Knowing What We're Talking About: Facilitating Decentralized, Unequivocal Publication of and Reference to Psychological Construct Definitions and Instructions

Gjalt-Jorn Peters and Rik Crutzen

2023-11-08

Abstract

A theory crisis and measurement crisis have been argued to be root causes of psychology's replication crisis. In both, the lack of conceptual clarification and the jingle-jangle jungle at the construct definition level as well the measurement level play a central role. We introduce a conceptual tool that can address these issues: a Decentralized Construct Taxonomy (DCT). This consists of comprehensive specifications of construct definitions and corresponding instructions for quantitative and qualitative research. We discuss how researchers can develop DCT specifications as well as how DCT specifications can be used in research, practice, and theory development. Finally, we discuss the implications and potential for future developments to answer the call for conceptual clarification and epistemic iteration. This contributes to the move towards a psychological science that progresses in a cumulative fashion through discussion and comparison.

Introduction

Recent large-scale efforts to replicate psychological studies have shown disappointing results. Results of replications of 100 experimental and correlational studies published in three psychology journals, for example, led to a clear conclusion: a large portion of replications produced no or weaker evidence for the original findings (Open Science Collaboration 2015). This warrants substantial skepticism regarding findings of psychological studies and the way in which they are conducted (Earp and Trafimow 2015). The ensuing period of self-reflection for psychological science (sometimes dubbed the "replication crisis", "crisis of confidence", or more optimistically, "credibility revolution") was associated with large-scale promotion and adoption of Open Science principles to decrease risk of bias contributing to this replication crisis (Vazire 2018; Pashler and Wagenmakers 2012; Nosek and Bar-Anan 2012; Nosek, Spies, and Motyl 2012; Uhlmann et al. 2019).

Open Science principles include, for example, Open Data (Murray-Rust 2018; Miyakawa 2020) and pre-registrations (Nosek et al. 2018; Lakens 2019). Although such practices yield many benefits, they have also been criticized as only being potential partial solutions to the replication crisis. This is because there are more fundamental problems underlying the replication crisis (Szollosi et al. 2020; Fried 2020; Van Rooij and Baggio 2020). These problems are referred to as the "measurement crisis" and the "theory crisis" (Eronen and Bringmann 2021; Oberauer and Lewandowsky 2019). The former concerns the validity of measurement instruments used to measure psychological constructs, and the latter to the underlying cause: the nature of much psychological theory.

The measurement crisis

The measurement crisis in a nutshell: Critical examination of measurement instruments demonstrates — over and over again — that measurement instruments that supposedly measure a given construct often turn out to measure different things (i.e. the jingle fallacy), and that measurement instruments for different constructs often turn out to measure the same thing [i.e. the jangle fallacy; Castanho Silva et al. (2020); Reschly and Christenson (2012); Fried and Nesse (2015); Fried (2017a); Hagger (2014); Lonsdorf, Merz, and Fullana (2019); Ponnock et al. (2020); Siegling and Petrides (2016); Skinner (1996); Warnell and Redcay (2019); Weidman, Steckler, and Tracy (2017); Williams and Rhodes (2016)]. As one of the causes of these fallacies, the popularity of the construct validity approach has been identified.

Based on a critique of how construct validity is employed in practice, Alexandrova and Haybron (2016) conclude that construct validity is not valid, asserting that it substitutes theory with statistics. Similarly, Borsboom et al. (2009) extensively discuss the logical and epistemological problems of construct validity theory, and propose to abandon construct validity in favor of test validity: whether a test measures what it should measure [note that this is not what construct validity is; see Borsboom et al. (2009)]. In short, test validity is based on the assumption that if a measurement instrument measures what it should measure, the measured attribute should be the cause of the variation in the item scores. A claim that a measurement instrument is valid for a given construct should therefore be based on a theory of the item response processes and the role that the construct plays in those processes. Similarly, Borgstede and Eggert argue for theory-based measurement: "performing a standardized application of a theory that is suitable to provide enough information to calculate a hitherto unknown value for a theoretical term" (Borgstede and Eggert 2022, 14).

Such approaches provide a promising and robust foothold to address the measurement crisis, provided that researchers have sufficiently comprehensively specified theories about the constructs they want to measure. However, theoretical underpinning for measurement models and methods often remain latent at best (Fried 2020). This brings us to the second, arguably more fundamental, crisis: the theory crisis.

The theory crisis

The mostly qualitative, narrative nature of much theory in psychology is central to the lack of robustness in psychological science. A proposed partial solution could be formalizing theory and moving towards computational models (Guest and Martin 2021; Borsboom et al. 2021; Smaldino 2008; but also see Oude Maatman 2021; Eronen and Romeijn 2020). Formalizing theory enables more accurate representation of the theories in scientists' minds and renders them more widely accessible. Within health psychology, a formal system for describing theories was recently proposed (West et al. 2019). This system can facilitate the transition to formal modeling as it places more formal theory specification within the grasp of all psychological researchers regardless of their computational modeling skills. However, this formal system is primarily concerned with specifying postulated relationships (e.g., "edges") between psychological constructs (e.g., "nodes"). Formalizing these relationships is a valuable step forward, but the precise definitions of the constructs that form the nodes remain external to this specification as qualitative descriptions using natural language.

These precise definitions are crucial to build better theories. As construct definitions become more comprehensive and precise, the value of theories involving those constructs increases as both holistic and contrastive theoretical underdetermination decrease (Oude Maatman 2021). As Smaldino (2016, 331) put it, "attempting to understand relationships among interacting systems of such startling complexity as human beings is a daunting task. Tackling it requires that we make the greatest attempt to specify precisely what we mean." Unfortunately, the current state of construct definitions is in dire straits (Leising and Borgstede 2019).

In practice, construct definitions are typically relatively short. Before going into why this brevity is problematic, we will illustrate it and speculate as to potential causes. A recent list of 1164 construct definitions was published as a part of an ontology-based modeling system for behavior change theories (Hale et al. 2020) and contains construct definitions ranging from 3 to 72 words in length (with a median of 15 words and a mean of 16 words). This typical brevity does not seem to be a matter of original extensive definitions having eroded over time (with original comprehensive definitions having become "lost in temporal translation"), because definitions in the list published by were formulated in careful collaboration with the original theorists (Hale et al. 2020). Thus, these brief definitions are both common and generally accepted in psychological theory.

One of the reasons for this may be what has been termed implicit realism (Yarkoni 2020). This means that researchers tend to treat constructs as if they represent natural kinds (Fried 2017b). In other words, constructs are treated as if they represent more or less discrete, modular parts of the natural world that are neatly delineated. This tendency is reflected by, for example, the widespread (and not always appropriate) reliance on reflective measurement models (Gruijters, Fleuren, and Peters 2021) and the neglect of ongoing construct validation (Flake, Pek, and Hehman 2017). At the same time, in discussions about philosophy of science, when explicitly prompted with this realist ontological perspective on psychological attributes, few researchers subscribe to it in our experience. The argument that, for example, attitude exists as a bounded psychological entity that is neatly delineated from other constructs such as risk perception or self-efficacy has to our knowledge never been seriously made. On the contrary, overlap among psychological constructs is postulated by many theories (for example, theories on behavior: Crutzen and Peters 2023a).

This discrepancy between what seemingly implicitly informs researchers' methodological and operational decisions and the position they take when explicitly reflecting on their ontological stance may be explained by what has been called an essentialist bias (Brick et al. 2022). In short, this bias describes people's tendency to implicitly infer from the use of construct labels such as "attitude" or "depression" that some "attitude essence" or "depression essence" exists (Brick et al. 2022). The existence of such an essence would justify accepting brief and general definitions. However, from an empirical point of view, essences are unidentifiable (Borgstede and Eggert 2022). If we use essences to define constructs, the definitions become circular (e.g., "attitude is something that has attitude essence") — and the other way around, corresponding constructs become vacuous (e.g., "everything that has attitude essence is attitude").

Hidden heterogeneity and epistemic diversity

These brief and incomplete definitions of psychological constructs require elaboration before being usable in empirical research. Given the variation in researchers' cultures, histories, academic training, and theories from psychology and other disciplines that researchers are familiar with, it makes sense that different groups of researchers arrive at different elaborated construct definitions. It is perhaps not surprising that these implicitly different elaborated construct definitions yield a plethora of measurement instruments that purportedly measure the same construct, but on closer inspection turn out to measure different things (Lawson and Robins 2021; Hodson 2021).

To give a concrete example, a great variety of measurement instruments are used to assess depression severity (Santor, Gregus, and Welch 2006). A content analysis of 125 items across 7 of these commonly used measurement instruments revealed that they included 52 disparate depression symptoms (Fried 2017a). Of these 52 symptoms, 21 (40%) appear only in one single instrument and only 6 (12%) feature across all instruments. This implies that using these measurement instruments interchangeably poses a threat to the generalizability and replicability of, in this case, depression research. Similarly, a recent investigation of mindfulness measurement instruments showed that eight commonly used measurement instruments all measure different subsets of aspects of mindfulness (Altgassen, Geiger, and Wilhelm 2023). Unfortunately, these are not rare examples, and such jingle-jangle fallacies seem the rule rather than the exception in psychological research (Lawson and Robins 2021), so much so that Flake and Fried devote the first of their six questions to help prevent Questionable Measurement Practices to construct explication (Flake and Fried 2020).

Despite the manifold problems of this heterogeneity, this situation is not all bad. This is because despite potential essentialist biases and implicit realism, in explicit discourse, earnest arguments that psychological constructs are natural kinds are rare (as far as we could find, non-existent). Because psychological constructs are not more or less discrete, modular parts of the natural world that are neatly delineated, different definitions of the same construct can arise. In fact, in any set of different definitions of the same construct, one definition cannot be considered "correct" and others "incorrect". As components of psychological models, definitions can be considered to share a characteristic of models more generally. As Box (1979) famously stated: "All models are wrong, but some are useful", and a similar stance may be adopted towards definitions of psychological constructs.

This ontological perspective suggests that there is the potential for epistemic benefits through what Zollman (2010) calls transient diversity (epistemic diversity has been argued to have benefits more widely; see Devezer et al. 2019; Cartwright 2021). When applied to construct definitions, the implication is that scientific progress is served better by heterogeneity in construct definitions than by uniformity. The benefits of Feyerabend's theoretical pluralism (1965; i.e., emphasizing the value of conflicting theories for scientific progress; see Bschir and Lohse 2022 for a recent discussion that broadens that perspective), then, may apply not only to theories, but also on the slightly lower conceptual level of the constructs defined therein. In other words, having multiple different definitions of psychological constructs may accelerate scientific progress.

However, if having many different implicitly elaborated definitions of constructs such as "attitudes", "introjected regulation", and "resilience" is in itself desirable in light of scientific progress, then what is the problem? The problem is the hidden nature of this heterogeneity. The exact (i.e. researcher-elaborated) definitions used in a given psychological study can vary considerably but typically remain unknown. Instead, reports of such studies tend to briefly (re)define the central constructs, or alternatively cite original formulations of the relevant theory.

This is problematic because there is no reason to assume that differences between different versions of the same construct (i.e. the actually employed, implicitly elaborated constructs) are trivial and nonconsequential. The different auxiliary assumptions implied by differences in definitions may explain part of the heterogeneity in empirical findings (Smaldino 2016), thus underlying replication failures as well as confounding different results in different populations and contexts. A recent large-scale collaboration highlighted these dynamics (Landy et al. 2020). Fifteen research teams independently designed studies and materials to test the same five hypotheses. The results illustrated both the variability in different research teams' design choices and the resulting heterogeneity in the observed effect sizes (Landy et al. 2020). In other words, this hidden heterogeneity likely contributes to both the measurement crisis and the replication crisis.

An illustration of hidden heterogeneity

To illustrate this hidden heterogeneity, we briefly reflect on the five open access studies with "attitudes" in their titles published in Psychology & Health[1]. None of the articles provides a definition of "attitude", suggesting that the authors assume that they share a common and well-known definition with the readership. However, each article did describe how attitude was measured (for quantitative studies) or which characteristics of qualitative data fit with the authors' attitude definition (for qualitative studies).

Huls et al. (2022) report on a validation study of an attitude measurement instrument, the Health-Risk Attitude Scale (HRAS-13). This measurement instrument contains items such as "my health means everything to me", "people say that I take risks with my health because of my habits", and "I think that I take good care of my body". These items have to be rated on a seven-point response scale ranging from "completely disagree" to "completely agree."

Taylor et al. (2022) report on a mixed method study where interviews were used to delve deeper into mothers' attitudes regarding HPV vaccination. They describe that "[…] all of the mothers attended screening because of its considered importance, with several expressing the opinion that it is something they had to do […]. One mother, Andrea, talked about how she attributed this attitude to an unquestioning trust in the importance of cervical screening […]. Belinda explicitly described her unquestioning attitude towards it: 'I didn't really question it, I just assumed it was a good thing to do' (FG1)."

Schnell et al. (2022) report on a study into COVID-19 measure compliance. They explain that "[a]greement with and opposition to public health guidelines measured the participants' attitudes towards these guidelines." Agreement with public health guidelines assessed the extent to which respondents found these measures appropriate, understandable, well-explained, reasonable and meaningful (five items). Opposition to public health guidelines measured the extent to which respondents found these measures oppressive, excessive, unreasonably restrictive of their civil rights and unlawful surveillance (four items). Participants were asked to position a slider on an eleven-point response scale (ranging from "not at all" to "very much so") to indicate for each of the nine adjectives how much

[1] This search can be repeated by visiting https://www.tandfonline.com/action/doSearch?field1=Title&text1=attitudes&SeriesKey=gpsh20&openAccess=18, persistently archived at https://web.archive.org/web/20230824103037/https://www.tandfonline.com/action/doSearch?field1=Title&text1=attitudes&SeriesKey=gpsh20&openAccess=18.

they describe how they felt about the authorities' public health decisions regarding the pandemic.

