1 Feasibility and desirability of a realist CMOC database: lessons
2 learned

6 Abstract

7 Central to realist evaluation (RE) is its focus on underlying generative mechanisms that cause
8 outcomes in particular contexts, often presented as context-mechanism-outcome configurations
9 (CMOCs). The first author brought together 22 RE experts in a Delphi technique inspired exercise to
10 explore the potential of creating a database from CMOCs identified in the empirical literature, to
11 enhance learning about what CMOCs are and to further the cumulation of knowledge. The exercise,
12 and the comments it evoked, showed why developing such a database is not straightforward and the
13 views on its desirability and utility are mixed. The main reasons are the lack of common concepts to
14 organize the database and the problem of the triple hermeneutic, different uses and non-uses of the
15 CMOC heuristic, and different perceptions of the quality of a CMOC. The discussion points raised in
16 this exercise offer valuable insights in realist reasoning and interesting avenues for further debate.

17 Key words
18 Realist evaluation, generative mechanisms, database, conceptualisation, CMO configuration

19 Introduction
20 Realist evaluation (RE) is a theory-driven evaluation approach that seeks to answer the question 'what
21 works for whom, in what circumstances, how, and why?' (Pawson and Tilley, 1997). It distinguishes
22 itself from other theory-driven approaches by its underlying realist philosophy inspired by the writings
23 of critical realist scholars[1] (like Archer, 1995; Sayer, 1992; Bhaskar, 2008). According to this realist
24 philosophy, events (including intervention outcomes or phenomena) are generated by mechanisms
25 that can be defined as 'underlying entities, processes, or structures which operate in particular
26 contexts to generate outcomes of interest' (Astbury and Leeuw, 2010: 368). By analysing and
27 identifying these mechanisms and the elements of context necessary for their operation (or which
28 prevent their operation), it aims to explain both the desired and undesired outcomes of an
29 intervention.

30 This context-specific, causal mechanism way of thinking is reflected in the context-mechanism-
31 outcome configuration (CMOC). Program theory can be defined as a detailed account of how we think
32 the intervention led or will lead to the observed outcome. It often comprises a combination of several
33 CMOCs. During a realist evaluation, different possible CMOCs are investigated to come to a refined
34 program theory that better explains why certain outcomes were (or were not) achieved in a specific
35 context. While the CMOC is only a heuristic to guide deeper reasoning of the realist evaluator, it
36 remains a very common way to explain and present the causal processes initiated by an intervention.

37 Since the seminal work by Pawson and Tilley (1997), RE has become increasingly popular (see Renmans
38 and Castellano Pleguezuelo, 2023; Nielsen et al., 2022; Lemire et al., 2020), showing its relevance for
39 the evaluation field. However, Pawson and Manzano-Santaella (2012) warn us of 'fake handbags':
40 evaluations that are realist in name, but do not or only partially adhere to realist principles set out by

---

[1] See Chapter 1 in Pawson's (2013) *The Science of Evaluation* for an overview of intellectual origins and influences on realist evaluation.

41 Pawson and Tilley (1997). These principles include, among others, that mechanisms should be
42 distinguished from actions or program components and are "causal powers of individuals and
43 communities" (p. 215), including stakeholders' reasoning; "the operation of mechanisms is contingent
44 on context" (p. 216) and "context refers to the spatial and institutional locations of social situations
45 together, crucially, with the norms, values and interrelationships found in them" (p. 216). Most of such
46 'flawed' studies are unintentionally imperfect. Indeed, it takes time to understand fully the
47 philosophical underpinnings of RE and their repercussions for evaluation practices. Moreover, despite
48 the very instructive RAMESES guidelines (Wong et al., 2016), there exists no detailed guidance on how
49 to do a realist evaluation or how to use its key concepts.

50 Indeed, several recent reviews show that the earlier mentioned principles are still not adhered to, too
51 often (Greenhalgh and Manzano, 2022; Lemire et al., 2020; Nielsen et al., 2022). Arguably, an
52 important issue appears to be that undertaking a realist evaluation necessitates skills, such as creative,
53 critical and flexible thinking, and abductive and retroductive reasoning (see Jagosh, 2020; Mukumbang
54 et al., 2021). These skills go beyond technical know-how and the ability to follow a prescriptive
55 protocol.

56 The lead author hypothesised that a database facilitating access to examples of realist causal
57 reasoning in the form of CMOCs might contribute to the development of these skills. Instead of simply
58 reading theoretical discussions, novice realist evaluators could quickly get 'a feel' for what realist
59 causal reasoning looks like in empirical work.

60 Another rationale for the development of a CMOC database can be found in one of the more
61 scientifically oriented objectives of realist evaluation, namely the accumulation of knowledge on
62 programs and their mechanisms. Pawson and Tilley (1997) discussed the need to go further than one-
63 off evaluations and accumulate knowledge across evaluations. It was hypothesised by the lead author
64 that the development of a CMOC database may contribute to accumulation of knowledge by
65 facilitating and promoting the use of existing realist causal explanations. The database may function
66 as a starting point to find similar realist explanations to inform development of initial program theory.

67 With these two main objectives of learning and knowledge accumulation in mind, and the
68 encouragements of other realist evaluators, the lead author set out to develop such CMOC database
69 by bringing together a group of seasoned realist researchers to contribute to this endeavour in a
70 Delphi-styled exercise. However, several important obstacles to the development of an easily
71 accessible CMOC database were identified. The process was therefore abandoned yet the experience
72 turned out to be very informative.

73 The focus of this paper is not to argue in favour or against a CMOC database, but to report on the most
74 interesting findings from the development process in the hope of stimulating further dialogue on the
75 merit of this exercise and to ensure that any future database development efforts consider the lessons
76 drawn from this initial attempt. The paper also aims to share with the wider evaluation community
77 key ideas and avenues for further discussion identified by participants during the development
78 process.

79 In the next section we discuss the approach taken to develop the database. The results section reports
80 on the main obstacles and arguments that came up during the process. We finally highlight some main
81 takeaways from the exercise and possible ways forward.

