Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Partial measurement limits the cumulative evidence for behavioural theories of climate adaptation

Sungju Han

sungju.han.1@uni-potsdam.de

University of Potsdam https://orcid.org/0000-0001-9386-8922

Samuel RUFAT

CY Cergy Paris University

Mariana de Brito

UFZ-Helmholtz Centre for Environmental Research https://orcid.org/0000-0003-4191-1647

Paul Hudson

University of York

Victor Santoni

Ecole polytechnique, Institut Polytechnique de Paris

Sofiia Skipor

Department of Psychology, University of Gothenburg

Lisa Köhler

Helmholtz Centre for Environmental Research - UFZ

Elsa Giffard

Institute for Sustainability Science and Technology, Universitat Politècnica de Catalunya

Emanuel Fusinato

Institute of Hydraulic Research, Universidade Federal do Rio Grande do Sul

Nuray Karanci

Department of Psychology, TOBB University of Economics and Technology

Christian Kuhlicke

Helmholtz Centre https://orcid.org/0000-0002-1193-228X

Analysis

Keywords:

Posted Date: August 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10821076/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Partial measurement limits the cumulative evidence for behavioural theories of climate adaptation

Sungju Han[1,2*], Samuel Rufat[3], Mariana Madruga de Brito[2,4], Paul Hudson[5], Victor Santoni[3], Sofiia Skipor[6], Lisa Köhler[2], Elsa Giffard[7,3], Emanuel Fusinato[8,2], Nuray Karanci[9], and Christian Kuhlicke[2,1]

[1] Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany
[2] Department of Urban and Environmental Sociology, Helmholtz Centre for Environmental Research—UFZ, Leipzig, Germany
[3] Ecole polytechnique, Institut Polytechnique de Paris, CREST, France
[4] Institute of Geography, Leipzig University, Leipzig, Germany
[5] Department of Environment and Geography, University of York, York, UK
[6] Department of Psychology, University of Gothenburg, Gothenburg, Sweden
[7] Institute for Sustainability Science and Technology, Universitat Politècnica de Catalunya, Spain
[8] Institute of Hydraulic Research, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[9] Department of Psychology, TOBB University of Economics and Technology, Ankara, Türkiye
[*] Corresponding author: sungju.han.1@uni-potsdam.de

Abstract

Individual and household protective actions such as retrofitting, evacuating, and buying insurance are central to climate-adaptation policy, yet how consistently the behavioural theories used to explain them are measured has gone unexamined. We reviewed 222 studies applying theory-guided surveys to risk perception and protective behaviour across natural hazards. Even within the two dominant theories most constructs go unmeasured: only 7% of Protection Motivation Theory studies and 4% of Protective Action Decision Model studies captured the full construct set, with easy-to-measure constructs favoured over policy-relevant ones like response cost and institutional trust. The best-documented theories are the least applied; only 29% of studies documented measures well enough to reuse or replicate. Because measurement is partial, much evidence may not test the theories it invokes and cannot be reliably pooled. We map where reusable instruments exist and where new ones are needed, a practical basis for building the toolbox this field lacks.

Introduction

Adapting to intensifying climate and natural hazards depends on more than infrastructure and institutional capacity: the protective actions individuals and households take (e.g., insuring, retrofitting homes, storing supplies, cooling homes, relocating) directly shape risk outcomes[1–3]. Behavioural measures are now treated in adaptation plans, policies and risk communication as complements to structural, planning and risk-transfer instruments[4,5], and identifying the factors that promote or impede protective action has become a central challenge for climate adaptation research and practice[5,6]. As climate change intensifies and shifts the distribution of these hazards[1], the demand for effective individual and household adaptation is rising, making the behavioural evidence base that informs it increasingly consequential.

The theoretical apparatus for this has long been available. For instance, the Protection Motivation Theory (PMT), the Protective Action Decision Model (PADM) and the Theory of Planned Behaviour (TPB) each specify how interconnected constructs, such as, risk perceptions, efficacy beliefs, social norms, response costs, jointly shape decisions, so that their explanatory value lies in the constructs operating together rather than in a checklist of predictors[7–9]. What has changed is the scale on which they are applied: Recent meta-analytic syntheses drawing on this accumulated work identify self-efficacy, response efficacy, descriptive norms and affective responses among the strongest correlates of adaptive behaviour, often outperforming knowledge and prior experience[9,10].

Yet this evidence is only as sound as the measurements beneath it. A systematic review of 4,432 vulnerability, resilience, and adaptation studies found that around 90% used no explicit theoretical framework[11]. That the majority skip theory altogether does not make partial measurement within the theory-guided minority a minor concern: even when a theory is named, studies typically measure some of its constructs while omitting others. It operates on two levels. First, which theory is applied and, second, which of that theory's constructs are actually measured, and how. Consequently, results built on the same theory are harder to compare than their shared vocabulary suggests[12,13].

This matters because meaningful synthesis depends on the comparability of the measures entering it. When primary studies attach the same label, such as 'self-efficacy', 'risk perception', to differently operationalised constructs, pooling their resulting estimates is correspondingly harder unless a synthesis explicitly defines and screens the constructs it combines[16–18]. Furthermore, as we show below, these omissions are not random: perceived severity and self-efficacy are typically captured with a few items, whereas response cost, institutional trust, situational constraints and maladaptive coping demand more elaborate measurement and are more often dropped, despite being the constructs that most directly explain why people who perceive a risk and feel able to act still do not. Where these constructs are measured inconsistently or not at all, the evidence reaching adaptation policy can overrepresent some pathways to action while obscuring others.

Here we assess how consistently and transparently behavioural theories are operationalised across natural hazards and climate adaptation research. We include both

climate-related and geophysical hazards by design: because instruments for risk perception and protective behaviour are largely hazard-general, measurement practice in. For example, earthquake research informs climate-hazard research directly, and a climate-only corpus would omit roughly a third of the relevant evidence without changing what the instruments look like (Methods). We quantify which constructs are measured, how measurement quality is reported, and whether instruments are documented well enough to reuse. We find that even in the studies that apply a theory, theories are often only partially measured; consequently, a large body of ostensibly theory-based evidence may not actually be testing the theories it invokes. A central obstacle to a cumulative adaptation evidence base is the inconsistent and largely undocumented measurement of widely used theories, which leaves studies invoking the same theory difficult to compare or combine. By mapping which constructs are measured, which are omitted, and how well instruments are documented, we provide the first systematic account of measurement practice in this field and translate it into a practical roadmap: which instruments can be reused now, which constructs and hazard contexts require new scales, and the reporting practices that would make future evidence comparable.

Results

A narrow, static evidence base: two theories, three hazards

We searched the Web of Science Core Collection using a Boolean query combining natural-hazard, adaptive-behaviour, population-level and survey/questionnaire terms, and screened the resulting records through a two-stage machine-learning-assisted and manual pipeline to identify empirical studies applying theory-guided survey instruments. We defined measurement documentation level (MDL) as a composite classification derived from three dimensions (i.e., survey instrument detail, psychometric quality, and measurement replicability), categorising each theory application as Well-Documented, Partially Documented, Minimally Documented or Not Documented. From 3,410 Web of Science records, our pipeline identified 222 empirical studies applying theory-guided survey instruments to risk perception or protective behaviour toward natural hazards, drawing on 57 distinct theoretical frameworks (Fig. 1a). We designate as *core theories* the ten frameworks that were applied in at least five studies; these account for 192 of the 222 studies (86%), while the remaining 47 frameworks appear too rarely for systematic comparison. Two dominate: Protection Motivation Theory (PMT; n = 86, 39%) and the Protective Action Decision Model (PADM; n = 59, 26%) together appear in 135 studies (61%), sometimes in combination (Fig. 1b); by adding the Theory of Planned Behaviour (TPB; n = 26, 12%) the share reaches 153 (69%). Such dominance may itself reflect the same operationalisation-ease logic that governs construct selection within a theory. The remaining seven core theories, the Psychometric Paradigm (11), Cultural Theory (9), Risk Information Seeking and Processing model (8), Extended Parallel Process Model (7), Health Belief Model (7), Expected Utility Theory (6) and Social Cognitive Theory (5), were each applied in 5% or fewer studies. In total, the 222 studies contained 285 theory applications, of which 224 were applications of a core theory; studies and theory

applications are distinct counting units throughout, and figures labelled N = 222 refer to studies unless stated otherwise (Methods). This concentration remained stable over time (Extended Data Fig. 2): PMT and PADM held roughly two-thirds of core-theory applications throughout (67% through 2016, 67% in 2017–2020, 63% in 2021–2025), even as output nearly tripled (42 studies through 2016; 116 in 2021–2025).

Hazard coverage was skewed toward rapid-onset events: flood dominates (*n* = 77, of which 65 single-hazard), followed by earthquake (31) and wildfire (25), which together account for 128 studies (58%), while slow-onset heatwave (12) and drought (11) remain marginal (Extended Data Fig. 5). Overall, the field has grown in volume without broadening in conceptual or hazard scope. 69 studies (26%) addressed more than one hazard category (Extended Data Fig. 5).

These two concentrations compound: of the 80 possible pairings between the ten core theories and eight major hazards, 24 (30%) are absent from the corpus (Fig. 2a). How the remaining pairings fare once documentation quality is required is taken up below (Fig. 2b).

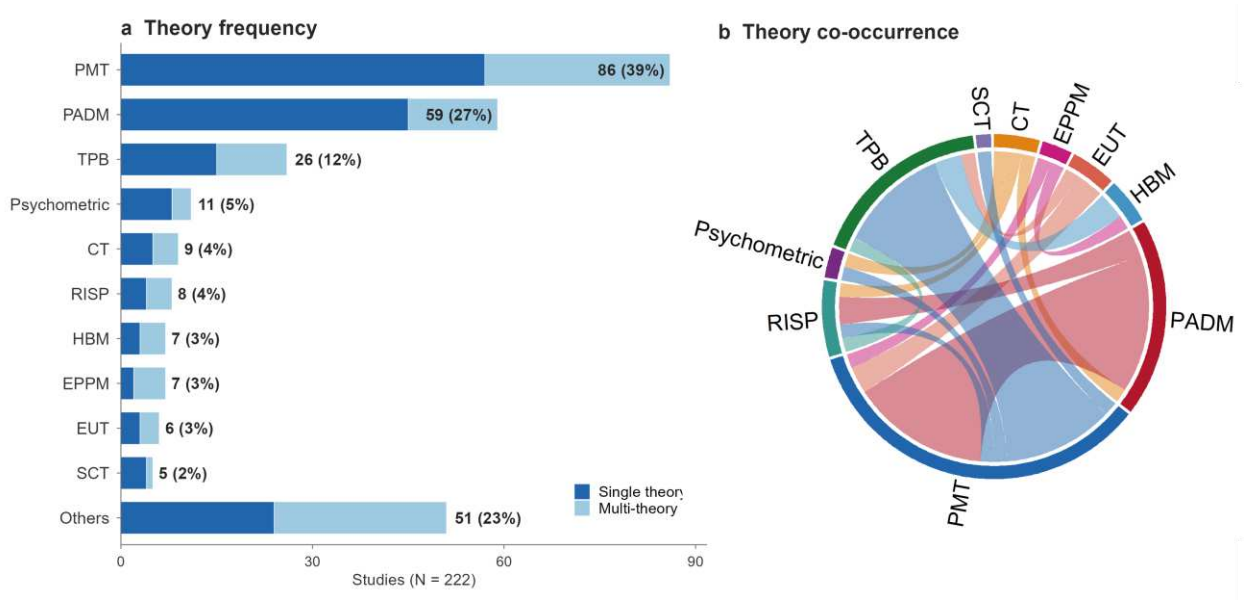


Figure 1. *The theory landscape of natural-hazards and climate-adaptation survey research. a, Number of studies applying each of the ten core theories (N = 222). The 47 non-core frameworks are grouped into 'Others'. Bars are not mutually exclusive so counts and percentages sum to more than 222 and 100%. Shading splits each bar into single-theory (dark) and multi-theory (light) studies; b, Co-occurrence of core theories. The ribbon width is proportional to the number of studies jointly applying each theory pair. Theory abbreviations: PMT, Protection Motivation Theory; PADM, Protective Action Decision Model; TPB, Theory of Planned Behaviour; Psychometric, Psychometric Paradigm; CT, Cultural Theory of Risk; RISP, Risk Information Seeking and Processing model; EPPM,*