De Graaf et al. (2017) report on the effects of a smoking education intervention. They measured attitude "[…] with five items, following the stem 'I find smoking …' measured on a four-point response scale ranging from, for example, very negative to very positive, or very unwise to very wise". Wood et al. (2014) report on a study into the potential role of attitude accessibility in the question-behavior effect, and used similar items. Specifically, they used three items with the stem 'For me, eating healthy foods in the next few weeks would be …' and seven-point response scales anchored by not worthwhile/worthwhile; bad/good; and not beneficial/beneficial).

Despite not having explicated attitude definitions available, these descriptions illustrate the heterogeneity in the implicit definitions of the attitude construct. The HRAS-13 items assess the importance of a person's health, the subjective evaluation of taking care of one's body, and the impression of other people's evaluations of one's risk taking behavior (Huls et al. 2022). The corresponding definition of attitude seems to contain elements of risk perception, social norms, and importance. These elements seem absent from the definition used by Taylor et al. (2022), which instead contains elements of trust and moral norms. The definition of attitude used by Schnell et al. (2022) appears to involve subjective evaluations of a policy as reasonable, well-explained, meaningful, oppressive, and illegal. Finally, both De Graaf et al. (2017) and Wood et al. (2014) use items close to the recommendations in the Reasoned Action Approach lineage theories (Fishbein and Ajzen 2010), where attitude is a generic evaluation of a behavior (but appearing to use different dimensions).

How exactly the implicitly elaborated definitions of these five groups of authors differ is hard to say based on the limited available information. Given the heterogeneity in how attitude was measured or coded, it seems unlikely that if they were to elaborate their implicit definitions into comprehensively described specifications this would yield similar definitions. This highlights three issues related to hidden heterogeneity. First, the fact that these articles report on studies into psychological constructs without transparently communicating to the scientific community how those constructs are defined is in itself problematic. We do not highlight this to accuse the authors, editors, or reviewers, but quite the opposite: to illustrate how widespread and accepted the practice of not explicitly defining psychological constructs is. Second, if explicit definitions were omitted based on an assumption that all health psychologists have the same definition of attitude in mind, this degree of hidden heterogeneity shows that assumption to be unwarranted. Third, this heterogeneity prohibits evidence syntheses. Science relies on knowledge accumulation, but a prerequisite to such accumulation is clarity about what is studied in a given study.

Towards a practical solution for hidden heterogeneity in construct definitions

The solution to the problem of hidden heterogeneity is making the heterogeneity explicit by publishing comprehensive construct definitions. These are ideally so comprehensive that they eliminate the need for further elaboration when researchers want to use the construct. By providing such comprehensive definitions, it becomes possible to identify subtle (or not so subtle) differences between construct definitions. The five studies described above would likely have at least four, and maybe even five different attitude definitions. Making these explicit enables critical reflection on and discussion of those differences, as well as empirical investigation of the conditions under which different definitions perform similarly or different.

This plea for explicit and comprehensive definitions is in line with more broadly shared needs concerning conceptual clarity. For example, Alexandrova and Haybron (2016) argue that when validating psychological measures, researchers often focus on statistical procedures at the expense of theory, largely forgoing comprehensive conceptual clarification. Vazire et al. (2022) review a number of studies that illustrate problems with conceptual clarity and validity. Bringmann et al. (2022) describe how a lack of conceptual clarity has held back scientific progress in psychology and call for dedicated efforts to remedy this. They emphasize that conceptual clarification enables epistemic iteration: iteratively improving theoretical construct definitions and measurement instruments (Chang 2007). Moreau and Wiebels (2022) argue that more attention for construct definitions and more explicit heterogeneity in construct definitions enables escaping "local optima": definitions and operationalizations that perform better than closely related alternatives, but do not necessarily represent the best possible version. Finally, Lawson and Robins (2021) include specific, comprehensive construct definitions as a prerequisite for establishing how constructs relate to each other.

Clearly, there is a consensus that psychological science should, as Bringmann et al. (2022) put it, go back to basics and establish conceptually clear and comprehensive construct definitions. At the same time, currently no tools exist to scaffold this practice in a manner that facilitates decentralized, unequivocal communication, epistemic iteration, and knowledge exchange and accumulation. This lack of concrete tools may be one of the causes for the observation that most psychological findings are "not even wrong" (Smaldino 2016; Scheel 2022). For example, Smaldino (2016) and Scheel (2022) argue that the verbal statements used by psychological researchers often leave much to be desired in terms of specificity: theories, predictions, and hypotheses are problematically underspecified. They argue for a shift towards formal modeling and machine-readable theory, prediction, and hypothesis specifications. At the same time, there is quite a chasm between the commonplace verbal, narrated descriptions of theories and constructs on the one hand and more precise specifications (Guest and Martin 2021). We here present Decentralized Construct Taxonomies as a bridge to start crossing that chasm from the narrative side: a concrete tool to help create comprehensive, explicit, transparent construct definitions.

Decentralized Construct Taxonomy specifications

Decentralized Construct Taxonomies are sets of one or more Decentralized Construct Taxonomy (DCT) specifications, each of which specifies a single construct definition, corresponding instructions for using the construct in primary and secondary research, and metadata. These accompanying instructions for using the construct in research and practice are so tightly tied to the construct definition itself to minimize the risk that using a construct still requires more implicit elaboration of a construct's definition. In the current contribution, we suggest that additionally, four sets of instructions are specified. First, instructions to measure the construct in a way consistent with the provided definition (e.g., to conduct quantitative research). Second, instructions to code or classify an existing measurement instrument as one that does or does not measure the construct as defined in the definition (e.g., to conduct literature reviews). Third, instructions to elicit qualitative data that is informative about the construct; and fourth, instructions to code qualitative data as indicative of content of the construct. For an example of a construct view in a construct repository, see Figure 1.

In the remainder of this text, we will first describe the components that make up a DCT specification followed by three examples. We will then describe ways to develop DCT specifications, and then describe ways to use them in research and practice. We will then briefly describe a number of open source technological tools that implement DCTs and that facilitate working with them: a machine- and human-readable standard to store DCT specifications in files; a way to generate Unique Construct Identifiers (UCIDs); the `{psyverse}` R package to work with DCTs; a psychological construct repository; and a Shiny app to import, edit, and submit DCT specifications to a psychological construct repository. We will then discuss strengths and limitations and avenues for further development.

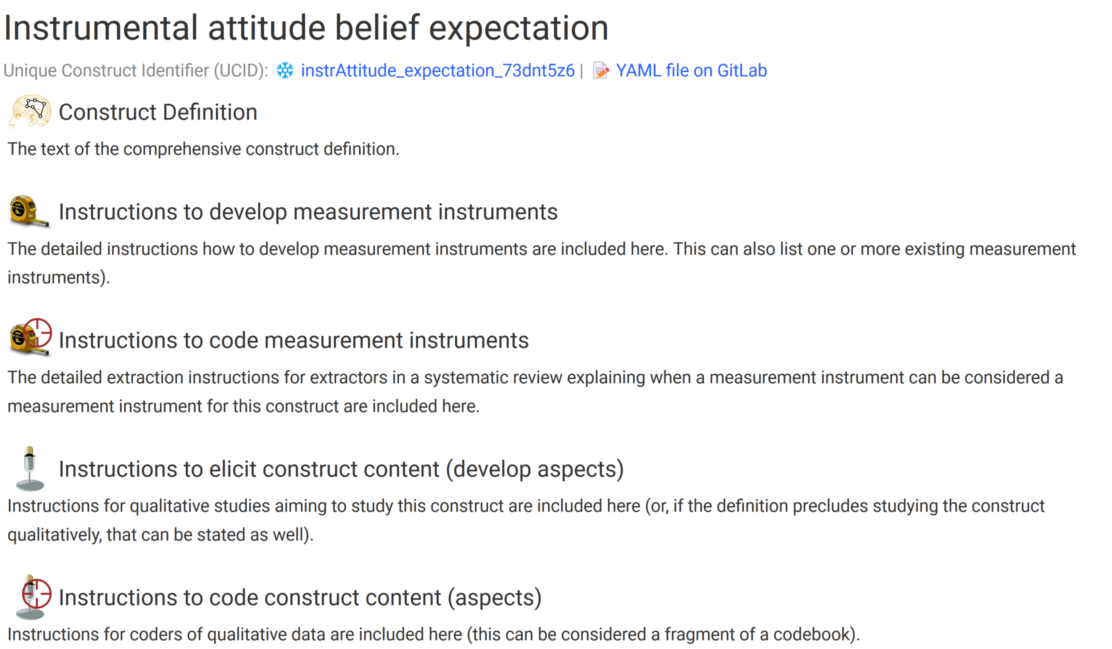

Figure 1: Figure 1. A screenshot of the Construct View of a single DCT specification in the PsyCoRe.one repository (specifically, https://psycore.one/instrAttitude_expectation_73dnt5z6). The contents of the definition and each of the four instructions has been replaced with a brief placeholder.

Components of a Decentralized Construct Taxonomy specification

The construct definition is the central part of a DCT specification. It describes, as comprehensively as possible, how a construct is defined. One way to think about this is that ideally, this definition is so comprehensive that it is easy for others to spot where they themselves would define the construct differently: it is that potential for heterogeneity in definitions that allows for epistemic iteration through empirically testing the implications of using the (potentially subtly) different definitions. For any given difference in definition, learning that it does not matter in practice is as valuable as learning that it matters, and if so, how. Therefore, the more comprehensive construct definitions are, the better: definitions are the bedrock of empirical research, and excessively brief or vague definitions prevent valid measurement [clarifying construct definitions are the first step in preventing Questionable Measurement Practices; Flake and Fried (2020)].

The instruction to develop measurement instruments describes how the construct can be measured. This can outline theoretical prerequisites and constraints, such as auxiliary assumptions and procedures that must be followed to develop a valid measurement instrument. It can also point to one specific measurement instrument and include the claim that that is the only possible way to validly measure that construct. It can also be a combination, referring to one or more existing instruments and including instructions on how to develop additional measurement instruments. Here, too, the adage is that more comprehensive instructions are better. Flake and Fried (2020) provide a useful overview of measurement-related issues to consider when writing these instructions.

The instruction to code measurement instruments describes when a measurement instrument can be considered a measurement instrument of the construct. The primary use case for such instructions is for conducting systematic reviews: the jingle-jangle jungle precludes blindly "trusting" that measurement instruments do indeed measure the construct as the original study authors labeled it. This necessitates critical examination of all measures to "reclassify" them in a way that is consistent with the construct definitions used by the systematic review authors and so try to prevent bias in the synthesized estimates. In addition, these instructions are useful when researchers or practitioners are looking for a measurement instrument to use in an empirical study, and they want to ensure the measurement instrument fits with the construct definition they use.

The instructions to elicit and code qualitative data are more closely related than those pertaining to measurement instruments. The former describes how qualitative data that is informative about the construct can be obtained. This can include, for example, questions to ask participants, procedures for observation, or procedures for autophotography and photo elicitation (Glaw et al. 2017). The instructions to code qualitative data then describe when data can be considered informative about the construct. For example, if qualitative data are elicited by asking people a question, the answers may not be informative about the corresponding construct; people may be distracted, misunderstand the question, or include irrelevant information. It is important to note that providing these instructions (e.g., clearly describing which types of expressions, behaviors, or interpersonal interactions can be coded as relevant to a given construct) does not require one to subscribe to any particular ontological or epistemological perspective regarding psychological attributes. For example, both from a hard realism perspective (e.g., assuming inherent properties of constructs) and from a constructivist perspective (e.g., assuming constructs are socially constructed), different constructs exist, and they can be distinguished from each other in empirical research.

Not every set of instructions may apply to each construct definition. For example, if researchers have a constructivist perspective and assume a given construct cannot be measured, the instructions for measurement could instead explain why the construct, as defined in the definition section, cannot be measured. Similarly, researchers may define a construct in a way that precludes qualitative investigation. In other scenarios, researchers might only require explicit instructions for coding measurement instruments (e.g., when they develop the DCT specification for a systematic review), and they may refrain from elaborating on other sets of instructions for the sake of efficiency. In such cases, instead of specifying all sets of instructions, researchers can describe why specific sets of instructions are not added. We want to emphasize that any reason can be valid, also depending on one's perspective: we propose DCTs as a tool for increasing transparency, not to enforce a given normative perspective (except perhaps the open science imperative to optimize transparency).

Examples of developed DCTs

Example 1: A DCT specification for Instrumental Attitude Belief Expectation

This first example is one of the products of a project where we aimed to produce DCT specifications for the entire Reasoned Action Approach (Fishbein and Ajzen 2010). We started off using a single theory approach, using two resources: the book on the Reasoned Action Approach (Fishbein and Ajzen 2010) and an article describing a formal system for representing behavior-change theories - among them the Reasoned Action Approach (West et al. 2019). Although we had this book available (comprising 518 pages about this specific theory) we were still not able to fully complete the DCT specifications for all constructs within the Reasoned Action Approach. Hence, we supplemented the single theory approach with the expert-based rationalist approach, — also based on our experience with applying this theory (e.g., Crutzen and Peters 2023b). When reporting about the development of DCT specifications, it is warranted to also describe which methods are used during the process. In our case, after extensive discussions in multiple meetings, we were able to complete DCT specifications that reflect our thinking about this theory (see Peters et al. 2020, which also visualizes the relationship between constructs, and see https://osf.io/me3h2 for all files with DCT specifications and the used R scripts).

