Platform

The Realist Foundations of Evidence-Based Medicine: A Review Essay

Evaluation
2018, Vol. 24(1) 42–50

DOI: 10.1177/1356389017746718
journals.sagepub.com/home/evi
SAGE

Ray Pawson
University of Leeds, UK

> Cancer, as we now know, is a disease caused by the uncontrolled growth of a single cell. This growth is unleashed by mutations – changes in the DNA that specifically affect genes to incite unlimited cell growth. In a normal cell powerful genetic circuits regulate cell division and cell death. In a cancer cell these circuits have been broken, unleashing a cell that cannot stop growing. That this seemingly simple mechanism – cell growth without barriers – can lie at the heart of this grotesque and multifaceted illness is a testament to the unfathomable power of cell growth. Cell division allows us . . . to recover, to repair – to live. And distorted and unleashed, it allows cancer to grow, to flourish, to adapt, to recover, and to repair… Cancer cells grow faster, adapt better. They are a more perfect version of ourselves. (Mukherjee, 2010, p. 6, The Emperor of All Maladies: A Biography of Cancer)

My thanks go to the editor for allowing me to mount the soap box, once again, in promoting the cause of realist inquiry. I want to use the opportunity to begin to make the case that evidence-based medicine (EBM) is profoundly realist in its methodological underpinnings. This, as readers will appreciate, is not the usual story.

The standard EBM narrative . . . and rival accounts

The received picture of EBM is of a singular, assured approach built upon the foundations of the 'gold-standard' scientific method, namely the randomised controlled trial (RCT). Reliable and valid knowledge on intervention outcomes is considered to reside in this particular research design. It is protected in formal pronouncements on hierarchies of evidence, by supplementary methods of bias elimination and by close policing of research proposals, protocols and publications. Certainty is further established by the successful replication of trials, as when EBM capitalises upon meta-analysis, considered the most objective way of synthesising evidence from a growing body of rigorous trials. All ends, the orthodoxy goes, in an evidence base identifying sturdy causal regularities, which provide authoritative guidance to the practice community on what works (and what doesn't).

Corresponding author:
Ray Pawson, University of Leeds, Leeds LS2 9JT, West Yorkshire, England LS2 9JT, UK.
Email: r.d.pawson@leeds.ac.uk

So goes the Cochrane Collaboration narrative. Returning home to evidence-based policy (EBP) for a moment, I would suggest that there has been a deeply divided response to our mighty sibling. To be sure, loyalty to the RCT persists. There is a powerful faction arguing that EBP should utilise exactly the same protocols and procedures and evidential warehouses as EBM, much of this promoted in the name of the Campbell Collaboration. Others perceive a significant contradistinction. Medical treatments are designed to work regardless of patients' hopes and expectations. Social programmes work though the volitions of their subjects and capturing these interpretations is moved to centre stage in many forms of evaluation research. The result is the grip of the RCT within EBP is much less secure and is replaced with something more diverse and quarrelsome (Alkin, 2012), much of it rejecting the cause of science in the evaluation of social programmes.

Here, I want to ruffle a few feathers by speaking up for the unity of science, at the same time as rejecting the Cochrane/Campbell credo. It transpires that there is also significant resistance to the 'gold-standard' narrative *within* EBM. Quietly but steadfastly, there has been an outgrowth in rival accounts of the methodological foundations of medical science (e.g. Hill, 1965; Russo and Williamson, 2007; Rawlins, 2008; La Caze, 2011; Greenhalgh et al., 2014). Several of these sources betray distinctly realist sympathies, bearing a strong family resemblance to the approach I have tried to foster in EBP (Pawson, 2013). Readers will spot an ambitious, quite possibly grandiose, thesis on the brew here. It would take a book or ten to adequately make the case for the ubiquity of realist science. So here I issue an initial call-to-arms, building my argument with a close interrogation of single source.