Extended Parallel Process Model; HBM, Health Belief Model; EUT, Expected Utility Theory; SCT, Social Cognitive Theory.

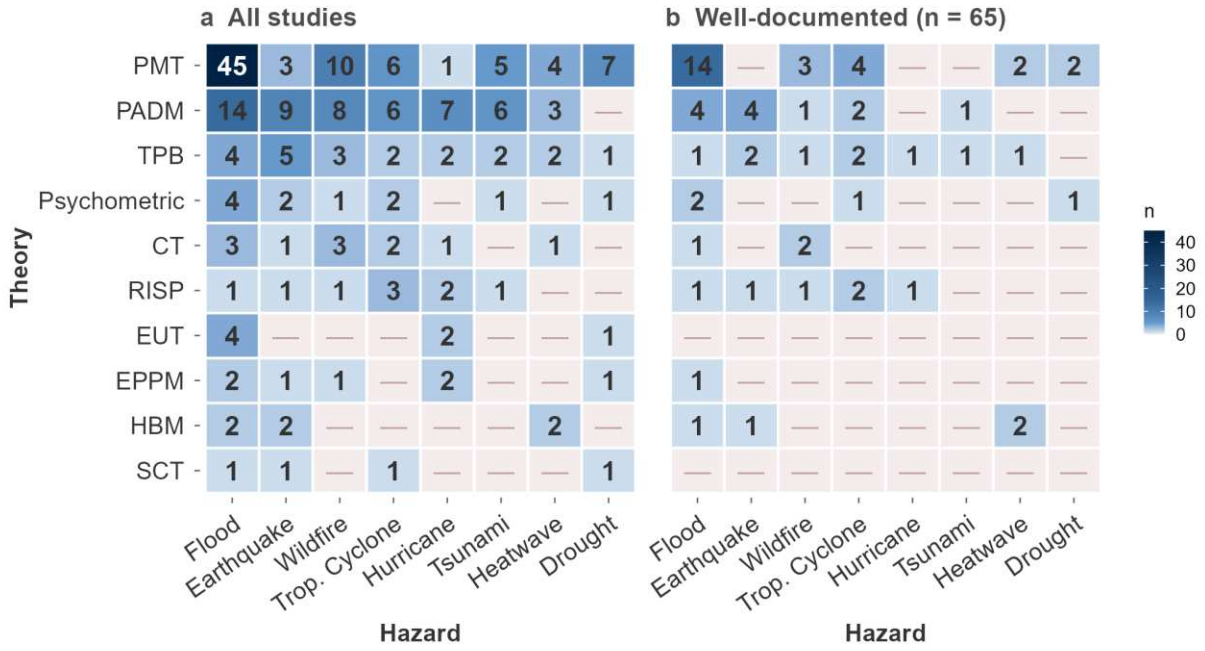


Figure 2. *Theory × hazard coverage of the evidence base. a, Number of studies applying each core theory to each of the eight hazards with at least ten studies (N = 222); Tornado, landslide, and general or multi-hazard studies are shown in Extended Data Fig. 5. b, The same matrix restricted to well-documented studies (n = 65). Cell colour intensity scales with the number of studies; multi-hazard studies contribute to every hazard they cover. Theory abbreviations as in Fig. 1.*

Researchers measure the constructs that are easiest to operationalise

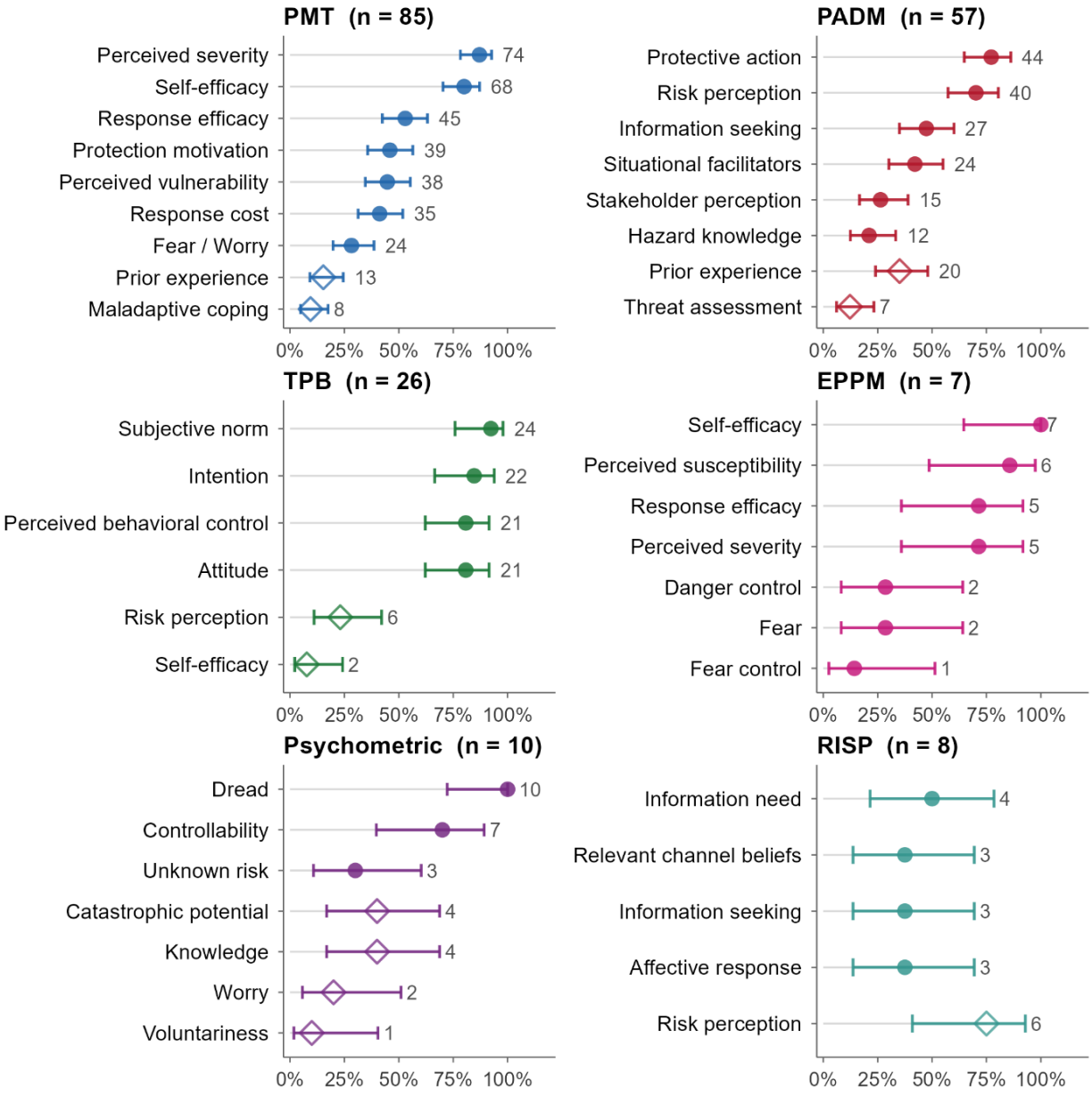
Applying a theory to design a survey implies, ideally, measuring the full set of constructs it specifies. In practice, this rarely happens: across all ten core theories, studies measured some constructs and omitted others, and the gap between the most- and least-measured construct within a single theory is large, 59 percentage points in PMT, 56 in PADM and 86 in EPPM (Fig. 3). The omissions are not random and follow a pattern; they track how much measurement effort and construct demands. A consistent ordering emerged across theories: risk perceptions, efficacy beliefs and the behavioural outcome itself were measured most often (roughly 70% of studies on average), emotional responses less (roughly 40%), contextual and institutional constructs less still (roughly 35%), and behavioural alternatives and process constructs least (roughly 25%). A possible explanation is that risk perception measurements require two or three survey items; a validated measure of institutional trust or maladaptive coping takes many more. PMT illustrates this divide: perceived severity (87%, a figure that includes studies reporting only the aggregate label ‘threat appraisal’) and self-efficacy (80%), each captured with a few items from established pools (Grothmann & Reusswig, 2006; Bubeck et al., 2012),

dominate, while response efficacy (51%), response cost (40%) and fear (27%) fall away, and maladaptive responses such as fatalism and denial, extended rather than core constructs, appear in just 9% of studies. Only 6 of 85 PMT studies (7%) measured all seven core constructs. Relaxing the threshold to any six of seven raises this to 20% (Extended Data Fig. 9a). This matters because response cost is the construct explaining why people who perceive a risk and feel able to act still do not, a study omitting it cannot test PMT's most policy-relevant prediction. PADM shows the same shape: the constructs any risk theory shares, protective action (77%) and risk perception (70%), are well covered, but the constructs distinctive to PADM (and hence not part of other theories) are not: information seeking (47%), situational facilitators and barriers (40%), stakeholder perception (26%) and hazard knowledge (19%). Only 2 of 57 PADM studies (4%) measured all six core well-defined constructs.

We identified two explanations for this selection bias: measurement cost, as harder-to-operationalise constructs are dropped; and intervention design, as constructs seen as easier to manipulate to change behaviour may be favoured over those that are not. Either way, the constructs most often omitted, namely response cost, institutional trust, situational barriers and maladaptive coping, are commonly argued to be precisely those describing why households fail to act. Several of these constructs are sometimes measured outside the named theories, so evidence on them is not wholly absent; the gap is specifically within *theory-guided* operationalisation.

