

Evaluation, Realism and Complexity

Nick Tilley, UCL
Edinburgh June 2008

The Nature of Programmes

- Programmes attempt to change a state of affairs or behavioural regularity
- Programmes are theories incarnate
- Programmes embody multiple theories, held by multiple stakeholders
- The theories embodied by programmes are often implicit
- The theories embodied by programmes are apt to change

How Programmes Work

- Programmes work by activating and deactivating causal mechanisms
- Social programme mechanisms involve:
 - Active participant interaction
- Clinical programme mechanisms may involve:
 - Passive participant interaction, and
 - Active participant interaction
- Programme theories include mechanisms

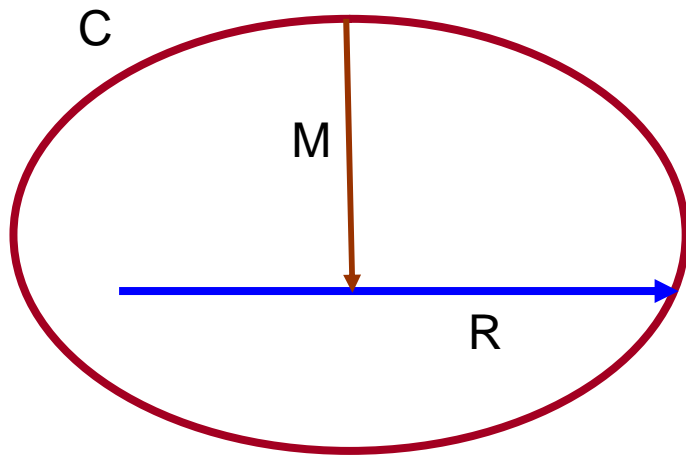
(Social) Programme Complexity

- Elements of programme complexity:
 - Programmes are active, not passive
 - Programmes have long implementation chains
 - Programmes have multiple stakeholders
 - Programmes are embedded in complex social systems
 - Programmes are implemented amidst other interventions
 - Programmes (and non-programmes) borrow and adapt
 - Programmes are the offspring of previous interventions
 - Programmes are open, not closed
 - Programmes change the conditions for their operation
- Programmes therefore are never delivered in the same way twice or in the same circumstances.
- Policy on/policy off comparisons are unhelpful.

Realist Evaluation

- Rooted in realist philosophy of science
- Has a generative account of causality
- Looks for and abstracts context, mechanism, outcome pattern configurations (CMOCs)
- Operates at the 'middle range'

Generative Causality: CMR



Mechanisms

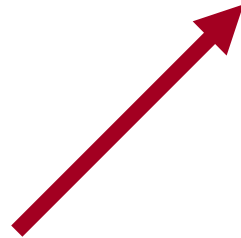
- Mechanisms lie in the 'black box'
- Mechanisms describe how effects are produced.
- Mechanisms are not sub-programmes or intermediary steps
- Mechanisms are often hidden.
- Mechanisms are part of the furniture of the sciences
 - Think, for example, gravity, magnetism, natural selection
- Social programmes activate and/or deactivate mechanisms
- Social programme mechanisms generally refer broadly to reasoning and resources
 - Think, for example, interests beliefs, affects, opportunities

Silent Alarm Theory Revised (Silence + Secrecy)

Thieves
enter
business
premises



Presence
of the alarm
is unknown
to the
intruders



Alarm is
triggered
– but
only at
police
control
centre



Patrol car
dispatched
ASAP



Intruders
caught
red
handed

Context

- Context describes the conditions relevant to the activation/deactivation of mechanisms.
- Programme outcomes depend on the conduciveness of the context to activation/deactivation of mechanisms.
- Most programmes operate in multiple contexts defining differing sub-groups
- What may be relevant to context is programme dependent and variable
 - Think, for example, culture, economy, technology, geography.

Mandatory Arrest Trial

- Minneapolis experiment
 - arrest 10% repeats in six months
 - advice 19% repeats in six months
 - sent away 24% repeats in six months
- Mandatory arrest advocated for all, 1984
- Cities over 100,000 with mandatory arrest policies
 - 10% 1984
 - 43% 1986
 - 90% 1988

Further Trials and Mixed Findings

- More experiments
 - repeat domestic violence went up in 3
 - repeat domestic violence went down in 3
- Post hoc conjecture that effects vary by employment status and community type
 - where up, anger triggered in context of high unemployment and community instability
 - where down, shame triggered, in context: low unemployment and community stability

