Linking Measures to Mechanisms of Action: An Expert Opinion Study

Abstract

Objective: This study sought to integrate the NIH Science of Behavior Change (SOBC) measures repository comprising measures of putative mechanisms with mechanisms of action (MoA) identified by the Human Behaviour-Change Project (HBCP).

Design: Participants were 30 international experts recruited from professional networks and societies. In three anonymous virtual rounds, experts established consensus on hypothesized links between 26 MoAs and 44 self-report measures.

Methods: In Round 1, experts completed a survey rating agreement with 84 pre-identified measure-MoA links and suggested new links. In Round 2, experts discussed 10 links in an online forum, including pre-identified links with <50% agreement and new links suggested by 20-50% of experts. In Round 3, experts completed a survey rating all links eligible for discussion in Round 2.

Results: Twenty-seven experts completed Round 1, 23 completed Round 2, and 18 completed Round 3. In Round 1, 82 of 84 pre-identified links reached >50% agreement and 14 new links were suggested by >50% of experts. In Round 2, experts discussed measure-MoA links and measurement quality. In Round 3, 71 of 96 links reached $\geq$50% agreement. A total of 167 links reached >50% expert agreement, 33 of which reached $\geq$90% agreement.

Conclusion: By identifying putative mechanisms (HBCP) for the 44 self-report measures (SOBC), this study advances the cumulation of scientific results and interoperability of resources to facilitate process research.

Keywords: behaviour, behaviour change, mechanisms, measurement

Background

To improve the rigor, reproducibility, and impact of behavioural interventions, research has been conducted into the hypothesized mechanisms that may underlie intervention success. This includes two groups of researchers working in complementary areas, the US-based NIH Science of Behavior Change (SOBC) initiative and the UK-based Human Behaviour-Change Project (HBCP). This paper reports a collaboration between the two groups to identify measures associated with behavioural mechanisms of action.

The SOBC has adopted a mechanism-focused, experimental medicine approach to behavioural research (Riddle & Science of Behavior Change (SOBC) Working Group, 2015; Sheeran, Klein, & Rothman, 2017; Sumner et al., 2019; Sumner et al., 2018), which emphasizes evaluating how changes in a mechanism lead to changes in the target behaviour. For example, a researcher might hypothesize a goal-setting intervention promotes physical activity by increasing self-efficacy and test whether increases in self-efficacy attributable to the intervention lead to increases in physical activity. One of the core pillars of this initiative was to develop an online, publicly available Measures Repository (Science Of Behavior Change (SOBC)) where researchers could upload candidate measures of hypothesized mechanisms that underlie behaviour change and share data on behavioural validation (e.g., does a change in the mechanism lead to a change in behaviour?). The first group of measures uploaded were focused on three, broad mechanistic domains: stress, self-regulation, and interpersonal processes (though the measures themselves were not necessarily developed to assess these specific mechanisms). Each uploaded measure includes supporting information (e.g., on scale development and previous validation), scoring procedures, Open Science Framework (OSF) documentation, and links to

relevant citations in multiple reference management software platforms via Google scholar (Science Of Behavior Change (SOBC)).

Whereas SOBC researchers aimed to curate measures of mechanisms as one scientific resource for behaviour change research, researchers at the Human Behaviour-Change Project (HBCP) aimed to link behaviour change techniques (BCTs) used in intervention research with the mechanisms through which these techniques are hypothesized to operate (mechanisms of action; MoA) (Johnston et al., 2021). To achieve this, researchers conducted literature analysis (Carey et al., 2019), expert consensus (Bohlen et al.), and triangulated studies to systematically develop the putative MoAs for BCTs. This work was conducted using the most frequently used 74 BCTs within the 93-item Behaviour Change Technique Taxonomy v1 (Michie et al., 2015), such as Goal-Setting or Salience of Consequences, and 26 coded MoAs (e.g., Beliefs About Capabilities, Self-Image; **Supplemental Appendix A**) (Carey et al., 2019; Johnston et al., 2020). The MoAs were developed to represent the most commonly occurring MoAs across 83 theories of behaviour and behaviour change. Based on this work, HBCP scientists developed an interactive online resource, the Theory and Techniques Tool (TaTT), which details the strength of the support for (or against) a given hypothesized BCT-MoA link (Human Behaviour Change Project (HBCP)). The TaTT includes functionality for users to comment and provide additional information regarding each link.

Advancing science cumulatively is facilitated by compiling systematic knowledge about all parts of the causal chain linking interventions, mechanisms, and behaviours. Lack of attention to measurement of hypothesized mechanisms of change presents a significant barrier to rigorous and reproducible science. Rather than selecting measures on the basis of scientific properties (e.g., content validity, sensitivity to change, appropriate timescale), researchers often choose

measures based on convenience, historical use, or face validity. Lack of attention to the correspondence between measures and mechanisms is a barrier to advancing behavioural science.

Creating a systematic set of links between the SOBC repository's measures and the HBCP's ontology of hypothesized mechanisms of behaviour change helps to address this gap between measures and mechanisms. However, the efforts of SOBC and HBCP researchers do not correspond perfectly to each other. The SOBC Measures Repository was initially designed to address three, broad mechanisms – stress, self-regulation, and interpersonal processes – with a goal of determining whether a given measure could capture change in hypothesized mechanisms through a rigorous behavioural validation process (Science Of Behavior Change (SOBC)). The measures uploaded to the repository did not have to correspond perfectly to any one category. They might fit well under one mechanistic domain, two, or none of these three domains. The true correspondence of these measures to any specific mechanism(s) is thus untested. The HBCP aims to develop a comprehensive Behaviour Change Intervention Ontology (BCIO), including ontologies of BCTs and MoAs. The BCIO intends to encompass the current state of knowledge on behaviour change, including frequently occurring concepts (i.e., MoAs) across theories of behaviour and behaviour change. Thus, it follows the measures should conceptually map to at least one of the HBCP mechanistic domains (Human Behaviour Change Project (HBCP)).

To build upon the work already completed within each project, meant the collaboration neither started from a bottom-up (e.g., creating measures from scratch to correspond to each HBCP mechanisms) nor top-down (e.g., creating a novel ontology of mechanisms assessed in the SOBC repository) approach. To bring the project data together, this study used a modified expert consensus methodology to compile preliminary evidence for potential linkages between 44 self-

report measures of mechanisms from the SOBC Measures Repository and 26 MoAs identified by the HBCP.

We aimed to identify (i) MoAs that best represent measures in the SOBC Measures Repository and (ii) potential measures of the MoAs that are targeted by BCTs in the TaTT.

Methods

Participants

Participants were interdisciplinary experts recruited via listservs of professional and scientific networks and societies that focus on behaviour change intervention development, including Society of Behavioral Medicine (SBM), European Health Psychology Society (EHPS), the SOBC listserv, and the HBCP listserv. Interested participants were directed to an online, Qualtrics-based informed consent procedure, followed by a brief screening survey that asked about professional discipline, country, and self-rated expertise in theories of behaviour change.

