were thus able to brush aside the constitutional challenges forwarded by minor lobbies’ (p. 39).

Insert figure 2 here

Here, Pawson uses contrastive theory building to identify patterns rather than laws (we would say ‘demi-regs’) about the potential of public disclosure policies to achieve their aims. The more tentative and less certain language here is also worth noting, especially in contrast with the ‘9%’ of meta-analysis detailed earlier.

‘Although popularly known as ‘naming and shaming’, public disclosure outcomes in these cases do not seem to depend, in the long term anyway, on the dishonour of the culprits. ... Public disclosure is meant to change behaviour – but seems effective only in relation to what organises that behaviour in the first place. What is more, in each of the [cases], it is the information providers rather than the public who are the key agents of change’. (p. 44).

In terms of synthesis, Pawson takes a comparative approach to identifying the mechanisms which link X and Y, and provides a ‘thick’ description of how and why these work in different circumstances.

Whilst no-one has yet carried out an explicit CRS, some CRs have implicitly started to go down this route. Three can be cited as examples. First, Fleetwood’s (2014, 2016) attempt to build a CR-oriented alternative model of labour markets draws upon a body of existing

theoretical and empirical research, that he refers to as 'the socio-economics of labour markets'. The key point to note is Fleetwood's rejection of existing quantitative empirical research that is rooted in quasi-empiricism because it contributes little or nothing to the generation of theoretically informed and empirically substantiated explanations of the way labour markets work. In contrast, Fleetwood accepts the 'socio-economics of labour markets' because it consists of existing qualitative empirical research that contributes to the generation of theoretically informed and empirically substantiated explanations of the way labour markets work. Whilst implicit, Fleetwood's work on labour markets might be thought of as a rudimentary CRS.

Second, Vincent's (2011) work, in this journal, focuses attention on emotional experiences at work, the organisation control mechanisms which seek to influence these experiences, and how different contextual conditions ( $M^c$ ) affect both organisational control systems and worker experiences. Whilst the paper is not explicitly either CR or CRS, it offers a form of analysis that is highly consonant with the approach outlined here. The paper maps the structural conditions and agential dispositions which affect emotional displays at work, and how these combine to explain experiences. It highlights, in particular, how employers' regulation and rewarding of workers' emotional displays interacted with workers conformity (or not) with organisational interests and rule systems. The paper then considers the contextual conditions that impel different types of control system and experience, for example, by highlighting the circumstances in which workers are rewarded for specific emotional displays at work. Overall, this paper contributes by developing theoretically informed and empirically substantiated insights about the way emotions are managed, experienced and enacted at work, offering another rudimentary CRS.

Third, Dirpal (2015) starts from the position that past quantitative empirical research on the HRM-Performance cannot *explain why* HRM practices are linked to performance. He *re-theorises* HRM practices to develop the concept of an 'HRMechanism' (i.e. HRM practice + causal mechanism) before applying *qualitative research techniques* to investigate what would normally be considered a *quantitative* research programme. Thus he offers a (meta) theoretically informed piece of qualitative research into six HRMechanisms: team working, corporate culture, empowerment, work-life balance, performance appraisal and reward. What makes Dirpal's research interesting for our purposes is how he uses past qualitative empirical research as a quasi-CRS. What he lacks, initially, is a sophisticated understanding of exactly

how HRM practices may or may not work to influence organisational performance. He turns to the existing literature to glean any theoretically informed and empirically substantiated insights, uses them to frame his interviews. He finds that team working, performance appraisal and work-life balance, generate powers/tendencies to *increase* organizational performance, whereas corporate culture, empowerment and rewards generate *neutral* powers/tendencies vis-à-vis organizational performance. Moreover, he generates *causal-explanations* of exactly what these HRMechanisms do to generate these powers/tendencies.

### *Aligning CRS and CR*

CRS is built upon the meta-physical claims of CR detailed in Section 4. In this section, we provide more detail about the alignment of our approach with specific methodological and theoretical applications of CR, namely Bhaskar's RRRE approach, and Lawson's contrastive explanation approach.