Selective measurement accumulates into fragmentation. Beyond the 6 PMT studies measuring all seven constructs, the remaining 79 applications use 33 distinct construct combinations, 20 of which appeared in only one study each (Extended Data Fig. 9a). A further 26% reported only aggregate labels ('threat appraisal', 'coping appraisal') without specifying which sub-constructs were included (Extended Data Fig. 10). This lack of consistency undermines comparability, and ultimately, synthesis: where studies attach the same construct label to differently operationalised measures, their estimates are not straightforwardly comparable and pooling them risks combining quantities that are not the same, a heterogeneity already documented[10]; for threat and coping appraisal specifically, Bamberg et al.,[19].

Complete measurement is nonetheless attainable. TPB (n = 26) breaks the pattern: its four core constructs (attitude, subjective norm, perceived behavioural control and intention) were measured at comparable rates (81–92%, an 11-point spread against 59 in PMT), and 73% of TPB studies measured all four, against 7% for PMT and 4% for PADM. On average, a TPB study covered 85% of its core construct set (median 100%), against 54% for PMT and 47% for PADM. Part of this gap is mechanical, since a four-construct set is easier to complete than a seven-construct one, but the uniformity of TPB's rates is not, and TPB is the only theory in this corpus whose foundational literature includes prescriptive item-construction guidance[20]. The coverage gaps in the dominant theories are therefore not inevitable features of survey research and could be narrowed by comparable item templates for the constructs now left out.



Filled circle = core construct. Hollow diamond = extended construct. Error bars: Wilson 95% CI. Constructs measured by <5% of studies omitted. PMT, Protection Motivation Theory; PADM, Protective Action Decision Model; TPB, Theory of Planned Behaviour; EPPM, Extended Parallel Process Model; Psychometric, Psychometric Paradigm; RISP, Risk Information Seeking and Processing model.

Figure 3. *Selective operationalisation of theoretical constructs. For the six theories discussed in the Results (PMT, PADM, TPB, EPPM, Psychometric Paradigm and RISP), points show the proportion of studies applying the theory that measured each construct; horizontal bars are Wilson 95% confidence intervals and numbers give the count of studies measuring the construct. Filled circles denote core constructs defined in the foundational literature; open diamonds denote extended constructs. Constructs measured by fewer than 5% of studies applying a theory are omitted. The remaining four core theories (CT, HBM, EUT and SCT) are shown in Extended Data Fig. 6. Theory abbreviations as in Fig. 1.*

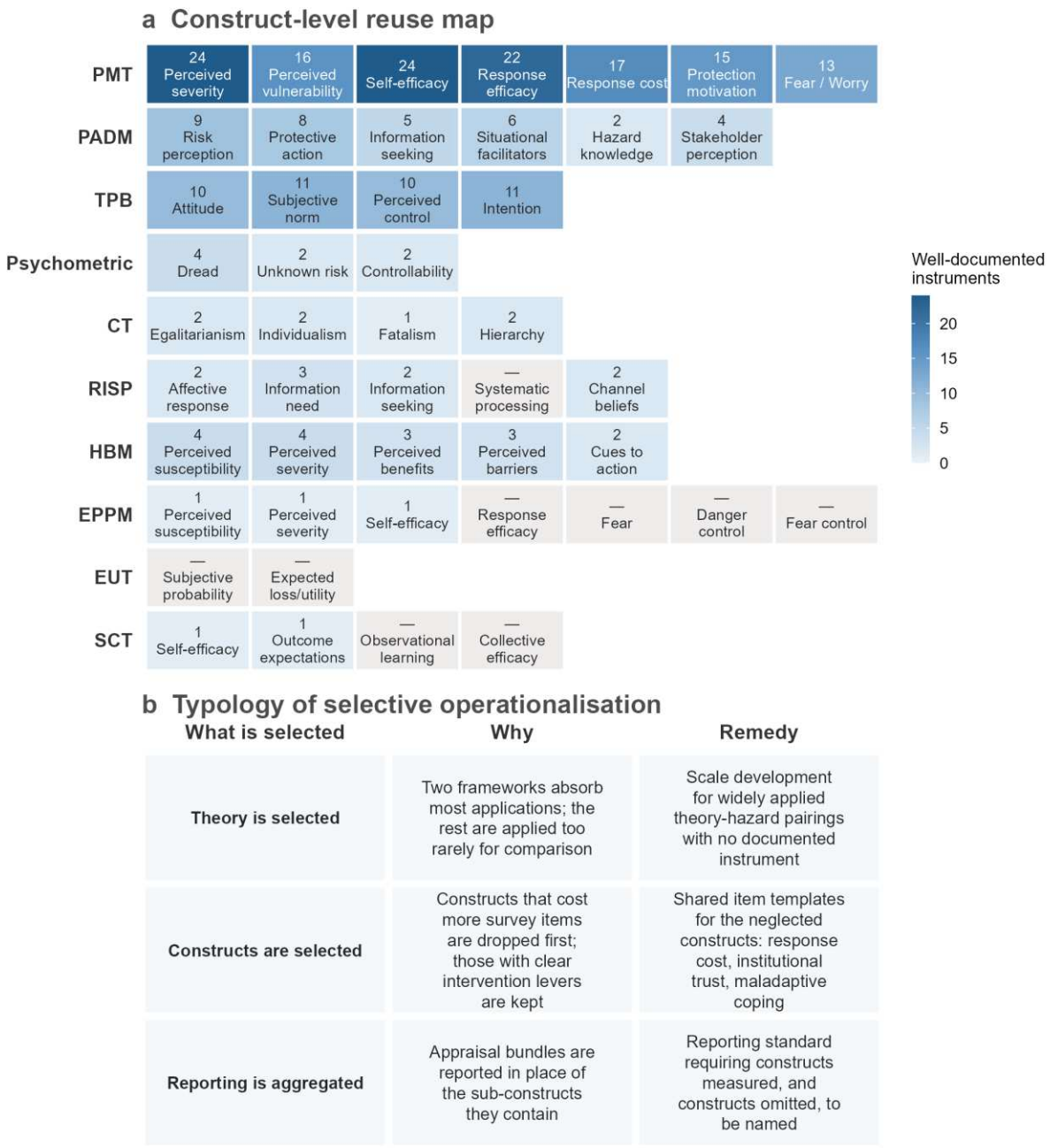


Figure 4. *Where reusable instruments exist, and why constructs go unmeasured. a, Construct-level reuse map. For each core theory and each of its core constructs, the cell gives the number of well-documented studies in the corpus that measured that construct, that is, the number of instruments a researcher could reuse directly. Em dashes on a grey ground mark constructs for which no well-documented study in the corpus provides an instrument. Constructs are theory-specific, so each row carries its own set of cells, and columns are not comparable across rows. Theories are ordered by the number of applications. b, Typology of selective operationalisation: what is selected, the mechanism*

that plausibly drives the selection, and the corresponding remedy. Theory abbreviations as in Fig. 1.

A popularity–quality paradox in measurement documentation

The most applied theories are not the best-documented. RISP, applied in just 8 studies, is well-documented in 62% of its applications; HBM (43%) and TPB (38%) are likewise strong. By contrast, only 30% of PMT applications are well-documented (52% report no or poor psychometric evidence) and 17% of PADM applications (54% inadequate) (Fig. 6). Thus, a researcher seeking a replicable PMT instrument faces a literature ten times larger than RISP's but only half as likely to yield a usable tool.

Overall, only 65 of 222 studies (29%) met the well-documented threshold; 48% reported no or poor psychometric evidence and 72% no qualitative validation (Fig. 5). Documentation did improve more recently: 24% of studies through 2016 were well-documented, 20% in 2017–2020 and 36% in 2021–2025, though the period samples are small, and whether this reflects changing norms or a shift in publishing venues cannot be resolved here.

These deficits compound the concentration described above: restricting the theory–hazard matrix to well-documented studies empties 49 of the 80 pairings (61%) (Fig. 2b): for three in five theory–hazard combinations, as an implication - the literature provides no replicable basis for how the relevant constructs are measured. Within this landscape, a minority of studies document their instruments well enough to serve as reference tools (Fig. 4a).

Three gaps warrant targeted investment: (i) several theory–hazard pairings lack well-documented instruments (SCT for any major hazard, EUT for any hazard, PMT for earthquake); (ii) three under-measured constructs have no validated climate-hazard scale (maladaptive coping, stakeholder perception, response cost); and (iii) a precision gap: the studies that report only aggregate labels ('threat appraisal', 'coping appraisal'; 26% of applications, Extended Data Fig. 10) cannot be disaggregated after the fact, so unlike (i) and (ii), this gap isn't one that new instrument development can close; it requires more precise reporting in future studies. At construct level the same thinness appears: of the 47 core theory–construct pairs, 9 (19%) have no well-documented instrument anywhere in the corpus and 15 (32%) have at most one, so for a third of the pairs a researcher has no reusable measure or no choice between measures (Fig. 4a). Figure 4b sets out what is being selected at each level, the mechanism that plausibly drives it, and the corresponding

remedy.