Mukherjee's epic volume, *The Emperor of All Maladies*

This book won the 2011 Pulitzer Prize for General Nonfiction. The author is an academic physician, with posts at Harvard and the Massachusetts General Hospital. The book has been celebrated as the 'defining history of the defining plague of our generation'. Commentators have lined up to praise Mukherjee's compassion and humanity as he describes his care for and his indebtedness to his patients. This account takes a different tack, drawing methodological lessons from the *longue durée* of cancer research. Take another glance at the short extract above, reproduced from the prologue to his book, and you will notice the subtitle, 'A Biography of Cancer', and also the stunning claim that cancer cells are 'a more perfect version of ourselves'. Mukherjee is preparing us for a tale of the cancer cell's anthropomorphic ability to survive and thrive in the face of a trillion dollar war for its obliteration. EBP often appears so difficult, so corrigible and so puny in the face of social complexity, human adaptation, unintended consequences, and ceaseless emergence. It may be of comfort to learn that we are not alone.

Mukherjee derives important principles on the nature of the evidence base for cancer therapy via reconstructions of the research record on scores of different interventions during the long history of cancer care. There is a remorseless pattern. A new therapy is devised, often demonstrating immediate, apparent success but which, over time, becomes qualified due to significant levels of relapse and disease spread. This state of affairs provokes revisions to the available clinical procedures, which in turn may generate more positive outcomes and longer remissions, but which in turn may become thwarted and so regenerate the research cycle once more. Precise navigation of this sequence of conjectures and refutations has proved to be vital. Sometimes inquiry has steered into blind alleys and sometimes it has generated clarity of

vision. In the space available here I feature a mere four of Mukherjee's case studies, which tell of the halting balance between progress and setback.

Lessons from early cancer interventions

An obvious starting point is the first major attack on breast cancer by an intervention that came to be known as radical mastectomy. The treatment was pioneered by Halstead in the late 19th century. The development of medical treatments of all kinds (from pharmaceuticals to physiotherapy) begins with an understanding of their mechanism of action (MoA). And the MoA of this form of surgery is seemingly very straightforward. The surgeon 'extirpates' (or cuts out) the cancerous tumour. 'Radical' surgery borrows a term from Latin referring to the idea of 'uprooting' cancer from its very source. Mammography of the time was beset with the problem of relapse, with reoccurrences accumulating precisely around the margins of the original surgery. The 'obvious' solution was to excise even more of the breast tissue. And thanks to technical advances in surgery at the time Halstead augmented the basic MoA by 'cleaning out' larger and larger areas of tissue, extending as far as the proximate shoulder and chest muscles. The resulting surgery was, of course, profoundly and infamously disfiguring.

The pressing question, as ever, is did it work? Mukherjee reminds us of the 'deep conceptual error':

> The tumours . . . demonstrate a spectrum of behaviour right from their inception. In some women, by the time the disease has been diagnosed the tumour has already been spread beyond the breast: there is metastatic cancer in the bones, lungs and liver. In other women, the cancer is confined to the breast . . . it is a truly local disease. Position Halsted now, with his scalpel in the middle of this population . . . The woman with the metastatic cancer is not going to be cured by a radical mastectomy, no matter how aggressively it extirpates the tumour in her breast; her cancer is not a local problem. In contrast the woman with the small confined cancer does benefit from the operation – but for her a far less aggressive procedure would have done just as well. Halstead's mastectomy is thus a misfit is both cases; it underestimates its target in the first case and overestimates in the second, p. 67.

Given its allure and status within the medical profession, and under Halstead's leadership, radical surgery was extended to cover more and more primary sites of cancer. But it was many years before the deep conceptual error was acknowledged. Cancer physicians only gradually came to understand that the effectiveness of any treatment depended upon, and needed to be developed in parallel with, investigation of the prior mechanism of the inception and spread of the disease. It is worth pausing to set down clearly the model of 'counteracting causation' at work here. Research begins with an understanding of the mechanisms that generate the disease and treatment is understood as the creation of a therapeutic mechanism that counters, obliterates, alleviates, or suppresses the disease mechanism. The causal logic, and the idea of the blocking mechanism, is illustrated, perhaps too playfully, in Figure 1.