Methods

83 The methods used in this exercise were continuously adapted, as one of the objectives was to come
84 to a workable method to develop a database. Hence, no general methodology can be described,

85 although Delphi method principles formed its basis (Helmer-Hirschberger, 1967). In the following
86 sections, we provide a step-by-step explanation of the process and reasons for adaptations.

The Delphi principles

88 The Delphi technique (Helmer-Hirschberger, 1967) is a method developed to reach decisions using
89 expert opinions while avoiding arguments from specious authority, the difficulty of abandoning
90 publicly expressed opinions, and the tendency for groupthink (Helmer-Hirschberger, 1967: 7). The
91 underlying idea is that by avoiding these influences, participants will come to more accurate
92 conclusions (Dalkey, 1969). While it was initially developed to make predictions about the future, it
93 has frequently been used to address other tasks such as identifying the state of the art, developing
94 tools and/or indicators, developing recommendations or creating standards and guidelines (Jorm,
95 2015; Niederberger and Spranger, 2020) – including the RAMESES guidelines for realist review (Wong
96 et al., 2013) and realist evaluation (Wong et al., 2016).

97 The technique involves experts being brought into an anonymous round-table discussion with the aim
98 of reaching a consensus. The process can be roughly summarized as follows. First, a facilitator
99 purposefully selects experts in the field of enquiry. The latter are unaware of who else is participating.
100 The facilitator then sends out a questionnaire on the topic of interest. Each expert answers the
101 questions using their knowledge of the topic and sends it back to the facilitator who summarises the
102 responses. In a second round, the facilitator shares the summary with the experts who are now asked
103 whether they would like to revise the draft summary and why or why not. Again, the facilitator
104 summarises the responses, and a third round is started. This goes on until a consensus is found or
105 stops after a pre-specified number of rounds (Jorm, 2015; Niederberger and Spranger, 2020).

106 The underlying approach of the database exercise was broadly similar. That is, experts' opinions about
107 CMOCs were solicited while making sure they could do this anonymously. It soon became clear that it
108 was not possible to go through continued rounds as responding to the surveys was time-consuming
109 and the number of CMOCs to analyse was too unwieldy. The process was therefore adapted at several
110 points to make it more feasible (see below).

The process

112 This section describes the different steps and the questionnaires that were sent out during the study.
113 For transparency reasons, we add the reports (including the main questions asked) that were sent to
114 the participants after every round in 'Supplementary Material 1'. The analysis of this process can be
115 found in the results and discussion section.

116 The study started by the lead author purposefully selecting a group of realist evaluators who could be
117 seen as experts given their publishing record and included some degree of diversity in terms of
118 backgrounds and work disciplines, geographical locations and years working in the field. First, 32
119 candidates were contacted of whom 24 responded positively. Eventually, 22 realist experts responded
120 to at least one survey. In relation to their expertise, 14 indicated to use RE 'always' or 'most of the
121 time', two indicated 'Sometimes' and three said 'About half the time'. Three did not respond. In
122 addition, 10 of the participants contributed to the 'Doing realist research' book (Emmel et al., 2018),
123 which is another indicator of expertise. Those who replied to at least three surveys were invited to
124 contribute to this paper as co-authors. Those who accepted, contributed to the manuscript, and
125 validated the final paper. The lead author took the role of facilitator for the process overall.

126 As Table 1 shows, participation strongly diminished towards the end of the study. This was probably
127 due to the heavy workload linked to the process for the participants and the lead author (which caused
128 delays), but possibly also to the lack of tangible results, as described in the results section.

Nevertheless, several participants indicated that it was a meaningful experience and made them reflect on their own work and the principles of realist evaluation.

Participant	Survey number						
	1	2	3	4	5	6	7
1	■	■	■	■			
2	■	■	■			■	■
3	■	■	■				
4		■					
5	■						
6				■			
7	■	■					
8	■						
9	■	■	■				
10	■	■					
11	■	■	■	■	■		■
12	■	■		■			
13	■	■	■	■	■	■	
14	■	■	■				
15	■	■			■		
16	■	■	■	■	■	■	■
17	■						
18	■	■					
19	■	■	■	■	■	■	■
20	■	■	■	■	■	■	
21				■			
22						■	■
Total	17	16	10	9	7	5	5

Table 1: Overview of the participation to the different surveys

The first survey that was sent out contained seven CMOCs that were purposively selected from several papers in an unrelated mapping review (Renmans and Castellano Pleguezuelo, 2023). The seven CMOCs aimed to demonstrate different forms, e.g., CMOCs presented in tables, full texts, and figures. For each CMOC the following questions were asked:

1. Do you feel this is a realist mechanism? (yes/no/don't know)
2. How would you term this mechanism? (Proposed term/other term)

After each question the participants had the opportunity to add comments. The responses and these comments were then summarized in a feedback report together with new questions in relation to the issues raised in the first survey (the feedback reports can be found in Supplementary Material 1).

The second survey did not have any new CMOCs. Instead, it presented questions on discussion points raised in the feedback report based on the responses to the first seven mechanisms. These questions concerned, among other things, whether the conceptualisation proposed by Dalkin et al. (2015) that explicitly distinguishes between resources and reasoning within the mechanism is useful or necessary; whether it is useful to adopt established scientific concepts and definitions; the best way to present CMOCs; and the use of multiple mechanisms in a single CMOC. A summary of the very insightful

147 arguments brought forward in response to the second survey can be found in the 'Supplementary
148 Material 1' and will be discussed later.

149 After the responses to this survey, an alternative third survey was sent out to understand whether the
150 response differences were due to a different attitude towards the purpose of realist evaluation. This
151 survey described three ideal types of realist evaluation - 'realist evaluation evaluation', 'realist
152 evaluation research' and 'realist evaluation science'. However, this distinction was not widely
153 supported and hence was abandoned. (This survey is not included in the supplementary material).