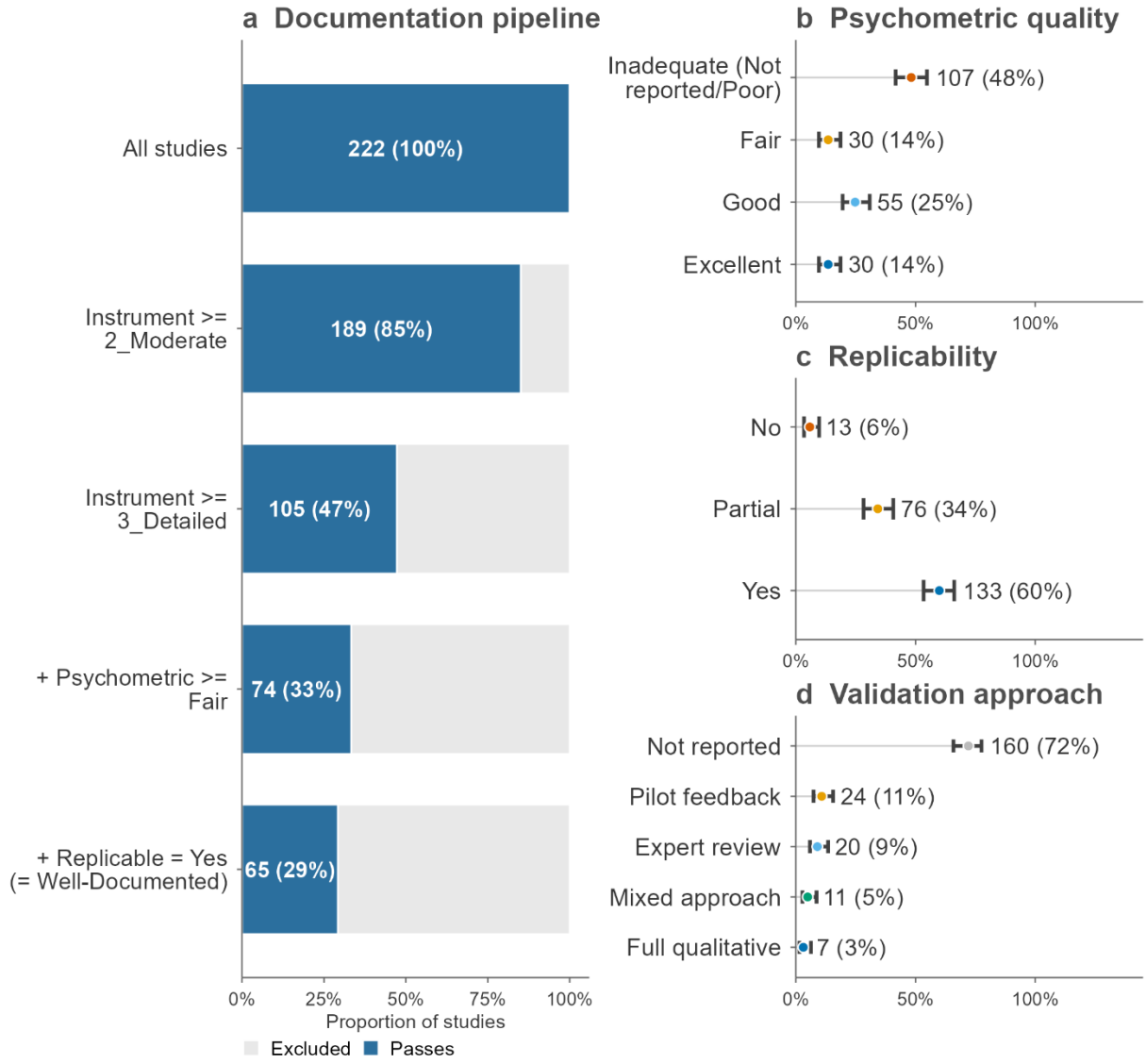


Figure 5. *The measurement documentation deficit (N = 222). a, Documentation pipeline showing the proportion of studies passing each cumulative criterion required for Well-Documented status under the rule-based decision rule (instrument ≥ 2_Moderate shown for reference; the rule requires instrument ≥ 3_Detailed, psychometric quality ≥ Fair and replicability = Yes); dark segments show studies passing all criteria up to that stage and grey segments show studies excluded. b, Psychometric quality (Not reported and Poor collapsed to Inadequate), c, replicability and d, qualitative validation approach across all studies; lollipops show proportions with Wilson 95% confidence intervals and absolute counts.*

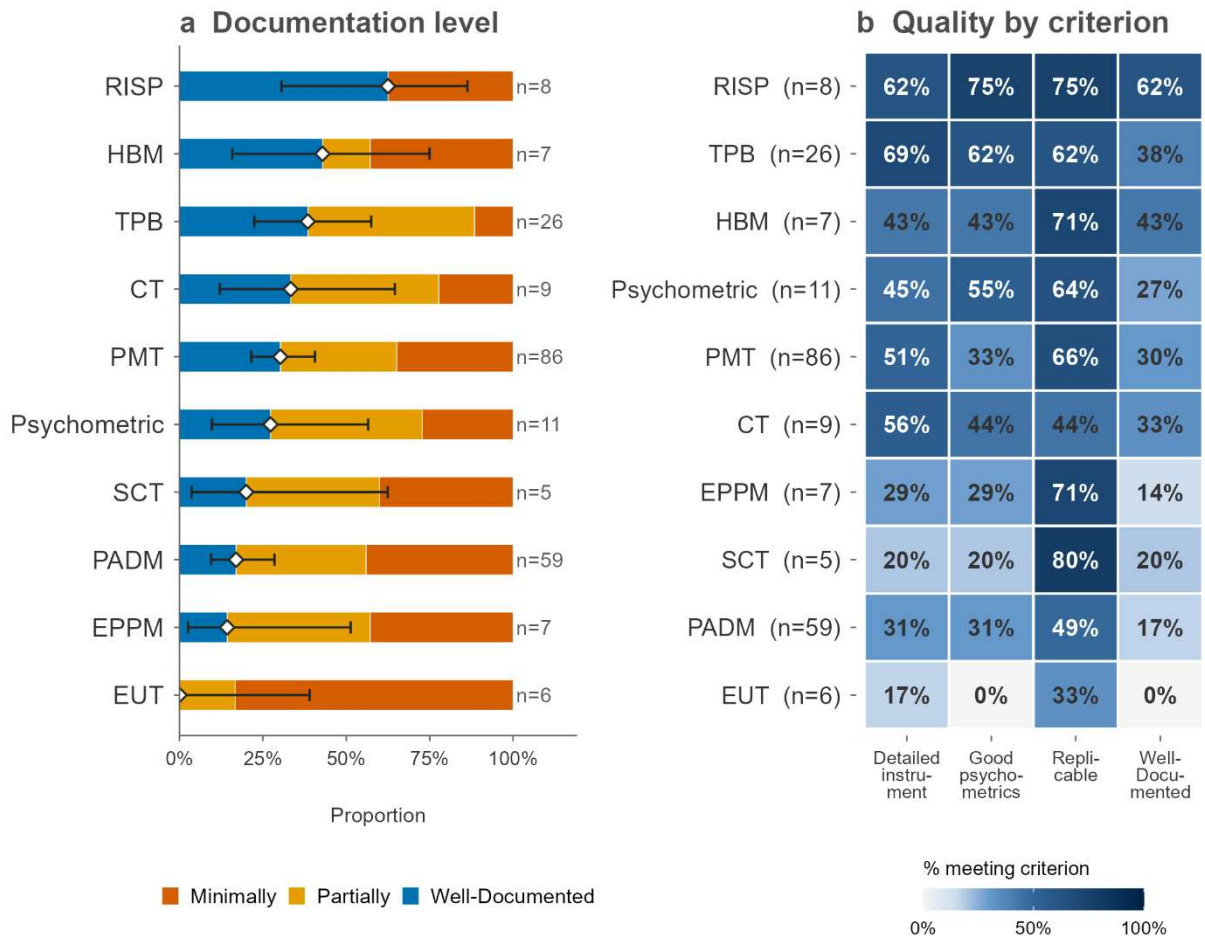


Figure 6. *The popularity–quality paradox. a, Measurement documentation levels across ten core theories, ordered by the proportion of well-documented studies (highest at top); diamonds with horizontal bars show the point estimate and Wilson 95% confidence interval for that proportion, with sample sizes at right. b, Proportion of studies per theory meeting each documentation criteria: detailed instrument (level ≥ 3), good or excellent psychometric reporting, full replicability, and overall Well-Documented status; theories ordered by mean proportion across criteria. Theory abbreviations as in Fig. 1.*

Discussion

Two theories, PMT and PADM, account for over half of all theory applications in natural-hazards and climate-adaptation research. Within them, researchers consistently measure the constructs that are easiest to operationalise and selectively omit others. Only 7% of PMT studies and 4% of PADM studies measured the full set of core constructs specified by the theory they applied, and PMT's 85 codable applications produced 34 distinct construct combinations. As behavioural evidence is increasingly drawn into adaptation planning and risk communication[4], this selectivity has practical consequences: it shapes the evidence

available for adaptation decisions. Extending the finding that roughly 90% of vulnerability, resilience, and adaptation studies engage in no explicit theory[11], our results show that even the theory-guided minority engage selectively and document their measures unevenly.

This selectivity is not primarily a matter of researcher negligence but reflects structural features of how the field is produced. Measurement infrastructure is unevenly distributed with theories offering prescriptive item-construction guidance, such as TPB[20], showing far greater operationalisation consistency than those defining constructs without measurement specifications. This mirrors broader evidence from psychology that construct proliferation thrives in fields lacking shared standards and incentives for reuse[21]. For some frameworks, the measurement boundary is also less clearly specified: PADM encompasses multiple stages, perceptions, and situational factors, and its application has continued to evolve.

Disciplinary conventions matter too: the reviewed Expected Utility Theory studies, largely published in economics and decision-science outlets, rarely report psychometric evidence, a reflection of disciplinary norms rather than a specific shortcoming of EUT applications.

Publication incentives reward breadth of application over depth of operationalisation, while much of this work is produced under the constraints of short-term funding, conditions that favour shorter surveys measuring three or four constructs over the fuller instruments that comparability requires.

A common counter-argument is that a 'reduced-form' model measuring only the constructs of immediate interest is legitimate when the research question does not need the whole theory. That is defensible within a single study, but it externalises a cost onto the collective evidence base. When studies each measure a different subset, their results cannot be pooled or compared. We cannot presently tell whether this evidence accumulates at all, because the studies are not measured comparably enough to establish whether a synthesis would be meaningful. We cannot show directly that omitting a construct changes a study's substantive conclusion, which would require re-analysis of primary data, but the comparability cost does not depend on it. Internal validity for one study does not guarantee comparability across studies, and it is comparability that a cumulative science depends on.

Several of the most commonly omitted constructs are also directly relevant to policy. Response cost, institutional trust, situational barriers and maladaptive coping describe why households fail to act despite recognising a threat; they are the mechanisms through which governments, agencies and adaptation programmes can actually influence behaviour, and they parallel the 'barrier-first' emphasis recently argued for pro-environmental behaviour[5,22–24]. This asymmetry reproduces a policy framing oriented toward prompting adaptation rather than understanding its absence; we propose a symmetry principle for future measurement, in which the motivation to adapt is measured

as consistently as the barriers that prevent it. This is a bias that is invisible within any single study and visible only once the literature is examined as a whole. This asymmetry can become self-reinforcing: constructs that are repeatedly measured are repeatedly available for testing, whereas constructs that are rarely operationalised generate little evidence with which to evaluate their role[12,21]. In this sense, selective operationalisation is less an oversight than a systematic pattern of what the field chooses to ignore[25], and it warrants attention as a source of bias in the knowledge systems that increasingly inform adaptation policy.

These patterns point to three concrete priorities, in ascending order of effort. The most immediate is a quick win: reusing the well-documented instruments the field already has rather than developing new ad hoc measures. For the common theories this is internally consistent with their rationalist core: Studies applying PMT can adopt Duijndam et al.[26] whose extended framework covers all seven core constructs in a climate-driven flood and migration context, or Lim[27], whose 34-item battery offers the most fully itemised core-construct set in the corpus across hurricane, flood and wildfire contexts, or the Babcicky–Seebauer[28] protocol for replicated flood instruments, PADM researchers the Terpstra–Lindell[29] inventory, and TPB researchers Ajzen's construction guidelines[20], which already yield consistent operationalisation across nine hazard contexts in our subset. These are named as exemplars of documentation quality rather than as endorsed standards; the generalisable output is the catalogue itself (Fig. 4a). Reuse does not preclude adaptation to local hazard contexts; it provides the shared foundation that makes findings comparable. Where a study's own scoping work indicates the template items don't fit the local hazard context, adaptation should be documented rather than treated as a deviation from best practice.

