Institution: University College London		
Unit of Assessment: 30 Philosophy		
Title of case study: Pluralistic evidence for successful policymaking about reactive systems		
Period when the underpinning research was undertaken: 2013–2018		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
Brendan Clarke	Associate Professor of History and Philosophy of Medicine	2010-August 2020
Phyllis Illari	Professor of Philosophy of Science (2020-present) Lecturer and then Associate Professor (2012-2020)	2012-present
Period when the claimed impact occurred: 2014-2020		
Is this case study continued from a case study submitted in 2014? N		

1. Summary of the impact (indicative maximum 100 words)

Public bodies, such as health boards or government departments, must reliably interpret evidence to properly inform their decision-making. Illari and Clarke's work explores the diverse types of evidence obtained from biological and social mechanisms and systems, and how these are used for multiple ***purposes***. This has impacted on **1) international methodologies for evidence assessment in health** by increasing the ***plurality*** of evidence they use, particularly ***evidence of mechanisms*** (NICE, IARC), and on **2) UK ethical frameworks for AI and data science** by improving their evidence use, particularly their attention to anticipating and monitoring how systems including populations ***react*** to new ethics frameworks (Cabinet Office, DCMS, West Midlands Police, the NHS). The beneficiaries are patients who need high quality medical advice both in England and Wales and internationally, and all people in the UK affected by government and NHS data projects such as the UK Cabinet Office Framework for Data Ethics and the NHS Code of conduct for data-driven health and care technology.

2. Underpinning research (indicative maximum 500 words)

Phyllis Illari's research has focused on causality, mechanism, and information quality in clinical contexts and the policymaking sphere, arguing in each of these areas that sound approaches to evidence and methods need plural, diverse forms of evidence for multiple *purposes.* For example, the clinic serves a diverse range of purposes, and clinicians benefit from a variety of types of evidence tailored for those different purposes. Policymakers, likewise, require different kinds of evidence to inform their interventions, depending on what policy goal they are pursuing. Hence evidence cannot be graded and ranked in abstraction from those purposes.

Illari argues that information quality is fitness for *purpose* (**[R1]**, particularly Illari's sole-authored chapter), and hence that information quality is purpose-specific rather than global. Illari's co-authored book with Russo **[R2]** argues for a *pluralist* approach to causality: multiple approaches to causality should be accepted, because they are each useful for different purposes. Illari's co-edited book on mechanisms **[R3]**, shows how the idea of mechanisms can and should extend far beyond biological mechanisms such as DNA replication to include also social mechanisms such as stigmatisation; Illari's co-authored chapter with Glennan **[R3]** examines the varieties of mechanisms.

Illari collaborated with Brendan Clarke as Co-Is on the AHRC-funded 'Evaluating Evidence in Medicine' (EBM+) project (**[i],** PI Williamson, University of Kent, 2015-2018). This project studied diverse ways of getting evidence of mechanism, such as of how diseases progress in the body, or how drugs act on targets to stop or inhibit disease progression. The project studied how evidence of the entities and activities of such mechanisms, and their organisation, is gained and used, and created the EBM+ network of researchers interested in these problems.

The core of Clarke and Illari's joint work (in two books **[R3, R4]**, and several papers including **[R5, R6]**) is in what they are arguing *against*. Clarke and Illari oppose a dominant trend in public agencies' official use of evidence, which prioritises a restricted range of evidence methods, so-called 'gold standard' evidence such as randomised controlled trials (RCTs). Instead, they argue that one single 'best' form of evidence cannot address the *plurality* of purposes for which evidence is used. Such purposes range from, for example, identifying subpopulations of patients who have different treatment risks, to deciding whether it is justifiable to use a surrogate outcome (an outcome that is easier to measure than the real target outcome) in a randomised clinical trial. This focus on purpose and context of use makes even characterising good evidence difficult, but one key finding is that proper *evidence of mechanism* (of disease causation, mechanisms of actions of drugs, and social mechanisms of health policy) is needed.

Clarke and Illari's work **[R4]** meets this challenge by providing detailed guidance on evaluating evidence of mechanism alongside the already universally used RCT evidence. It also raises a central problem for mechanisms in the context of policy: reactivity—the fact that when mechanisms and systems are acted on, they change. Illari first writes about changing experimental systems as a central scientific problem in **[R2]**. Understanding this is vital to any use of evidence of mechanism and of system by policy bodies, because policy decisions (including publishing evidence assessment guides or data ethics frameworks) are actions that the system being governed *reacts* to. After Clarke and Illari's joint work, public bodies should use diverse evidence of mechanism or evidence of system to make good policy decisions, but they also need to take account of the fact that any policy decision they make will change that system.