Bringing in context and patient characteristics

A more successful encounter with this causal model is to be found in Mukherjee's account of Kaplan's research, conducted in the early 1960s, on the use of 'extended field radiation' to treat Hodgkin's disease (cancer of the lymph glands). As ever, Kaplan's research did not start from scratch and was inspired by the work of a Canadian surgeon, Peters, who had devised an

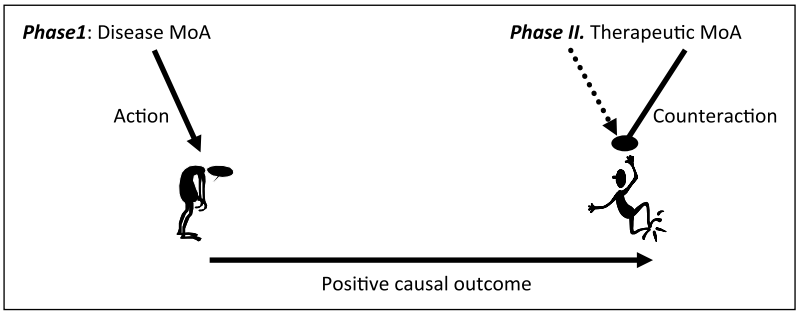

Figure 1. Clinical causation as a counteracting mechanism.

apparently successful treatment called 'extended field radiation', which delivered X-rays not to a single swollen node but to the entire area of the lymph gland. Note the initial thinking here (viz. extend the target area under treatment) has much in common with radical surgery. The vital difference, as we will come to see is the synchronous development of medical tests to identify the precise context in which this MoA applies. Returning to the historical account, we note that Peters's findings were only supported flimsily by retrospective comparisons with unmatched patients treated by other means, and so Kaplan devised an improved test using a pioneering version of what would now be recognised as an RCT. A pilot trial compared randomly allocated groups undergoing 'extended' versus 'involved' radiation therapy. Most promisingly, the former therapy appeared to significantly diminish the Hodgkin's relapse rate.

This research programme was, however, still in its infancy – 'a diminished relapse rate was not a cure' (Mukherjee, 2010: 253). Kaplan's team pressed on devising further radical techniques to deliver a still wider field of radiation covering additional nodes and blood vessels. More crucially and quite exceptionally, Kaplan also began to focus on the patient characteristics that might be particularly responsive to this form of radiation therapy. There was no repetition of Halstead's conceptual error. By this time it was common practice to attempt to differentiate patients with localised disease from those experiencing more disseminated forms. However, testing for a means to establish this difference was in its infancy. More research was needed, not in form of additional trials, but in identifying the patients most likely to benefit. Hence Kaplan's development of: i) blood tests, ii) detailed clinical examinations, and iii) lymphangiography (a 'primitive ancestor of a CT scanner') to assess the type and stage of cancer. Not content with devising these, Kaplan introduced simple forms of pre-intervention surgery to ensure only patients with a locally confined disease were entered in a subsequent trial.

Careful recruitment proved vital. The following, closely targeted trial was able to deliver a further substantial improvement in survival rates stretching into months and then years. In Mukherjee's words:

> this simple principle – the meticulous matching of a particular therapy to a particular form and stage of cancer – would eventually be given its due merit in cancer therapy . . . even if Kaplan understood it in 1963 . . . it would take decades for a generation of oncologists to come to the same realisation, p. 161.

Again we pause to formalise Kaplan's methodological strategy, which is summarised in Figure 2. It is wedded, as we have already seen, to a generative model of causation which provides an

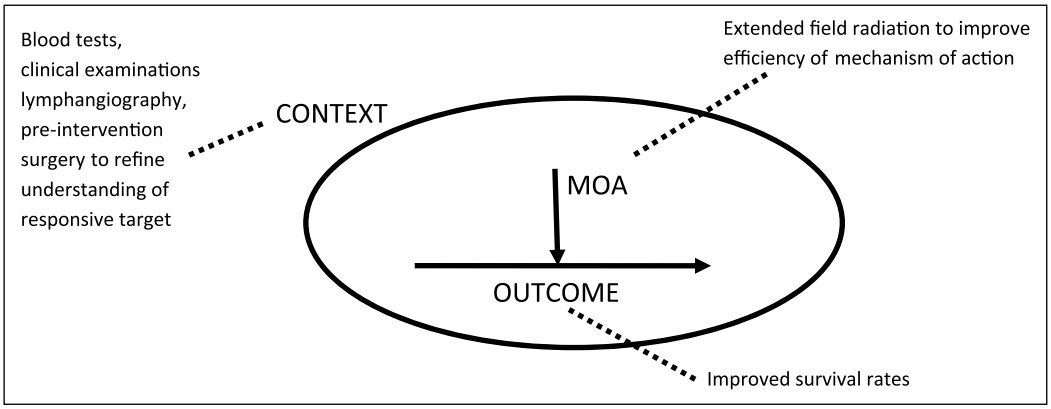

Figure 2. Kaplan's Method: Matching Therapy to the Cancer Context.

understanding and refinement of the mechanism (M) underlying the treatment, to which he adds further research providing an understanding of the patient context (C) in which that mechanism will act to produce optimal outcomes (O). Understanding 'what works' requires basic biological science, laboratory tests, medical imaging, clinical examinations and experimental research to reveal the optimal CMO configurations.