The second priority is targeted scale development for the constructs and theory–hazard combinations where no adequate instrument exists (Fig. 4a), chiefly institutional trust, response cost and maladaptive coping, a gap that individual researchers cannot close within normal publication incentives and that funders, journals and professional societies are better placed to address.

The third is reporting reform: at a minimum, authors stating which constructs they measured and, equally, which they omitted and why. An emerging reporting checklist adapted for hazard and adaptation surveys, building on templates such as SOBER (Systematic Operationalization and Bounds for Empirical Research, which mandates explicit construct boundaries, item-selection rationales, and disclosure of dropped variables), would give this a practical form, as would minimum psychometric-reporting standards at journals covering the field[12,17,21].

Improving climate-adaptation research therefore does not primarily require more applications of existing theories; it requires more complete and comparable measurement of the theories already in wide use. Building that measurement infrastructure, meaning shared instruments, validated scales for the neglected barriers, and reporting norms that make omissions visible, may prove as important for adaptation science as developing new

theoretical frameworks, and increasingly so as climate risks intensify and adaptation decisions grow more complex. Our mapping provides a diagnostic baseline to track progress on all three priorities.

Methods

Definitions, counting units and scope

Throughout, *theory* and *framework* are used interchangeably for named models of individual protective decision-making; *construct* denotes a latent concept that a theory specifies (for example, self-efficacy); and *variable* denotes a measured indicator of a construct in a survey instrument. A study *applies* a theory when the theory guided questionnaire design, not when it is cited only as background. *Study* (used interchangeably with *paper*) and *theory application* are distinct counting units: the corpus contains 222 studies and 285 theory applications, of which 224 are applications of one of the ten core theories. A theory application is a distinct study × theory pair, not a coding row; where a study was coded in several rows for the same theory, those rows were consolidated into a single application before analysis. Counts labelled N = 222 in figures refer to studies; construct-level denominators (85 for PMT, 57 for PADM, 26 for TPB) refer to theory applications with a codable construct description.

The frameworks we compare are not all of one kind, and the comparison should be read with that in mind. PMT, TPB, EPPM, HBM and SCT specify a determinate construct set and a causal structure among its elements. PADM and RISP specify a staged process whose full construct set is less sharply bounded. The Psychometric Paradigm and Cultural Theory are measurement traditions that define a dimensional space rather than a behavioural model. Expected Utility Theory is a normative model whose empirical implementation varies by discipline. We therefore apply the completeness metric only where the founding literature specifies an enumerable construct set, and we report per-application coverage alongside all-or-nothing completeness so that theories with small core sets are not read as more complete by construction.

Our corpus spans climate-related and geophysical hazards. We retained both because survey instruments for risk perception and protective behaviour are largely hazard-general: construct definitions and often the items themselves are transferred across hazards, so measurement practice in earthquake or tsunami research bears directly on climate-hazard research. Restricting the corpus to climate-related hazards alone would remove roughly a third of the evidence on how these theories are operationalised without changing what the instruments look like.

Literature search

We searched the Web of Science Core Collection on 24 April 2025 using a Boolean query combining four concept groups: natural hazards, adaptive behaviours, population level and data-collection methods. The natural-hazards group included terms for flood, storm

and tidal surge, drought, heatwave, cold spell, landslide, earthquake, tsunami, wildfire, volcanic eruption, tropical cyclone, tornado, hurricane, avalanche and multi-hazard contexts. The adaptive-behaviours group captured protective, mitigation, coping, preparedness, evacuation and insurance behaviours and their variants. The population-level group restricted results to studies of individuals, households, residents, citizens or communities. The data-collection group restricted results to survey and questionnaire research. The full query string is provided in Supplementary Methods. Non-article document types were removed, leaving 3,344 records for screening.

Two-stage machine-learning-assisted screening

We used a two-round Random Forest classifier implemented in R 4.3.1 to triage records for full-text review. Text preprocessing combined titles, abstracts, keywords and research-area fields, applied lowercasing and punctuation removal, filtered English stop-words and constructed a document–term matrix weighted by term frequency–inverse document frequency, supplemented by structural features such as title and abstract length. In round one, 8 members of the author team screened 1,055 records independently, without sight of one another's decisions, against predefined inclusion criteria (individual- or household-level empirical survey of risk perception or protective behaviour in a natural-hazard context), labelling each as relevant (1) or irrelevant (0). Records on which screeners disagreed were resolved by discussion. Screening decisions were not double-coded across the full set, so we do not report an inter-screener agreement statistic for this stage; reliability was instead assessed at the measurement-coding stage, where the substantive coding decisions are made (see below). A Random Forest model trained on this labelled set was applied to the remaining 2,289 records using decision thresholds of ≥0.6 (relevant), ≤0.4 (irrelevant) and 0.4–0.6 (uncertain). Round two extended manual screening to the 970 uncertain records and applied a refined classifier to the residual 1,227 records, followed by a final pass of manual adjudication that reclassified 5 machine-labelled relevant records as irrelevant. Model performance on a held-out set of 422 manually screened records was: accuracy 81.3% (95% CI 77.2–84.9%), sensitivity 89.6%, specificity 69.4%, precision 80.8%, F1 85.0%, Cohen's $\kappa = 0.60$. Manual screening covered 2,117 records in total (1,280 irrelevant, 837 relevant) and machine-assisted screening the remaining 1,227, producing 1,214 records for full-text assessment.

Full-text retrieval and automated theory detection

PDFs were retrieved through institutional access and organised in a Zotero library. We extracted text from 1,103 of the 1,214 records (91%) using the pdftools R package; the remaining 111 were inaccessible because of paywalls or missing PDFs. Extracted text was linked to bibliographic metadata through a four-strategy matcher combining filename parsing, author–year tokens, DOI extraction from the first page and title-similarity scoring; unlinked cases were resolved manually. Theory identification followed a three-tier procedure: an exact-match pass against a curated database of theoretical frameworks (confidence 9–10); an acronym-recognition pass (PMT, PADM, TPB, EPPM, HBM, TRA, SCT, RISP and others; confidence 8); and a discovery pass using regular expressions of the form

‘Capitalised Phrase + (Theory|Model|Framework|Paradigm|Approach)’ with exclusion filters for statistical and methodological terminology (confidence 6–7), applied to the title, abstract and first 4,000 characters of main text. The pipeline identified 1,053 theory mentions (430 unique strings) in 544 of the 1,103 documents. Manual validation by two authors removed pattern-discovered false positives, records without a natural-hazard context and records where the theory appeared only in reference lists, leaving 428 papers with at least one validated theoretical framework and hazard context.

Manual coding of measurement information

11 trained coders (authors with at least graduate-level training in risk-perception or behavioural research) extracted measurement information from each validated paper using a structured spreadsheet protocol. For each theory × paper combination they recorded: the location of the theory in the paper (methods / introduction / discussion / references only); whether the theory guided questionnaire design or was cited only as background; which constructs were measured; the number of items shown; the response-scale format; psychometric evidence (reliability statistics, factor analyses, validity tests); any qualitative validation (pilot testing, cognitive interviews, expert review); and whether the measurement could be replicated from the information provided. The full protocol is in Supplementary Methods and Supplementary Data 1. Of the 428 validated papers, 205 contained theories that did not actually guide measurement (’Excluded – no guiding theory’) or lacked measurement detail (’Excluded – no measurement details’). One further study was excluded after full-text check because its respondents were members of organisational disaster-response teams rather than individuals or households, and its outcome was performance in a cyber-incident tabletop exercise rather than protective behaviour toward a natural hazard. This left 222 papers with 285 theory–paper applications, published in 80 journals between 2000 and 2025, as the final analytical corpus. This step removes papers that named a theory without measuring it and papers reporting no measurement detail at all, so the documentation statistics we report are conditional on a corpus that has already cleared a minimum reporting bar. They should be read as upper bounds on documentation quality in the wider theory-citing literature, not as estimates for it.

Inter-coder reliability

Inter-coder reliability was assessed by independent re-coding of a random subset of 43 theory–paper rows (15% of the 285 applications), each re-coded by a different team member than the original coder and drawn from a rotating pool of seven coders. Because the design uses a rotating pool rather than two fixed raters, Krippendorff’s α is reported as the primary statistic, with the ordinal metric for the three ordinal fields and the nominal metric for psychometric quality, whose ‘Not reported’ category is not a point on the quality scale. The subset was drawn at random from the full set of coded theory–paper rows. Agreement was α = 0.69 for survey instrument level (95% CI 0.46 to 0.84; 58% exact agreement), 0.38 for psychometric quality (0.17 to 0.56; 53%), 0.35 for measurement replicability (0.03 to 0.63; 72%) and 0.64 for the derived documentation level (0.45 to 0.77;

51%). Confidence intervals are percentile bootstrap intervals from 10,000 resamples, resampling theory–paper rows; with 43 rows they are wide, and the point estimates should be read with that in mind. Two features qualify these figures. Replicability shows the prevalence paradox: 33 of 43 originals are 'Yes', so chance agreement is high and α is penalised despite the highest raw agreement of the four fields. Documentation level is derived mechanically from the other three, so its agreement is conditional on theirs rather than independent evidence. Most disagreements on the ordinal fields are single-step boundary calls rather than fundamental disagreement: on instrument level, 25 of 43 rows agree exactly and 8 more differ by one level. Full agreement statistics and the row-level discrepancy list are in Supplementary Table 3.

Construct operationalisation analysis

The construct-level analysis operates at the level of theory × study applications. Of the 224 core-theory applications, 220 provided a codable free-text description of measured constructs; four rows containing coder annotations rather than construct lists were excluded, giving per-theory denominators of 85 for PMT, 57 for PADM and 26 for TPB (denominators for all theories are shown in Fig. 3).

For each of the ten frameworks with at least five applications, we defined a canonical construct set from foundational literature (Rogers[29,7] for PMT; Lindell & Perry[8] for PADM; Ajzen[31] for TPB; Slovic[32] for the Psychometric Paradigm; and equivalent sources for the remainder; full list in Supplementary Table 1). Within each set, we distinguished core constructs, those the founding statement of the theory treats as necessary components of the model, from extended constructs added in later applications. Core sets differ in size (seven for PMT, six for PADM, four for TPB), so all-or-nothing completeness rates are not directly comparable across theories; we therefore report both the proportion of applications measuring every core construct and the mean proportion of the core set covered per application, the latter being independent of set size. For PMT the core set is perceived severity, perceived vulnerability, self-efficacy, response efficacy, response cost, fear/worry and protection motivation; prior experience and maladaptive coping are treated as extended. Fear is the most contestable inclusion, since Rogers's 1983 revision treats fear as an indirect influence operating through severity appraisal, and Supplementary Table 1 therefore reports completeness with and without it.