3. References to the research (indicative maximum of six references)

[R1] Illari, P. & Floridi, L. (eds): *The Philosophy of Information Quality*, Springer (August 2014). (Published by respected peer-reviewed academic publisher; chapters were peer-reviewed.)

[R2] Illari, P. & Russo, F.: *Causality: Philosophical theory meets scientific practice*, OUP (October 2014). (Top quality rigorously peer-reviewed academic press.)

[R3] Glennan, S. & Illari, P. (eds.): *Routledge Handbook of Mechanisms and the Mechanical Philosophy*, Routledge (2017). (Outcome of grant **[i]**, high quality peer reviewed academic press; chapters peer reviewed.)

[R4] Parkkinen, V-P., Wallmann, C., Wilde, M., Clarke, B., Illari, P., Kelly, M.P., Norell, C., Russo, F., Shaw, B., Williamson, J. *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures*. Springer (2018). doi:10.1007/978-3-319-94610-8 (Outcome of grant **[i]**, published by respected peer-reviewed academic publisher.)

[R5] Vineis, P., Illari, P., & Russo, F. (2017). 'Causality in cancer research: a journey through models in molecular epidemiology and their philosophical interpretation'. *Emerging Themes in Epidemiology*, 14 (1). doi:10.1186/s12982-017-0061-7 (Peer reviewed scientific journal.)

[R6] Clarke, B., Gillies, D., Illari, P., Russo, F. and Williamson, J. (2014). Mechanisms and the Evidence Hierarchy. *Topoi.* 33(2): 339-60. doi:10.1007/s11245-013-9220-9 (Respected peer-reviewed journal.)

Grants [i] Illari and Clarke (Co-Is) *Evaluating Evidence in Medicine* (2015-18, AHRC, AH/M005917/1). PI: University of Kent, Co-Is: UCL and University of Ferrara. Amount to UCL: GBP281,771.20. Outputs available upon request.

4. Details of the impact (indicative maximum 750 words)

There is strong public pressure for agencies to make transparent and publicly justifiable decisions about complex questions such as when the identification of potential cancer hazards should trigger a public health response, which rapidly-emerging technologies and treatments to fund, and how to regulate AI and data driven technologies. The idea that a single, narrow, 'gold standard' of evidence is available, and can help public decision-making is tempting, but illusory **[R2, R4, R6].** Clarke and Illari worked alongside practitioners and policy-makers to impact evidence assessment methods in healthcare, and ethics frameworks for data science and AI in public bodies. Their research **[R4, R6]** increased the **plurality** of evidence taken into account in formal evidence assessment methods in healthcare, including or improving attention to **evidence of mechanism**, increasing the fitness of evidence for multiple **purposes [R1]**. It also **[R3**, **R1]** illuminated where knowledge is crucial, particularly when systems, including populations, **react** to new AI and data science ethics frameworks **[R2].**

1) Shaping international methodologies for evidence assessment in health
From 2015-18, as part of the 'Evaluating Evidence in Medicine' (EBM+) **[i]** project team at UCL and Kent, Clarke and Illari worked with stakeholders from the International Agency for Research on Cancer (IARC) and the National Institute for Health and Care Excellence (NICE). Beth Shaw, senior technical adviser in the Centre for Guidelines, Michael Kelly, retired Director of Public Health, NICE, and Kurt Straif, Head of the Section of Evidence Synthesis and Classification, IARC, were associate members of the project. This team developed a methodology to show how formal evidence assessment frameworks can take account of evidence of mechanism for multiple purposes. Project members then sat on formal consultations with IARC and NICE on their methodologies for evidence assessment, and submitted written material to them recommending inclusion of evidence of mechanism. This led to both organisations changing their methods to reflect the project findings; as Kelly explains "changes to the methods officially adopted by these extremely influential health agencies are very significant and have global reach" **[A]**. These are illustrated below:

IARC: IARC is the specialized cancer research agency of the World Health Organization (WHO), responsible for assessing all carcinogenic hazards. IARC's *Preamble to the Monographs* is the methodology manual governing how these assessments are performed, and as IARC's Head of the Section of Evidence Synthesis and Classification explained "WHO and National health agencies use this information as scientific support for their actions to prevent exposure to potential carcinogens" **[B.1]**. Before consulting with EBM+, IARC included mechanistic evidence in the *Preamble*, but it was tentative and secondary to other evidence; by 2017 they confirmed: "the EBM+ programme has already led us to think in new ways about the nature and kinds of evidence of carcinogenicity and to scrutinize our own approach" **[B.1].** This was reflected in the update to the 2006 *Preamble* which was published in 2019 and is likely to be used for many years **[B.2]**. The project PI represented the project on the international Advisory Group recommending revisions to the *Preamble*, which "led to significant changes in the way mechanistic evidence is evaluated in the new *Preamble*, with mechanistic evidence now treated on a par with epidemiological studies on humans and with animal studies. Mechanistic evidence thus plays a much more substantial role in the new *Preamble* than in previous versions of the *Preamble*." **[B.3]** This improves treatment of **evidence of mechanism**, and **plurality** of evidence used. The *Preamble* explicitly refers to **[R4]**, including its systematic review methods for evidence of mechanism. It now includes much more information about kinds of mechanistic evidence that might be relevant and how to get it. There are many more extended treatments through the second half of the *Preamble*, and attention to grading of overall mechanistic evidence in **[B.2],** section 6c, reflecting **[R4]**. This contribution is deeply and internationally significant: as the *Preamble* puts it, "Cancer prevention is needed as much today as it was when IARC was established, because the global burden of cancer is high and continues to increase" **[B.2]**, p1. By changing IARC's methodology to improve treatment of evidence of mechanism and plurality of evidence, Illari and Clarke have thus contributed to robust and responsible public decision-making in the assessment of cancer risk and prevention.