154 The fourth, fifth, sixth and seventh surveys returned to the original structure, with five new
155 mechanisms each. They did however entail some important changes. First, the initial question
156 changed from "Do you feel this is a realist mechanism?" to "How would you evaluate the overall
157 quality of this CMOC? (i.e., structure of the causal explanation, mechanism, etc.)". This was done
158 because some participants highlighted that a mechanism cannot be seen apart from the CMOC as a
159 whole. Second, the response possibilities were expanded from a yes/no question to a five-point Likert
160 type scale (very poor, poor, average, good, excellent) to suit the question stem and allow more
161 variation in the responses. Finally, in the last surveys, when justifying their response, respondents
162 were given the option to choose among a selection of often mentioned critiques on the CMOCs in the
163 earlier surveys to make the process more streamlined and efficient. The focus of the analysis moved
164 away from a mainly quantitative driven search for consensus to a more qualitative appreciation of
165 differences. These last four surveys were summarised, and feedback was provided to participants, but
166 the surveys were no longer sent back to the participants for follow-up questions. Hence, it departed
167 from some of the central principles of the Delphi technique for reasons discussed in the next section.

168 Overall, 27 CMOCs were evaluated (see 'Supplementary Material 1'). Because of the heavy workload
169 for both the lead author and the participants, the diminishing participation, and the limitations
170 encountered in relation to the database, the intended development of a CMOC database was
171 abandoned.

Results and discussion

173 Although the development of the database was unsuccessful, the lessons learned from the exercise
174 may help to inform future realist evaluation debates, and/or to inform (or deter) efforts to develop
175 similar databases. In providing this article, we also avoid the well documented publication bias against
176 negative results (Franco et al., 2014).

177 In this section we will discuss these lessons and the different positions taken by the respondents. We
178 use anonymous quotes from the surveys to illustrate our findings and depict the different views.

Organising the database

180 The lead author reflected on several ways to organise the database and it became clear that not all
181 options are equally useful, as each had their own difficulties and concerns.

182 First, organising the database solely according to the setting of the study, e.g. by country or by sector,
183 would defeat the purpose of learning across domains, contexts and sectors which is central to the
184 realist evaluation goal of middle range theory building (Pawson, 2013). Second, organising according
185 to the outcome would assume that only a limited set of mechanisms could possibly create a particular
186 outcome – something that is by no means established in social sciences.

187 A third possibility would be to organise according to the intervention family under evaluation, as
188 envisaged by Pawson (2013) when talking about 'generic conceptual platforms' and also Leeuw (2023)
189 when he called for 'subsummation' - "[learning] from earlier evaluations and (other) research by

190 bringing the intervention or program (and preferably its mechanisms) that they are working on under
191 a more general 'umbrella.'"(p.416). While no two interventions are the same and they may differ in
192 small but significant ways, the realist approach emphasises that the intervention and its peculiarities
193 are not the focus of the study; rather, the programme theory underlying the intervention is (Pawson,
194 2006). Therefore, the database could be organized according to intervention families and their related
195 program theories, based on the 'necessary and internal components' that make up the type of
196 intervention (Pawson, 2013). (Note that 'internal components and their necessary relations' have
197 particular meanings in realism. See Pawson, 2013, pp 92 or Sayer, 1992: 89). While a promising idea
198 for the future of RE, the main organising elements are still lacking. Very few studies have focused on
199 developing the 'generic conceptual platforms' (some notable exceptions being Burrows (2020) and
200 Pearson et al. (2015)) and, hence, few or no studies explicitly mention the intervention family to which
201 the intervention under study belongs. While this might be deduced from the description of the
202 intervention, the lack of explicit intervention families and generic conceptual platforms leaves us
203 without a framework to work with, although existing taxonomies may help in this endeavour (e.g., Kok
204 et al., 2016).

205 Finally, it could be useful to organize the database according to the mechanism embedded in the
206 CMOC. To do so, the mechanisms would need to be described using common concepts across
207 evaluations. The need for a process of abstraction and conceptualization is clear, but whether this is
208 feasible or desirable remains open for debate, as the discussion in the next section shows.

209 The lead author eventually decided to use mechanisms as the organising principle of the CMOC
210 database, in line with a previously developed approach related to realist evaluation that puts
211 mechanisms at the centre of the knowledge accumulation process, rather than whole programme
212 theories (Renmans, 2023) .

Lack of common concepts

214 To organise the database according to mechanisms, we needed a common language to describe
215 mechanisms that were essentially the same. The lead author hypothesized that scientific concepts
216 could fulfil this function. These are terms that have a specific definition in the scientific literature
217 (although not fixed forever) and serve as the building blocks of scientific theories such as: intrinsic
218 motivation (self-determination theory (Deci and Ryan, 2000)), self-efficacy (social cognitive theory
219 (Bandura, 1997)), or social comparison (social comparison theory (Festinger, 1954)). While some
220 participants agreed with the idea of using scientific concepts to describe studied mechanisms, the
221 process and the comments to the surveys highlighted several obstacles.

222 Foremost, many realist evaluations do not use scientific concepts to describe mechanisms. This may
223 be because programmes themselves are rarely designed around such concepts, and evaluations are
224 often designed primarily to inform decisions and actions, rather than contribute to general scientific
225 theory-building. Moreover, they often involve multiple stakeholders from various backgrounds,
226 making the use of layperson's concepts to describe the mechanisms more common.

227 In the absence of such concepts within the empirical RE literature, it would have been up to the
228 database developers to link mechanisms explained in lay terminology to existing scientific or newly
229 created realist concepts. However, participants highlighted associated problems.

230 First, appropriate concepts are often lacking to describe the mechanisms found in many realist
231 evaluations. They may not yet have been defined as concepts in the scientific literature or be known
232 by the realist evaluator, or the realist mechanisms may be more fine-grained than the scientific

concepts at our disposal. Hence, using existing scientific concepts may weaken the explanatory power of the CMOC:

> "It depends on how refined the mechanism has become (after cycles of theory testing). Therefore, we should be open to the use of descriptive "ungrounded" mechanisms as well as those that are already grounded in existing concepts."

Second, concepts may mean different things in different disciplines, theories and/or studies, have definitions that are not in line with the realist way of thinking (for example, they may lack ontological depth) or be defined inaccurately, hence using them may lead to more confusion instead of clarity.