We will describe one DCT specification in more detail: that of Instrumental Attitude Belief Expectation (https://psycore.one/instrAttitude_expectation_73dnt5z6). In its definition ("the expectation of how probable (i.e. unlikely versus likely) it is that engaging in the target behavior will cause one specific instrumental potential consequence to come about"), we explicitly distinguish the construct from Experiential Attitude Belief Expectation (https://psycore.on/expAttitude_expectation_73dnt5z1), which captures expectations relating to experiences and sensations (as opposed to Instrumental Attitude Belief Expectation, which captures expectations as to how instrumental potential consequences of a behavior are given an individual's goals). We also explicitly describe how this construct relates to related constructs, such as Instrumental Attitude Belief Evaluations (https://psycore.one/instrAttitude_evaluation_73dnt5z7) and the compound construct https://psycore.one/instrAttitude_belief_73dnt5z8.

The instructions for developing measurement instruments are extensive (almost 800 words). This is meant to minimize heterogeneity by constraining researcher degrees of freedom in developing measurement instruments. For example, take the following fragment: "Try to always be consistent in the scale valence; in languages that are read from left to right, always place the most passive/low/less/weak/unlikely scale extreme (anchor) on the left, and the most active/high/more/strong/likely scale extreme (anchor) on the right. Do not reverse this order for one or more items." This fragment leaves relatively little room for variation, and simultaneously likely causes disagreement by some researchers (e.g. who believe that it is important to include reverse-coded items). Note that researchers can always copy the DCT specification, make adjustments, and then use their own version. However, this still serves to make exercise of those degrees of freedom salient, and explicates the differences in DCT specifications.

Example 2: A DCT specification for Problematic Hypersexuality

In their critical discussion of the theoretical and methodological issues surrounding research into hypersexuality, Van Tuijl et al. (2023b) present an elaborated definition of Problematic Hypersexuality (https://psycore.one/problematicHypersexuality_7mm4hr4f). In their definition, they draw on three diagnostic perspectives on problematic hypersexuality (i.e. sex addiction, hypersexual disorder, and compulsive sexual behavior disorder) and ten characteristics based on criteria for Problematic Hypersexuality retries from these three perspectives. They postulate that these ten characteristics manifest as trait-like components of Problematic Hypersexuality, and state that their relative importance can vary per population. In addition, they describe two drivers as part of their definition (high sexual desire and emotion dysregulation with regard to sexuality), and state that these are necessary but not sufficient for hypersexuality to fall within the definition of Problematic Hypersexuality. In addition, these drivers would need to interact with the ten characteristics in a causal network where the drivers manifest as a predisposition to engage in sex in response to negative experiences, while engaging in sex strengthens the ten characteristics, which then further augment the drivers. They explicitly describe their ongoing research and explicitly state the preliminary nature of the construct.

In their instructions, they describe how they are working on measurement instruments, but do not provide explicit instructions. In their instructions for coding measurement instruments (e.g., for systematic reviews), they explain that in their view, as yet, no measurement instruments exist that can measure Problematic Hypersexuality consistent with the definition they provided. They do provide instructions for qualitative research, linking to the open materials they use in a specific study. In addition, they have since developed an updated DCT specification (https://psycore.one/ProblematicHypersexuality_7qfdty7g), where they describe ongoing work to arrive at measurement instruments.

Example 3: A DCT specification for Life Satisfaction

A third example is a DCT specification for Life Satisfaction (https://psycore.one/life_satisfaction_7pk39d16), developed by Panasiuk (2023). Because this is a relatively brief DCT specification, we include it here in full. This DCT specification contrasts with the Problematic Hypersexuality DCT specification by including a brief construct definition ("A global assessment of a person's satisfaction with their life according to their chosen criteria."). It includes longer instructions for developing measurement instruments ("Life satisfaction is assessed using a single or multi-item measure that prompts a cognitive evaluation of one's life according to their own perspective that can be interpreted as an attitude, assessment, belief, evaluation or judgment. Items are, ideally, ordinal 5- 7- 10- or 11-point Likert scales with verbal anchors at the bottom and top. Typically, finer-grained scales are more reliable. Items may or may not prompt to think about a certain timeframe (e.g., 'last two weeks', 'last year') and scope (e.g., 'overall', 'all things considered'). Anchors at the bottom represent the lowest levels of satisfaction with life (e.g., 'worst possible life', 'completely dissatisfied', 'dissatisfied', 'not at all'). Anchors at the top represent the highest levels of satisfaction with life (e.g., 'best possible life', 'completely satisfied', 'satisfied', 'completely'). Care should be taken when including life satisfaction in a larger survey. There is evidence that answers to life satisfaction scales are sensitive to question order effects (Schwarz & Strack, 1999).").

The instructions to code measurement instruments are "Life satisfaction is measured with a Likert scale with anchors at the top and bottom indicating dissatisfaction and satisfaction with life respectively. To avoid ambiguity, items usually include a timeframe (e.g., As a whole, how satisfied are you with your life in the past year?). Items measuring life satisfaction could either measure one's satisfaction with their life in the present moment or with their life as a whole, as far back as they can recall or is relevant to inform their judgment.", the instructions to elicit qualitative data are "Conduct a qualitative study where participants are interviewed and interviews are either recorded and transcribed, or notes are kept. These sources are then coded using the instruction for aspect coding.

These questions may be formulated in the following way: 'How do you feel about your life?' and 'How is your life going?' or 'What do you think about your life'.", and the instructions to code qualitative data are "Expressions that demonstrate either positively or negatively valenced feelings towards one's life express one's satisfaction with life. For example, 'my life is going well' is an expression of satisfaction with life. There can also be statements that could be informing life satisfaction, but it is not apparent from the outset whether they positively or negatively contribute to one's life satisfaction. For example, 'I have a lot of responsibilities in life' can be an expression signifying satisfaction or dissatisfaction with this fact. Ideally, the interviewer would follow-up to identify whether this is a positive or negative contributor to the individual's life satisfaction."

Discussion of the examples

These three examples illustrate how varied DCT specifications can be. The first one, Instrumental Attitude Belief Expectation (https://psycore.one/instrAttitude_expectation_73dnt5z6), starts with the explicit description provided by the authors of the related theory, but includes elaboration — and as such, decisions that others may want to take differently. The second one, Problematic Hypersexuality (https://psycore.one/problematicHypersexuality_7mm4hr4f and https://psycore.one/ProblematicHypersexuality_7qfdty7g), describes a much more recently postulated construct and therefore contains more original foundational work. The third one, Life Satisfaction (https://psycore.one/life_satisfaction_7pk39d16), does the opposite, and resembles closely how researchers who study Life Satisfaction often approach the construct, but making this approach explicit.

The instructions to develop measurement instruments for Life Satisfaction are consistent with the most commonly used measurement instruments (e.g., Diener et al. 1985), suggesting that perhaps of the three DCT specifications, this is the one other researchers will agree with most. On the other hand, in the DCT specification for Instrumental Attitude Belief Expectation, many decisions were taken in the process to substantially elaborate the definition and corresponding instructions. This means that it is easier for researchers to spot aspects of this DCT specification they disagree with. The Problematic Hypersexuality DCT definitions have a different profile again. Given their clear deviation from existing definitions of Problematic Hypersexuality, as evidenced by their explanation in the first DCT specification that no existing measurement instruments can measure the construct as they define it, it is likely other researchers will disagree with their definitions. This is partly by virtue of how comprehensive and explicit the definition of Problematic Hypersexuality by Van Tuijl et al. (2023b) is. The ensuing empirical work to compare definitions and the accompanying measurement models (Van Tuijl, Verboon, and Van Lankveld 2023a), and the subsequent updating of definitions (and assignment of new UCIDs) is a good example of what epistemic iteration can look like in psychology (https://psycore.one/ProblematicHypersexuality_7qfdty7g).

Note that from a scientific point of view, disagreements about DCT specifications are valuable. The point of DCT specifications is not to codify consensus. This may seem counterintuitive, so to be clear: we consider this production of comprehensive construct definitions tied to explicit instructions for using those constructs in primary and secondary research a worthwhile accomplishment in itself. They represent opportunities to resolve such disagreements with empirical research and conceptual delineation of constructs, and they enable unequivocal reference to constructs, decrease hidden heterogeneity, and facilitate coherent research in separate projects. However, it is important to realize that although those benefits are real, this act of creating these DCT specifications does not suggest that these specific construct definitions (and accompanying instructions) are necessarily correct or useful. As for the example of the DCT specification we created (Instrumental Attitude Belief Expectation): it is our hope that other researchers will disagree with some or all of the decisions we took when developing this DCT specification. As such, we welcome suggestions for improvement, and consider these definitions an explicit starting point for reflections, discussions, and empirical work on the foci of such disagreements.

Methods to develop Decentralized Construct Taxonomy specifications

There are several methods to develop these DCT specifications. These methods differ in terms of, for example, point of departure (e.g., theory, expertise) and number of sources. We describe nine methods in Appendix 1, “Methods to develop Decentralized Construct Taxonomy specifications”, and will describe five of them here. It is important to keep in mind that these methods are ways to elaborate construct definitions into comprehensive definitions with corresponding instructions — they are not ways to arrive at those constructs. To that end, we direct readers to procedures to develop theory (e.g., Hawkins-Elder and Ward 2020; Borsboom et al. 2021; Podsakoff, MacKenzie, and Podsakoff 2016).

Transparent descriptive approach

An initial step to move away from implicitly different construct definitions is to transparently describe constructs and the way in which they are used in a specific study. This transparency enables consumers of research (e.g., readers of an article) to understand how producers of research (e.g., authors of an article) defined and used the constructs in that study. An advantage is that this approach is very easily applicable — it is in line with the imperative for transparent reporting that is also stimulated by, for example, various reporting guidelines concerning what aspects to describe when reporting a specific study (e.g., CONSORT for randomized trials, COREQ for qualitative research). Moreover, it can in principle also be applied in hindsight after initial reporting of a study. On the one hand, this approach can be seen as an initial step towards being more explicit, even though the other methods described below are expected to result in more rigorous elaboration of construct definitions. On the other hand, the enhanced transparency achieved by simply making explicit what would otherwise remain implicit makes even this most straightforward approach an important step forward.

Single theory approach

The single theory approach consists of studying existing documentation about a specific theory. The primary source is often the work of the original author(s) of the theory. A benefit is that this approach requires limited resources. Moreover, this approach is probably easily applicable, because in theory, a theory should also specify how to develop a measurement instrument (and perhaps also how to conduct qualitative research using that theory). So, there is a more direct link between the construct definitions and instructions on how to use them in research. Over time, DCT specifications based on a single theory can also be improved by the original author(s) — or other researchers using this specific theory (Kok and Ruiter 2014). A downside is that in practice, many theories have very brief construct definitions, let alone detailed instructions on how to use them in research.

This may force researchers themselves to “fill in the blanks” that exist in documentation about a specific theory. And, more in general, relying on a single theory might result in a narrow view on the psychological construct of interest. Still, even if the resulting DCT specifications end up just describing how a team of researchers decided to “fill in blanks” in the construct definitions in the original theory formulation, that would provide valuable context to other researchers.

Meta-empirical approach

Where the single theory approach and the multiple theories approach (see Appendix 1) can be considered deductive, the meta-empirical approach is instead inductive. It consists of conducting a scoping review of empirical contributions in which psychological constructs have been used. This results in an overview of various (aspects of) definitions that are used in different empirical contributions as well as their operationalizations (e.g., questionnaires, items, tasks). It casts a wide net of how the psychological construct of interest is used. Similar to conducting a systematic review within the multiple theories approach, the resources required for conducting a scoping review are relatively high. However, a benefit is that an overview of measurement instruments and applications in qualitative research is obtained, which can help the researchers by grounding the four sets of instructions in the existing literature. For an example of this approach, see Snippe et al. (M. H. M. Snippe, Peters, and Kok 2021), and for a subsequent application, see Snippe et al. (Marwin H. M. Snippe, Peters, and Kok 2023). Like the multiple theory approach, this approach still requires that choices need to be made about whether and how to integrate into one DCT specification or multiple DCT specifications.

Reverse engineering approach

It is not always feasible or necessary to conduct a systematic review to develop a DCT specification. Researchers may also already have a measurement instrument they typically use. In that case, they can derive a construct definition from that instrument by closely inspecting the items. In this process, the researchers explicate those aspects of the construct that they consider most relevant, as well as their interpretation. Such DCT specifications may start out relatively brief if the item content does not provide much footing. However, even in such scenarios, brief initial DCT specifications still provide a convenient tool to enhance the transparency about authors’ elaborated construct definition as it guided their theorizing and study design, and it serves as a concrete starting point for further development.

Expert-based Delphi approach

The expert-based Delphi approach is similar to the expert-based rationalist approach (see Appendix 1), although it usually involves a larger number of experts. This approach is characterized by a systematic process used to enable collective input of a panel of experts. Usually this process is characterized by multiple iterative rounds in which input from previous rounds is fed back in an anonymous way (Nasa, Jain, and Juneja 2021). The benefit of this anonymity is that experts might be more inclined to share their thinking about the psychological construct of interest and feel less vulnerable. Furthermore, the Delphi approach is geared towards reaching consensus (based on pre-specified criteria), which might also contribute to broadly supported DCT specifications. A drawback of such consensus-based broad support is that it may carry the risk of stagnating epistemic iteration (though awareness of that risk may allow mitigating it). A second drawback is that optimizing for consensus does not simultaneously optimize for accuracy, comprehensiveness, validity, or conceptual clarity, while still imbuing the result with a certain authority that may suggest such merits. Therefore, it is important to bear in mind that this support on its own is not informative regarding the quality of the DTC specification.