NICE: NICE improves outcomes for all people using the NHS, by providing assessments of the value for money of new technologies and medicines proposed for adoption in the NHS, and evidence-based guidance for clinicians. NICE's *Methods Manual* **[C.1]** is, as Kelly puts it: "core to what it does and how it does it", **[A]** and is used in all assessments of evidence NICE runs for medical treatments in England and Wales. NICE "is a leading health agency internationally, with many European countries simply adopting NICE guidelines, and NICE influencing pharmaceutical companies", **[A]** and it is copied across the world because, "It is internationally recognized for its independence and methodological rigour" **[C.2]**. Illari and Clarke's research **[R2, R4, R6]** changed the methodology recommended by NICE in the methods manual. To change these methods is to shape social policy interventions, domestically through NICE's role in evidence assessment, and internationally through the wider impact of NICE's methods of appraisal.

In 2017, NICE's Senior Technical Advisor stated that "the work of the EBM+ group is helping us to explore the question around the value and use of different types of evidence (including the use of evidence of mechanisms), from a multi-disciplinary perspective, in order to make practice recommendations that are robust and justifiable" **[C.3]**. After further consultation by the project team **[i]**, evidence of **pluralistic** types, particularly **evidence of mechanism [R2**, **R4**, **R6]**, were explicitly recognised in the 2018 manual **[C.4],** and remain in the 2020 version **[C.1]**. As Kelly writes, Clarke and Illari "worked particularly intensively" on the evidence assessment tables in section 4 [of **R.4**] which "connect[ed] their theoretical work to evidence assessment processes, particularly those used at NICE" **[R4]**. Kelly explains that "These showed how to incorporate better assessment of evidence of mechanism into existing evidence assessment processes, particularly focusing on the justification of intervention methods and surrogate outcomes used" **[A]**. The manual has changed to include "explicit recognition of the importance of combining statistical evidence and evidence of mechanisms" **[C.2]**. It now recognises key purposes **[R1]** for which evidence of mechanism is required, including for identifying relevantly different sub-populations and for assessing potential interventions or risks. The former recognises that people do not always react in the same way to a treatment, necessitating assessment of whether trial evidence gained in one population will apply to a different group or a sub-group (**[C.1]**, p26) while the latter has led to explicit recognition in the manual of causal mechanisms linking factors to outcomes, and "mechanism of action" of potential treatments as two of nine recommended review questions (**[C.1]**, p58). Mechanism as important to understanding how the medical intervention being assessed works now appears in four other places (**[C.1]**, p62, p63, p72, p113). Finally, there is implicit recognition of evidence of mechanism as important in justifying use of a surrogate outcome in trials in two places (**C.5 Appendix A**, 2020, p7), with reference to evidence of process, and of intermediate outcomes. As Kelly puts it, "The work of the project increased the presence of evidence of mechanism in the manual from very close to zero, thereby increasing the variety of evidence taken into account, and explicitly acknowledging the importance of evidence of mechanism for particular purposes". These research-led changes to IARC and NICE methodologies reach an extremely broad population globally: "it includes anyone exposed to a carcinogen, or treated with a pharmaceutical intervention or other treatment, or subject [to] a public health improvement" **[A]**.

2) Influencing UK ethical frameworks for AI and data science

Cabinet Office and the Department for Digital, Culture, Media and Sport (DCMS): In 2014, due to the expertise demonstrated in **[R1],** Illari was invited onto a months-long consultation by the Cabinet Office on the first major ethical framework for government data science. As the co-chair of the Committee writes: "Illari's expertise in information and evidence, *and* in causality and causal inference, is almost unique" [**D**]. This consultation created the 2016 Cabinet Office Framework for Data Ethics **[E.1],** which applied to all government data projects. It was the precursor of the framework still in force with DCMS (2020) **[E.2]**. Using attention to purpose **[R1]**, Illari was "responsible for Point F" **[D]**, presenting **reactivity [R2]** as one of 11 key checks that frame the whole document (**[E.1]**, p5, elaborated p12, remaining in force now (**[E.2],** p14). Illari also fed into Point D on data quality **[R1]**, (**[E.1]**, p5; (**[E.2],** p27). The co-Chair of the Committee explains that "this work probably prevented significant problems in government data science and public sector data projects are supposed to apply this framework to guide their data projects" **[D]**.