More Realist Hypotheses for Mandatory Arrest and Domestic Violence

Mechanism	Context	Expected outcome
Offender shame	Membership of 'respectable' knowing community	Reduced repeat incidents
Offender anger	Poorly integrated, unemployed men	Repeat assaults
Women's shame	Membership of 'respectable' knowing communities	Reduced reporting
Women's fear of reprimination	History of violence; alcoholic offenders	Reduced reporting
Women's fear of loss of partner	Emotional or financial dependency on partner	Reduced reporting
Women's fear of children being taken into care	Pattern of general domestic violence against whole family	Reduced reporting
Women's empowerment	Availability of refuges; support for women; financial resources of women	Increased separation
Changed norms about domestic violence	Positive publicity	Longer term reduced violence

Outcome-patterns

- Outcomes are generated by mechanisms according to contexts
- Contexts and mechanisms are always plural
- Therefore outcome patterns are always plural
- Realist evaluation focuses on outcome 'footprints'
- Realist evaluation does not produce pass/fail verdict
- Realist evaluation begins with theory and ends with more refined theory

Summary: The Language of Realist Evaluation

- *Context*: the contingencies for the generation of regularities in context
- *Mechanism*: the generation of regularities in context
- *Regularity*: what is generated by mechanisms in context
- *Outcome*: changes in regularity produced by altered mechanisms effected by modifications in context
- *Context-mechanism-outcome pattern configuration (CMOC)*: the product of realist evaluation

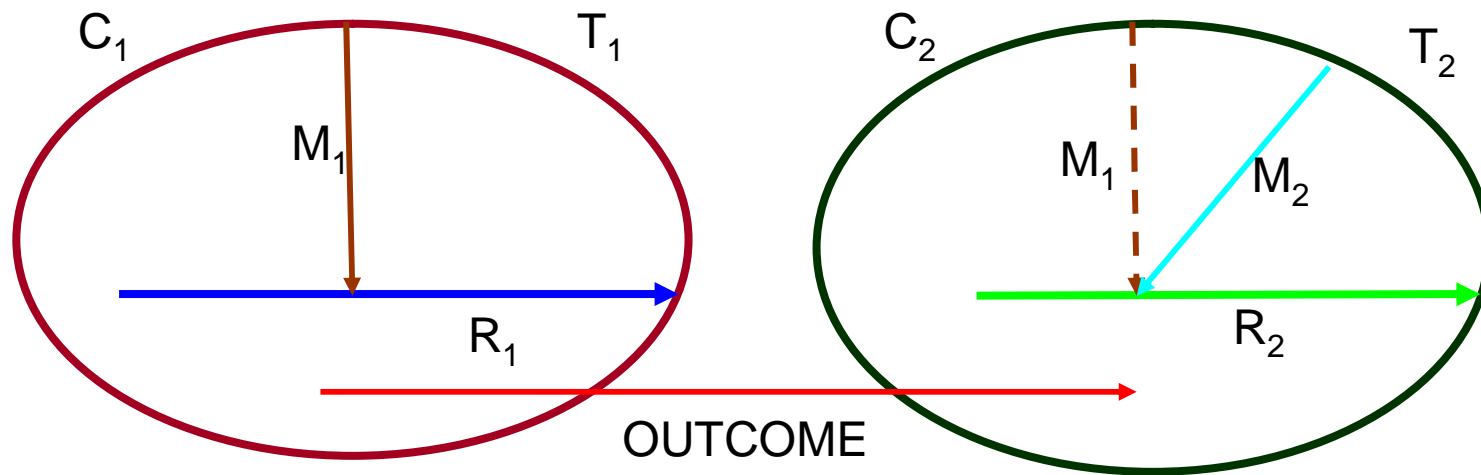
Realist evaluation questions

- Do not ask:
 - What works?
 - Does this work?
- Do ask:
 - What works for whom in what circumstances?
 - What is it about x which may lead it to have a particular outcome in a given context?
 - What conditions are needed for x to trigger mechanisms to produce particular outcome patterns?
 - What sorts of mechanism are triggered in what sorts of context to produce what sorts of outcome?