A total of 80 individuals consented to be screened and completed the screening questionnaire. We selected 30 experts—15 based in the United States, 15 non-US-based—using criteria that maximized self-rated expertise in theories of behaviour change, and ensured a broad range of represented disciplines, and geographical regions. Selections based on expertise were made according to two self-ratings: (i) a response of *Extensively* for at least one of: use of behaviour change theory, publications on the topic, reviews on the topic, or design of a behaviour change intervention, where the scale was *Extensively* to *Not at all*; (ii) a response of 5 or over on knowledge of behaviour change theories, mechanisms, or intervention evaluation/design, where the scale was 1 [*No knowledge/expertise*] to 7 [*Profound*

knowledge/expertise]. Experts who met both of these criteria and elected to participate in the study could earn up to $200 ($100 for Round 1, $50 for Round 2, and $50 for Round 3).

Procedure

All procedures were virtual, and experts remained anonymous to each other. This study followed similar procedures to those conducted by HBCP researchers in which experts identified links between BCTs and MoAs (Bohlen et al.). The procedures in the HBCP studies were based on a modified Nominal Group Technique (NGT) (Van de Ven & Delbecq, 1972) spread over three rounds and two weeks to complete each round. Participating experts received an introductory email and two emails prior to each round. One email was a reminder, and the other notified experts the round had begun. Three further modifications to the NGT were made. First, a consensus benchmark of 50% was selected, which is lower than the traditional cutpoint of 70 to 80%. This was determined *a priori* as this is an exploratory developmental study intended to capture a large number of potential measure-MoA links for further study (Nair, Aggarwal, & Khanna, 2011). It also allows flexibility due to the lack of perfect correspondence between the labels given to mechanisms targeted by SOBC and the comprehensive MoAs identified by the HBCP. Second, to reduce participant burden, not all links were re-rated in Round 3 (i.e., those that reached 50% consensus in Round 1 were not presented again in Round 3). Third, summary statistics were not provided for participants after Round 1.

Study procedures were approved by the [BLINDED] Institutional Review Board, and participants provided electronic informed consent prior to completing study procedures.

Preliminary Work

Two members of the SOBC Resource and Coordinating Center ([BLINDED] and [BLINDED]) independently coded links between 44 self-report measures from the SOBC

Measures Repository and the 26 MoAs used in the HBCP studies. Each rater considered all possible links between each measure and MoA, then selected those they considered to be "related." Some examples of measure-MoA links easily excluded at this stage are *10-Item Personality Inventory-Knowledge* and *Zimbardo Time Perspective Inventory-Social/Professional Role & Identity*. An example that was more "on the fence" is *Domain Specific Risk Taking Survey - Expected Benefits-Behaviour Regulation*. Although the scale does concern behavioural choices, coders reasoned that expecting a beneficial outcome for a given risky behaviour did not correspond to the MoA definition of "Behavioural, cognitive and/or emotional skills for managing or changing behaviour" (Human Behaviour Change Project (HBCP)).

[BLINDED] and [BLINDED] met to discuss preliminary codes. Of the 44 self-report measures, seven were flagged for discussion with [BLINDED] to provide additional expertise on the MoA definitions. After this discussion, codes were reconciled. To reduce participant burden, we pre-identified links rather than asking experts to rate all 1,144 possible measure-MoA links. Because decisions to exclude certain measure-MoA links may have differed with a larger group of coders during this preliminary round, experts were provided the opportunity to suggest any missed measure-MoA links that were not pre-identified and presented in Round 1. This ensured no widely suggested measures were missed, because all links suggested by 50% or more of experts were rated by all participants in Round 3.

Round 1

In Round 1, experts completed an online survey via Qualtrics, rating each of the 84 pre-identified measure-MoA links based on whether each measure was related to the coded MoA(s). MoAs were defined as "the processes through which a behaviour change technique (BCT) affects behaviour" and definitions were provided for each (see **Supplemental Appendix A**).

Experts were informed that each measure may be linked to more than one MoA, some MoAs may be linked to no measures at all, and that these links were intended to be broad and useful. An email prior to the launch of Round 1 was sent to provide experts with MoA definitions, a link to the Measures Repository, general instructions (i.e., they would be asked to rate agreement with measure-MoA links and have the opportunity to suggest new links between measures and MoAs that were not pre-identified by the research team), and the study timeline (see **Supplemental Appendix B**).

Each measure-MoA link was presented as an individual question to rate, and experts were provided with the measure name, measure description, and a link to the Measures Repository to access all measure items (see **Supplemental Appendix C**). Each question was phrased: "Is this measure related to [MoA] as a MoA?" Response options ranged from 1 (*Definitely no*) to 5 (*Definitely yes*). The consensus threshold was set low to be highly inclusive of potential measure-MoA links because this study constitutes a preliminary step toward the larger goal of establishing measure-MoA links in future research and because measures were not explicitly designed to assess specific MoAs. Specifically, the threshold was set such that at least 50% of experts must agree the measure and MoA were related (rating of 4 [*Probably yes*] or 5 [*Definitely yes*]) to be included for further consideration as a potential link. Using the term "related" was selected for similar reasons (i.e., to allow for a broad capture of potential measure-MoA links). Experts were given the opportunity to suggest any additional links between measures and MoAs that had not been pre-identified.

Round 2

Results from Round 1 were tabulated and used to inform Round 2. In Round 2, experts were invited to participate in an anonymous, asynchronous discussion occurring over a two-week

period via an online platform ('Courseworks') to share feedback regarding their experiences in Round 1, and to discuss a selection of 10 measures that did not reach consensus in Round 1. This number was selected to reduce participant burden and allow deeper conversation about a smaller number of measures rather than only a few comments on a much larger number. Measures eligible for discussion were those pre-identified links with less than 50% agreement (i.e., a response of 4 [*Probably yes*] or 5 [*Definitely yes*]) as well as new links suggested by 20-50% of respondents. Selections were determined to ensure representation of a range of different measures and MoAs.

Round 3

In Round 3, experts rated the links discussed in Round 2, plus all other measure-MoA links that met criteria for discussion in Round 1 (see **Supplemental Appendix C**). As in Round 1, experts were asked to rate agreement for each measure-MoA link. Experts were provided with descriptions of each MoA and the measure name, description, and a link to the measure on the Measures Repository. Each question was phrased thus: "Is this measure related to [MoA] as a MoA?" Response options ranged from 1 (*Definitely no*) to 5 (*Definitely yes*).