Third, scientific concepts are in constant development and may change over time, proving the initial definition inaccurate or incomplete. Finally, some respondents – in line with the focus in many evaluations on practical results – highlighted that a focus on the operationalization of concepts to fit a specific study might be more fruitful than discussing definitions.

> "The same concept may actually mean quite different things in different studies."

> "You can't keep knowledge still, and you can't hold language still. Realist epistemology says that knowledge develops over time - so does language. So too do concepts and definitions."

> "The challenge that may arise is searching for such literature, in which set of social or psychological theories without losing the relevance and also by avoiding the cherry picking of theories."

> "In line with Karl Popper, definitional issues are 'what is-questions' that run the risk of leading to essentialism. Operationalization I prefer over big discussions about definitions."

Several of these issues surfaced when participants were asked to link existing scientific concepts to the mechanisms in the CMOCs. Box 1 shows one of the main problems of such an approach: participants assigned very different concepts to the same mechanism.

Box 1: Assigning concepts to a mechanism.

Participants received the CMOC below and were asked to describe it with a scientific concept. They could either choose 'social comparison' or give another scientific concept.

"Feelings of inadequacy prevailed coupled sometimes with isolation (CONTEXT). The SMA [shared medical appointment] created social contact amongst a group of people with similar illness experiences. This exposure helped to correct misperceptions about their capabilities and the capabilities of others in self-efficacy (MECHANISM). The social contact combined with people sharing similar experience contributed to esprit de corps which promoted self-efficacy (OUTCOME)." (Kirsh et al., 2017)

Out of ten respondents, 4 agreed with social comparison, one did not know and five gave the following other scientific concepts: peer support, reciprocity, social exposure-driven self-help, social identification, and social solidarity.

Clearly, in this sense, developing a CMOC database based on mechanisms is as much a theoretical and interpretative endeavour as it is an empirical mapping. This can be called a 'triple hermeneutic' in which the database developers interpret the interpretation of people (researchers) who have interpreted the interpretation of other people (study participants). This may contribute to the misidentification of the mechanisms and partly explain the different opinions observed in Box 1.
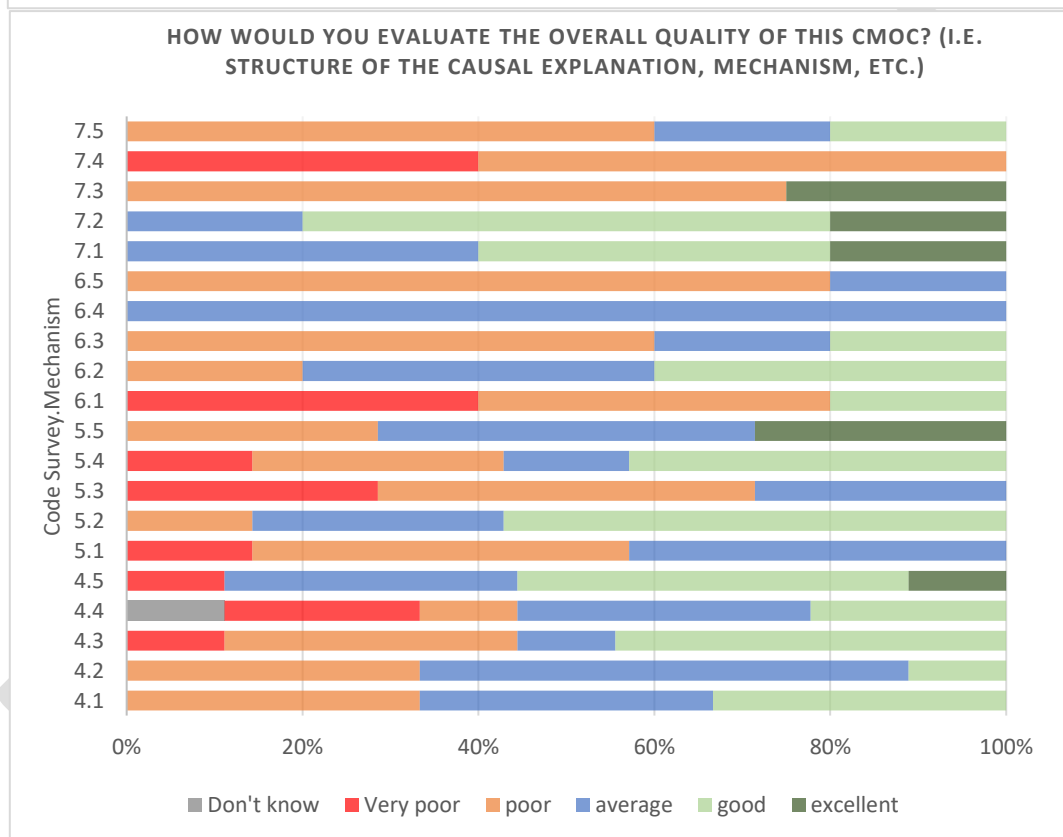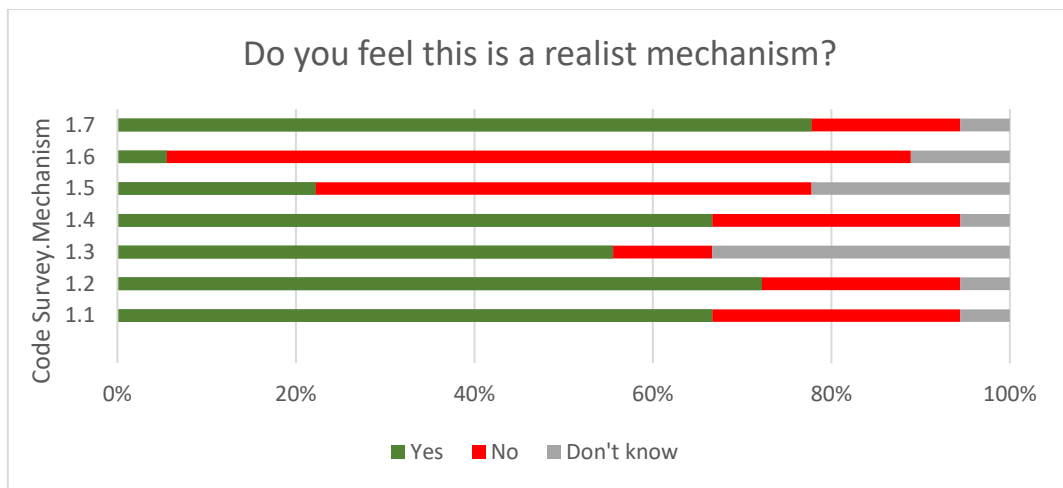
The quality of the CMOC