Construct presence was detected by case-insensitive regular-expression matching of curated synonym sets against the recorded construct descriptions (Supplementary Table 2). The synonym sets were revised once, after an inventory of coder language across the corpus showed that the initial theory-derived terms missed vocabulary the coders actually used; the revision was specified against that inventory rather than against construct rates, and Supplementary Table 2 reports rates under both the initial and revised term sets so that the effect of the revision is visible. Two aggregate labels required an explicit rule. Coders frequently recorded 'threat appraisal', 'coping appraisal' or 'risk appraisal' rather than the sub-constructs those bundles contain. We assigned each aggregate label to a single representative construct rather than to every construct it spans, because assigning

it to all of them would make the constituent constructs artefactually correlated: in PMT, 'threat appraisal' and 'risk appraisal' were assigned to perceived severity and 'coping appraisal' to self-efficacy; in PADM, threat appraisal was assigned to risk perception. One consequence is that a study reporting only the two PMT appraisal bundles cannot be classified as complete, so the completeness rate is in part a property of this rule. The disaggregation analysis below quantifies how many studies are affected.

To distinguish studies operationalising components discretely from those reporting only aggregate scores, each PMT application was classified as 'fully disaggregated', 'partially disaggregated' or 'aggregate only' (the same applied to PADM using 'risk appraisal'); this classification underpins Extended Data Fig. 10.

Measurement documentation level

Three coded dimensions determine measurement documentation level (MDL): survey instrument level, psychometric quality, and measurement replicability. 'Well-Documented' requires instrument level ≥ 3_Detailed (item wordings shown), psychometric quality of Fair, Good or Excellent, and replicability of Yes; 'Partially Documented' applies when one of these is missing; 'Minimally Documented' when only minimal items or weak psychometric evidence are provided; 'Not Documented' when no items or operationalisation are shown. MDL is derived mechanically from the three component fields through a fixed decision table rather than entered as a coder's judgement. Exact decision rules, worked examples, and the reconciliation of the 76 reclassified boundary cases are in Supplementary Methods.

Coders recorded the three component fields for each theory a study applied, so a study that documents one applied theory thoroughly and another minimally is counted once at each standard. Twenty of the 222 studies carry more than one distinct documentation level across their theory rows. The by-theory rates in Fig. 6 use these theory-specific values, and are read as '30% of PMT applications are well-documented'. Computing the same rates from the study-level value instead, that is, by assigning each study's overall documentation level to every theory it applies, changes no conclusion and leaves the ordering of theories intact (PMT 31% rather than 30%, PADM 20% rather than 17%, TPB 42% rather than 38%, RISP 62% unchanged); the two sets differ for 23 of the 285 applications and both are reported in Supplementary Table 3.

Reporting and statistical analysis

All descriptive statistics are reported with 95% confidence intervals computed using the Wilson score interval for proportions. Analyses and figures were produced in R 4.3.1[33] using ggplot2 3.4.4, dplyr, tidyr, stringr, forcats, scales, patchwork, and circlize. No formal hypothesis tests are reported; all percentages are descriptive summaries of the coded corpus, and differences between theories or periods should be read against the confidence intervals shown rather than as tests. Reliability was computed on the independent codes as first recorded; rows on which the two codings disagreed were subsequently adjudicated against the original papers, and the adjudicated values are

those carried in the published dataset. The study follows PRISMA 2020 reporting standards where applicable to systematic reviews of measurement[33] (Extended Data Fig. 1).

Limitations

Several limitations qualify these findings. The corpus is restricted to English-language articles indexed in Web of Science; research in other languages or regional databases may show different patterns, and because academic teams are rarely professional translators, some cross-study inconsistency could reflect translation quality rather than measurement practice, though this is unlikely to account for the systematic ordering we observe. The search strategy will also have missed genuine applications, since a study can apply a theory without using our screening keywords in its title or abstract; a PADM-based drought study not retrieved by our search illustrates this known search-boundary effect, and the theory–hazard coverage gaps we report should therefore be read as properties of this corpus rather than proof that no such study exists. We omitted grey literature, which may bias the corpus if some theories appear disproportionately in unpublished work. Construct presence was detected from coder-recorded descriptions rather than re-extracted mechanically from full texts; this human coding is in one respect an advantage, capturing interpretive judgement a keyword extractor would miss, but it also means our measurement rates are lower bounds, since a construct a coder did not record is treated as absent. The classification of psychometric quality involves coder judgement; although the protocol was standardised across twelve coders, formal inter-rater reliability was assessed on a random subset rather than the full corpus. Our conclusions rest on a corpus of correlational survey research, which limits causal claims, and we cannot say whether the same practices hold in experimental adaptation research. The ten core theories were selected on frequency; other frameworks may show different operationalisation patterns. Finally, and most fundamentally, our assessment captures what researchers report, not necessarily what they measured: poor documentation may conceal adequate practice, just as detailed reporting does not guarantee good measurement. Despite these limitations, the patterns are consistent across theories, hazards and time periods and point to specific, addressable gaps.

Data availability

The list of included studies and their structural metadata (bibliographic details, hazard classification and theory identification), an empty coding template, and the coding scheme are available at Zenodo (DOI at acceptance). These materials allow the analysis to be reproduced by independently re-coding the included studies. The coded measurement assessments that underlie the reported figures and statistics constitute evaluative judgements of individual third-party studies; to ensure they are interpreted only in the context of the full coding protocol, they are available from the corresponding author on reasonable request.

Code availability

All analysis and figure-generation code is available at Zenodo (DOI at acceptance) as the R script ncc_submission_figures_published.R. The script reproduces every figure and its source data from the two input files once their assessment fields have been coded according to the published coding scheme. Analyses were run in R version 4.3.1 using ggplot2, dplyr, tidyr, stringr, forcats, scales, patchwork, tibble, circlize and png.

Supplementary Materials

Supplementary Methods. Full Web of Science search string; hazard-classification and harmonisation procedure; machine-learning screening procedure and performance; PDF text-extraction and theory-detection pipeline; manual coding of measurement information; measurement documentation level decision rules and the reconciliation of the 76 reclassified boundary cases; construct detection procedure and aggregate-label assignment rules; inter-coder reliability and adjudication procedure.

Supplementary Table 1. Canonical constructs and foundational references for the ten theories, with the core/extended designation of each construct and completeness rates under alternative core-set definitions. **Supplementary Table 2.** Construct-detection search terms for all ten theories, with measurement rates under both the initial and the revised term sets. **Supplementary Table 3.** Inter-coder reliability for the four measurement-coding fields (n = 43).

Reference

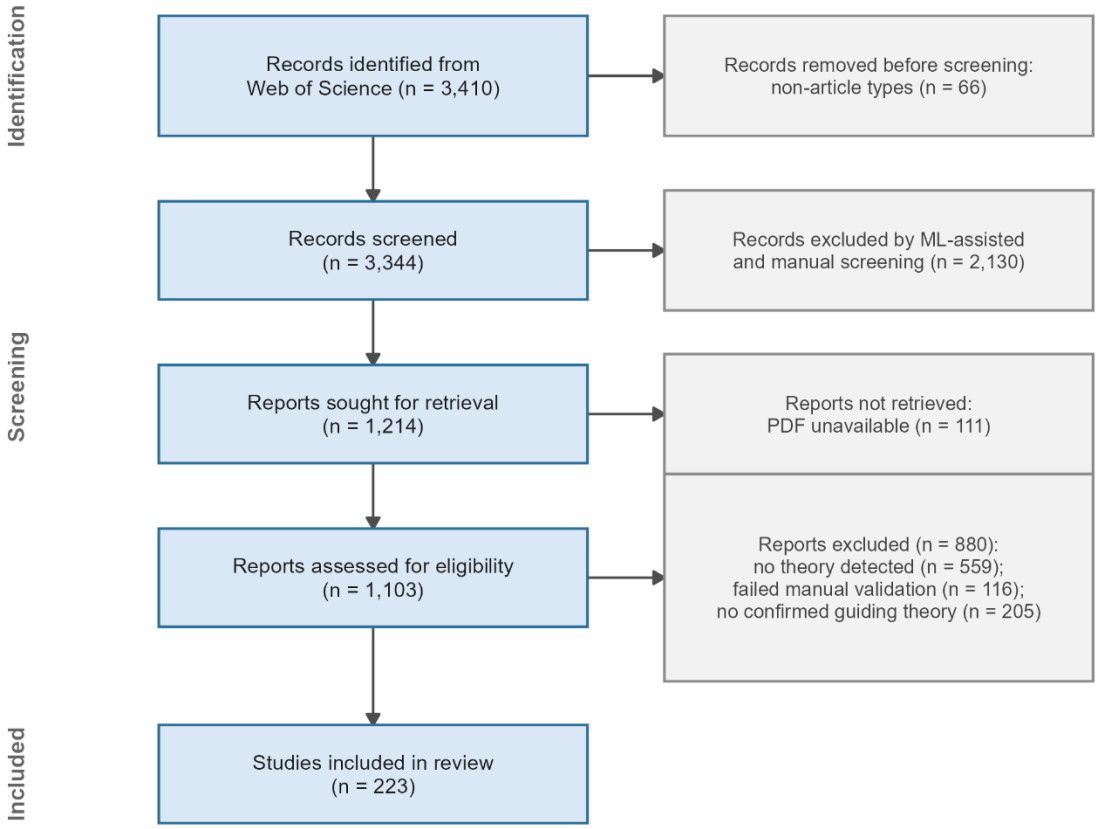
1. Calvin, K. et al. IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (Eds.)]. IPCC, Geneva, Switzerland. https://www.ipcc.ch/report/ar6/syr/ (2023) doi:10.59327/IPCC/AR6-9789291691647.
2. Aerts, J. C. J. H. et al. Integrating human behaviour dynamics into flood disaster risk assessment. Nat. Clim. Change **8**, 193–199 (2018).
3. Dottori, F., Mentaschi, L., Bianchi, A., Alfieri, L. & Feyen, L. Cost-effective adaptation strategies to rising river flood risk in Europe. Nat. Clim. Change **13**, 196–202 (2023).
4. Kuhlicke, C. et al. The behavioral turn in flood risk management, its assumptions and potential implications. Wiley Interdiscip. Rev. Water **7**, e1418 (2020).
5. Noll, B., Filatova, T., Need, A. & Taberna, A. Contextualizing cross-national patterns in household climate change adaptation. Nat. Clim. Change **12**, 30–35 (2022).
6. Berrang-Ford, L. et al. A systematic global stocktake of evidence on human adaptation to climate change. Nat. Clim. Change **11**, 989–1000 (2021).
7. Rogers, R. W. Cognitive and psychological processes in fear appeals and attitude change: A revised theory of protection motivation. in Social psychophysiology: A