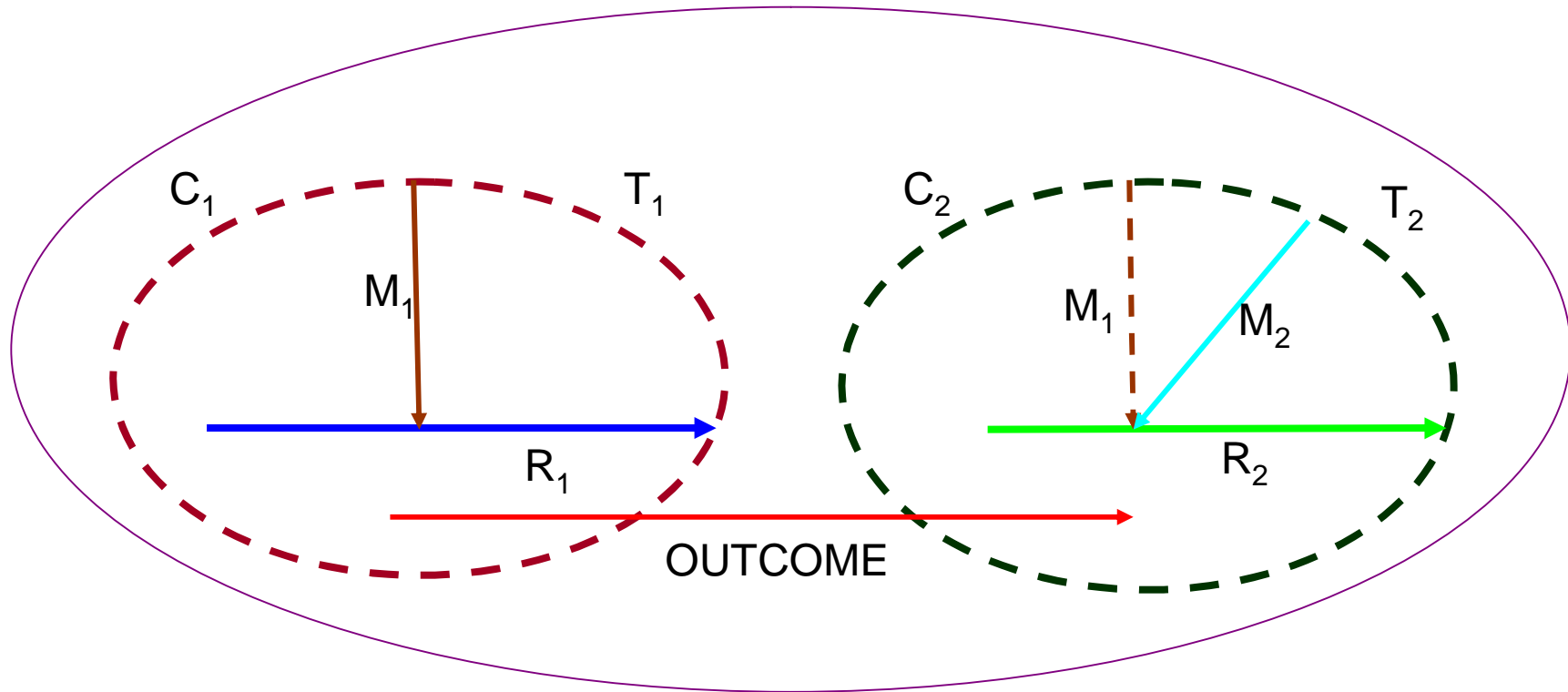
To Do Realist Evaluation

- Elicit/identify 'working theories' about how the programme may work
- Select out promising theories that may have reach beyond the specific case under study
- Formalise theories identifying hypothetical CMO propositions
- Derive empirical test of these propositions identifying the 'footprint' of success and failure of the programme
- Collect data (quantitative, qualitative, comparative etc.) that will help test, refine CMO conjectures or arbitrate between competing theories
- Present improved theory at a middle-level of abstraction that will inform future programmes