As well as a theoretical endeavour, establishing a database is arguably also a normative undertaking. Even though a database might be just a simple collection of published CMOCs, it is difficult to escape the implied authority that would automatically come with such a database: 'what is in the realist CMOC database must be a realist CMOC'. Moreover, one of the main drivers behind the development of the database was the idea that it would help novice realist evaluators to get a grip on what a mechanism and a CMOC are. Consequently, this means that the database would need a procedure for quality control. Yet the quality of a CMOC is not just dependent on its formulation but also the context in which it was developed, i.e., the methods used, the evaluand, and the evidence to support it. This broader evaluation context cannot be considered in an assessment of the CMOC alone. As a result, only a relatively superficial quality control focused on the logical connection between the C, M and O and of the conceptualization of the mechanism was attempted.

In Figure 1, each bar in the two diagrams represents a CMOC that was presented to the panel, while the colours represent the percentage of respondents that gave a certain answer. The upper diagram shows the CMOCs presented in the first survey, while the lower diagram concerns the CMOCs presented during surveys 4 until 7, when the question was slightly different.

289

Figure 1: Visualization of divergence of opinions

Do you feel this is a realist mechanism?

Legend: Yes, No, Don't know

290

HOW WOULD YOU EVALUATE THE OVERALL QUALITY OF THIS CMOC? (I.E. STRUCTURE OF THE CAUSAL EXPLANATION, MECHANISM, ETC.)

Legend: Don't know, Very poor, poor, average, good, excellent

291

We observed very different opinions about what a good realist mechanism and CMOC are. The comments from the respondents that accompanied the rating showed some lack of consensus, and/or of clarity, on how a realist mechanism and CMOC should be defined, operationalized and/or conceptualized. While everyone agrees that different visions may co-exist and are even necessary to allow for a fit-for-purpose approach, some discussion points sparked interesting and constructive comments for further debate. (See also Supplementary Material 1.)

Since a final judgment about the quality of a CMOC could not be made, it was decided not to hold back CMOCs from the database but to add a short paragraph on the main discussion points for each of the CMOCs (see Box 2).

301	Box 2: Example of a commentary on a CMOC
302 303 304 305	The panel members rated this CMOC as average but more leaning towards poor. The main criticism here is that it consists of two CMOCs merged into one and it would have been better to split them. One member felt that "perceptions of a more appropriate skill mix within the clinical and managerial teams" is the central mechanism and not the outcome.

306 Although very instructive for the practice of realist evaluation, it remained a very time-consuming and
307 tedious process that was difficult to do continuously, even across the limited number of CMOCs
308 considered. Efforts to streamline this by providing predefined options of critique failed, due to the
309 specificity of the critique of each of the individual CMOCs.

310 Another important aspect that emerged in relation to the quality assessment concerned the CMOC's
311 comprehensiveness and accuracy. The idea of the CMOC database was based on the premise that a
312 CMOC can supply a sound causal and understandable explanation of how some outcome is achieved
313 – or not – and the necessary circumstances for it, in a way that is transferrable to other realist
314 investigations. However, many CMOCs do not honour the wealth of information that is present in the
315 other parts of the paper. Indeed, they were sometimes a heavily summarised version of the
316 sometimes very nuanced and rich causal explanations described elsewhere in the paper. The final
317 CMOC was in such case perceived as a shortened summary of the work or an aide-mémoire,
318 sometimes in the form of a table or a small figure. Some saw this as a limitation inherent to the CMOC,
319 which they suggested was not always the most appropriate formula to capture a complex realist
320 explanation.

321 Others pointed out that a CMOC should not be seen as a truncated summary but needed all the details
322 necessary to understand the causal explanation. This could include the development of as many
323 CMOCs as necessary to explain long and non-linear causal chains. In part, the debate revolved around
324 whether a CMOC can – or should – be seen as a standalone product (meaning that it is useful and
325 understandable without the context of the full paper) or not. If not, it seems difficult to develop a
326 CMOC database.

327 Whereas the quality assessment proved to be too onerous for our objective, it did bring forward some
328 important insights on current realist practice and potential 'requirements' for CMOCs. We mention
329 those insights that are relevant in relation to the database.

330 ## Form 'requirements' of CMOCs
331 We put the term 'requirements' in quotation marks because it is important to acknowledge that the
332 realist approach is one of methodological and to some extent also of conceptual openness. This leaves
333 room for different fit-for-purpose approaches. Indeed, the following discussion shows that the CMOC
334 heuristic in particular is used in different ways and can (and perhaps should) be adapted according to
335 the needs of the evaluator and context in which it is being used.

336 Some panel members believed that multiple mechanisms and outcomes in a single CMOC made it
337 difficult to understand the causal claim being made. However, opinions about whether a CMOC should
338 contain only one mechanism and one outcome were divided. To some, a CMOC can, and maybe even
339 should, have multiple mechanisms. They stated that mechanisms may combine and depend on one
340 another to lead to a certain outcome.

341 "I fully disagree [that CMOCs should have just one mechanism] because simply it may be a
342 combination of mechanisms in a specific set of conditions, set of mechanisms, and so on..."