Data Analysis Strategy

In Round 1, experts who did not answer any questions ($n = 2$), or who began but did not complete the survey ($n = 1$) were excluded from further analysis. Consensus with measure-MoA links was calculated as percent agreement that the measure and MoA were related (rating of 4 [*Probably yes*] or 5 [*Definitely yes*]) using pairwise deletion (i.e., the denominator for 50% consensus was always computed out of the total number of experts responding to a given question). This technique allowed for accidental missingness on a given question. Newly suggested measure-MoA links were considered to meet consensus if more than 50% of experts

suggested a given link; the denominator for this calculation was 27 experts. If links were suggested by at least 20% of experts but were below the 50% benchmark for agreement, then these were considered eligible for discussion and for rating in Round 3. Round 3 employed identical criteria to Round 1. In Round 3, all 18 remaining experts began and completed the survey.

Heat Maps. All results which reached at least 50% consensus are represented in "heat maps." Heat maps present data values as color gradients within a matrix to provide an additional means of interpreting the results. The cells within the heat map reflect the percentage of agreement among experts and shaded to reflect the relative strength or "heat" of that value (in this case, the extent of agreement for that particular link). All heat maps were produced using ggplot2 (Wickham, Chang, & Wickham, 2016) in R version 4.1.2. The heat map groups the rows (i.e., measures) and columns (i.e., MoAs) by similarity, such that measures linked to similar MoAs are closer together, and MoAs linked to similar numbers of measures are closer together.

Data Availability Statement. Data tables with all percent consensus ratings are available in the **Supplemental Appendix D**, and raw data are available upon reasonable request from the first author ([BLINDED]).

Results

Most experts were from the United States (15; 50%), followed by Canada (6; 20%), and Ireland and the United Kingdom (5; 16.7%). Others (4; 13.3%) lived in Australia, Brazil, Israel, or the Netherlands. Most (25; 83.3%) worked in a university setting, followed by the public sector, charity/voluntary sector, or academic medical center. Most (13; 43.3%) described their discipline as psychology, 7 as public health (23.3%), and the remaining as other (e.g., nursing,

kinesiology; 10; 30.0%). All experts had extensive experience in behaviour change interventions, theories, and mechanisms of behaviour change (**Figure 1**).

Of the 30 experts, 27 completed Round 1, 23 completed Round 2, and 18 completed Round 3. Dropout was monotonic (i.e., no experts who missed Round 2 returned for Round 3). Only one participant formally withdrew due to external circumstances. There were no differences between the 18 experts completing all 3 rounds and those who dropped out in terms of location (US v. non-US) or self-rated expertise. Sensitivity analyses examining agreement with pre-identified measure-MoA links data from only the 18 experts completing all three rounds did not substantively impact consensus ratings. In Round 1, 13 items had one skipped response, and two items had two skipped responses. In Round 3, only one item had a single skipped response. The median time spent on the survey for experts who completed Round 1 was 6.30 hours (*IQR* 89.97, *Range* 0.54, 323.48), and for Round 3, the median was 30.95 hours (*IQR* 112.47, *Range* 0.16, 225.41). Time spent completing each survey varied widely because experts were able to complete the survey in multiple rounds, meaning they could start and stop the survey as many times as necessary over a two-week period. Thus, hours spent reflects the total amount of time elapsed from when an expert first started the survey, and the time it was completed.

Round 1

Of 84 pre-identified measure-MoA links presented to the expert panel, 82 reached at least 50% agreement, and only two did not. Experts suggested 109 new links, 14 met the 50% threshold level of consensus (i.e., were suggested by 50% or more participants) and 95 of which had 20-50% agreement. All links are displayed in **Supplemental Appendix D, supplemental Table 1** and are separated by type (i.e., pre-identified v. suggested).

Round 2[1]

There were 16 posts in the general discussion thread, and the 10 measure-MoA-specific posts had a mean of 17 replies (range: 12-19). There was debate over the MoAs themselves, including whether MoAs must be modifiable, by definition, along with some acknowledgement that some theoretically stable MoAs may be modifiable after all (e.g., Optimism). There was also discussion of whether a MoA could be a moderator or a tailoring variable rather than exclusively a mediator. Experts raised important points about the boundaries of these measures (e.g., invariance across cultures).

In terms of the task, there were some discrepancies in how experts interpreted "related to" when assessing measure-MoA links. As one expert stated, *"I also found that it was not clear what was meant by 'related'. Some experts interpreted the task as judging to which degree the measure assesses the MOA, others did not."* This theme recurred in the measure-MoA-specific boards, and some experts cautioned against overstating the measure-MoA links, drawing attention to the need for further examination of validity. Specifically, as one expert pointed out, "*Although a "link" between a measure and MOA exists (the links we are drawing, in this consensus exercise), this does not mean that that measure assesses the entire construct of interest and ONLY that construct of interest, for the intended population, context, behavior of interest, etc. Therefore, regardless of a conceptual "link", the measures are not valid for use as measures of the MOAs*."

[1] Qualitative data will be comprehensively coded and analyzed, and results will be published in a future manuscript.

The primary concern was over the quality of the measures (construct validity, convergent/divergent validity, double-barrel items, etc.). One expert suggested that linking measures-MoAs would necessitate developing measures from scratch for the sole purpose of measuring a specific MoA, and another pointed out that these flaws in the measures themselves would hinder the ability to detect meaningful associations (e.g., due to measurement error). Experts additionally acknowledged lack of expertise in a specific theoretical domain for a given measure, recommended a focus on the wording of measure items (rather than measure titles alone), and drew attention to measures-outcome correspondence (e.g., to measure a MoA such as Intention, Skills, or Knowledge, a measure must be content-specific; the target in dyadic scales may focus on only one individual or on both). Experts also noted some of the MoA definitions were quite broad (e.g., Self-Image), such that almost any measure might be "related."

Round 3

The two pre-identified links that did not reach agreement and the 94 new links with 20-50% agreement were rated in Round 3 (note: one newly suggested link was left out of the survey in error). Of the two pre-identified links, one reached 72.2% consensus after Round 3. Of the 94 newly suggested links, 70 reached at least 50% consensus. All ratings are displayed in **Supplemental Appendix D, Supplemental Table 2**.

Combining results from Rounds 1 and 3, a total of 167 measure-MoA links reached consensus: 82 pre-identified links from Round 1, 14 newly suggested links from Round 1, and 71 from Round 3. These links included all 44 self-report measures from the SOBC Measures Repository and 20 of the 26 HBCP MoAs. The MoAs Knowledge, Norms, Subjective Norms, Needs, Feedback Processes, Social Learning/Imitation, and Perceived

Susceptibility/Vulnerability were not linked to any measures. All links meeting consensus are displayed in **Table 1**.

Out of the 167 links meeting the 50% threshold, 33 had consensus between 90-100%. Mean agreement was 74.21% (*SD* = 15.62%; *Range*: 50%-100%). Across the 44 measures, a single measure was linked to a mean of 3.80 MoAs (*SD* = 1.36; *Median* = 4.0; *Range*: 1-6), and the mode was 5 MoA linkages for a single measure (15; 34.1%). Across the 20 MoAs, a single MoA was linked to a mean of 8.35 measures (*SD* = 9.18; *Median* = 3.5; *Range*: 1-33). Of all MoAs, Behavioural Regulation had the greatest number of linked measures (*n* = 33). **Figure 2** depicts all final links between measures and MoAs, and **Figure 3** shows links with 90% or more agreement.