sourcebook (ed. John T. Cacioppo, R. E. P.) 153–176 (Guilford Publications, New York, 1983).
8. Lindell, M. K. & Perry, R. W. The protective action decision model: theoretical modifications and additional evidence. Risk Anal. Int. J. **32**, 616–632 (2012).
9. Skipor, S. & Bergquist, M. Meta-analytic evidence identifies efficacy beliefs and social norms as key correlates of public emergency preparedness. Commun. Earth Environ. **7**, 132 (2026).
10. Van Valkengoed, A. M. & Steg, L. Meta-analyses of factors motivating climate change adaptation behaviour. Nat. Clim. Change **9**, 158–163 (2019).
11. Kuhlicke, C. et al. Spinning in circles? A systematic review on the role of theory in social vulnerability, resilience and adaptation research. Glob. Environ. Change **80**, 102672 (2023).
12. Flake, J. K. & Fried, E. I. Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. Adv. Methods Pract. Psychol. Sci. **3**, 456–465 (2020).
13. Kothe, E. J. et al. Protection motivation theory and pro-environmental behaviour: A systematic mapping review. Aust. J. Psychol. **71**, 411–432 (2019).
14. Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P. & Billiet, J. Measurement equivalence in cross-national research. Annu. Rev. Sociol. **40**, 55–75 (2014).
15. Leitgöb, H. et al. Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. Soc. Sci. Res. **110**, 102805 (2023).
16. Borsboom, D. The Attack of the Psychometricians. Psychometrika **71**, 425–440 (2006).
17. Parsons, S., Kruijt, A.-W. & Fox, E. Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. Adv. Methods Pract. Psychol. Sci. **2**, 378–395 (2019).
18. Ke, Z. & Tong, X. Correcting for the multiplicative and additive effects of measurement unreliability in meta-analysis of correlations. Psychol. Methods **28**, 21 (2023).
19. Bamberg, S., Masson, T., Brewitt, K. & Nemetschek, N. Threat, coping and flood prevention – A meta-analysis. J. Environ. Psychol. **54**, 116–126 (2017).
20. Ajzen, I. Constructing a theory of planned behavior questionnaire: Conceptual and methodological considerations. (2006).
21. Elson, M., Hussey, I., Alsalti, T. & Arslan, R. C. Psychological measures aren't toothbrushes. Commun. Psychol. **1**, 25 (2023).
22. Sharpe, E. J. & Steg, L. A critical reflection on behavioural difficulty: proposing a barrier-first approach. Curr. Opin. Behav. Sci. **62**, 101500 (2025).
23. Eisenack, K. et al. Explaining and overcoming barriers to climate change adaptation. Nat. Clim. Change **4**, 867–872 (2014).
24. Fielding, K. S. & Hornsey, M. J. Understanding and reducing the intention–behaviour gap in climate action. Nat. Clim. Change **16**, 638–647 (2026).
25. Wilson, R. S., Herziger, A., Hamilton, M. & Brooks, J. S. From incremental to transformative adaptation in individual responses to climate-exacerbated hazards. Nat. Clim. Change **10**, 200–208 (2020).

26. Duijndam, S. J. et al. A look into our future under climate change? Adaptation and migration intentions following extreme flooding in the Netherlands. Int. J. Disaster Risk Reduct. **95**, 103840 (2023).
27. Lim, J. R. Why People Adopt Climate Change Adaptation and Disaster Risk Reduction Behaviors: Integrated Model of Risk Communication and Results from Hurricanes, Floods, and Wildfires. Bull. Am. Meteorol. Soc. https://doi.org/10.1175/BAMS-D-21-0087.1 (2022) doi:10.1175/BAMS-D-21-0087.1.
28. Babcicky, P. & Seebauer, S. Unpacking Protection Motivation Theory: evidence for a separate protective and non-protective route in private flood mitigation behavior. J. Risk Res. **22**, 1503–1521 (2019).
29. Terpstra, T. & Lindell, M. K. Citizens' Perceptions of Flood Hazard Adjustments: An Application of the Protective Action Decision Model. Environ. Behav. **45**, 993–1018 (2013).
30. Rogers, R. W. A protection motivation theory of fear appeals and attitude change. J. Psychol. **91**, 93–114 (1975).
31. Ajzen, I. The theory of planned behavior. Theor. Cogn. Self-Regul. **50**, 179–211 (1991).
32. Slovic, P. Perception of risk. Science **236**, 280–285 (1987).
33. R Core Team, R. R: A Language and Environment for Statistical Computing. (Vienna, Austria, 2024).
34. Page, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. bmj **372**, (2021).

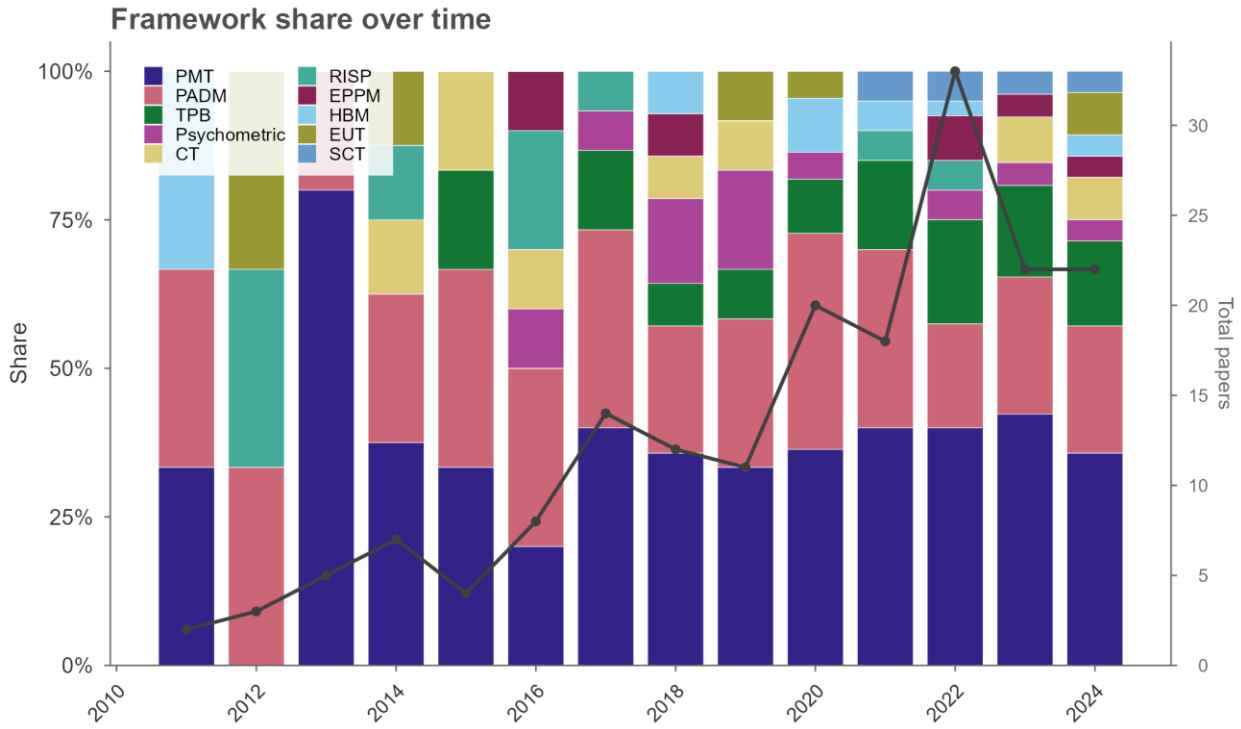
Extended Data

Extended Data Fig. 1



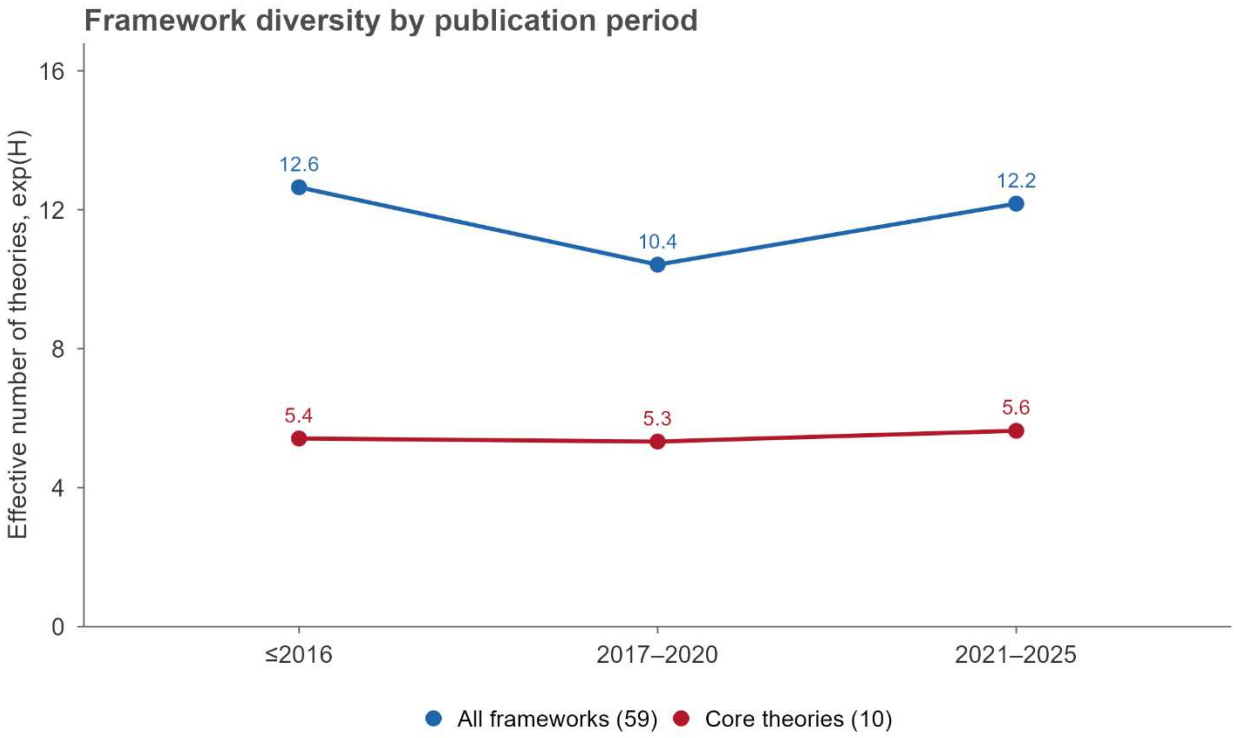
Study selection. PRISMA 2020 flow diagram for the identification, screening and inclusion of studies, from 3,410 Web of Science records to the 222 studies in the analytical corpus. Counts at each stage distinguish manual from machine-assisted screening decisions.