343 Others opposed the use of multiple mechanisms. As the mechanism is the main unit of analysis, it is
344 central to a causal explanation that shows how an intervention leads to a specific outcome when
345 implemented in a context containing certain specific conditions. If the CMOC is sufficiently detailed
346 then it can be expected that the specific linkage between the three elements is unique to a specific
347 mechanism, outcome, and group of contextual conditions. Any other mechanism would be expected
348 to have other relevant contextual conditions and/or another outcome. Therefore, when more than
349 one mechanism is mentioned, it may be unclear to which of the mechanisms the relevant contextual
350 conditions refer or how the different mechanisms relate to each other: do they work side-by-side,
351 strengthen each other, are they contingent upon each other, and so on?

352 "Mechanisms are the 'unit of analysis' and as such should be analysed individually, and
353 where they are contingent on other mechanisms, this should be noted. "

354 Similarly, when multiple outcomes are mentioned in a CMOC, the link between them may not be clear:
355 are they causally linked, conceptually linked or are they outcomes at different levels? Moreover, it
356 may be questionable whether all conditions and all mechanisms were equally relevant for each
357 outcome. Therefore, according to this way of thinking, each CMOC should include one set of
358 contextual conditions, one mechanism and one outcome. Clearly, this approach may require several
359 different CMOCs to honour the realist assumption that many outcomes require multiple mechanisms
360 to operate concurrently, each affected by their own sets of conditions.

361 Other comments concerned the conceptualisation and operationalisation of mechanisms. Pawson and
362 Tilley (1997) suggested that a program mechanism is, among other things, a "[demonstration of] how
363 program outcomes follow from the stakeholders' choices (reasoning) and their capacity (resources) to
364 put these into practice" (p. 66). This distinction between reasoning and resources within the
365 mechanism was strengthened and made explicit in the heuristic proposed by Dalkin et al. (2015), as
366 in the following example:

367 "[t]he palliative care register [resource] which, when used with older adults who had
368 unpredictable illness trajectories (context), resulted in anxiety in registering these patients
369 (reasoning), which meant that less older patients in care homes were registered (outcome)"
370 (Dalkin et al., 2015)

371 Within the panel, very different positions were taken regarding this framing of mechanisms. Nearly
372 everyone agreed that this was just one way of depicting a mechanism (see also Westhorp, 2018) that
373 "could be useful in some contexts / to some researchers, and not useful in others." However, some
374 found it confusing when it was used while others felt it was missing whenever it was not used. Still
375 others put into question the usefulness of discussions on specific heuristics and warned against a too
376 recipe-like approach to RE.

377 "I think people get a bit too caught up in trying to think what's a C, what's an M and what's
378 an O - at the end of the day, we are trying to build explanations of causality and that is
379 what matters."

380 Those who were more critical towards this 'reasoning-resources heuristic' highlighted an
381 inconsistency between realist philosophy and the singling out of resources, arguing that "resources
382 [are] things which are in principle measurable, whereas mechanisms are not". Although resources as
383 defined within RE are not necessarily observable (and indeed often are not), some argued that when
384 this heuristic is used within the literature, 'resources' often refers to intervention components (which

385 by definition are not mechanisms, or part of mechanisms). Others found the division between the
386 context and the resources in the mechanism confusing.

387 However, some proponents emphasized the fact that Dalkin et al's heuristic is merely an (albeit just
388 one) operationalisation of how mechanisms were described by Pawson and Tilley (1997).

389 A similar argument was used in relation to factors added to the CMOC, such as 'intervention' (CIMO),
390 'strategies' (SCMO) or 'actors' (ICAMO) (see De Weger et al., 2020). One respondent highlighted that
391 "the very detailed account of a configuration [makes] it harder to move to the next layer of
392 abstraction". These factors were often added to facilitate the use of the CMOC in practice, yet De
393 Weger et al. (2020) state that while they can help to achieve the objectives of individual evaluations,
394 the "ontological 'status' of [for example] a strategy as an additional explanatory factor remains
395 unclear" (p.5). For example, if the strategy is financial incentives, they may be relevant because they
396 increase the salary of employees (i.e., change the context) which in turn may trigger higher motivation
397 or a feeling of appreciation. However, the financial incentives can also trigger mechanisms in
398 themselves like a feeling of achievement. Hence, it is unclear how describing the strategy contributes
399 to the deeper understanding of the workings of the mechanism or if it is used to clarify the relationship
400 between the intervention (which may be multi-faceted) and the 'resources' that the intervention
401 provides - i.e. "this aspect of the intervention provides 'x' resource, in response to which some
402 participants reasoned 'y'".

403 These different positions and/or different uses are in themselves unproblematic as they may be useful
404 in different situations depending on the objectives of the evaluation. However, they further
405 complicate the task of developing a workable database.

Ways forward

407 This attempt to develop a realist CMOC database to support learning and knowledge accumulation
408 encountered several important obstacles. However, the value of the exercise lies in the insights it
409 provided on these obstacles and the arguments surfacing of issues that may further debate.

410 An important conclusion is that the rationales underlying individual realist evaluations differ from the
411 rationale for the development of a CMOC database . As a result, what is useful for the former is
412 inconvenient or even detrimental for the latter - for example, the use of different heuristics (CIMO,
413 ICAMO, etc.). This should not be a surprise as RE is a relatively flexible approach that allows for fit-for-
414 purpose adaptations rather than strict methods prescription. Objectives and underlying rationales of
415 realist evaluations influence the methodological and conceptual choices made.

416 As De Weger et al. (2020) emphasize, the choice for a specific heuristic (CMO, ICAMO, CSMO or other
417 variations) should be guided by the objectives of the evaluation. There is no reason why this should
418 be any different for how we approach the concept of a mechanism – so long as the concept used
419 remains consistent with a realist philosophy of science. Hence, some realist evaluations will use
420 established scientific concepts, while others will stick to a more empirical description; some will have
421 clearly distinguishable CMOCs, while others will have a dense narrative. Strict guidelines are not
422 necessarily desirable, which means that this diversity will remain an important obstacle for the
423 development of a coherent CMOC database and at the same time an important asset for the wider
424 applicability of RE.