Discussion

This study aimed to link measures from the Science of Behavior Change (SOBC) Measures Repository to hypothesized mechanisms of action (MoAs) identified by the Human Behaviour-Change Project (HBCP). A consensus exercise involving three rounds generated 167 measure-MoA links that reached at least 50% agreement (i.e., experts agreed these measures were “related”), 33 of which had 90% or greater agreement. This is a first step in obtaining data to increase interoperability of scientific resources designed to facilitate process research. By linking measures to a comprehensive ontology of MoAs, this study also enables consistency of measurement of MoAs, so that research on how interventions exert their effects on behavioural outcomes can accumulate systematically. Compiling evidence for or against these measure-MoA links can inform researchers in selecting optimal measures from the SOBC repository when designing studies that utilize the experimental medicine approach. Further, for example, just because a measure is purportedly a stress measure, does not mean it truly measures stress, or

purely measures stress, and it may be associated with other MoAs. This study allows us to tease apart this conceptual issue for measures currently housed within the SOBC measures repository.

Research building on this preliminary study should further investigate these putative measure-MoA links and test the utility of these links for use in behavioural research. For example, is a measure valid and reliable for capturing change in the hypothesized MoA? Is one measure more readily influenced than another measure of that same MoA? If so, does this depend on choice of BCT? On what timescale can these measures capture change? Evidence to this effect may benefit research by highlighting the differences in performance of measures of MoAs across studies. This line of research could shed valuable light on a variety of measurement issues, including measurement quality, timescale of measurement (e.g., short-term v. long-term change), jingle-jangle fallacies (Marsh et al., 2019), and nuances in the operationalization of MoAs. Evidence would also accumulate in support of BCT-MoA linkages. As an example, one may find that all of the SOBC measures deemed to be of Beliefs About Capabilities are modified by a Goal-Setting intervention. It is also important to acknowledge that poor measurement practices remain a serious issue in the broader field of psychological science, and there have been other calls to action for more attention to the validity of measures. A recent manuscript (Flake & Fried, 2020) highlights a number of these issues, including differing results across studies that utilize different measures of the same ostensible construct. Flake and Fried (2020) additionally provide some helpful guidance on considerations for researchers when selecting measures to optimize the validity of their study.

This study had several strengths. First, the broad expertise of the participants allowed for informed ratings, fruitful consensus discussions, and valuable study feedback. Second, the multiple rounds allowed for correction and refinement after each round. Third, selecting a low

threshold for consensus (50%) and using a general term for potential linkage (“related”) is appropriate for exploratory developmental work in this new area, allowing us to capture a large number of potential measure-MoA links and reducing the possibility that important links were missed.

Several challenges arose, and expert feedback highlighted important areas for improvement in the field. The lack of a one-to-one correspondence between an operational measure and a conceptual mechanism can be attributed to multiple factors, including poor practices in measurement, lack of shared definitions of constructs between different research groups or fields, or different levels of specificity in a measure or mechanism, such as measuring the construct of self-regulation as the tendency to engage in self-regulation generally v. the tendency to engage in self-regulatory abilities around a specific behaviour. When designing behavioural research studies, rather than starting investigations with a measure, one often begins with the conceptual structure (mechanisms). Measures that are conceptually appropriate for a specific study are critical to test these mechanistic hypotheses, but appropriate measures may not exist. Barriers to the creation of new measures, or comprehensive validation of previously created measures, include continuing with past practice, poor uptake of new measures, and lack of resources (e.g., time, funding), among others.

In the present study, the 44 self-report measures from the SOBC repository were not designed to measure the coded MoAs. Rather, they were originally created to assess specific conceptual constructs that each can be categorized as pertaining to one or more of the broad domains of stress, self-regulation, and interpersonal processes. Similarly, the MoA definitions were not developed in reference to the repository measures. Despite this inherent challenge, we were able to bridge two research approaches by forming initial links between the useful

repository resource and the valuable MoA definitions. The experts reached consensus on many measure-MoA links, some of which reached 100% agreement. Future work should consider how to bring entity definitions together with measures of constructs across all the entities in the Behaviour Change Intervention Ontology (Michie et al., 2020), especially behaviour itself, and should incorporate non-self-report measures, such as tasks or observational coding. This would help in the design of future research as well as the curation of extant knowledge. The SOBC repository is not a comprehensive collection of measures, whereas the HBCP MoAs were developed to represent the most commonly occurring constructs across theories of behaviour and behaviour change. Because the goal was to link the SOBC repository and the HBCP TaTT tool, this study was limited to the available measures; however, the SOBC repository is evolving and growing, and there are plans to incorporate methods to allow for linkages between newly uploaded measures and the HBCP MoAs, as well as methods for the refinement of existing links.

The measure-MoA links were intended to be useful for researchers who want to identify measures of MoAs that can be targeted for change in behavioural interventions. The heat maps and associated consensus data can be viewed as a summary of behavioural health researchers' beliefs about measure–MoA links and can be used as a starting point for intervention designers and evaluators. These data have been integrated into the user interface for both the SOBC Measures Repository and the Theory and Techniques Tool (TaTT) (Human Behaviour Change Project (HBCP)), allowing users to access these data and navigate between the TATT and the Measures Repository with ease. However, knowing that a measure and MoA are "related" should not be understood as a recommendation to use a specific measure for a specific MoA. Experts may have operationalized "related" in different ways when considering measure-MoA links. Much like the TaTT (Human Behaviour Change Project (HBCP)), the current set of links

between MoAs and measures represents only hypothesized associations that should be evaluated further before use. Furthermore, experts may have focused on MoA labels rather than definitions (or measure titles rather than item content). Future research should investigate and assess the utility of these links, the strength of these links, and whether better alternatives exist for measurement, particularly given the varied quality of the self-report measures in terms of validity and reliability. This could be accomplished by seeking feedback from a wide range of researchers interested in the development of behavioural interventions, for example, users of the SOBC Measures Repository or the HBCP TaTT. In fact, the TaTT already includes a feature such that users can comment and discuss the different BCT-MoA links (Human Behaviour Change Project (HBCP)), and a similar feature may be useful for the measure-MoA links identified in the present study. Additional studies, using different study designs to assess this information, as well as triangulation techniques, may also be a fruitful next step (e.g., gamification techniques that allow "non-expert" repository users to assess the strength of various measure-MoA links, including links and non-links identified by the experts in the present study). Multiple replicating methodologies and triangulation efforts were used to generate the results of BCT-MoA links in the TaTT project (Bohlen et al.; Carey et al., 2019; Johnston et al., 2020). A similar approach, and/or data addressing the validity and reliability of new measures uploaded to the SOBC Measures Repository could also be considered.