I return to setback in the next vignette, for in a perhaps surprising sense it is evidence on the precise anatomy of failure that drives medicine forward. Failures, I might briefly interject, are not cherished under the arithmetic of meta-analysis. Chemotherapy works by poisoning cancer cells, the great art and science of that treatment being to do so in a manner which can be tolerated by healthy cells. So-called 'high-dosage combination chemotherapy' became the norm in the 1960s in search of a precise treatment with this delicately balanced MoA. The VAMP trial, led by Freireich, was one such attempt, the acronym standing for the first letter of the fourfold drug package. This admixture was considered particularly toxic, indeed possibly life-threatening in its own regard. Moreover, VAMP was designed to attack childhood leukaemia and mounting even a small trial proved ethically problematic. I omit the background politics here in order to get to Mukherjee's account of the preliminary results of the gruelling treatment:

> At the end of three excruciating weeks, a few of Freireich's patients somehow pulled through. Then, unexpectedly . . . there was a payoff . . . the leukemia went into remission. The bone marrow biopsies came back one after another – all without leukemia cells . . . The Clinical Center was now filled with the familiar chatter of children in wigs and scarves who had survived two or three sessions of chemotherapy. Critics were slowly turning to converts. Other clinical centers around the nation joined . . . the experimental regime, p. 145.

Optimism was relatively short lived, however, and the trial was destined to prove ineffectual. Over the coming months patients returned to the clinic with minor symptoms, most notably persistent headaches. Unintended outcomes multiplied with serious neurological complaints of 'vision speckles, seizure, facial paralysis and then coma'. But even under distressing failure, vital lessons were gleaned about the behaviour of cancer cells. Blood and bone marrow biopsies were performed on the relapse cases and no cancer was found. The cancer, however, had relapsed explosively in the central nervous system. Mukherjee provides the explanation:

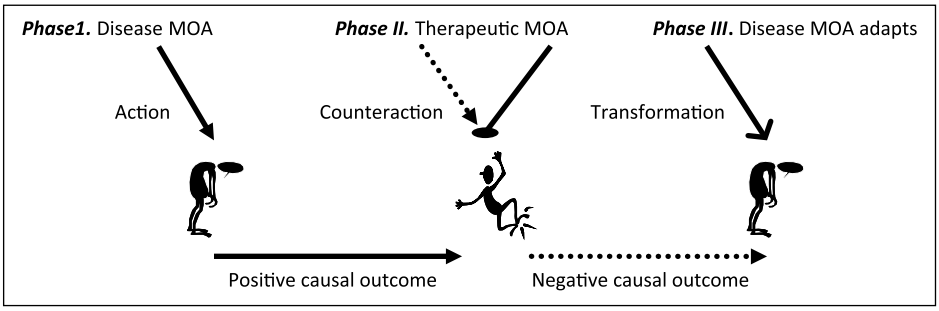

Figure 3. Fluctuating causal outcomes under self-transformative mechanisms.

> The brain and the spinal cord are insulated by a tight cellular seal called the blood-brain barrier that prevents foreign chemical from easily getting into the brain. It is an ancient biological system that has evolved to keep poisons from reaching the brain. But that same system had likely also kept VAMP out of the nervous system . . . The leukaemia, sensing an opportunity in that sanctuary, had furtively climbed in, colonising the one place that is fundamentally unreachable by chemotherapy, p. 147.

Discovered in this episode is a very early example of what became known as the cancer cell's 'acquired capability' to transform – the ability of a disease to change its form in the face of treatment. The VAMP trials were mounted in the midst of the 'War on Cancer', with its obsession to discover the most powerful dosages and drug combinations to obliterate the disease. But as more and more (partial) remissions and (partial) relapses accumulated across diverse therapies for different cancers, a paradigm shift was inspired, articulated by Mukherjee as follows:

> Even targeted therapy was a cat-and-mouse game. One could direct endless arrows at the Achilles' heel of cancer, but the disease might simply shift its foot, switching one vulnerability for another. We were locked into a perpetual battle with a volatile combatant, p. 443.