425 This exercise also revealed some more fundamental questions about the role and form of the CMOC
426 in realist evaluation. Should a mechanism explicitly distinguish between resources and reasoning?
427 Should a CMOC entail only one mechanism and/or outcome? Is a particular format of CMOCs desirable

428 in realist explanations? And if so in what circumstances, when, how, and why? These questions go to
429 the essence of RE and the appropriate forums (conferences, webinars, mailing lists, etc.) should be
430 mobilized to discuss them in order to foster mutual learning and understanding, as also emphasized
431 by Pawson and Manzano-Santaella (2012).

432 Even after the failure of this attempt, the question remains whether a CMOC database is feasible
433 and/or desirable and if it could achieve the original objectives. In accordance with realist practice, the
434 answer is: 'it depends', with responses varying according to the position each practitioner takes on
435 the above-mentioned questions and discussion points.

Position 1: A database is not feasible and undesirable
437 Based on the above discussion, it can be argued that developing a CMOC database is not feasible nor
438 is it ever likely to be desirable. This is firstly because a common language of concepts is missing for
439 mechanisms, outcomes, and intervention families for the reasons elicited above. Secondly, it is too
440 time consuming to assess quality; or even, as some may argue, inherently impossible as no quality
441 standards for CMOCs or descriptions of mechanisms exist. Finally, CMOCs are not used in the same
442 way across evaluations: they can be presented in tables, figures or in text, be extensive or a summary,
443 have just one or multiple mechanisms and/or outcomes, use the CMO heuristic, another heuristic
444 (CIMO, ICAMO, CSMO) or no heuristic at all. These limitations do not only affect the feasibility of a
445 CMOC database, but also its desirability. If a CMOC database tries to overcome these limitations, it
446 may force causal explanations into a CMOC straitjacket which may entail a loss of detail and clarity.
447 Moreover, such a database may effectively push towards a possibly undesirable exacerbation of a
448 recipe-like use of CMOCs. In the end, the CMOC remains just a heuristic and not the end-goal of a
449 realist evaluation.

450 Such costs might be worthwhile if the database contributed significantly to its two main objectives of
451 learning and knowledge accumulation. However, the limitations mentioned clearly hamper the role a
452 CMOC database could play in the realist evaluation cycle of knowledge accumulation. Similarly, the
453 database failed to show its utility for the learning objective, because differing ideas about the quality
454 of CMOCs limit the ability of a CMOC database to present typical or good quality examples. Yet, what
455 did appear to be useful were the arguments clarifying positions on the quality of certain CMOCs. It
456 allowed several participants to clarify their assumptions and ideas about CMOCs and realist
457 evaluation. These can be important learning moments for both the discussants and the spectators of
458 the discussion.

Position 2: A database is feasible and desirable
460 Researchers who take a somewhat different position may be more optimistic (or stubborn) about the
461 feasibility and desirability of a CMOC database. While a database comes with its challenges and
462 limitations, it can be argued that it could nonetheless contribute to the two predefined objectives.
463 First, while there is a diversity of views on what quality means, certain basic rules (e.g., mechanisms
464 are not program components) do apply and could serve as a gatekeeper for CMOCs in the database.
465 This would leave intact the diversity of other positions for novice realist evaluators to discover without
466 leading them down a wrong path. A guiding text may help them navigate this diversity. Second,
467 different realist evaluators may use different heuristics and concepts but are eventually trying to
468 describe the same underlying reality. Developers of a CMOC database could try to develop a common
469 language (i.e., novel realist concepts or adapted scientific concepts) to enable them to identify CMOCs
470 that have different wordings but essentially refer to the same mechanism and facilitate the
471 accumulation of knowledge across realist evaluations. Importantly, the unit of analysis and organizing
472 element would in this case be the mechanism and its link to contexts and outcomes (the CMOC). In

473 this view, a CMOC database is not only feasible, but also desirable, and could contribute to the stated
474 objectives.

475 However, it remains important to acknowledge that the earlier discussed 'triple hermeneutic' will
476 always remain an obstacle.

Position 3: Another database is feasible and desirable

478 One could also take the position that a database is feasible but not in the way that was envisaged in
479 this exercise by the lead author. Indeed, while this exercise focused primarily on mechanisms as the
480 organising principle, other principles are also possible.

481 Probably the most promising alternative is to organise the database according to intervention families,
482 i.e. interventions with very similar necessary and internal sets of relations and components and,
483 hence, underlying program theory (e.g., vouchers, pay-for-performance, cash-on-delivery can be seen
484 as belonging to the 'incentivization family') (see Pawson, 2013; Leeuw, 2023). The unit of analysis
485 changes from the individual CMOCs to CMOCs linked together in program theories. Again, different
486 options present themselves.

487 First, it could simply list the different CMOCs and/or program theories and organise them according
488 to intervention family. In this way the task of ad hoc abstraction is left to the user of the database. A
489 second option is to go a step further and develop a realist evidence-based program registry (EBPR).
490 These EBPR show the evidence related to certain types of interventions and are becoming increasingly
491 common. However, they often use the standard hierarchy of evidence that puts randomized control
492 trials on top (Magura et al., 2023). A realist version of such a registry may analyse the realist evidence
493 in relation to a certain intervention family and put together a refined theory with that evidence (i.e.,
494 performing realist reviews). This would also fit nicely within a broader trend towards middle-range
495 theories in evaluation in general (Cartwright, 2020; Guerzovich et al., 2022).

496 While some level of abstraction and interpretation remains necessary (invoking the triple
497 hermeneutic), more theoretical coherence within the intervention families would make the
498 abstraction exercise more feasible. However, both approaches require a clear development of family
499 interventions, which could be based on existing taxonomies (e.g., Kok et al., 2016). In the beginning it
500 might also be useful to not structure it according to intervention families but to stay one step lower
501 on the ladder of abstraction and to organize it according to intervention types (e.g., treat vouchers
502 and pay-for-performance separately).

Conclusion

504 While the feasibility and potential value of a CMOC database remains uncertain, the value of this initial
505 attempt lies in its illumination of the rich diversity of positions on the role, form and use of CMOCs.
506 The paper also highlights the main obstacles in relation to the development of a CMOC database
507 aiming to prevent future attempts to make the same mistakes or even to deter some from getting
508 initiated.