In the long-term, it will be important to identify measure-MoA linkages not just for self-report measures but also behavioural measures (e.g., Stroop task). Research should address not only whether or not a measure taps into a putative mechanism but also whether it taps that *specific* mechanism and *not* other similar mechanisms (i.e., content validity and discriminant content validity (Burrell, Allan, Williams, & Johnston, 2018; Johnston et al., 2014)). Moreover, a

critical next step will be to link measures of mechanisms to BCTs that are associated with particular interventions (i.e., the early-stage steps of the causal chain) and particular behavioural outcomes (i.e., the late-stage steps of the causal chain). This latter research direction is currently being addressed in the HBCP (Michie et al., 2020).

Measures may not generalize to certain populations or settings, and data on measurement invariance for many of these measures are lacking (Milfont & Fischer, 2010). There may be limits for certain measures when considering behavioural domains (e.g., the measure may be too specific, or not specific enough) or target individuals (e.g., dyadic scales may concern behavioural regulation in one person, the other, or both). Some measures may not demonstrate strong construct validity, convergent validity, or divergent validity (e.g., a measure may be equally related to two different MoAs, but a researcher may desire a pure measure of only one of these MoAs). Indeed, construct validity was a major concern raised by experts in Round 2.

Additional features of the study design are important to consider when evaluating the findings. Participants may have been biased to agree with the pre-identified measure-MoA links. Because no questions with non-links were included as attention checks, this possibility cannot be tested. That said, reducing participant burden from rating over 1,000 links to only 84 outweighs the limitations of this approach. The consensus ratings of "related" in Rounds 1 and 3 are not directly comparable to those links suggested by more than 50% of experts in Round 1. Not all links rated in Round 1 were re-rated in Round 3. Summaries of Round 1 results were also not provided prior to Rounds 2 and 3, which could have reduced the ability of experts to reach consensus. Another question raised in this study included whether the MoAs were believed to be modifiable or stable personality traits (e.g., Optimism) and whether some MoAs might serve as moderators rather than mediators (i.e., mechanisms) of intervention effects on behavioural

outcomes. However, recent evidence indicates that personality may prove to be more malleable to behavioural intervention than previously considered (Allemand & Flückiger, 2022; Roberts et al., 2017). As such, there may be value in examining personality as both a mediator and moderator of behaviour change. Finally, there was substantial participant attrition of 33.3% from Round 1 to Round 3, which may have increased bias in consensus ratings. Although this reduced the sample size from 30 to 18, a panel size of 5-11 is considered more than sufficient for a consensus study (Sinha, Smyth, & Williamson, 2011; Waggoner, Carline, & Durning, 2016). Furthermore, sensitivity analyses comprising only the 18 participants who completed Round 3 did not alter the Round 1 study conclusions.

Conclusion

This study contributes to behaviour change research through a preliminary integration of putative measures of mechanisms from the SOBC Measures Repository with the MoAs used within the HBCP. A total of 167 potential measure-MoA links were identified for refinement in future studies, with agreed upon links based on expert consensus rather than assumptions about the measure, or traditions for using the measure within the field. These results can be used to facilitate the accumulation of scientific evidence for process pathways and can also be used to increase the interoperability of scientific resources in the larger mission of understanding how interventions can effectively change behaviours.

Acknowledgements

Declaration of Interest

The authors declare that they have no conflicts of interest.

References

Allemand, M., & Flückiger, C. (2022). Personality change through digital-coaching interventions. *Current Directions in Psychological Science*, 09637214211067782.

Bohlen, L. C., Carey, R., de Bruin, M., Rothman, A., Johnston, M., Kelly, M. P., & Michie, S. Links between behaviour change techniques and mechanisms of action: an expert consensus study. *Annals of Behavioral Medicine, 53*(8), 708-720.

Burrell, A. M., Allan, J. L., Williams, D. M., & Johnston, M. (2018). What do self-efficacy items measure? Examining the discriminant content validity of self-efficacy items. *British Journal of Health Psychology, 23*(3), 597-611.

Carey, R. N., Connell, L. E., Johnston, M., Rothman, A. J., De Bruin, M., Kelly, M. P., & Michie, S. (2019). Behavior change techniques and their mechanisms of action: a synthesis of links described in published intervention literature. *Annals of Behavioral Medicine, 53*(8), 693-707.

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science, 3*(4), 456-465.

Human Behaviour Change Project (HBCP). Theory and Techniques Tool. Retrieved from https://theoryandtechniquetool.humanbehaviourchange.org/

Johnston, M., Carey, R. N., Connell Bohlen, L. E., Johnston, D. W., Rothman, A. J., de Bruin, M., . . . Michie, S. (2020). Development of an online tool for linking behavior change techniques and mechanisms of action based on triangulation of findings from literature synthesis and expert consensus. *Translational Behavioral Medicine*.

Johnston, M., Carey, R. N., Connell Bohlen, L. E., Johnston, D. W., Rothman, A. J., de Bruin, M., . . . Michie, S. (2021). Development of an online tool for linking behavior change techniques and mechanisms of action based on triangulation of findings from literature synthesis and expert consensus. *Translational Behavioral Medicine, 11*(5), 1049-1065.

Johnston, M., Dixon, D., Hart, J., Glidewell, L., Schröder, C., & Pollard, B. (2014). Discriminant content validity: A quantitative methodology for assessing content of theory-based measures, with illustrative applications. *British Journal of Health Psychology, 19*(2), 240-257.

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of educational psychology, 111*(2), 331.

Michie, S., West, R., Finnerty, A. N., Norris, E., Wright, A. J., Marques, M. M., . . . Hastings, J. (2020). Representation of behaviour change interventions and their evaluation: Development of the Upper Level of the Behaviour Change Intervention Ontology. *Wellcome Open Research, 5*.

Michie, S., Wood, C. E., Johnston, M., Abraham, C., Francis, J., & Hardeman, W. (2015). Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). *Health technology assessment, 19*(99).

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of psychological research, 3*(1), 111-130.

Nair, R., Aggarwal, R., & Khanna, D. (2011). *Methods of formal consensus in classification/diagnostic criteria and guideline development.* Paper presented at the Seminars in arthritis and rheumatism.

Riddle, M., & Science of Behavior Change (SOBC) Working Group. (2015). News from the NIH: using an experimental medicine approach to facilitate translational research. In: Oxford University Press.

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychol Bull, 143*(2), 117.

Science Of Behavior Change (SOBC). The Measures. Retrieved from https://measures.scienceofbehaviorchange.org/

Sheeran, P., Klein, W., & Rothman, A. J. (2017). Health behavior change: Moving from observation to intervention. *Annual review of psychology, 68*(1), 573-600.

Sinha, I. P., Smyth, R. L., & Williamson, P. R. (2011). Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS Med, 8*(1), e1000393.