Figure 3 adapts a previous diagram to illustrate the acquired capability of a disease to transform itself under therapy.

How normal cells are transformed: Complexity in open systems

This brings us to our fourth vignette, taken from the concluding years of Mukherjee's history. By the early 1980s the attention of cancer scientists had begun to switch to the inside of cancer cells. Evidence was constructed about the precise manifestations of cell alteration, mutation, invasion and survival that generated malignant growth. Here the account turns to the labours of the pioneers of cancer biology and work done in the laboratory, in petri dishes, and on mouse cancers. Much of this inquiry originated in Weinberg's MIT laboratory and here many of the basic building blocks of cancer were discovered, beginning with the identification of oncogenes (which drive normal cell growth to become out of control) and tumour suppressor genes (which normally slow down excessive cell division but if malfunctioning may allow cells to grow out of control).

Box 1. The Hallmarks of Cancer.

1. *Self-sufficiency in growth signals*. Cancer cells do not need stimulation from external signals (in the form of growth factors) to multiply. Normal cells require external growth signals (growth factors) to grow and divide. These signals are transmitted through receptors that pass through the cell membrane. When the growth signals are absent, they stop growing. Cancer cells can grow and divide without external growth signals. Some cancer cells can generate their own growth signals.
2. *Insensitivity to anti-growth signals*. Cancer cells are generally resistant to growth-preventing signals from their neighbours. The growth of normal cells is kept under control by growth inhibitors in the surrounding environment, in the extracellular matrix and on the surfaces of neighboring cells. These inhibitors act on the cell cycle clock, by interrupting cell division (mitosis) in the interphase.
3. *Evading apoptosis*. Apoptosis is a form of programmed cell death (cell suicide), the mechanism by which cells are programmed to die in the event they become damaged. Cancer cells characteristically are able to bypass this mechanism.
4. *Limitless replicative potential*. Non-cancer cells die after a certain number of divisions. Cancer cells escape this limit and are apparently capable of indefinite growth and division (immortality). But those immortal cells have damaged chromosomes, which can become cancerous.
5. *Sustained angiogenesis*. Angiogenesis is the process by which new blood vessels are formed. Cancer cells appear to be able to kick-start this process, ensuring that such cells receive a continual supply of oxygen and other nutrients.
6. *Tissue invasion and metastasis*. Cancer cells can break away from their site or organ of origin to invade surrounding tissue and spread to distant body parts. Primary tumor masses spawn "pioneer cells" that invade adjacent tissues, and may then travel to distant sites, and establish metastases. The newly formed metastases arise as amalgams of cancer cells and normal supporting cells conscripted from the host tissue.

These herculean efforts are now synthesised in a work by Hanahan and Weinberg (2000). At the time of writing Google Scholar records more than 26,000 citations for this *Hallmarks of Cancer* article as well as 10,000+ for a more comprehensive and updated paper (2011). This vital repository of evidence does not come in the form of a totalising meta-analysis of effectiveness research but takes the form of an account of 'six biological capabilities acquired during the multistep development of human tumors'. All cancers share common traits (hallmarks) that govern the transformation of normal cells to cancer cells. They are paraphrased (in lay terms) in Box 1.

Mukherjee's point about this episode is that vital evidence comes increasingly from the bottom-up – 'from the cancer cell to its therapy' – to which end he quotes Chabner, a former director of the US NCI's Division of Cancer Treatment:

> It was as if the whole discipline of oncology, both prevention and cure, had bumped up against a fundamental limitation to knowledge. We were trying to combat cancer without understanding the cancer cell, which was like launching rockets without understanding the internal combustion engine, p. 304.

Cancer, in other words, exemplifies what complexity theory understands as the challenge of conducting research on self-transformative, open systems. The consequence, opines Mukherjee, has been a step-change in the way this 'emperor of all maladies' is studied. The failed *war on cancer* (the search for 'the monolithic hammer to demolish the monolithic disease') has been turned into an increasingly successful *cat and mouse game* (in which 'biological heterogeneity demands therapeutic heterogeneity').