Methodological and human considerations

Even though the approaches discussed here may appear distinct, in practice we expect that they will often be combined. Another methodological consideration, and one that holds for all methods, is that while efforts to develop a DCT specification may result in one DCT specification, they may also result in multiple DCT specifications for each construct. The latter can happen when multiple researchers have different views on a specific construct, but also when an individual researcher develops their thinking about a construct. Intuitively, this might initially be perceived as not being helpful and resulting in more heterogeneity. However, the problem to address is not heterogeneity itself but its hidden nature. DCT specifications foster openness and these specifications are not in and of themselves right or wrong. In fact, all DCT specifications might be wrong, but their usefulness to psychology is based on their fostering of transparency, not on a pretense to be correct. Their explicitness is a crucial step towards a psychological science that progresses in a cumulative fashion that is fueled by discussion, comparison, and epistemic iteration.

At the same time, we realize that researchers might feel vulnerable to develop and publish DCT specifications because their openness might lead to critique on and even falsification of aspects of their construct definition or operationalization. However, if this vulnerability is a barrier to transparency and self-correction, it might actually hinder scientific progress (Rohrer et al. 2021). In this sense, being transparent about construct definitions and how they change over the course of subsequent studies can contribute to changing the research culture (a need also emphasized by Rohrer et al. 2021). To contribute to this, we provided the examples above based on our own efforts and those of others.

Using existing DCT specifications in research and practice

Now that we have described methods that can be used to develop comprehensive construct definitions, we will describe how to work with DCTs in research and in practice.

Using Decentralized Construct Taxonomies in research

DCT specifications can be used in research on two levels: that of the lab, institution, or learned society and that of the individual researcher or student. Organizations can maintain dedicated Decentralized Construct Taxonomies, which can then be used by the affiliated researchers and students. This minimizes hidden heterogeneity within the lab, institution, or learned society, ensuring that the results can be synthesized relatively effortlessly and with minimal risk of bias as a consequence of differences in construct definitions, measurement, or qualitative elicitation or coding. Note, however, that when larger organizations provide standardized Decentralized Construct Taxonomies, these run the risk of being used by large groups of researchers, which would decrease epistemic diversity (and potentially slow down scientific progress).

Individual researchers and students can use such repositories to obtain DCT specifications. For example, if their lab, institution, learned society, or other organization maintains a psychological construct repository, they can use it to locate the DCT specifications for the constructs they aim to study. If researchers or students have no access to a repository of "preferred construct definitions", they can consult the literature and obtain the DCT specifications that are linked to in other repositories or added in the supplemental materials. They can then either use those as-is or adapt them. If no preexisting DCT specifications are available that they would like to use, they can develop their own (for example, using the transparent descriptive approach as discussed above).

It is important that researchers and students either include the DCT specifications for the constructs they study in the supplemental materials, or post them in a repository and provide a link to the construct page in that repository. Without explicating the elaborated, comprehensive definitions, others cannot be certain how constructs were defined exactly in a given study, and how this definition was taken to result in guidelines and constraints for quantitative or qualitative research.

Using DCT specifications in primary empirical quantitative research

In quantitative research, DCT specifications enable answering the first question that researchers have to answer to avoid Questionable Measurement Practices (Flake and Fried 2020): what is the construct that a researcher wants to measure and how is it defined exactly? The other five questions build upon this first question, firmly placing comprehensive construct definitions at the core of empirical research. Similarly, DCT specifications can be used to document the epistemic iteration that occurs when studying participants' response processes using Response Process Evaluation (Wolf et al. 2023).

Furthermore, DCT specifications can ensure that multiple studies in a lab or organization use the same construct definition, and when different measurement instruments are used, that these are nonetheless all consistent with that definition (as they all fit with the instruction for measurement instruments). This ensures that studies produced by researchers and students in a given lab or organization are internally coherent. In addition, DCT specifications provide a tangible, concrete prompt for reflection on the used definitions and instructions, encouraging regular evaluation and revision as a team. In a way, this achieves epistemic iteration on a small scale.

In addition, DCT specifications are a prerequisite for applying the first of the ten criteria for establishing to which degree constructs are sibling constructs [specifically, the extent to which constructs are defined in a conceptually similar way; Lawson and Robins (2021)]. This, in turn, is a necessary step before researchers can determine whether they can safely include variables as covariates in multivariate analyses. Including two or more sibling constructs (or constructs that otherwise describe overlapping parts of the human psyche) results in biased estimates (Crutzen and Peters 2023a; Lawson and Robins 2021), so for all analyses including more than two variables, it is vital that researchers first obtain sufficient conceptual clarification. DCT specifications facilitate this process.

Primary empirical quantitative research can be, for example, a cross-sectional survey to assess stress among students or an experimental study investigating the effect of various role model stories on self-efficacy. Regardless of study design, there should be a comprehensive and accurate definition of all psychological constructs measured in the study (e.g., stress, self-efficacy) as well as the way in which they are measured. This information is also valuable to include in the pre-registration of a study or the application form for ethical approval. DCT specifications concerning measurement instruments can be developed from scratch (e.g., when using measurement instruments specifically developed for the study), but can also be completed based on existing material. For example, when measuring self-efficacy, there might be an existing measurement instrument available. In fact, there are multiple measurement instruments available (Panc, Mihalcea, and Panc 2012; Romppel et al. 2013). In that case, it must be clear how the choice for a specific measurement is in line with the construct definition used in that specific study and how it adheres to instructions regarding measurement instruments for that construct. Over time, these DCT specifications can be re-used in other studies by the same researchers or others. This ensures consistency among studies when using the same DCT specification - and also makes explicit when different specifications are used for what appears to be the same construct (e.g., self-efficacy).

Using DCT specifications in primary empirical qualitative research

Primary empirical qualitative research can be, for example, an interview study about patients' beliefs concerning medication adherence. DCT specifications provide instructions regarding two important aspects of such studies: (1) how to elicit construct content and (2) how to code qualitative data as informative about construct content. The former can be used to develop study materials that are used during data collection (e.g., an interview guide), the latter can be used during data analysis (e.g., in the codebook). For example, the Reasoned Action Approach provides instructions on how to elicit construct content by means of a belief elicitation procedure (Fishbein and Ajzen 2010). For attitude beliefs, this is done by means of asking the following questions: 'What do you see as the advantages of you engaging in [target behavior]?', 'What do you see as the disadvantages of you engaging in [target behavior]?', and 'What else comes to mind when you think about [target behavior]?' These questions can be included in an interview guide that is used during data collection. In other words, this is helpful to elicit construct content, but not informative to guide decisions during data analyses.

Therefore, when using a DCT specification that would include those instructions for eliciting qualitative data, it should also be clear what to code as a belief *evaluation* and what as a belief *expectation*. And, in the case of expectation, whether it concerns an *experiental* attitude belief expectation or an *instrumental* one. In this example, the experiential nature of expectations means that these expected consequences must concern experiences and sensations, such as pleasure or pain. These experiential attitude belief expectations cover acute hedonic expectations, disconnected from potential long-term consequences the target behavior may have. Experiential attitude belief expectations always refer to immediately experienced consequences of a behavior. Expressions of expectations of consequences that render the target behavior less or more desirable without the immediate experiential effects of the behavior playing a role are captured in instrumental attitude belief expectations. Similar to empirical quantitative research, sometimes DCT specifications need to be developed from scratch, but there might also already be instructions available for certain aspects (such as the belief elicitation procedure in this example). For both aspects, elicitation and coding, instructions need to be in line with the construct definition.

Not all constructs can be studied qualitatively; and for some constructs that can be studied qualitatively, there is no way for researchers to elicit qualitative data that lends itself to studying the construct. In such cases, researchers can explain this in the relevant instruction, and for example only describe how qualitative data can be coded as informative about a given construct.

Using DCT specifications in literature reviews

DCT specifications also facilitate literature reviews (e.g., scoping reviews, meta-analyses, and systematic reviews without meta-analysis). The instructions for coding measurement instruments are directly linked to these research methods: researchers or students engaging in a literature review can use these in their instructions for data extraction from included primary sources. This will ensure that all univariate and multivariate results that are extracted pertain to the constructs as defined by the research team. That will in turn manifest as lower heterogeneity among effect sizes, and so in a higher probability that meta-analysis can be justified. Another way in which DCT specifications can facilitate literature review is analogous to the way they can be used in primary empirical qualitative research. We have provided the example of coding attitude beliefs based on data obtained by means of an interview study. These instructions on how to code qualitative data as informative about construct content can also be used to code data from multiple studies identified in a literature review. Such a review might be useful to create a higher level overview of beliefs and might also facilitate comparison between, for example, different subgroups.

For example, fear of falling was initially only used to describe the fear that older people could develop after falling (Kruisbrink 2022). The Falls Efficacy Scale (FES) was one of the first measurement instruments specifically developed for the construct "fear of falling." The FES conceptualizes fear of falling as low perceived self-efficacy at avoiding falls during essential non-hazardous activities of daily living. In other words, self-efficacy has been added to the construct of fear of falling. After this, measurement instruments for worry about falling and balance confidence were developed (e.g., The Survey of Activities and Fear of Falling in the Elderly and The Activities-specific Balance Confidence scale). All these measurement instruments and the constructs they intend to measure were put under the same umbrella and fear of falling became the umbrella term. This example clarifies the need for clear instructions on which measurement instruments to code when conducting a literature review on fear of falling. This does not mean that a review needs to be limited to a specific measurement instrument. In fact, it might even be of interest to compare data obtained by means of various measurement instruments. In that case, it is even more important to be explicit about the measurement instruments of interest and have DTC specifications available for all of them.

Using DCT specifications in theory development

DCT specifications are also valuable in the process of theory development. This process does not happen overnight and requires epistemic iteration. In *Inventing Temperature*, Chang (2007) examines how scientists created thermometers and how they managed to assess the reliability and accuracy of these instruments without a circular reliance on the instruments themselves. This process of iteratively improving construct definitions and measurement instruments took centuries of work by many scientists — for a construct which in hindsight seems straightforward to measure — and relied heavily on theory development during that process. This intertwined development of measurement instruments and theory manifested in a variety of different measurement instruments and corresponding theories having been in use simultaneously for decades, until enough evidence accumulated to allow abandoning some instruments — and theories.

If we consider 1879, the year that Wilhelm Wundt founded the first laboratory of psychology, the birth year of psychology as a science, then it is no wonder that research concerning many psychological constructs is still in its early stages in terms of construct definitions and measurement instruments. However, in the case of psychology, it appears the way measurement is approached (e.g., heavily relying on construct validity and quantitative psychometric methods) is, as Alexandrova and Haybron (2016) put it, theory-avoidant. This effectively stifles the potential for epistemic iteration, stagnating further theory development. Therefore, psychology is simultaneously in the situation that it finds itself in its infancy as a science in terms of theoretical maturity and the corresponding maturity of its measurement instruments; and where the customary approach to measurement seems to stunt the potential for further maturation. Specifying DCTs, as an explicit instrument to share and discuss construct definitions and corresponding measurement implications, may help to break out of this stalemate.

For example, analogous to the multiphase approach for theory construction in the psychopathology domain specifically (Hawkins-Elder and Ward 2020), for most psychological constructs in general we are in the second phase. In this phase, the key aim is to first develop a rich compositional description of the construct of interest. DCT specifications help to document these descriptions and their development over time and foster discussion about these developments within and across research teams. In this process, it might be that certain aspects of a DCT specification cannot be completed yet. For example, because it is too soon to provide instructions about coding qualitative data because there is no definition yet or an attempt to develop a measurement instrument failed. It is important to document this as well, because in the former case this might be informative to others in terms of what to prioritize, while in the latter case this contributes to learning by means of trial and error.

Using DCT specifications in practice

DCT specifications are also valuable in terms of knowledge translation from research to practice. Professionals working in practical settings often have to make decisions concerning psychological constructs. For example, a prevention worker that develops a behavior change intervention needs to know which determinants to target in an intervention. These determinants often concern psychological constructs such as attitude or self-efficacy. Assessment of employees in an organization, as another example, also often concerns psychological constructs, such as flexibility or decisiveness. DCT specifications can help professionals to apply the latest insights regarding a psychological construct that are often not described in detail in scientific publications. For example, repositories with construct definitions and corresponding instructions can help a prevention worker set up a survey for their priority population, or improve assessment of employees by having instructions on which questions to ask in a job interview to gain insight into specific psychological constructs.

Using DCT specifications in large-scale collaborations

In recent years, large-scale collaborations have been set up to progress psychological science (e.g., Open Science Collaboration 2015). DCT specifications can be helpful in such large-collaborations in three ways. First, they can document heterogeneity in construct definitions and the corresponding measurements and investigations using qualitative methods. Landy et al. (2020) showed that research teams studying the same hypothesis independently make many different design decisions, partly based on hidden heterogeneity in their elaborated construct definitions and implications for measurement. Letting each team use DCT specifications to document these elaborations allows studying how these differences manifest in different study outcomes. This approach can be used to provide insight into the amount of heterogeneity that exists "in the wild" for a given construct or research domain.

Second, instead of transparently documenting existing heterogeneity, DCT specifications can also be used to reduce heterogeneity within a large-scale collaboration. If all research teams use the same elaborated construct definitions and measurement instructions (for example developing those collaboratively prior to engaging in the design of the studies and data collection), the degrees of freedom in design choices are decreased in a theory-driven manner.