Sumner, J. A., Birk, J. L., Cornelius, T., Derby, L., Edmondson, D., & Davidson, K. W. (2019). *The NIH Science of Behavior Change Mechanism-Focused Approach to Behavior Change Research.* Paper presented at the PSYCHOSOMATIC MEDICINE.

Sumner, J. A., Carey, R. N., Michie, S., Johnston, M., Edmondson, D., & Davidson, K. W. (2018). Using rigorous methods to advance behaviour change science. *Nature human behaviour, 2*(11), 797.

Van de Ven, A. H., & Delbecq, A. L. (1972). The nominal group as a research instrument for exploratory health studies. *American journal of public health, 62*(3), 337-342.

Waggoner, J., Carline, J. D., & Durning, S. J. (2016). Is there a consensus on consensus methodology? Descriptions and recommendations for future consensus research. *Academic Medicine, 91*(5), 663-668.

Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2'. *Create elegant data visualisations using the grammar of graphics. Version, 2*(1), 1-189.

Table 1. Ratings for the 167 measure-MoA links meeting the consensus cutpoint.

Mechanism	**Measure**	**Agreement**
Attitudes towards Behavior	Domain Specific Risk Taking Survey - Expected Benefits	55.6
Attitudes towards Behavior	Domain Specific Risk Taking Survey - Risk Perceptions	64.7
Attitudes towards Behavior	Domain Specific Risk Taking Survey - Risk Taking	50.0
Behavioral Cueing	Three Factor Eating Questionnaire R-18	72.2
Behavioral Regulation	Barratt Impulsiveness	96.3
Behavioral Regulation	Behavior Rating Inventory of Executive Function (BRIEF) - Adults	96.3
Behavioral Regulation	BIS/BAS	77.8
Behavioral Regulation	Brief COPE	100.0
Behavioral Regulation	Brief Risk-Resilience	81.5
Behavioral Regulation	Brief Self-Control	96.3
Behavioral Regulation	Cognitive Reflection Test	55.6
Behavioral Regulation	Consideration of Future Consequences Scale	85.2
Behavioral Regulation	Couple Coercion Scale	61.1
Behavioral Regulation	Deferment of Gratification	92.6
Behavioral Regulation	Dickman Functional and Dysfunctional Impulsivity Survey	77.8
Behavioral Regulation	Domain Specific Risk Taking Survey - Expected Benefits	55.6
Behavioral Regulation	Emotion Regulation Questionnaire	88.9
Behavioral Regulation	Emotion Regulation Strategies Scale	92.6
Behavioral Regulation	Five Facts of Mindfulness Survey	88.9
Behavioral Regulation	Future Orientation Scale of Time Perspective Study	100.0

Behavioral Regulation	Generalized Self-Efficacy Scale	72.2
Behavioral Regulation	Grit-S	73.1
Behavioral Regulation	I-7: Impulsiveness and Venturesomeness Questionnaire	81.5
Behavioral Regulation	Mindful Attention Awareness Scale	92.3
Behavioral Regulation	Multidimensional Assessment of Interoceptive Awareness	76.9
Behavioral Regulation	Multidimensional Personality Questionnaire: Control vs. Impulsivity Scale	96.3
Behavioral Regulation	NIH Self-Efficacy Scale	81.5
Behavioral Regulation	Parent-rated Stress (NIH Perceived Stress Scale)	55.6
Behavioral Regulation	Pearlin Mastery Scale	50.0
Behavioral Regulation	Selection-Optimization-Compensation Questionnaire	92.0
Behavioral Regulation	Short Self-Regulation Questionnaire	100.0
Behavioral Regulation	SIDES Affect Dysregulation Scale (Child-Reported)	88.9
Behavioral Regulation	Theories of Willpower Scale	88.5
Behavioral Regulation	Three Factor Eating Questionnaire R-18	96.3
Behavioral Regulation	UPPS-P Impulsivity Scale Survey	96.3
Behavioral Regulation	Zimbardo Time Perspective Inventory	69.2
Behavioral Regulation	Zuckerman Sensation Seeking Survey-V	63.0
Beliefs about Capabilities	Barratt Impulsiveness	50.0
Beliefs about Capabilities	Behavior Rating Inventory of Executive Function (BRIEF) - Adults	51.9
Beliefs about Capabilities	Brief COPE	72.2
Beliefs about Capabilities	Brief Risk-Resilience	72.2
Beliefs about Capabilities	Brief Self-Control	100.0

Beliefs about Capabilities	Dickman Functional and Dysfunctional Impulsivity Survey	50.0
Beliefs about Capabilities	Emotion Regulation Questionnaire	50.0
Beliefs about Capabilities	Generalized Self-Efficacy Scale	100.0
Beliefs about Capabilities	Grit-S	72.2
Beliefs about Capabilities	NIH Self-Efficacy Scale	74.1
Beliefs about Capabilities	Pearlin Mastery Scale	96.3
Beliefs about Capabilities	Short Self-Regulation Questionnaire	51.9
Beliefs about Capabilities	SIDES Affect Dysregulation Scale (Child-Reported)	70.4
Beliefs about Capabilities	Theories of Willpower Scale	55.6
Beliefs about Capabilities	Three Factor Eating Questionnaire R-18	72.2
Beliefs about Capabilities	UPPS-P Impulsivity Scale Survey	55.6
Beliefs about Consequences	Consideration of Future Consequences Scale	92.6
Beliefs about Consequences	Domain Specific Risk Taking Survey - Expected Benefits	92.3
Beliefs about Consequences	Domain Specific Risk Taking Survey - Risk Perceptions	100.0
Beliefs about Consequences	Domain Specific Risk Taking Survey - Risk Taking	81.5
Beliefs about Consequences	Selection-Optimization-Compensation Questionnaire	73.1
Beliefs about Consequences	Theories of Willpower Scale	76.9
Beliefs about Consequences	Three Factor Eating Questionnaire R-18	63.0
Beliefs about Consequences	Zimbardo Time Perspective Inventory	76.9
Emotion	10-Item Personality	55.6
Emotion	Barratt Impulsiveness	50.0
Emotion	Brief COPE	88.9

Emotion	Brief Risk-Resilience	95.8
Emotion	Couple Coercion Scale	66.7
Emotion	Daily Inventory of Stressful Events (DISE)	63.0
Emotion	Ecological Momentary Assessment of Stressful Events	81.5
Emotion	Emotion Regulation Questionnaire	92.6
Emotion	Emotion Regulation Strategies Scale	96.3
Emotion	Five Facts of Mindfulness Survey	81.5
Emotion	I-7: Impulsiveness and Venturesomeness Questionnaire	59.3
Emotion	Kessler Psychological Distress Scale (K6+)	96.3
Emotion	Multidimensional Assessment of Interoceptive Awareness	81.5
Emotion	Parent-rated Stress (NIH Perceived Stress Scale)	96.3
Emotion	Positive and Negative Affect Scheduled (PANAS)	92.6
Emotion	Positive and Negative Affect Scheduled (PANAS)- Child	100.0
Emotion	SIDES Affect Dysregulation Scale (Child-Reported)	96.3
Emotion	Three Factor Eating Questionnaire R-18	61.1
Environmental Context & Resources	Daily Inventory of Stressful Events (DISE)	88.9
Environmental Context & Resources	Ecological Momentary Assessment of Stressful Events	88.9
General Attitudes/Beliefs	Future Time Perspective Scale	70.4
General Attitudes/Beliefs	Theories of Willpower Scale	73.1
Goals	BIS/BAS	88.9
Goals	Consideration of Future Consequences Scale	72.2
Goals	Future Orientation Scale of Time Perspective Study	55.6