Following this, Illari was appointed from Sept 2016-20 as a founder member of the Data Ethics Group (DEG) of the Alan Turing Institute, which is the UK's flagship Institute for Data Science and AI. The DEG has been consulted by many significant UK public bodies, so that, with the DEG's interdisciplinary team, Illari has improved attention to **plural evidence** and **reactivity** in national approaches to data ethics. This included informal consultations with DCMS on setting up the Centre for Data Ethics and Innovation, with the Information Commissioner's Office, and with Police frameworks on predictive policing. Illari as part of the DEG particularly pressed for serious consultations with stakeholders from minoritized groups, and for the need for all projects to monitor for ongoing evidence of impact on them.

In one particularly important case in 2019, NHS England consulted the DEG on a Code of Conduct for data-driven health and care technology ethics **[F.1]**. As the resulting DEG report put it, “The Code is intended to form the basis for the NHS’s long-term strategy of developing, procuring and deploying data-driven tools.” **[F.2]** Illari and her UCL colleague Melanie Smallman argued that the Code’s focus on individual providers or technologies provided no protection against system-level harms (**reactivity** of a system to a policy intervention) that NHS England had a duty to prevent. Instead, there needed to be thorough and sustained attention to “wider social implications of data-driven and AI technologies on health” **[F.2].** Instead of the consultation on the Code of Conduct closing, as planned, they were successful in maintaining a consultative status following the 2019 launch. As the Chair of the DEG explains, “The civil servants involved are to be commended in taking seriously Illari and Smallman’s representations, as DEG members, that NHSX had a duty to understand and monitor the potential collective effects of the use of technologies which, even if individually ethically reasonable, might collectively create serious harms, especially for vulnerable groups.” **[D]** Given current COVID-19 circumstances this work is not complete. The Chair notes that “the Cabinet Office, and NHSX, wish to keep a high standard – and *demonstrate* that they are keeping to a high standard – so that public confidence in government and the NHS can be maintained. Illari played an important role in advancing this in these consultations” **[D]**. Furthermore, “She has helped to establish robust practices in public bodies that are more sensitive to evidence, taking into account reactivity, that have raised the quality of data projects in the UK and prevented harm” **[D]**.

The significance of this contribution extends beyond even all UK government data science and the NHS use of data-driven tools. These have been crucial years for data ethics and good work is rapidly copied. For example, recent documents on such strategies for the Scottish Government cite links to the Turing Institute work. Illari’s focus on **pluralistic evidence** of what the system is doing and how we *continue* to know this as it changes is vitally important. Both here and in medical methodology this research has wide-ranging and long-lasting impact in encouraging and enabling robust and responsible public decision-making.

5. Sources to corroborate the impact (indicative maximum of 10 references)

- **A.** NICE testimonial from Former Director of Public Health, NICE, and associate member of the project team on grant [i].
- **B.** Medical evidence (IARC Preamble) [B.1] IARC testimonial from the Head of the Section of Evidence Synthesis and Classification at IARC; [B.2] Jan 2019 IARC Preamble to the Monographs; [B.3] IARC testimonial from the Acting Head of the Monographs Group in the Section of Evidence Synthesis and Classification at IARC.
- **C.** Medical Evidence (NICE Methods Manual) [C.1] Current NICE manual 2020; [C.2] NICE testimonial from Programme Director of the Guidance Transformation Programme, NICE [C.3] NICE testimonial from Senior Technical Adviser in the Centre for Guidelines at NICE; [C.4] NICE manual updated October 2018; [C5] NICE manual 2020 appendices.
- **D.** Data Ethics (All Impact) testimonial from the Co-chair of the Cabinet Office committee that created the Framework for Data Science, and Chair of the Data Ethics Group of the Alan Turing Institute, The UK’s Institute for Data Science and AI.
- **E.** Data Ethics (Cabinet Office Framework for Government data projects) [E.1] Cabinet Office Framework Illari fed directly into; [E.2] DCMS framework which is the version currently in force showing Illari’s impact still existing despite the shift in government department.
- **F.** Data Ethics (NHS Code of conduct for data-driven health and care technology) [F.1] The NHS Code of conduct for data-driven health and care technology, July 2019; [F.2] Turing Institute report of the DEG consultation on the NHS Code of Conduct.