Third, DCT specifications can also be useful to systematically explore the effects of differences between construct definitions. In this approach, DCT specifications provide the infrastructure required to conduct conceptual multiverse studies as recommended by Bringmann et al. (2022). Instead of all research teams creating their own DCT specification or using the same DCT specification, sets of parallel DCT specifications are developed for the same construct in advance. This allows systematic investigation of the effects of the differences in construct definitions and measurement instructions or instructions for qualitative research.

The technological implementation of DCTs

The development of DCTs involved developing two types of tools: conceptual tooling (logical structures or procedures that aid certain intellectual or epistemic tasks, e.g., Theory Construction Methodology, Borsboom et al. 2021) and technological tooling [the implementation of such procedures in, often digital, infrastructure; e.g., Holcombe et al. (2020)]. The latter facilitates the former, but also requires a number of additional decisions that, in the case of scientific endeavors, have their own epistemic consequences (also see Es 2023). The conceptual tooling consists of the structure of a DCT specification (i.e. the construct label, identifiers, and four sets of instructions). As technological tooling to implement this conceptual tooling, we have developed five open source products: a machine- and human-readable standard to store DCT specifications in files; a way to generate Unique Construct Identifiers (UCIDs); the `{psyverse}` R package to work with DCTs; a psychological construct repository; and a Shiny app to import, edit, and submit DCT specifications to a psychological construct repository.

In developing these technological implementations, we aimed to keep the infrastructure both open (i.e. open source, and so consistent with open science practices) and lightweight, so as to make it as easy as possible for anybody to set up their own repository, adapt the Shiny app or R package, or develop additional tooling. We aimed to make these implementations practically useful to as many researchers as possible (regardless of their philosophical (of science), theoretical, or methodological perspective). Related, we designed them to be easy (and free or cheap) to create, share, and update. Furthermore, we designed the technical implementation of DCTs with open standards to facilitate findability, accessibility, interoperability, and re-usability.

To explain why we prioritized this high level of openness and inclusiveness, contrast DCTs with the use of ontologies that are gaining popularity Spadaro et al. (2022). Such ontologies (i.e. uniformized and interoperable standards to describe concepts and their relationships) are powerful and flexible tools for data modeling, and as such, for formalizing knowledge. However, this power and flexibility comes at a literal price: developing and updating ontologies requires specialized training which psychological researchers often lack, excluding researchers without such training or without sufficient funding from participating in epistemic efforts scaffolded by ontologies. This effectively centralizes epistemological power around more privileged institutions and researchers, which is both not straightforward to reconcile with open science principles (UNESCO 2021), and epistemologically inefficient (Zollman 2010; Devezer et al. 2019; Cartwright 2021). Instead of designing DCTs around a central authority on which use would then rely, we designed this implementation to allow anybody to set up the required infrastructure. We mean this to do justice to the UNESCO Open Science recommendation to build community-owned, open infrastructure for scientific endeavors (UNESCO 2021). However, just like epistemological and methodological diversity benefits science, we believe the same to be true for organizational and operational diversity. The technological infrastructure that implements DCTs is also easy to implement in a centralized manner: for example, a learned society or a journal can copy and adapt the PsyCoRe.one psychological construct repository and impose whichever procedures and policies they deem beneficial to their aims. This still allows others to implement alternative instances, similar to how decentralized social media such as Mastodon operate (and which has been proposed as a superior model for publishing scientific results: Brembs et al. 2023).

The open source technical tools we developed are described in more detail in Appendix 2, “Technical tools to work with Decentralized Construct Taxonomies”. We hope these will enable individual researchers, labs, institutions, and societies to create their own repositories with construct definitions to achieve unequivocal communication. With this design for decentralization we aim to as much as possible prevent the reliance of researchers on a small number of curated databases of construct definitions (which would again impede epistemic iteration in favor of sub-optimal local optima; Moreau and Wiebels 2022).

Discussion

A strength of the concept of Decentralized Construct Taxonomy (DCT) specifications is that the value of explication is mostly orthogonal to one’s ontological, epistemological, and methodological perspectives (excluding e.g., pure operationalism or idealism). Clearly describing what you study in a way that lends itself well to re-use or improvement by other researchers (or your future self) is useful regardless of whether one adopts a hardcore relativist or a moderate realist stance towards psychological constructs, or whether one believes that psychological processes are interaction-dominant or component-dominant; whether one has a pragmatic, constructivist, or positivist epistemological stance; and whether one finds oneself primarily using complex systems approaches, thematic analysis, clustered randomized trials, Mokken scaling, implicit association tasks, grounded theory approaches, or experience sampling methods.[2]

This design for ontological, epistemological, and methodological orthogonality facilitates epistemic iteration by enabling a single DCT specification to be invoked in an epistemically diverse set of studies. It has been argued that mono-method research is the biggest threat to the advancement of the social sciences (Onwuegbuzie and Leech 2005), and by attempting to remain untethered to a specific ontological, epistemological or methodological approach, DCT specifications can facilitate multidisciplinary collaboration.

[2] Only if one’s conviction is that all concepts one studies are entirely ephemeral and do not persist over time, striving for unequivocality and minimal hidden heterogeneity has no benefit, since then there is no pretense that the results of a study have any value in understanding the world.

Despite that generic nature and associated versatility, the implementation of DCT specifications as we present them here entailed a number of decisions. In these decisions, we aimed to approach a sweet spot in terms of trade-offs of simplicity and ease of use, required IT infrastructure and costs, and versatility and flexibility. However, their central functions (enabling unique reference to construct definitions without central curation; attaching specific additional instructions; and easily re-using or changing existing specifications) can be fulfilled in myriad other ways. For the time being, we believe that consistent use of DCTs would be a good first step to achieve decentralized unequivocality, but we hope DCTs will be superseded by better technology. For example, in its current technical implementation, the Unique Construct Identifiers (UCIDs) generated by the `{psyverse}` package are not absolutely guaranteed to be unique. In addition, they are not centrally registered. A community-owned not-for-profit organization could provide a central register where UCIDs could be "minted", similar to how Crossref manages Digital Object Identifiers (DOIs).

Existing infrastructure can also be leveraged to obtain these benefits. When posting a DCT specification in a psychological construct repository (such as PsyCoRe.one) researchers can publish the DCT specification and its provenance in a journal, which couples the construct's UCID to a DOI. Other services, such as Zenodo, enable researchers to deposit academic resources (such as construct definitions) and offer version management, with separate DOIs for each subsequent version as well as one canonical DOI that always resolves to the latest version. This enables researchers to curate a series of construct definition iterations, where each iteration has its own unique UCID and DOI, but also allows easy reference to the entire lineage of construct definitions with a single DOI (see e.g., Panasiuk 2023). This is one example of how DCTs can already be combined with existing infrastructure to augment its functionality.

A risk of using DCT specifications is that while they were designed to enable easy iteration, they carry the risk of researchers simply reusing a specific set without reflection on its quality and using the methods described above to improve the set. Although this would partly ameliorate the amount of hidden heterogeneity, epistemic iteration would stagnate. On the other hand, DCT specifications do lend themselves well for achieving explicit homogeneity in the used construct definitions within a lab or department, or large-scale collaborations, so that all studies conducted there will be consistent in terms of how the investigated constructs are defined, measured, and coded. To try to minimize this risk, we designed DCTs to be easy to implement in a decentralized manner (by only requiring open source software and infrastructure that is ubiquitous and often free). If for example a large learned society (e.g., the American Psychological Association) would set up and maintain the only construct repository, this would likely nudge researchers towards using the construct definitions stored there, especially if that society would decide to moderate or curate the definitions in their repository. This would centralize and likely constrain which epistemological considerations underlie the accepted construct definitions, limiting epistemic iteration and bestowing considerable epistemological power on that society.

Note that this design for decentralized (and easy and free) implementation does not *preclude* a landscape with a number of large construct repositories. We believe it would be beneficial if a plethora of construct repositories evolves. Because DCTs native format is machine-readable (i.e. YAML), this provides a straightforward starting point for harvesting the data and providing access through other interfaces. This would retain the DCTs' findability, thereby helping researchers to see the construct forest for the trees. However, because it's relatively easy for single researchers or small groups to set up their own repository, this safeguards against the tendency towards centralization that would result from a less lightweight infrastructure.

A related consideration is that the fields that make up a DCT specification (UCID, label, definition, and the four instructions) have been selected based on our experience of conducting research in health psychology and what would be needed in that domain. However, in other domains of psychology, other fields may be more relevant for a DCT specification, and some may not be useful. For example, in some domains, there may not be any constructs that lend themselves to study using qualitative methods. Given that different fields may be relevant to different domains, and that it will not be possible to establish a universally optimal set of fields, we would see a proliferation of different DCT-like standards as scientific progress. We do not consider this a discouraging prospect. The goal is to decrease hidden heterogeneity, and as long as unique identifiers are used and specifications can easily be re-used or changed by other researchers, that goal gets closer.

Similar to how DCT-like standards could proliferate, various procedures for establishing DCT specifications can also be developed. Different domains, societies, and labs may have different ideas about what, for example, quality control of DCT specifications could look like. A society such as the Society for Mathematical Psychology could maintain a repository where they require specification of formal models when describing instructions to develop measurement instruments (Van Der Maas et al. 2011), whereas the editors of a journal such as Qualitative Research in Psychology could require extensive instructions for coding qualitative data. Given the variety in ontological and epistemological stances taken by psychological researchers, the way different communities envision the path of epistemic iteration can vary, and DCTs as presently designed (i.e. lightweight) allow communities to devise their own procedures and additional infrastructure.

Another avenue for further development is the shift towards measurement. In their current form, DCT specifications facilitate both a more operationalist approach (by specifying in the instructions for developing measurement instruments that only one specific measurement instrument can be used to measure the target construct) and more epistemological flexibility (by specifying instructions to develop measurement instruments). A next step could be to develop another standard for decentralized operationalization specifications: a medium-independent format for specifying a measurement instrument. By attaching unique identifiers to such specifications and specifying the UCID they are designed to measure, it becomes possible to automatically trace columns in datasets to constructs through their decentralized operationalization specifications. This would eliminate the need for coding measurement instruments when conducting systematic reviews, facilitating living individual participant data meta-analyses and systematic reviews. Data management standards such as the PsychDS[3] standard can then integrate such functionality.

Taking this one step further, adding specifications of the measurement theory seems promising. For example, the hypothesized response processes could be described in a decentralized operationalization specification. Documenting these in a transparent manner would facilitate parallel epistemic iteration over variations of DCT specifications for a given construct by making auxiliary assumptions required for valid measurement explicit. This would yield at least some of the benefits described by Robinaugh et al. (2021) even when formal modeling is not possible or desirable. Essentially, this would provide a unified operational tool for implementations of approaches such as test validity (Borsboom et al. 2009) and theory-based measurement (Borgstede and Eggert 2022).

Adding UCIDs to dataset metadata has the potential to make data more findable, as it becomes possible to search for all data investigating a given construct using a given definition. When decentralized operationalization specifications are added as well, findability increases further, as searches for datasets collected using specific measurement instruments

[3] See the project's landing page at https://github.com/psych-ds/psych-DS

and items will become possible. In addition, datasets become truly interoperable, not only for humans, but also for machines. If each data series in a dataset (e.g., each column if the dataset is in wide format) corresponds to an item-level decentralized operationalization specification, and each decentralized operationalization specification lists the corresponding UCID, it becomes possible to partly automate dataset aggregation. Combining this with machine-readable hypothesis tests as described by Lakens and De Bruine (2021) will go a long way towards enabling automated living meta-analyses, all without requiring (but still being compatible with) curation by a central authority.

In some domains of psychology, it could be beneficial to direct this same meta-approach (i.e. creating means to enable decentralized unequivocal comprehensive specification) towards the description of processes and mechanisms. While West et al. (2019) provide accessible means to describe theoretical relationships, the definitions of these relationships are again brief, and also leave room for hidden heterogeneity. A similar tool could be developed for relationships. In fact, this apparent distinction between "nodes" and "edges" as core theory constituents is itself a choice.

Regardless of which elements one deems sensible to distinguish, a decentralized way to document extensive definitions, share those, and unequivocally refer to them can be a powerful method to decrease hidden heterogeneity. From this perspective, DCT specifications can be considered just one incarnation of a meta-approach that can be applied to other theoretical elements to provide a standard for specifying, updating, and unequivocally referring to theoretical terms: in other words, to make sure we know what we're talking about, which is arguably the least that society can expect from scientists.

COI, funding and author contributions

Conflict of Interest and Funding

The authors are not aware of any potential conflict or interest.

Funding

No funding was received to conduct this work.

Acknowledgements

We would like to thank editor Matt Williams and reviewers Jessica K. Flake and Sam Parsons, for very helpful comments.

Author Contributions

Conceptualization: G.-J.P. and R.C.; Methodology: G.-J.P. and R.C.; Software: G.-J.P.; Writing - original draft: G.-J.P. and R.C.; Writing - review & editing: G.-J.P. and R.C.

References

Alexandrova, A., and D. M. Haybron. 2016. "Is Construct Validation Valid?" *Philosophy of Science* 83 (5): 1098–1109. https://doi.org/gg33wb.

Altgassen, Elisa, Mattis Geiger, and Oliver Wilhelm. 2023. "Do You Mind a Closer Look? A Jingle-Jangle Fallacy Perspective on Mindfulness." *European Journal of Personality*, May, 089020702311745. https://doi.org/kcrc.