... And putting it together in a diagram 1



... And putting it together in a diagram 2

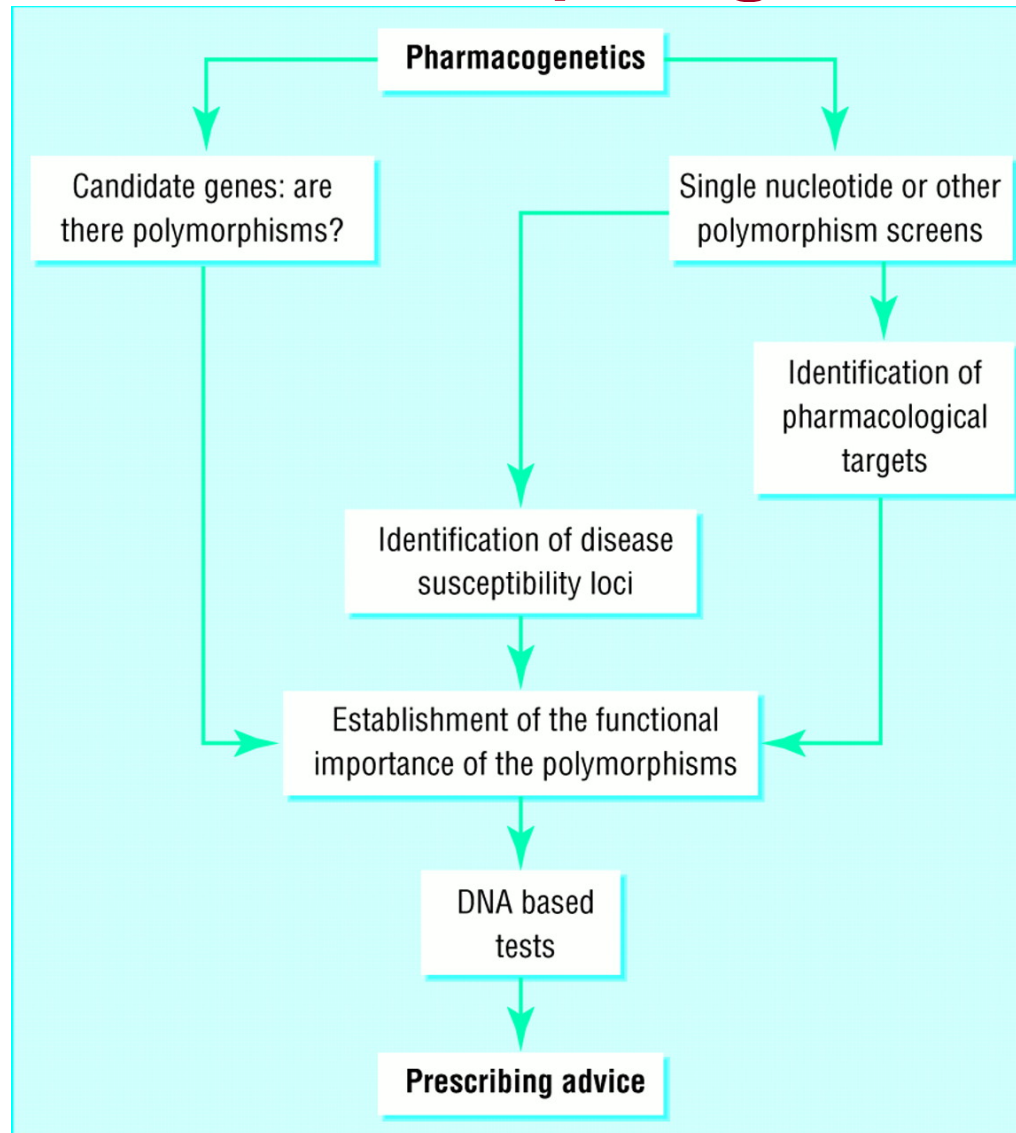


Conclusion: Programmes and Realist Evaluation

- Programmes work through mechanisms
- Programmes work in multiple contexts
- Mechanisms depend on contexts
- Programmes are theories of context and mechanism
- Programmes embody multiple theories
- Programmes are delivered in open systems
- Programmes theories are fallible
- Evaluation elicits, specifies, tests and refines programme theories

Medicine, drugs and realist
evaluation?

An implicit realist programme?



Source: Wolf, C, Smith, G, and Smith, R. (2000) 'Pharmacogenetics', *BMJ* 320: 987-990.

‘Our increasing knowledge of the mechanisms of drug action, the identification of new drug targets and the understanding of genetic factors that determine our response to drugs may allow us to design drugs that are specifically targeted towards particular populations or that avoid genetic variability in therapeutic response. The extent of genetic polymorphism in the human population indicates that pharmacogenetic variability will probably be an issue for most new drugs.

The development of pharmacogenetics provides at least one mechanism for taking prescription away from its current empiricism and progressing towards more “individualized” drug treatment.’

Improvements in healthcare and realist evaluation

Example: Rapid Response Teams (RRTs)

Source: Berwick, D. (2008) 'The Science of Improvement' *JAMA* 299(10): 1182-1184

- Theory
 - 10%-15% of inpatients resuscitated outside intensive care units survive to hospital discharge
 - Early warning signs are often present amongst those eventually experiencing cardiac arrest
 - RRTs bring expert clinicians to bedsides of deteriorating patients before arrest occurs

RRTs

- Findings
 - Local experience strongly suggested that RRTs often associated with improved outcomes: reduced anxiety, increased interdisciplinary teamwork, decreased cardiac arrests outside intensive care units, and some declines in mortality
 - Cluster randomised prospective trial (Medical Early Response Interventions and Therapy – MERIT) undertaken
 - No beneficial effects found ‘on several primary outcomes’

RRTs

- MERIT problems
 - Lack of statistical power (assumed rate of events 30 per 1,000 admissions; actual 7 per 1,000)
 - Cross contamination abounded: control hospitals implemented RR protocols; some study hospitals didn't
 - Wide variation in outcome events varied: at 95% confidence interval range of 4.37 events per 1,000 admissions, 80% of the total event rates in both groups
- Some critics therefore urged caution on spread of RRTs, refusing to entertain other evidence of successes

Berwick's suggestions

- Use wider range of methods than just RCTs, considering mechanisms and contexts
- Reconsider thresholds for action on evidence, such as 'canonical' $P < 0.05$
- Rethink views about trust and bias, as they risk neglecting consideration of 'local wisdom' about contexts and mechanisms
- Ensure mutual respect between academicians and frontline caregivers.

Questions?