The aims of CRS are of course compatible with CR empirical or applied research. Bhaskar's (1998: 129) RRRE model, for example, suggests the following four steps for undertaking such work:

1. *Resolution* of a complex event into its components (causal analysis).
2. *Redescription* of component causes.
3. *Retroduction* to possible (antecedent) causes of components via independently validated normic statements.
4. *Elimination* of alternative possible causes of components

According to Collier, (1994: 163) 'RRRE has redescription as its second stage, indicating the presence of an already established stock of concepts, well enough defined....to justify using them for revisionary description'. We would add that the second and third step definitely, and perhaps the first and fourth also, would be impossible to take without existing knowledge and, therefore, without drawing upon existing research.

There are, however, two potential problems that we want to eliminate before proceeding. First, are we not simply 'making a virtue out of a necessity'? After all, RRRE or otherwise,

almost all empirical researchers start with existing research. What sets CRS apart, however, is that a great deal of meta-theoretical thought goes into identifying precisely the kinds of existing research that will be accepted and rejected; not anything 'goes'. Second, the same could be said of MA: not anything 'goes' for MA's either. Indeed, they accept existing quantitative research, and reject existing qualitative research. This is not, however, because MA's hold that quantitative research delivers theoretically informed and empirically substantiated explanations, but simply because only quantitative research can be analysed with MA's statistical toolkit. Thus, CR eliminates research that it holds to be theoretically flawed (and for other reasons) whereas MA is driven by a desire to employ specific set of statistical techniques. With these potential problems dealt with, we can turn to the issue of illustrating how CRs might use CR methods to guide CRS.

Those new to CR often complain about the abstractness of retrodution and retrodiction, and so it is important to consider how we can more easily deploy these approaches to extract new understanding in the context of CRS. In our view, and whilst far from being a point of departure, those wishing to employ CRS can aim towards Lawson's (1997, 2003, 2009) contrastive method. This approach compares ostensibly similar cases (e.g. specific countries, such as the UK or China; old or young workers; corporations or charities) to identify different or surprising demi-regs, generated by similar causal mechanisms, but calling attention to specific contextual features ( $M^c$ ) that interact to affect outcomes differently in otherwise similar circumstances. Thus, rather than explaining a single outcome (set of events  $E^n$ ), the objective is to account for some contrast " $P^n$ " rather than " $Q^n$ " and to use retrodution and retrodiction to identify the particular conditions which drove the outcome in particular direction. Arguably, by identifying our analytical target in terms of particular forms of difference, in worlds that are otherwise similar, the process of working out the particular mechanism that is causal, in one instance or another, becomes much simplified. This way, knowledge of causal mechanisms can develop incrementally by reflecting on unexpected contrasts in the existing stock of research.

Whilst contrastive explanation offers a viable strategy for knowledge development, as it focuses attention on the particular, getting any CRS inspired project to the point at which a contrastive strategy is possible typically involves a lot of ground-work (as illustrated in Table 4). However, as any CRS project assimilates the existing body of knowledge, in CR compatible terms, and approaches the point of analytical saturation (when it the review

exhausts what we know), it becomes increasingly possible to deploy a contrastive explanatory method. At this point, the project will understand the stock of related qualitative described cases and examples and the different conditions that explain demi-regs within these. As a consequence, CRS scholars will find themselves in a position to explain novel causal mechanisms that give rise to unexplained and unexpected events.

## **Conclusions**

The appeal of methods that allow for an integration and synthesis of existing research to produce more robust and, even, novel findings is obvious. It is in this context that MA has grown in popularity. Yet, even on its own terms, MA has a number of technical, procedural and practical problems that can limit its usefulness. More significantly for our argument, the meta-theoretical foundations of MA, which have attracted little, if any, critical comment, are flawed. We have argued that the lack of explanatory power that characterises individual quantitative studies, rooted in quasi-empiricist meta-theory, is the result of their commitment to a particular chain of meta-theoretical concepts. Unfortunately, this problem carries over into MA, meaning the explanations contained in MA are as lacking in explanatory power as the individual quantitative studies upon which they are based.