Borgstede, M., and F. Eggert. 2022. "Squaring the Circle: From Latent Variables to Theory-Based Measurement." *Theory & Psychology*. https://doi.org/grhv9p.

Borsboom, D., A. Cramer, R. A. Kievit, A. Z. Scholten, and S. Franić. 2009. "The End of Construct Validity." In *The Concept of Validity: Revisions, New Directions, and Applications*, edited by R. W. Lissitz, 135–70. IAP Information Age Publishing.

Borsboom, D., H. L. J. Van der Maas, J. Dalege, R. A. Kievit, and B. D. Haig. 2021. "Theory Construction Methodology: A Practical Framework for Building Theories in Psychology." *Perspectives on Psychological Science* 16 (4): 756–66. https://doi.org/gh65nr.

Box, G. E. P. 1979. "Robustness in the Strategy of Scientific Model Building." In *Robustness in Statistics*, edited by R. L. Launer and G. N. Wilkinson, 201–36. Academic Press, Inc. https://doi.org/gd32bv.

Brembs, Björn, Adrian Lenardic, Peter Murray-Rust, Leslie Chan, and Dasapta Erwin Irawan. 2023. "Mastodon over Mammon: Towards Publicly Owned Scholarly Knowledge." *Royal Society Open Science* 10 (7): 230207. https://doi.org/gstkx9.

Brick, C., B. Hood, V. Ekroll, and L. De-Wit. 2022. "Illusory Essences: A Bias Holding Back Theorizing in Psychological Science." *Perspectives on Psychological Science* 17 (2): 491–506. https://doi.org/gmdj5f.

Bringmann, L. F., T. Elmer, and M. I. Eronen. 2022. "Back to Basics: The Importance of Conceptual Clarification in Psychological Science." *Current Directions in Psychological Science* 31 (4): 340–46. https://doi.org/gqmqg5.

Bschir, Karim, and Simon Lohse. 2022. "Pandemics, Policy, and Pluralism: A Feyerabend-Inspired Perspective on COVID-19." *Synthese* 200 (6): 441. https://doi.org/gq5cz9.

Cartwright, N. 2021. "Rigour Versus the Need for Evidential Diversity." *Synthese* 199: 13095–119. https://doi.org/gstkzb.

Castanho Silva, B., S. Jungkunz, M. Helbling, and L. Littvay. 2020. "An Empirical Comparison of Seven Populist Attitudes Scales." *Political Research Quarterly* 73 (2): 409–24. https://doi.org/ghx8nm.

Chang, H. 2007. *Inventing Temperature: Measurement and Scientific Progress*. Oxford Studies in Philosophy of Science. Oxford University Press.

Crutzen, Rik, and Gjalt-Jorn Ygram Peters. 2023a. "The Regression Trap: Why Regression Analyses Are Not Suitable for Selecting Determinants to Target in Behavior Change Interventions." *Health Psychology and Behavioral Medicine* 11 (1). https://doi.org/10.1080/21642850.2023.2268684.

———. 2023b. "A Lean Method for Selecting Determinants When Developing Behavior Change Interventions." *Health Psychology and Behavioral Medicine* 11 (1): 2167719. https://doi.org/js9b.

De Graaf, A., B. Van den Putte, M.-J. Nguyen, S. Zebregs, J. Lammers, and P. Neijens. 2017. "The Effectiveness of Narrative Versus Informational Smoking Education on Smoking Beliefs, Attitudes and Intentions of Low-Educated Adolescents." *Psychology & Health* 32 (7): 810–25. https://doi.org/gstkzd.

Devezer, B., L. G. Nardin, B. Baumgaertner, and E. O. Buzbas. 2019. "Scientific Discovery in a Model-Centric Framework: Reproducibility, Innovation, and Epistemic Diversity." *PLOS ONE* 14 (5): e0216125. https://doi.org/gf86cs.

Diener, Ed, Robert A. Emmons, Randy J. Larsen, and Sharon Griffin. 1985. "The

Satisfaction With Life Scale." *Journal of Personality Assessment* 49 (1): 71–75. https://doi.org/fqqbmr.

Earp, B. D., and D. Trafimow. 2015. "Replication, Falsification, and the Crisis of Confidence in Social Psychology." *Frontiers in Psychology* 6: 621. https://doi.org/gfpn82.

Eronen, M. I., and L. F. Bringmann. 2021. "The Theory Crisis in Psychology: How to Move Forward." *Perspectives on Psychological Science* 16 (4): 779–88. https://doi.org/ghw2x3.

Eronen, M. I., and J.-W. Romeijn. 2020. "Philosophy of Science and the Formalization of Psychological Theory." *Theory & Psychology* 30 (6): 786–99. https://doi.org/ghqcr8.

Es, Karin van. 2023. "Unpacking Tool Criticism as Practice, in Practice." *Digital Humanities Quarterly* 017 (2). https://digitalhumanities.org/dhq/vol/17/2/000692/000692.html.

Feyerabend, P. 1965. "Problems of Empiricism." In *Beyond the Edge of Certainty. Essays in Contemporary Science and Philosophy.*, edited by R. G. Colodny, 145–260. Prentice-Hall.

Fishbein, M, and I Ajzen. 2010. *Predicting and Changing Behavior: The Reasoned Action Approach*. Taylor & Francis Group.

Flake, Jessica Kay, and Eiko I. Fried. 2020. "Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them." *Advances in Methods and Practices in Psychological Science* 3 (4): 456–65. https://doi.org/ghnbdg.

Flake, Jessica Kay, Jolynn Pek, and Eric Hehman. 2017. "Construct Validation in Social and Personality Research: Current Practice and Recommendations." *Social Psychological and Personality Science* 8 (4): 370–78. https://doi.org/gbf8nx.

Fried, E. I. 2017a. "The 52 Symptoms of Major Depression: Lack of Content Overlap Among Seven Common Depression Scales." *Journal of Affective Disorders* 208: 191–97. https://doi.org/gcsk3p.

———. 2017b. "What Are Psychological Constructs? On the Nature and Statistical Modeling of Emotions, Intelligence, Personality Traits and Mental Disorders." *Health Psychology Review* 11 (2): 130–34. https://doi.org/gfj8tz.

———. 2020. "Lack of Theory Building and Testing Impedes Progress in the Factor and Network Literature." *Psychological Inquiry* 31 (4): 271–88. https://doi.org/ghsbbk.

Fried, E. I., and R. M. Nesse. 2015. "Depression Sum-Scores Don't Add up: Why Analyzing Specific Depression Symptoms Is Essential." *BMC Medicine* 13 (1): 72. https://doi.org/f67sz8.

Gibbs, G. 2007. *Analyzing Qualitative Data*. Sage Publications Ltd.

Glaw, Xanthe, Kerry Inder, Ashley Kable, and Michael Hazelton. 2017. "Visual Methodologies in Qualitative Research: Autophotography and Photo Elicitation Applied to Mental Health Research." *International Journal of Qualitative Methods* 16 (1). https://doi.org/gfvmq8.

Gruijters, S. L. K., B. P. I. Fleuren, and G.-J.Y. Peters. 2021. "Crossing the Seven Cs of Internal Consistency: Assessing the Reliability of Formative Instruments." *PsyArXiv*. https://doi.org/g5sh.

Guest, O., and A. E. Martin. 2021. "How Computational Modeling Can Force Theory Building in Psychological Science." *Perspectives on Psychological Science* 16 (4): 789–802. https://doi.org/ghvcwr.

Hagger, M. S. 2014. "Avoiding the 'Déjà-Variable' Phenomenon: Social Psychology Needs More Guides to Constructs." *Frontiers in Psychology* 5: 52. https://doi.org/gd7fkx.

Hale, J., J. Hastings, R. West, C. E. Lefevre, A. Direito, L. C. Bohlen, C. Godinho, et al. 2020. "An Ontology-Based Modelling System (OBMS) for Representing Behaviour Change Theories Applied to 76 Theories." *Wellcome Open Research* 5: 177. https:

//doi.org/ghgnb5.

Hawkins-Elder, H., and T. Ward. 2020. "Theory Construction in the Psychopathology Domain: A Multiphase Approach." *Theory & Psychology* 30 (1): 77–98. https://doi.org/ghvmvg.

Hodson, G. 2021. "Construct Jangle or Construct Mangle? Thinking Straight about (Nonredundant) Psychological Constructs." *Journal of Theoretical Social Psychology* 5 (4): 576–90. https://doi.org/gng6pw.

Holcombe, Alex O., Marton Kovacs, Frederik Aust, and Balazs Aczel. 2020. "Documenting Contributions to Scholarly Articles Using CRediT and Tenzing." Edited by Cassidy R. Sugimoto. *PLOS ONE* 15 (12): e0244611. https://doi.org/gk347c.

Huls, S. P. I., S. M. C. Van Osch, W. B. F. Brouwer, J. Van Exel, and A. M. Stiggelbout. 2022. "Psychometric Evaluation of the Health-Risk Attitude Scale (HRAS-13): Assessing the Reliability, Dimensionality and Validity in the General Population and a Patient Population." *Psychology & Health* 37 (1): 34–50. https://doi.org/gstkzf.

Kok, G., and R. A. C. Ruiter. 2014. "Who Has the Authority to Change a Theory? Everyone! A Commentary on Head and Noar." *Health Psychology Review* 8 (1): 61–64. https://doi.org/gstkzg.

Kruisbrink, Marlot. 2022. "Towards Enhanced Management of Fear of Falling in Older People: Unravelling Interventions and Measuring Related Avoidance of Activity." PhD thesis, maastricht university. https://doi.org/kxkf.

Lakens, D. 2019. "The Value of Preregistration for Psychological Science: A Conceptual Analysis." *Japanese Psychological Review* 62 (3): 221–30. https://doi.org/gk9jpb.

Lakens, D., and L. M. DeBruine. 2021. "Improving Transparency, Falsifiability, and Rigor by Making Hypothesis Tests Machine-Readable." *Advances in Methods and Practices in Psychological Science* 4 (2): 251524592097094. https://doi.org/gmbq55.

Landy, J. F., M. Jia, I. L. Ding, D. Viganola, W. Tierney, A. Dreber, M. Johannesson, et al. 2020. "Crowdsourcing Hypothesis Tests: Making Transparent How Design Choices Shape Research Results." *Psychological Bulletin* 146 (5): 451–79. https://doi.org/ggj87k.

Lawson, K. M., and Richard W. Robins. 2021. "Sibling Constructs: What Are They, Why Do They Matter, and How Should You Handle Them?" *Personality and Social Psychology Review* 25 (4): 344–66. https://doi.org/gm62mr.

Leising, D., and M. Borgstede. 2019. "Hypothetical Constructs." In *Encyclopedia of Personality and Individual Differences*, edited by V. Zeigler-Hill and T. K. Shackelford, 1–6. Springer. https://doi.org/kxkb.

Lonsdorf, T. B., C. J. Merz, and M. A. Fullana. 2019. "Fear Extinction Retention: Is It What We Think It Is?" *Biological Psychiatry* 85 (12): 1074–82. https://doi.org/ggw7x6.

Marques, Marta M., Alison J. Wright, Elizabeth Corker, Marie Johnston, Robert West, Janna Hastings, Lisa Zhang, and Susan Michie. 2023. "The Behaviour Change Technique Ontology: Transforming the Behaviour Change Technique Taxonomy V1." *Wellcome Open Research* 8 (July): 308. https://doi.org/kxj9.

Maul, A. 2017. "Rethinking Traditional Methods of Survey Validation." *Measurement: Interdisciplinary Research and Perspectives* 15 (2): 51–69. https://doi.org/gf5stn.

Metz, G., G.-J.Y. Peters, and R. Crutzen. 2022. "Acyclic Behavior Change Diagrams: A Tool to Report and Analyze Interventions." *Health Psychology and Behavioral Medicine*. https://doi.org/hrz4.

Miyakawa, T. 2020. "No Raw Data, No Science: Another Possible Source of the Reproducibility Crisis." *Molecular Brain* 13: 24. https://doi.org/ggmgw6.

Moreau, D., and K. Wiebels. 2022. "Psychological Constructs as Local Optima." *Nature Reviews Psychology* 1 (4): 188–89. https://doi.org/grhkxq.

Murray-Rust, P. 2018. "Open Data in Science." *Nature Precedings*. https://doi.org/gnkn6x.

Nasa, P., R. Jain, and D. Juneja. 2021. "Delphi Methodology in Healthcare Research: How to Decide Its Appropriateness." *World Journal of Methodology* 11 (4): 116–29. https://doi.org/gmpjvz.

Nosek, B. A., and Y. Bar-Anan. 2012. "Scientific Utopia: I. Opening Scientific Communication." *Psychological Inquiry* 23 (3): 217–43. https://doi.org/gcsk27.

Nosek, B. A., C. R. Ebersole, A. C. DeHaven, and D. T. Mellor. 2018. "The Preregistration Revolution." *Proceedings of the National Academy of Sciences* 115 (11): 2600–2606. https://doi.org/gc6xk8.

Nosek, B. A., J. R. Spies, and M. Motyl. 2012. "Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth over Publishability." *Perspectives on Psychological Science* 7 (6): 615–31. https://doi.org/f4fc2k.

Oberauer, K., and S. Lewandowsky. 2019. "Addressing the Theory Crisis in Psychology." *Psychonomic Bulletin & Review* 26 (5): 1596–1618. https://doi.org/gf8dr6.

Onwuegbuzie, A. J., and N. L. Leech. 2005. "On Becoming a Pragmatic Researcher: The Importance of Combining Quantitative and Qualitative Research Methodologies." *International Journal of Social Research Methodology* 8 (5): 375–87. https://doi.org/bwmj6n.

Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349 (6251): aac4716. https://doi.org/68c.

Oude Maatman, F. 2021. "Psychology's Theory Crisis, and Why Formal Modelling Cannot Solve It." *PsyArXiv*. https://doi.org/jspz.

Panasiuk, Sofia. 2023. "Life Satisfaction: DCT Specification," August. https://doi.org/kqtb.

Panc, T., A. Mihalcea, and I. Panc. 2012. "Self-Efficacy Survey: A New Assessment Tool." *Procedia - Social and Behavioral Sciences* 33: 880–84. https://doi.org/gk6fwr.

Pashler, H., and E.-J. Wagenmakers. 2012. "Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?" *Perspectives on Psychological Science* 7 (6): 528–30. https://doi.org/gckf56.

Peters, Gjalt-Jorn Ygram, Rik Crutzen, Sylvia Roozen, and Gerjo Kok. 2020. "The Reasoned Action Approach Represented as a Decentralized Construct Taxonomy (DCT)." https://a-bc.gitlab.io/dct-raa/.

Podsakoff, Philip M., Scott B. MacKenzie, and Nathan P. Podsakoff. 2016. "Recommendations for Creating Better Concept Definitions in the Organizational, Behavioral, and Social Sciences." *Organizational Research Methods* 19 (2): 159–203. https://doi.org/f8fqdb.

Ponnock, A., K. Muenks, M. Morell, J. Seung Yang, J. R. Gladstone, and A. Wigfield. 2020. "Grit and Conscientiousness: Another Jangle Fallacy." *Journal of Research in Personality* 89: 104021. https://doi.org/ghcrsz.

Reschly, A. L., and S. L. Christenson. 2012. "Jingle, Jangle, and Conceptual Haziness: Evolution and Future Directions of the Engagement Construct." In *Handbook of Research on Student Engagement*, edited by S. L. Christenson, A. L. Reschly, and C. Wylie, 3–19. Springer. https://doi.org/gmxbks.

Robinaugh, D. J., J. M. B. Haslbeck, O. Ryan, E. I. Fried, and L. J. Waldorp. 2021. "Invisible Hands and Fine Calipers: A Call to Use Formal Theory as a Toolkit for Theory Construction." *Perspectives on Psychological Science* 16 (4): 725–43. https://doi.org/gh65nn.

Rohrer, J. M., W. Tierney, E. L. Uhlmann, L. M. DeBruine, Tom Heyman, B. Jones, S. C. Schmukle, et al. 2021. "Putting the Self in Self-Correction: Findings from the Loss-of-Confidence Project." *Perspectives on Psychological Science* 16 (6): 1255–69.

https://doi.org/gh6f6r.
Romppel, M., C. Herrmann-Lingen, R. Wachter, F. Edelmann, H.-D. Düngen, B. Pieske, and G. Grande. 2013. "A Short Form of the General Self-Efficacy Scale (GSE-6): Development, Psychometric Properties and Validity in an Intercultural Non-Clinical Sample and a Sample of Patients at Risk for Heart Failure." *GMS Psycho-Social-Medicine* 10: Doc01. https://doi.org/ggbbw8.
Santor, D. A., M. Gregus, and A. Welch. 2006. "Eight Decades of Measurement in Depression." *Measurement: Interdisciplinary Research and Perspectives* 4: 135–55. https://doi.org/fcmm9d.
Scheel, Anne M. 2022. "Why Most Psychological Research Findings Are Not Even Wrong." *Infant and Child Development* 31 (1). https://doi.org/gn8sks.
Schnell, T., D. Spitzenstätter, and H. Krampe. 2022. "Compliance with COVID-19 Public Health Guidelines: An Attitude-Behaviour Gap Bridged by Personal Concern and Distance to Conspiracy Ideation." *Psychology & Health* 37 (12): 1680–1701. https://doi.org/gm35dt.
Siegling, A. B., and K. V. Petrides. 2016. "Zeroing in on Mindfulness Facets: Similarities, Validity, and Dimensionality Across Three Independent Measures." *PLOS ONE* 11 (4): e0153073. https://doi.org/f8w3hz.
Skinner, E. A. 1996. "A Guide to Constructs of Control." *Journal of Personality and Social Psychology* 71 (3): 549–70. https://doi.org/d9q7cv.
Smaldino, P. E. 2008. "Models Are Stupid, and We Need More of Them." In *Computational Social Psychology*, 311–31. Routledge.
———. 2016. "Not Even Wrong: Imprecision Perpetuates the Illusion of Understanding at the Cost of Actual Understanding." *Behavioral and Brain Sciences* 39: e163. https://doi.org/gjfx4j.
Snippe, M. H. M., G.-J.Y. Peters, and G. Kok. 2021. "The Operationalization of Self-Identity in Reasoned Action Models: A Systematic Review of Self-Identity Operationalizations in Three Decades of Research." *Health Psychology and Behavioral Medicine* 9 (1): 48–69. https://doi.org/gmzj98.
Snippe, Marwin H. M., Gjalt - Jorn Ygram Peters, and Gerjo Kok. 2023. "Determining the Conceptual Independence of Self-Identity in Reasoned Action Models," June. https://doi.org/kdwx.
Spadaro, Giuliana, Ilaria Tiddi, Simon Columbus, Shuxian Jin, Annette Ten Teije, CoDa Team, and Daniel Balliet. 2022. "The Cooperation Databank: Machine-Readable Science Accelerates Research Synthesis." *Perspectives on Psychological Science* 17 (5): 1472–89. https://doi.org/gstkzh.
Szollosi, A., D. Kellen, D. J. Navarro, R. Shiffrin, I. Van Rooij, T. Van Zandt, and C. Donkin. 2020. "Is Preregistration Worthwhile?" *Trends in Cognitive Sciences* 24 (2): 94–95. https://doi.org/ggf3pd.
Taylor, J., E. Nailer, C. R. Cohen, C. W. E. Redman, and S. M. Sherman. 2022. "HPV Vaccination and Cervical Screening: The Knowledge and Attitudes of Mothers of Adolescent Girls." *Psychology & Health*. https://doi.org/gstkzj.
Uhlmann, Eric Luis, C. R. Ebersole, C. R. Chartier, T. M. Errington, M. C. Kidwell, C. K. Lai, R. J. McCarthy, A. Riegelman, R. Silberzahn, and B. A. Nosek. 2019. "Scientific Utopia III: Crowdsourcing Science." *Perspectives on Psychological Science* 14 (5): 711–33. https://doi.org/gf4jhq.
UNESCO. 2021. "UNESCO Recommendation on Open Science." UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000379949.
Van Der Maas, Han L. J., Dylan Molenaar, Gunter Maris, Rogier A. Kievit, and Denny Borsboom. 2011. "Cognitive Psychology Meets Psychometric Theory: On the Relation Between Process Models for Decision Making and Latent Variable Models for Individid-

ual Differences." *Psychological Review* 118 (2): 339–56. https://doi.org/drf5nw.
Van Rooij, I., and G. Baggio. 2020. "Theory Development Requires an Epistemological Sea Change." *Psychological Inquiry* 31 (4): 321–25. https://doi.org/gjzbzq.
Van Tuijl, Piet, Peter Verboon, and Jacques Van Lankveld. 2023a. "Initial Development and Validation of Item Banks to Measure Problematic Hypersexuality." *Open Research Europe* 3 (August): 129. https://doi.org/gstkzk.
Van Tuijl, Piet, Peter Verboon, and Jacques J. D. M. Van Lankveld. 2023b. "Three Quarks for Hypersexuality Research." *Sexes* 4 (1): 118–32. https://doi.org/kqjb.
Vazire, S. 2018. "Implications of the Credibility Revolution for Productivity, Creativity, and Progress." *Perspectives on Psychological Science* 13 (4): 411–17. https://doi.org/gd59dt.
Vazire, S., S. R. Schiavone, and J. G. Bottesini. 2022. "Credibility Beyond Replicability: Improving the Four Validities in Psychological Science." *Current Directions in Psychological Science* 31 (2): 162–68. https://doi.org/grht5z.
Warnell, K. R., and E. Redcay. 2019. "Minimal Coherence Among Varied Theory of Mind Measures in Childhood and Adulthood." *Cognition* 191: 103997. https://doi.org/gf3787.
Weidman, A. C., C. M. Steckler, and J. L. Tracy. 2017. "The Jingle and Jangle of Emotion Assessment: Imprecise Measurement, Casual Scale Usage, and Conceptual Fuzziness in Emotion Research." *Emotion* 17 (2): 267–95. https://doi.org/f9w6ff.
West, R. 2021. "PAT: An on-line Paper Authoring Tool for Writing up Randomized Controlled Trials." *Addiction* 116 (8): 1938–40. https://doi.org/gstkzm.
West, R., C. A. Godinho, L. C. Bohlen, R. N. Carey, J. Hastings, C. E. Lefevre, and S. Michie. 2019. "Development of a Formal System for Representing Behaviour-Change Theories." *Nature Human Behaviour* 3 (5): 526. https://doi.org/gf5fmd.
Williams, D. M., and R. E. Rhodes. 2016. "The Confounded Self-Efficacy Construct: Review, Conceptual Analysis, and Recommendations for Future Research." *Health Psychology Review* 10 (2): 113–28. https://doi.org/ggzktz.
Wolf, Melissa Gordon, Elliott Daniel Ihm, Andrew Maul, and Ann Taves. 2023. "The Response Process Evaluation Method." Preprint. PsyArXiv. https://doi.org/rbd2x.
Wood, C., M. Conner, T. Sandberg, G. Godin, and P. Sheeran. 2014. "Why Does Asking Questions Change Health Behaviours? The Mediating Role of Attitude Accessibility." *Psychology & Health* 29 (4): 390–404. https://doi.org/gf83n2.
Yarkoni, T. 2020. "Implicit Realism Impedes Progress in Psychology: Comment on Fried (2020)." *Psychological Inquiry* 31 (4): 326–33. https://doi.org/ghzsjk.
Zollman, K. J. S. 2010. "The Epistemic Benefit of Transient Diversity." *Erkenntnis* 72 (1): 17–35. https://doi.org/dnjk7f.

Appendix 1: Methods to develop Decentralized Construct Taxonomy specifications

There are several methods to develop these DCT specifications. These methods differ in terms of, for example, point of departure (e.g., theory, expertise) and number of sources. We introduce these different methods as well as their benefits and drawbacks. It is important to keep in mind that these methods are ways to elaborate construct definitions into comprehensive definitions with corresponding instructions — they are not ways to arrive at those constructs. To that end, we direct readers to procedures to develop theory (e.g., Hawkins-Elder and Ward 2020; Borsboom et al. 2021; Podsakoff, MacKenzie, and Podsakoff 2016).

Transparent descriptive approach

An initial step to move away from implicitly different construct definitions is to transparently describe constructs and the way in which they are used in a specific study. This transparency enables consumers of research (e.g., readers of an article) to understand how producers of research (e.g., authors of an article) defined and used the constructs in that study. An advantage is that this approach is very easily applicable — it is in line with the imperative for transparent reporting that is also stimulated by, for example, various reporting guidelines concerning what aspects to describe when reporting a specific study (e.g., CONSORT for randomized trials, COREQ for qualitative research). Moreover, it can in principle also be applied in hindsight after initial reporting of a study. On the one hand, this approach can be seen as an initial step towards being more explicit, even though the other methods described below are expected to result in more rigorous elaboration of construct definitions. On the other hand, the enhanced transparency achieved by simply making explicit what would otherwise remain implicit makes even this most straightforward approach an important step forward.

Single theory approach

The single theory approach consists of studying existing documentation about a specific theory. The primary source is often the work of the original author(s) of the theory. A benefit is that this approach requires limited resources. Moreover, this approach is probably easily applicable, because in theory, a theory should also specify how to develop a measurement instrument (and perhaps also how to conduct qualitative research using that theory). So, there is a more direct link between the construct definitions and instructions on how to use them in research. Over time, DCT specifications based on a single theory can also be improved by the original author(s) — or other researchers using this specific theory (Kok and Ruiter 2014). A downside is that in practice, many theories have very brief construct definitions, let alone detailed instructions on how to use them in research. This may force researchers themselves to “fill in the blanks” that exist in documentation about a specific theory. And, more in general, relying on a single theory might result in a narrow view on the psychological construct of interest. Still, even if the resulting DCT specifications end up just describing how a team of researchers decided to “fill in blanks” in the construct definitions in the original theory formulation, that would provide valuable context to other researchers.

Multiple theories approach

The multiple theories approach consists of conducting a systematic review of primary sources of various theories. This results in a comprehensive overview of theoretical contributions. For such an overview to be comprehensive, the search strategy should not be focused on specific theories, but on the psychological construct of interest. A benefit similar to the single theory approach is the direct link between the construct definitions and instructions on how to use them in research (if they are specified). The resources required are relatively high, given the need to conduct a systematic review. This investment likely pays off through obtaining various construct definitions and instructions on how to use them in research. The researchers can then choose to integrate these into one comprehensive DCT specification, or alternatively, to develop multiple DCT specifications. This choice between producing one or multiple DCT specifications is a methodological consideration that does not only hold for this specific method to develop DCT specifications.