Reprise: Lessons for evaluation

I commenced this essay in search of mutual enlightenment between EBP and EBM, promising some less orthodox learning points if the point of comparison is with historical accounts of development in clinical interventions. Mukherjee's chronicle, quite inadvertently, presages some surprising parallels. So what lessons might be drawn across from *The Emperor of all Maladies*?

1. *Evaluation should be mechanistic.* We should follow EBM in always beginning research by inquiring into an intervention's mechanism of action – what do we suppose it does to change behaviour? Evaluation, moreover, should be based on models of counteracting causation and centred on a close understanding of how effectively that programme mechanism blocks, obliterates, alleviates, or suppresses the underlying problem. In both programme development and outcome evaluation in EBP, such reasoning is often absent or, at best, vague. The inception of social programmes is still largely based on common sense responses to the need to 'do something'. The targeting of social programmes is often decided by local dictate and administrative convenience. Many social interventions remain improvised and atheoretical. We are still inclined to launch makeshift interventional rockets without a solid theoretical base in social and behavioural science.
2. *Evaluation should be configurational*. Mukherjee's central message is that treatments never work unconditionally and that the research task is to work painstakingly through the contingencies. The core challenge follows the realist mantra of discovering what works, for whom, in what circumstances, in what respects, over what duration and, above all, why. In the examples above and in EBM more generally, this undertaking requires a combination of basic biological and genetic research, diagnostic tests, medical imaging, and structured clinical examinations in addition to the clinical trial. EBP should follow exactly the same mission. The data collected will, of course, be collected from quite different sources using quite different techniques – but it will always require the construction of a comprehensive, multi-method evidence base.
3. *Evaluation should be adaptive*. Just as disease is adaptive, we need to understand the myriad ways in which a social problem can transform itself when a remedy is applied. There is an interesting parallel here with EBP in Mukherjee's description of clinical progress as a 'cat and mouse game'. Many social programmes, as for instance in the field of crime prevention, have a relatively short period of effectiveness. This particular 'hallmark' occurs when the criminal fraternity retreat in the face of a freshly implemented scheme but then gradually adapt, altering their tactics in such a way as to avoid the newly understood risks. Authorities respond with even smarter programmes but the adaptive cycle, of course, continues. More generally, we can say that very few social interventions are entirely durable. More generally, we need to understand that social interventions often change the conditions that made them work in the first place. Programme development and evaluation research confront an endlessly renewing challenge.

Author's Note

This paper is a prologue to Pawson's forthcoming but as yet untitled book on the relationship between evidence-based policy and evidence-based medicine. Other preparatory materials may be found at http://realism.leeds.ac.uk/ray-pawson-exaugural-lecture/

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Reviewed work

Mukherjee S (2010) *The Emperor of All Maladies: A Biography of Cancer*. Detroit, MI: Scribner Press.

Other references

Alkin M (ed.) (2012) *Evaluation Roots*, 2nd ed. Thousand Oaks, CA: SAGE.

Greenhalgh T, Howick J and Maskrey N (2014) Evidence based medicine: A movement in crisis. *BMJ* 348: g3725.

Hanahan D and Weinberg R (2000) The hallmarks of cancer. *Cell* 100(1): 57–70.

Hanahan D and Weinberg R (2011) Hallmarks of cancer: The next generation. *Cell* 144(5): 646–74.

Hill B (1965) The environment of disease: Association or causation. *Proceedings of the Royal Society of Medicine* 58: 295–300.

La Caze A (2011) The role of basic science in evidence-based medicine. *Biology and Philosophy* 26(1): 81–98.

Pawson R (2013) *The Science of Evaluation: A Realist Manifesto*. London: SAGE.

Rawlins M (2008) *De Testimonio: on the evidence for decisions about the use of therapeutic interventions: the Harvelian Oration*. Royal College of Physicians of London.

Russo F and Williamson J (2007) Interpreting causality in the health sciences. *International Studies in the Philosophy of Science* 21(2): 157–70.

Ray Pawson is Emeritus Professor of Social Research Methodology, University of Leeds, UK. Publications include *A Measure for Measures* (1989), *Realistic Evaluation* (1997 with Nick Tilley), *Evidence-Based Policy: A Realist Perspective* (2006) and *The Science of Evaluation: A Realist Manifesto* (2013).