What then is left for MA? We have argued that regression analyses, and thus MAs, are not suitable for the open, emergent systems that typify organisational studies, or indeed, the social world generally. This is because the interaction of complex, emergent mechanisms in different contexts does not give rise to regularities in relations between events. Yet, for critical realists this does not mean jettisoning MA altogether. Two alternatives are proposed here. The first is that if MAs are not suited to open systems, then they are suited to closed systems, such as IT or the physical sciences, where empirical regularities between events exist. This raises an interesting question as to ‘whether some disciplines can be classified as ‘less open’ / ‘more closed’ on the basis that they concern themselves with simpler or less emergent systems. The answer to this question is contested and cannot be explored in detail here, but Fleetwood (2016) provides an overview of the key issues, arguing that systems are either open or closed.

The second approach is implied by Porpora (2015: 62):

Demoting regression analysis and other statistical techniques from explanation to evidence, critical realism has no reason to reject them as such....Statistics are employed to indicate the contingent operation of a mechanism in a particular context.

This shift in this framing of MAs implies that well-designed regression analyses (and therefore MAs) *can* provide indications that causality may be at work, or at least that phenomena require investigation. For example, research claiming to identify a statistical association between bundles of HRM practices and improved organisational performance, have prompted authors to carefully investigate the mechanisms and contexts that might sustain such an association (Fleetwood and Hesketh, 2006). Importantly, subsequent investigation can, and sometimes does, undermine claims deriving from these statistical associations. In the MOS literature, for example, quantitative research claiming to have identified an association between HRM practices (e.g. TQM, BPR or Lean) in high performing organisations, have been exposed by qualitative studies revealing flawed assumptions. Some qualitative research, for example, has suggested that reporting of these practices has been exaggerated by respondents (e.g. Collinson et al., 1998).

Our critique of MA led us to develop an alternative, CRS, which is driven by the objective of creating theoretically informed and empirically substantiated explanations. CRS, rooted in CR meta-theory and predicated upon the claim that the social world is characterised by demi-regs, requires a conception of causality that is not exhausted by regularities in the flux of events, but is understood at the relative push and pull of powers or tendencies. As we demonstrate in part four, this allows CRs to make *tendential predictions* and, thereby, generate substantive implications. We explained how CRS allows for insights to be incorporated from the widest possible source material, including qualitative research, social constructionist-oriented research and, with caution, some quantitative, empirical research. CRS resonates with work on systematic reviews by other realist scholars, such as Pawson, and thus contributes to debates already existing in social science, more generally, about how realist philosophical commitments might shape analyses.

Although we hold that CRS is a superior approach to that of MA, we note here that CRS does have a number of problematic features. First, the method of CRS is less formulaic than that of MA, putting more emphasis on the intuition (via retrodution and retrodiction) of the researcher. Moreover, the outcome of CRS is more complex than the single number generated

by MA, and perhaps therefore less attractive to some managers or policy makers. We would hope, however, that our proposal places an emphasis on the expertise and experience of these people in helping *understand* the complexities of the world in which they are embedded. Secondly, tendential predictions are only that. As Pawson et al. note: social interventions are so complex that there is little hope of reproducing them, and even if one could, they are so context specific that the same ‘assemblage’ may go onto misfire’ (2005:21). Thirdly, although we have pointed to examples of good practice in *parts* of a CRS (e.g. O’Mahoney 2011; Pawson 2005; Fleetwood 2014; Vincent, 2011), and explained how we would approach a CRS, we have not found an examples of a complete CRS. This is a gap that we would urge researchers to explore.

## References

Ackroyd, S. and Karlsson J.C. (2014) Critical Realism, Research Techniques, and Research Designs. In Edwards, P.K., O'Mahoney, J. and Vincent, S. (eds.) *Studying Organizations Using Critical Realism A Practical Guide*. Oxford: OUP.

Aguinis, H., Edwards, J.R. and Bradley, K.J., (2016) Improving Our Understanding of Moderation and Mediation in Strategic Management Research. *Organizational Research Methods*.

Bhaskar, R., (1978) On the possibility of social scientific knowledge and the limits of naturalism. *Journal for the Theory of Social Behaviour* 8(1): pp.1-28.

Bhaskar, R. (1993) *Dialectic: the pulse of freedom*. New York: Verso.