Meta-empirical approach

Where the single theory approach and the multiple theories approach can be considered deductive, the meta-empirical approach is instead inductive. It consists of conducting a scoping review of empirical contributions in which psychological constructs have been used. This results in an overview of various (aspects of) definitions that are used in different empirical contributions as well as their operationalizations (e.g., questionnaires, items, tasks). It casts a wide net of how the psychological construct of interest is used. Similar to conducting a systematic review within the multiple theories approach, the resources required for conducting a scoping review are relatively high. However, a benefit is that an overview of measurement instruments and applications in qualitative research is obtained, which can help the researchers by grounding the four sets of instructions in the existing literature. For an example of this approach, see Snippe et al. (M. H. M. Snippe, Peters, and Kok 2021), and for a subsequent application, see Snippe et al. (Marwin H. M. Snippe, Peters, and Kok 2023). Like the multiple theory approach, this approach still requires that choices need to be made about whether and how to integrate into one DCT specification or multiple DCT specifications.

Reverse engineering approach

It is not always feasible or necessary to conduct a systematic review to develop a DCT specification. Researchers may also already have a measurement instrument they typically use. In that case, they can derive a construct definition from that instrument by closely inspecting the items. In this process, the researchers explicate those aspects of the construct that they consider most relevant, as well as their interpretation. Such DCT specifications may start out relatively brief if the item content does not provide much footing. However, even in such scenarios, brief initial DCT specifications still provide a convenient tool to enhance the transparency about authors' elaborated construct definition as it guided their theorizing and study design, and it serves as a concrete starting point for further development.

Expert-based rationalist approach

While the previously discussed approaches are based on what is available in either theoretical or empirical literature, DCT specifications can also be developed based on one's own thinking about the psychological construct of interest. Reasoning by experts is then the point of departure, which is why this is labeled as a rationalist approach. This can be done by a single expert or a (mostly small) team of experts. This approach requires limited resources. An important benefit is, in contrast to relying on what is available in existing literature, the level of detail in the DCT specifications can be as extensive as needed. This enables a detailed specification of the direct link between the construct definitions and instructions on how to use them in research (which is often lacking in existing literature). Given that this approach relies on a limited number of experts, this might result in a narrow view on the psychological construct of interest. A more important downside is that estimates of individual experts about psychological constructs are prone to error (Crutzen and Peters 2023b).

Expert-based Delphi approach

The expert-based Delphi approach is similar to the expert-based rationalist approach, although it usually involves a larger number of experts. This approach is characterized by a systematic process used to enable collective input of a panel of experts. Usually this process is characterized by multiple iterative rounds in which input from previous rounds is fed back in an anonymous way (Nasa, Jain, and Juneja 2021). The benefit of this anonymity is that experts might be more inclined to share their thinking about the psychological construct of interest and feel less vulnerable. Furthermore, the Delphi approach is geared towards reaching consensus (based on pre-specified criteria), which might also contribute to broadly supported DCT specifications. A drawback of such consensus-based broad support is that it may carry the risk of stagnating epistemic iteration (though awareness of that risk may allow mitigating it). A second drawback is that optimizing for consensus does not simultaneously optimize for accuracy, comprehensiveness, validity, or conceptual clarity, while still imbuing the result with a certain authority that may suggest such merits. Therefore, it is important to bear in mind that this support on its own is not informative regarding the quality of the DTC specification.

Primary qualitative research approach

Besides the approaches based on available literature or expertise, also primary research approaches can be used to develop DCT specifications. The primary qualitative research approach consists of interviewing people about a preliminary conceptualization of a construct. The transcripts of these interviews are then coded — a process of identifying passages in the text, searching and identifying concepts and finding relations between them (Gibbs 2007). This allows for integration of lived experiences of people in DCT specifications. A downside, however, is that the preliminary conceptualization is based on an initial construct definition that serves as a starting point for interview questions. This might steer interviewees when elaborating on the psychological construct of interest. It is recommended, therefore, to document this initial construct definition in a DCT specification and use the results from primary qualitative research to develop a new DCT specification. This contributes to transparency in the process of developing construct definitions. This approach can be seen as an example of epistemic iteration, albeit on a very small scale.

Primary psychometric research approach

Another method relying on data from participants is the primary psychometric research approach. Imagine that a content analysis of 7 commonly used measurement instruments reveals 52 disparate depression symptoms (Fried 2017a). A next step could then be to collate all these items from the existing measurements and run an exploratory factor analysis to identify (and subsequently define) one or more psychological constructs. At first glance, this might sound like an appealing approach because of two reasons. First, many researchers are already familiar with the required statistical procedures, and collecting data from large numbers of participants may be taken to imply generalizable findings. Second, a focus on psychometrics can be seen as legitimizing psychology as a "hard" science (Alexandrova and Haybron 2016). However, this approach also suffers from a number of drawbacks. First, the set of items that is used already narrows down what could potentially be part of those definitions (like in the primary qualitative research approach). Second, as mentioned before, sets of items designed to not measure any coherent psychological construct can still yield factor loading patterns consistent with being valid measures (Maul 2017). For psychometric analyses to be informative, a strong theory is first required as well as a well-justified measurement model, and when still working towards adequate construct definitions, this may not yet be available.

Appendix 2: Technical tools to work with Decentralized Construct Taxonomies

We developed five open source technical tools to support working with DCT specifications. The first is a way to store and share DCT specifications using the YAML file format. The second is the `{psyverse}` R package that supports creating and reading such files. The third is the generation of Unique Construct Identifiers (UCIDs). The fourth is the Constructor Shiny App that provides a graphical user interface to `{psyverse}` to make it as easy as possible to create files with DCT specifications. The fifth is an open source and open access repository that hosts DCT files; organizes them in a number of overviews as well as in a searchable table; and produces unique UCID-based URLs for a given DCT specification. We will briefly discuss each technical tool here.

The DCT YAML standard

It is important that published construct definitions are simultaneously human- and machine-readable. This imperative is included in the UNESCO Open Science recommendation (UNESCO 2021) and put into practice in a rapidly growing number of recent innovations (West 2021; Lakens and DeBruine 2021; Metz, Peters, and Crutzen 2022). In order to achieve this machine-readability, a data serialization file format needs to be chosen, and we have chosen YAML (a recursive acronym that stands for "YAML Ain't Markup Language"). This format is designed to be simultaneously human-readable and machine-readable. An additional advantage is that YAML is also a superset of the JavaScript Object Notation (JSON) standard, which means that DCTs can also be specified in JSON if researchers prefer (but note that YAML is more human-readable).

A DCT specification has to be stored with the key "`dct`", the value of which should be a list containing a set of values with keys "`version`" (containing one value: the version of the DCT specification format), "`id`" (containing the UCID), "`label`" (containing the human-readable label for the construct, which does not have to be unique), "`date`" (containing the date the DCT specification was created), as well as lists with keys "`definition`", "`measure_dev`", "`measure_code`", "`aspect_dev`", and "`aspect_code`". The "`definition`" list contains at least one key-value pair with key "`definition`" and as value the construct definition. The last four lists contain at least one key-value pair with key "`instruction`" and as value the instructions for, respectively, developing measurement instruments, coding measurement instruments, eliciting qualitative data, and coding qualitative data.

There are also a number of optional additional key-value pairs. For example, as metadata, directly within the "`dct`" list, "`dct_version`" can be specified, which holds the version of the DCT specification. Given the aim to establish a practice of epistemic iteration, being able to refer to a version of a construct definition can be more convenient than using the UCID or date. It is also possible to specify one or more sources the DCT specification was based on, as well as relationships to other DCT specifications. There is also an optional "`comments`" field which can be used to specify, for example, the authors, additional metadata, or link to repositories or web pages with more information. Finally, there are two fields that can be used to automatically retire older versions of DCT specifications (“`retires`”) and to document their lineage (“`ancestry`”). The “`retires`” field instructs repositories to no longer show DCTs with the specified UCIDs. This can be used if a DCT specification supersedes older version. The “`ancestry`” field can be used to document which DCT specifications a given DCT specification was based on, allowing a repository to show constructs’ “family trees”.

The `{psyverse}` R package

We implemented functionality to create and read DCT specifications in the `{psyverse}` R package. Notable functions in this package are “`psyverse::dct_object()`” (to directly create a DCT specification object), “psyverse::dct_object_to_html()” (to show a human-readable HTML representation of the DCT specification), “`psyverse::save_to_yaml()`” (to save a DCT specification to a file), “`psyverse::dct_from_gs()`” and “`psyverse::dct_from_xlsx()`” (to create a DCT from a Google sheet or Office Open XML spreadsheet, respectively, and optionally save them as YAML), and “`psyverse::generate_id()`” (to generate a UCID; see the following section).

When importing a DCT specification from a spreadsheet format, the spreadsheet must contain one or more worksheets where each worksheet contains at least two columns labeled “`field`” and “`content`”. Each subsequent row contains one key-value pair, and at least the following fields must be specified: “`label`” (with as content, the human-readable label), “`definition`” (with as content, the construct definition), “`measure_dev`” (with as content, the instruction for developing measurement instruments), “`measure_code`” (with as content, the instruction for coding measurement instruments), “`aspect_dev`” (with as content, the instruction for eliciting qualitative data), and “`aspect_code`” (with as content, the instruction for coding qualitative data).

Optionally, “`prefix`” and “`id`” can be specified. If content is specified for field “`id`”, that is used as the UCID. If content is specified for field “`prefix`”, a quasi-unique string is appended to create a UCID. If both “`prefix`” and “`id`” are specified, the prefix must be exactly the prefix as used in the identifier. The last optional field is the “`comments`” field.

Unique Construct Identifiers (UCIDs)

Unequivocal reference to constructs requires unique construct identifiers. However, without a central authority that implements procedures to prevent duplicate identifiers (as is the case for digital object identifiers (DOIs) and the identifiers used by the Open Science Framework (OSF)), this is not trivial to accomplish. Most decentralized implementations of universally unique identifiers combine local identifiers such as network adapter MAC addresses with the exact time, but have the drawback of being relatively long (unlike, for example, short DOIs and OSF identifiers).

In `{psyverse}`, we solve this problem by letting users specify a prefix and combining this with an already reasonably unique brief text string. That text string is the local time accurate to one hundredth of a second represented as the number of seconds since the UNIX epoch (1970-01-01 00:00:00 UT), multiplied by 100, and then converted to a 30-character number system to shorten it. The resulting identifier has two advantages over most universally unique identifiers: it is relatively brief and its prefix conveys human-readable information about the relevant construct.

These UCIDs can be used in publications to efficiently and unequivocally refer to a given DCT specification. In addition, they can be used in URLs to create URIs that can be resolved to the corresponding DCT specification. For example, the open source repository at https://psycore.one exposes the DCT specifications it contains at https://psycore.one/x, where x represents a UCID (e.g., https://psycore.one/instrAttitude_expectation_73dnt5z6).

The Constructor Shiny App

The Constructor Shiny App facilitates constructing DCT specifications. It is a web application that allows users to type in or copy-paste the different parts of a DCT specification and download the resulting YAML file. It is also possible to convert a DCT specification in a spreadsheet to the YAML format, or to upload an existing YAML file so it can be edited and downloaded again. Finally, it is possible to submit a DCT specification to the PsyCoRe.one repository, after which it becomes publicly available. The Constructor Shiny App is available at https://psycore.one/constructor.

An open source repository of construct definitions: PsyCoRe

The final tool we developed to support implementing DCTs in research workflows is an open source repository for psychological constructs: Psychological Construct Repository or PsyCoRe. It leverages the Hugo static site generator and its ability to integrate data from YAML files in a site to create a repository that can be hosted using GitLab Pages (or even closed source solutions such as GitHub Pages, for projects where applying open science principles is not desirable). All components are open source software, which means the repository can easily be cloned and adapted to the needs of different researchers, labs, institutions, and societies.

PsyCoRe also creates unique URLs using the UCID of each included DCT specification, allowing unequivocal reference to construct definitions. These URLs show the construct's UCID, label, definition, and the four sets of instructions. It also links to the DCT specification's YAML file on GitLab, allowing researchers to easily download it (and then, for example, adjust it and upload the adjusted version, which will then obtain a unique UCID). In addition to creating a webpage for each DCT specification, PsyCoRe also creates a number of overviews: a searchable table showing all constructs and which instructions are available for each as well as overviews with all definitions and an overview for each type of instruction.

Because both the Constructor Shiny App and the PsyCoRe repository are fully open source software, it is relatively easy for learned societies, institutions, labs, or individual researchers to set up their own repository and so provide both a user-friendly interface to add new DCT specifications and to browse existing specifications. Because the repository is a static site generator, it can be hosted on a number of free services, including an open source GitLab instance administered by the learned society, institution, lab, or individual researcher. However, a drawback of these design choices is that there is no backend running server-side code, which prohibits, for example, user management or verification. Implementing verification could be added to the current project architecture by adapting the Constructor Shiny App to require users to specify a username, e-mail address, and personal access token, but this would considerably decrease user friendliness. Hence, to lower barriers for initial use, we decided not to do this in the instance of PsyCoRe that we run.

We run an instance of PsyCoRe at https://psycore.one, and it is this instance to which the Constructor Shiny App submits new DCT specifications. This submission process consists of a pushed Git commit which triggers the execution of a number of commands using GitLab's continuous integration / continuous deployment functionality. GitLab Pages runs Hugo to rebuild the website. Hugo then integrates all data from the DCT specifications in the YAML files (including any newly committed files) in the overviews and construct pages. In our implementation, the Git commit also causes a webhook on another server to be called, which then pulls the repository and also runs Hugo. This server also includes an .htaccess file which is used to rewrite shortened UCID URLs (such as https://psycore.one/instrAttitude_expectation_73dnt5z6) to the URL created by Hugo (such as https://psycore.one/construct?ucid=instrAttitude_expectation_73dnt5z6). This instance of PsyCoRe does not yet support automatically retiring DCT specifications or showing DCT specification ancestry (see section "The DCT YAML standard" above).