Goals	Grit-S	76.9
Goals	Selection-Optimization-Compensation Questionnaire	51.9
Goals	Short Self-Regulation Questionnaire	88.9
Intentions	Domain Specific Risk Taking Survey - Risk Taking	55.6
Memory, Attention, & Decision Processes	Barratt Impulsiveness	96.3
Memory, Attention, & Decision Processes	Behavior Rating Inventory of Executive Function (BRIEF) - Adults	100.0
Memory, Attention, & Decision Processes	Cognitive Reflection Test	96.3
Memory, Attention, & Decision Processes	Consideration of Future Consequences Scale	74.1
Memory, Attention, & Decision Processes	Deferment of Gratification	70.4
Memory, Attention, & Decision Processes	Dickman Functional and Dysfunctional Impulsivity Survey	70.4
Memory, Attention, & Decision Processes	Domain Specific Risk Taking Survey - Expected Benefits	55.6
Memory, Attention, & Decision Processes	Domain Specific Risk Taking Survey - Risk Perceptions	50.0
Memory, Attention, & Decision Processes	Emotion Regulation Questionnaire	61.1
Memory, Attention, & Decision Processes	Five Facts of Mindfulness Survey	77.8
Memory, Attention, & Decision Processes	Future Orientation Scale of Time Perspective Study	77.8
Memory, Attention, & Decision Processes	I-7: Impulsiveness and Venturesomeness Questionnaire	61.1
Memory, Attention, & Decision Processes	Mindful Attention Awareness Scale	96.3
Memory, Attention, & Decision Processes	Multidimensional Assessment of Interoceptive Awareness	85.2
Memory, Attention, & Decision Processes	Multidimensional Personality Questionnaire: Control vs. Impulsivity Scale	70.4
Memory, Attention, & Decision Processes	Selection-Optimization-Compensation Questionnaire	84.6
Memory, Attention, & Decision Processes	UPPS-P Impulsivity Scale Survey	74.1
Motivation	BIS/BAS	63.0

Motivation	Consideration of Future Consequences Scale	61.1
Motivation	Deferment of Gratification	50.0
Motivation	Grit-S	55.6
Optimism	Future Time Perspective Scale	63.0
Optimism	Generalized Self-Efficacy Scale	66.7
Optimism	NIH Self-Efficacy Scale	66.7
Optimism	Pearlin Mastery Scale	50.0
Optimism	Zimbardo Time Perspective Inventory	50.0
Perceived Susceptibility /Vulnerability	Daily Inventory of Stressful Events (DISE)	50.0
Perceived Susceptibility /Vulnerability	Domain Specific Risk Taking Survey - Risk Taking	66.7
Perceived Susceptibility /Vulnerability	Domain Specific Risk Taking Survey - Risk Perceptions	55.6
Reinforcement	BIS/BAS	63.0
Reinforcement	Deferment of Gratification	70.4
Self-Image	10-Item Personality	88.9
Self-Image	Barratt Impulsiveness	88.9
Self-Image	Behavior Rating Inventory of Executive Function (BRIEF) - Adults	88.9
Self-Image	BIS/BAS	83.3
Self-Image	Brief COPE	72.2
Self-Image	Brief Risk-Resilience	85.2
Self-Image	Brief Self-Control	77.8
Self-Image	Dickman Functional and Dysfunctional Impulsivity Survey	77.8
Self-Image	Domain Specific Risk Taking Survey - Risk Taking	77.8

Self-Image	Five Facts of Mindfulness Survey	77.8
Self-Image	Generalized Self-Efficacy Scale	96.3
Self-Image	Grit-S	59.3
Self-Image	I-7: Impulsiveness and Venturesomeness Questionnaire	88.9
Self-Image	Kessler Psychological Distress Scale (K6+)	61.1
Self-Image	Multidimensional Assessment of Interoceptive Awareness	55.6
Self-Image	Multidimensional Personality Questionnaire: Control vs. Impulsivity Scale	70.4
Self-Image	NIH Self-Efficacy Scale	92.3
Self-Image	Pearlin Mastery Scale	88.9
Self-Image	Selection-Optimization-Compensation Questionnaire	56.0
Self-Image	Short Self-Regulation Questionnaire	88.9
Self-Image	SIDES Affect Dysregulation Scale (Child-Reported)	72.2
Self-Image	Three Factor Eating Questionnaire R-18	77.8
Self-Image	UPPS-P Impulsivity Scale Survey	55.6
Self-Image	Zimbardo Time Perspective Inventory	72.2
Self-Image	Zuckerman Sensation Seeking Survey-V	88.9
Skills	Behavior Rating Inventory of Executive Function (BRIEF) - Adults	66.7
Skills	Brief COPE	55.6
Skills	Brief Risk-Resilience	55.6
Skills	Brief Self-Control	55.6
Skills	Cognitive Reflection Test	61.1
Skills	Emotion Regulation Questionnaire	72.2

Skills	Emotion Regulation Strategies Scale	61.1
Skills	Five Facts of Mindfulness Survey	72.2
Skills	Future Orientation Scale of Time Perspective Study	55.6
Skills	Generalized Self-Efficacy Scale	66.7
Skills	Mindful Attention Awareness Scale	55.6
Skills	Multidimensional Assessment of Interoceptive Awareness	66.7
Skills	NIH Self-Efficacy Scale	72.2
Skills	Short Self-Regulation Questionnaire	61.1
Skills	SIDES Affect Dysregulation Scale (Child-Reported)	50.0
Social Influences	Couple Coercion Scale	77.8
Social Influences	Parent-Child Coercion Scale	66.7
Social Influences	Parent Cognition Scale	81.5
Social/Professional Role & Identity	10-Item Personality	55.6
Social/Professional Role & Identity	Parent-Child Coercion Scale	50.0
Values	Consideration of Future Consequences Scale	77.8

Figure 1. Self-rated expertise by study participants (N = 30).