Bhaskar, R. (1998) *The Possibility of Naturalism*, London: Routledge

Bornmann, L. and Mutz, R., (2015) Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*. 6(11): pp.2215-2222.

Collier, A. (1994) *Critical Realism: An Introduction to Roy Bhaskar's Philosophy*, London: Verso.

Collinson, M., Rees, C., Edwards, P. K. & Inness, L. (1998) *Involving Employees in Total Quality Management: Employee Attitudes and Organisational Context in Unionised Environments*. Department of Trade and Industry: London.

Cowton, C.J., (1998) The use of secondary data in business ethics research. *Journal of Business Ethics*. 17(4): pp.423-434.

Coyne, J.C., Hagedoorn, M. and Thombs, B., (2011). Most published and unpublished dissertations should be excluded from meta analyses: comment on Moyer et al. *Psycho Oncology*. 20(2): pp.224-225.

Danermark, B., Ekstrom, M. and Jakobsen, L., (2002). *Explaining society: an introduction to critical realism in the social sciences*. Abingdon: Routledge.

Dirpal, G. (2015) *Human resource management practices and performance link: Applying critical realist meta-theory*. Unpublished doctoral dissertation, University of the West of England, United Kingdom.

Donaldson, L., 1996. *For positivist organization theory*. London: Sage.

Donaldson, L., 2003. Organization theory as a positive science. In Knudsen, C. and Tsoukas, H., (eds.) *The Oxford handbook of organization theory*. Oxford: OUP pp.39-62.

Donaldson, L., (2005) Vita Contemplativa Following the Scientific Method: How I Became a Committed Functionalist and Positivist. *Organization Studies*, 26(7): pp.1071-1088.

Downward, P. (2016). Regression analysis: a review. In Lee, F. and Cronin, B. *Handbook of Research Methods and Applications in Heterodox Economics*, 210-256. Elgar: London.

Eden, D., (2002). From the editors. Replication, MA and Scientific Progress, and AMJ's Publication Policy *Academy of Management Journal*. 45(5): pp.841-846.

Edwards, P., (2005) The challenging but promising future of industrial relations: developing theory and method in context sensitive research. *Industrial Relations Journal*, 36(4): pp.264-282.

Edwards, P. K., O'Mahoney, J. and Vincent, S. [Eds.] (2014) *Studying Organizations Using Critical Realism: A Practical Guide*, Oxford: Oxford University Press.

Elder-Vass, D. (2010). *The Causal Power of Social Structures: Emergence, Structure and Agency*. Cambridge: Cambridge University Press.

Ellis, P.D., (2010) *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge: Cambridge University Press.

Fiske, D.W., (1982) Convergent-discriminant validation in measurements and research strategies. *New Directions for Methodology of Social & Behavioral Science*, 12: pp77-92

Fleetwood, S. (2005). Ontology in organization and management studies: A critical realist perspective. *Organization*, 12(2): pp197-222.

Fleetwood, S., (2009) The ontology of things, properties and powers. *Journal of Critical Realism*, 8(3): pp.343-366.

Fleetwood, S., (2011) Powers and tendencies revisited. *Journal of Critical Realism*, 10(1): pp.80-99.

Fleetwood, S., (2012) Laws and tendencies in Marxist political economy. *Capital & Class*, 36(2): pp.235-262.

Fleetwood, S. (2014) 'Bhaskar and Critical Realism', Adler, P. Paul Du Gay, P. Morgan, G. Reed, M. *The Oxford Handbook of Sociology, Social Theory and Organisation Studies: Contemporary Currents*, Oxford University Press: Oxford.

Fleetwood, S. (2016) 'The Critical Realist Conception of Open and Closed Systems', forthcoming in *Journal of Economic Methodology*,

Fleetwood, S. & Hesketh, A. (2006) 'HRM-performance research: Under-theorized and lacking explanatory power.' *The International Journal of Human Resource Management*, 17(12): pp1977-1993.

Fleetwood, S. & Hesketh, A. (2010) *Explaining the Performance of Human Resource Management*. Cambridge: Cambridge University Press.

Franke, G.R., (2001) Applications of meta-analysis for marketing and public policy: A review. *Journal of Public Policy & Marketing*, 20(2): pp.186-200.