509 Importantly, we do not argue in favour or against a CMOC database, as some disagreement will
510 probably remain for some time. Instead, the paper should be seen as an invitation for further
511 constructive debates about 'which methodological choices work best, when, for whom, in what
512 circumstances and why'.

513

References

Archer MS. (1995) *Realist Social Theory: The Morphogenetic Approach,* Cambridge: Cambridge University Press.

Astbury B and Leeuw FL. (2010) Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation. *American Journal of Evaluation* 31: 363-381.

Bandura A. (1997) *Self-efficacy: The exercise of control,* New York, NY, US: W H Freeman/Times Books/ Henry Holt & Co.

Bhaskar R. (2008) *A realist theory of science,* London: Verso books.

Burrows L. (2020) More than tea and cake: a realist evaluation of memory cafés in Cornwall. *School of Nursing and Midwifery.* Plymouth, UK: University of Plymouth.

Cartwright N. (2020) Using middle-level theory to improve programme and evaluation design. *CEDIL Methods Brief.* Oxford: CEDIL.

Dalkey N. (1969) An experimental study of group opinion: The Delphi method. *Futures* 1: 408-426.

Dalkin SM, Greenhalgh J, Jones D, et al. (2015) What's in a mechanism? Development of a key concept in realist evaluation. *Implementation Science : IS* 10: 49.

De Weger E, Van Vooren NJE, Wong G, et al. (2020) What's in a Realist Configuration? Deciding Which Causal Configurations to Use, How, and Why. *International Journal of Qualitative Methods* 19.

Deci EL and Ryan RM. (2000) The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry* 11: 227-268.

Emmel N, Greenhalgh J, Manzano A, et al. (2018) *Doing realist research,* London: Sage.

Festinger L. (1954) A theory of social comparison processes. *Human Relations* 7: 117-140.

Franco A, Malhotra N and Simonovits G. (2014) Publication bias in the social sciences: Unlocking the file drawer. *Science* 345: 1502-1505.

Greenhalgh J and Manzano A. (2022) Understanding 'context' in realist evaluation and synthesis. *International Journal of Social Research Methodology* 25: 583-595.

Guerzovich F, Aston T, Levy B, et al. (2022) *'How do we shape and navigate pathways to social accountability scale? Introducing a middle-level Theory of Change,* London and Oxford: Centre of Excellence for Development Impact and Learning (CEDIL).

Helmer-Hirschberger O. (1967) Analysis of the future: The Delphi method. Santa Monica, California, USA: Rand Corporation.

Jagosh J. (2020) Retroductive theorizing in Pawson and Tilley's applied scientific realism. *Journal of Critical Realism* 19: 121-130.

Jorm AF. (2015) Using the Delphi expert consensus method in mental health research. *Australian & New Zealand Journal of Psychiatry* 49: 887-897.

Kirsh SR, Aron DC, Johnson KD, et al. (2017) A realist review of shared medical appointments: How, for whom, and under what circumstances do they work? *BMC Health Services Research* 17: 113.

Kok G, Gottlieb NH, Peters G-JY, et al. (2016) A taxonomy of behaviour change methods: an Intervention Mapping approach. *Health Psychology Review* 10: 297-312.

Leeuw FL. (2023) John Mayne and Rules of Thumb for Contribution Analysis: A Comparison With Two Related Approaches. *Canadian Journal of Program Evaluation* 37: 403-421.

Lemire S, Kwako A, Nielsen SB, et al. (2020) What is this thing called a mechanism? Findings from a review of realist evaluations. *New Directions for Evaluation* 2020: 73-86.

Magura S, Lee-Easton MJ, Abu-Obaid RN, et al. (2023) The influence of evidence-based program registry websites for dissemination of evidence-based interventions in behavioral healthcare. *Evaluation and Program Planning* 97: 102214.

Mukumbang FC, Kabongo EM and Eastwood JG. (2021) Examining the Application of Retroductive Theorizing in Realist-Informed Studies. *International Journal of Qualitative Methods* 20.

Niederberger M and Spranger J. (2020) Delphi Technique in Health Sciences: A Map. *Frontiers in Public Health* 8.

565 Nielsen SB, Lemire S and Tangsig S. (2022) Unpacking context in realist evaluations: Findings from a
566   comprehensive review. *Evaluation* 28: 91-112.
567 Pawson R. (2006) *Evidence-Based Policy: A Realist Perspective,* London: SAGE Publications.
568 Pawson R. (2013) *The science of evaluation: A realist manifesto,* London: Sage publications.
569 Pawson R and Manzano-Santaella A. (2012) A realist diagnostic workshop. *Evaluation* 18: 176-191.
570 Pawson R and Tilley N. (1997) *Realistic evaluation,* London: Sage.
571 Pearson M, Brand SL, Quinn C, et al. (2015) Using realist review to inform intervention development:
572   methodological illustration and conceptual platform for collaborative care in offender
573   mental health. *Implementation Science* 10: 134.
574 Renmans D. (2023) The ResQ approach: theory building across disciplines using realist evaluation
575   science and QCA. *International Journal of Social Research Methodology* 26: 469-482.
576 Renmans D and Castellano Pleguezuelo V. (2023) Methods in realist evaluation: A mapping review.
577   *Evaluation and Program Planning* 97: 102209.
578 Sayer A. (1992) *Method in Social Science: A Realist Approach,* London, New York: Routledge.
579 Westhorp G. (2018) Understanding mechanisms in realist evaluation and research. In: Emmel N,
580   Greenhalgh J, Manzano A, et al. (eds) *Doing realist research.* London: Sage Publications, 41-
581   57.
582 Wong G, Greenhalgh T, Westhorp G, et al. (2013) RAMESES publication standards: realist syntheses.
583   *BMC Medicine* 11: 21.
584 Wong G, Westhorp G, Manzano A, et al. (2016) RAMESES II reporting standards for realist
585   evaluations. *BMC Medicine* 14: 96.

586