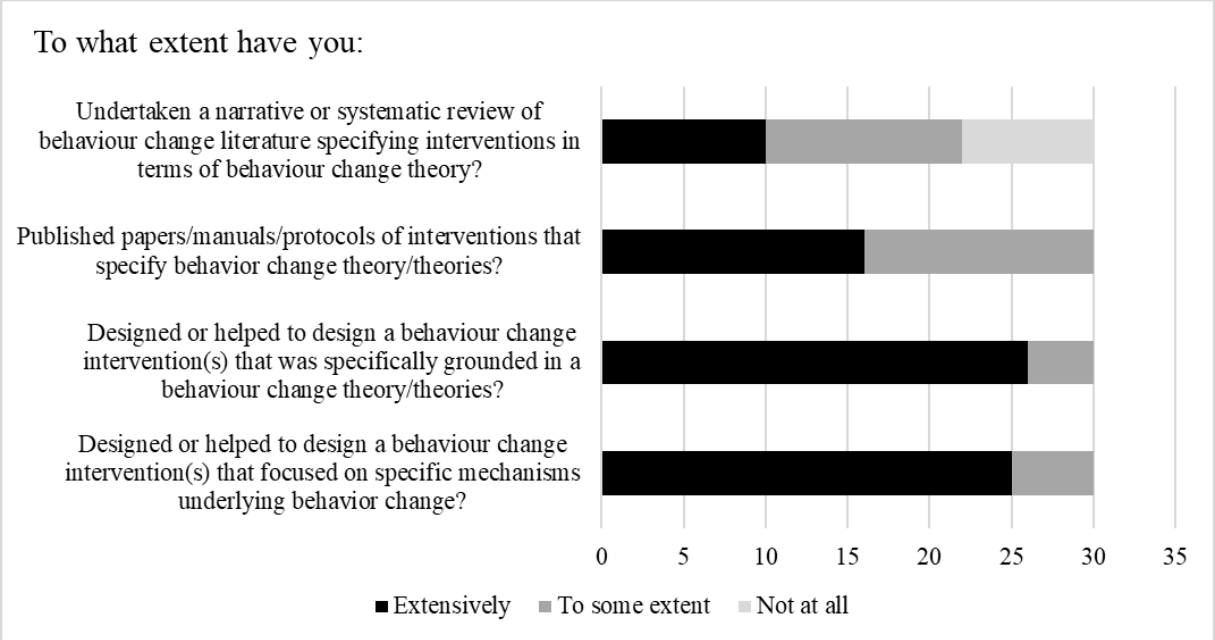

Figure 2. Heat map depicting links and percent agreement on relatedness of SOBC Measures and HBCP Mechanisms of Action (MoAs) reaching ≥ 50% agreement. Each cell represents a numerical value (i.e., percent agreement) and is shaded to indicate the relative "heat," i.e., the relevant percent agreement of a particular link. To view the data for the percent agreement, see Supplemental Appendix D.

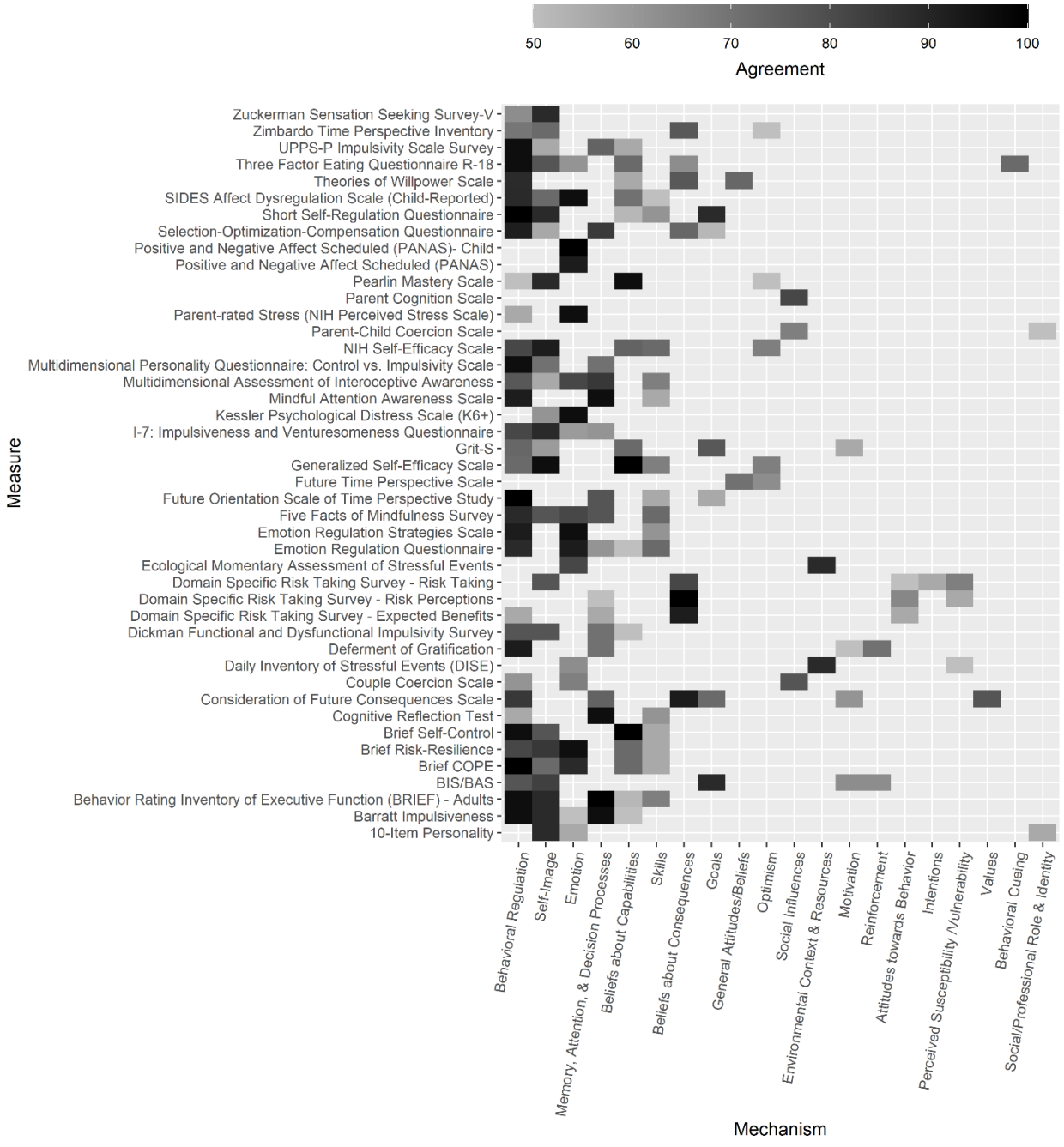

Figure 3. Heat map depicting links and percent agreement on relatedness of SOBC Measures and HBCP Mechanisms of Action (MoAs) reaching ≥ 90% agreement. Each cell represents a numerical value (i.e., percent agreement) and is shaded to indicate the relative "heat," i.e., the relevant percent agreement of a particular link. To view the data for the percent agreement, see Supplemental Appendix D.