Gielen, A. C., Kerkhofs, M. J. & Van Ours, J. C. (2010) How performance related pay affects productivity and employment. *Journal of Population Economics*. 23(1): pp291-301.

Glass, G.V., (2000) *Meta-analysis at 25*. Available at: <http://www.gvglass.info/papers/meta25.html>

Hodgkinson, G.P. and Rousseau, D.M., (2009) Bridging the rigour–relevance gap in management research: it's already happening! *Journal of Management Studies*, 46(3): pp.534-546.

Hough, S.L. and Hall, B.W., (1994) Comparison of the Glass and Hunter-Schmidt meta-analytic techniques. *The Journal of Educational Research*, 87(5): pp.292-296.

Johnson, P. and Duberley, J., (2000) *Understanding management research: An introduction to epistemology*. London: Sage.

Kersley, B., Alpin, C., Forth, J., Bryson, A., Bewley, H., Dix, G. and Oxenbridge, S., (2013) *Inside the workplace: findings from the 2004 Workplace Employment Relations Survey*. London: Routledge.

Lawson, T. (1995) A realist perspective on contemporary “economic theory”. *Journal of Economic Issues*, 29(1): pp.1-32.

Lawson, T. (1997) *Economics and reality*. London: Routledge.

Lawson, T. (2003) *Reorienting Economics*. London: Routledge.

Lawson, T. (2009) 'Applied economics, contrast explanation and asymmetric information'. *Cambridge Journal of Economics*, 33: 405-419.

Linden, B. and Priestley, J., (2009) Critical reading of meta-analyses. *British Journal of Cardiac Nursing*, 4(7): pp.335-337.

Maslach, C., Schaufeli, W.B. and Leiter, M.P., (2001). Job burnout. *Annual review of psychology*, 52(1): pp.397-422.

Miller, G.A., (1987) Meta-analysis and the culture-free hypothesis. *Organization Studies*, 8(4): pp.309-326.

Mitchell, T.R., (1985) An evaluation of the validity of correlational research conducted in organizations. *Academy of Management Review*, 10(2): pp.192-205.

O'Mahoney, J. (2012) Embracing essentialism: a realist critique of resistance to discursive power. *Organization*, 19(6): pp723-74

Pawson, R. (2004). 'Evidence based policy: a realist perspective.' In B. Carter & C. New (eds.) *Making Realism Work: Realist Social Theory And Empirical Research*: 26-50. London: Routledge.

Pawson R, Greenhalgh T, Harvey G, Walshe K. (2004) *Realist synthesis: an introduction*. RMP Methods Paper 2/2004, Manchester, UK: ESRC Research Methods Programme, University of Manchester

Pawson, R., (2016) The ersatz realism of critical realism: A reply to Porter. *Evaluation*.

Porpora, D. V. (2015). *Reconstructing Sociology: The Critical Realist Approach*. Cambridge: Cambridge University Press.

Porter, S. (2015) The uncritical realism of critical realism. *Evaluation*, 21(1): pp65–82.

Psillos, S., (2002) *Causation and explanation*. Chesham: Acumen Press.

Pullen, A. and Simpson, R., (2009) Managing difference in feminized work: Men, otherness and social practice. *Human relations*, 62(4): pp.561-587.

Ragin, C. (1987) *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*, Los Angeles: University of California Press

Reichl, C., Leiter, M.P. and Spinath, F.M., (2014) Work–nonwork conflict and burnout: A meta-analysis. *Human Relations*, 67(8): pp. 979-1005.

Rossi, J.S., (1987). How often are our statistics wrong? A statistics class exercise. *Teaching of Psychology*, 14(2): pp.98-101.

Roth, P.L., Purvis, K.L. and Bobko, P., (2012) A meta-analysis of gender group differences for measures of job performance in field studies. *Journal of Management*, 38(2): pp.719-739.

Rousseau, D.M., Manning, J. and Denyer, D., (2008) 11 Evidence in Management and Organizational Science: Assembling the Field's Full Weight of Scientific Knowledge Through Syntheses. *The academy of management annals*, 2(1): pp.475-515.

Sayer, A., (2010) *Method in Social Science: Revised 2nd Edition*. Abingdon: Routledge.

Shirom, A., Gilboa, S.S., Fried, Y. and Cooper, C.L., (2008) Gender, age and tenure as moderators of work-related stressors' relationships with job performance: A meta-analysis. *Human Relations*, 61(10): pp.1371-1398.

Sims-Schouten, W. and Riley, S., (2014) 'Employing a form of critical realist discourse analysis for identity research: an example from women's talk of motherhood, childcare and employment' in P. Edwards, S. Vincent & J. O'Mahoney (Eds.) *Using critical realism to study organisations*: 109-31. Oxford: Oxford University Press.

Smith, C. & Elger, T. 2014. 'Critical Realism and Interviewing Subjects.' In P. Edwards, S. Vincent & J. O'Mahoney (Eds.) *Using critical realism to study organisations*: 109-31. Oxford: Oxford University Press.

Subramony, M., (2009) A meta analytic investigation of the relationship between HRM bundles and firm performance. *Human Resource Management*, 48(5): pp.745-768.

Thomas, R. and Davies, A., (2005) Theorizing the micro-politics of resistance: New public management and managerial identities in the UK public services. *Organization studies*, 26(5): pp.683-706.

Tourish, D., (2013) 'Evidence Based Management', or 'Evidence Oriented Organizing'? A critical realist perspective. *Organization*, 20(2): pp.173-192.

Tranfield D, Denyer D and Smart P. (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review'. *British Journal of Management* 14(13) pp207.

Vincent, S. (2011) 'The emotional labour process: An essay on the economy of feelings', *Human Relations*, 64(10): 1369-1392.

Wanous, J.P., Sullivan, S.E. and Malinak, J., (1989) The role of judgment calls in Meta-Analysis. *Journal of Applied Psychology*, 74(2): p.259-264.

Weed, M., (2005) " Meta Interpretation": A Method for the Interpretive Synthesis of Qualitative Research. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 6(1).

Figure 1

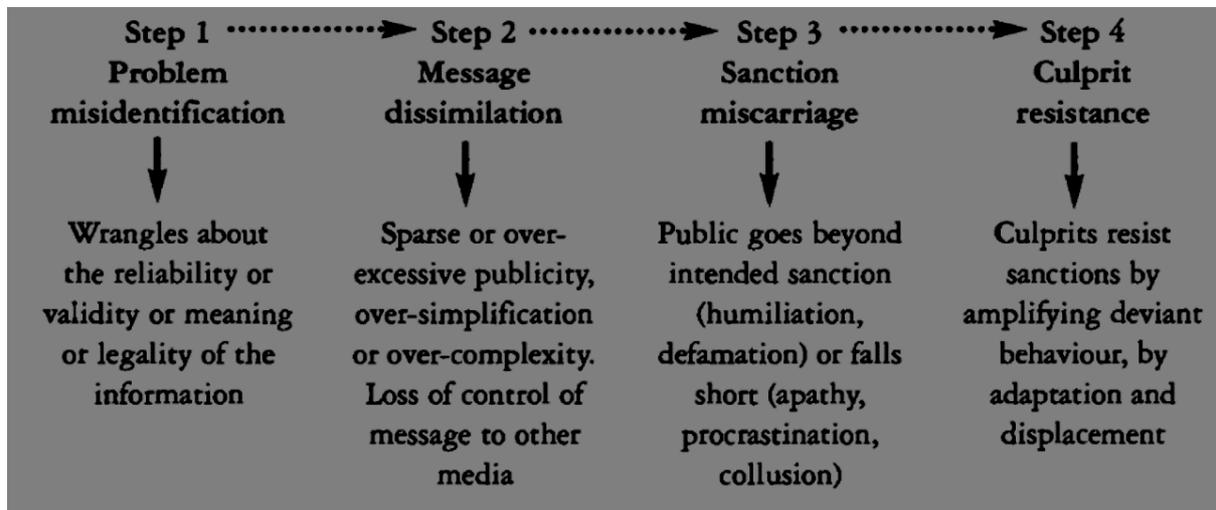


Figure 2