Extended Data Fig. 2



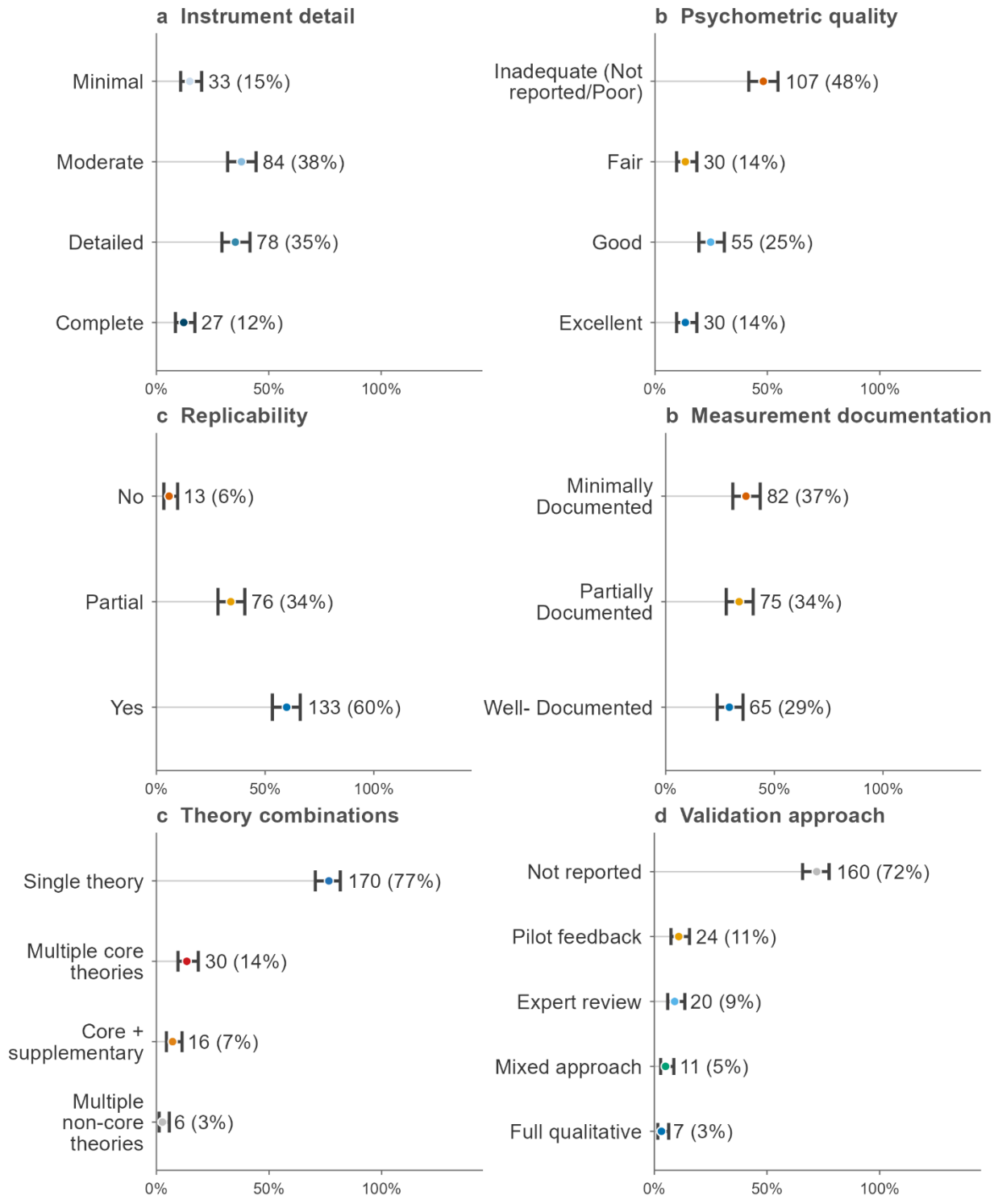
Share of core theories over time. Proportion of core-theory applications accounted for by each of the ten core theories in three publication periods (through 2016, 2017 to 2020, 2021 to 2025). PMT and PADM together hold roughly two-thirds of applications in every period.

Extended Data Fig. 3



Diversity of theory use over time. Shannon diversity of the distribution of core-theory applications, computed within each publication year and shown with a smoothed trend. Higher values indicate a more even spread of applications across theories.

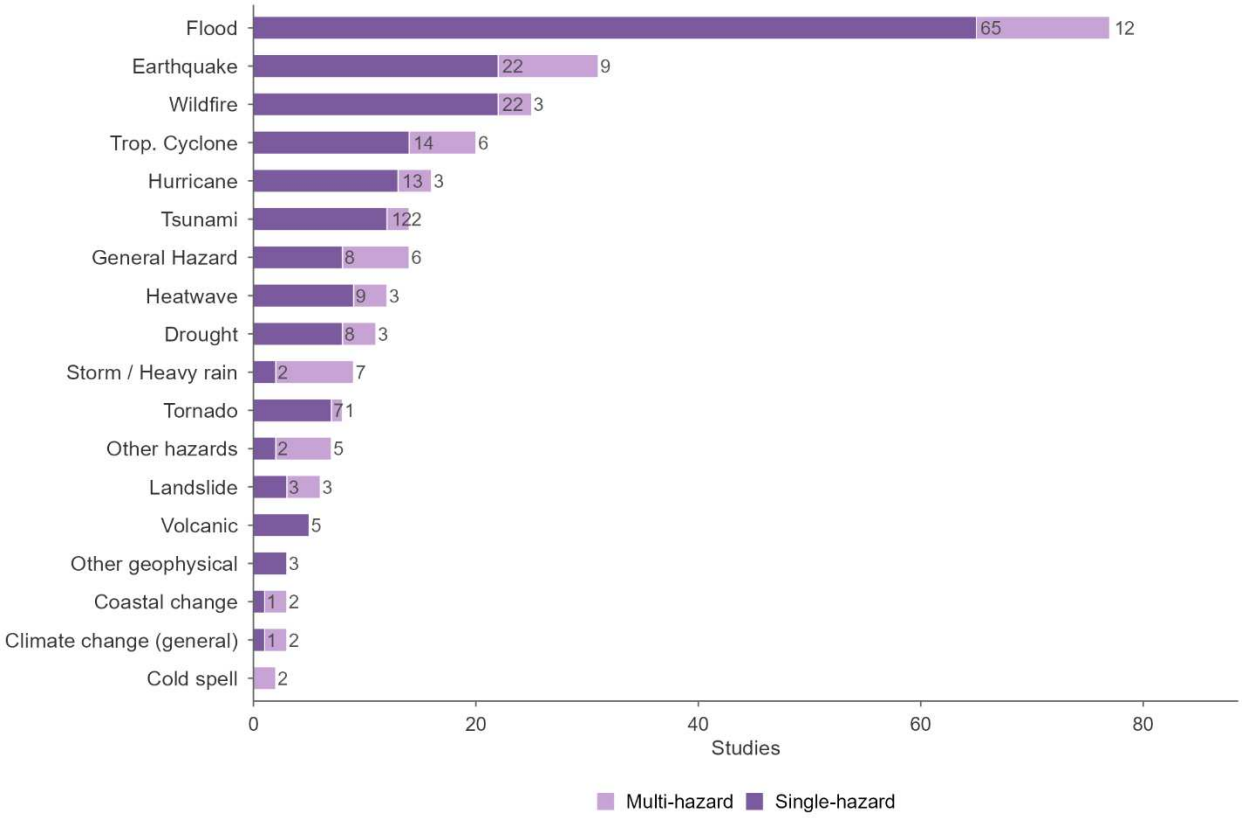
Extended Data Fig. 4



Composition of the corpus across six coding dimensions. Distribution of studies across survey instrument level, measurement documentation level, theory-combination type,

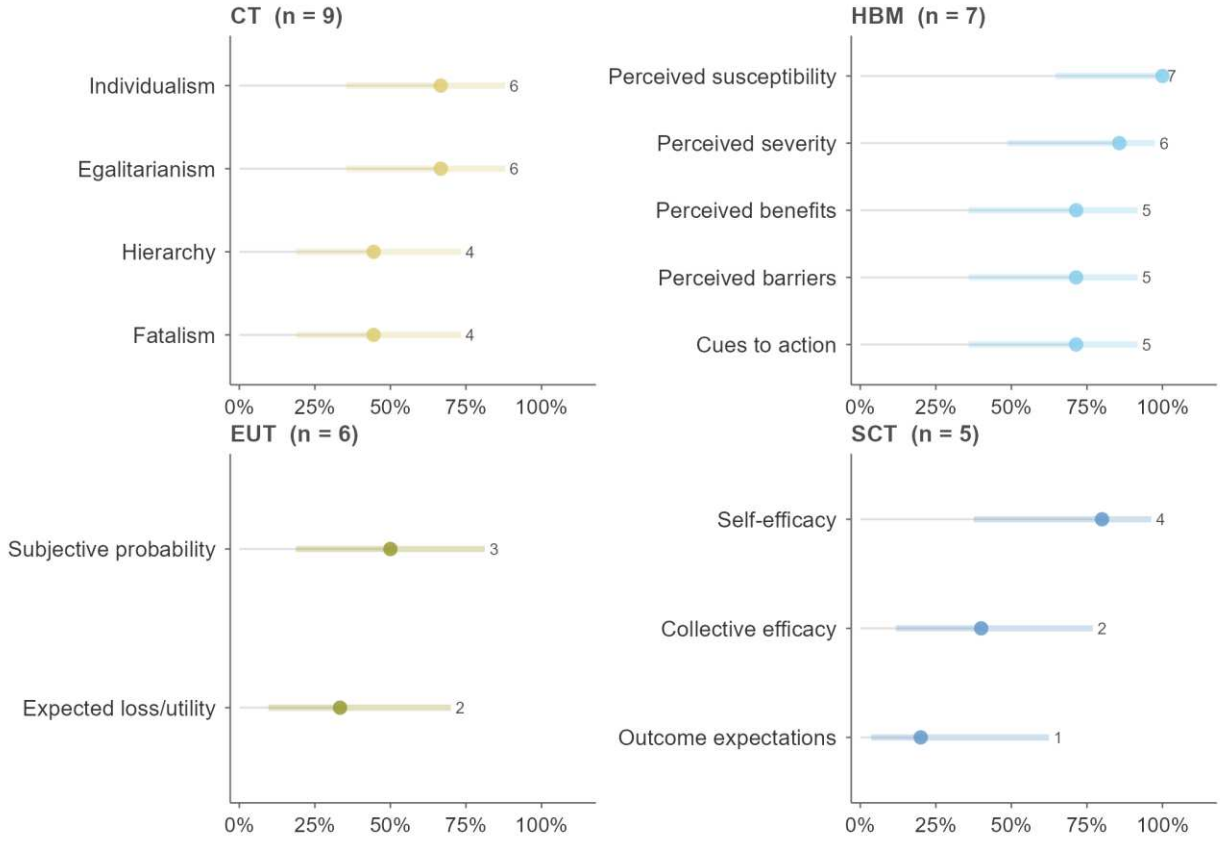
psychometric quality, replicability and qualitative validation approach. Proportions are shown with Wilson 95% confidence intervals and absolute counts.

Extended Data Fig. 5



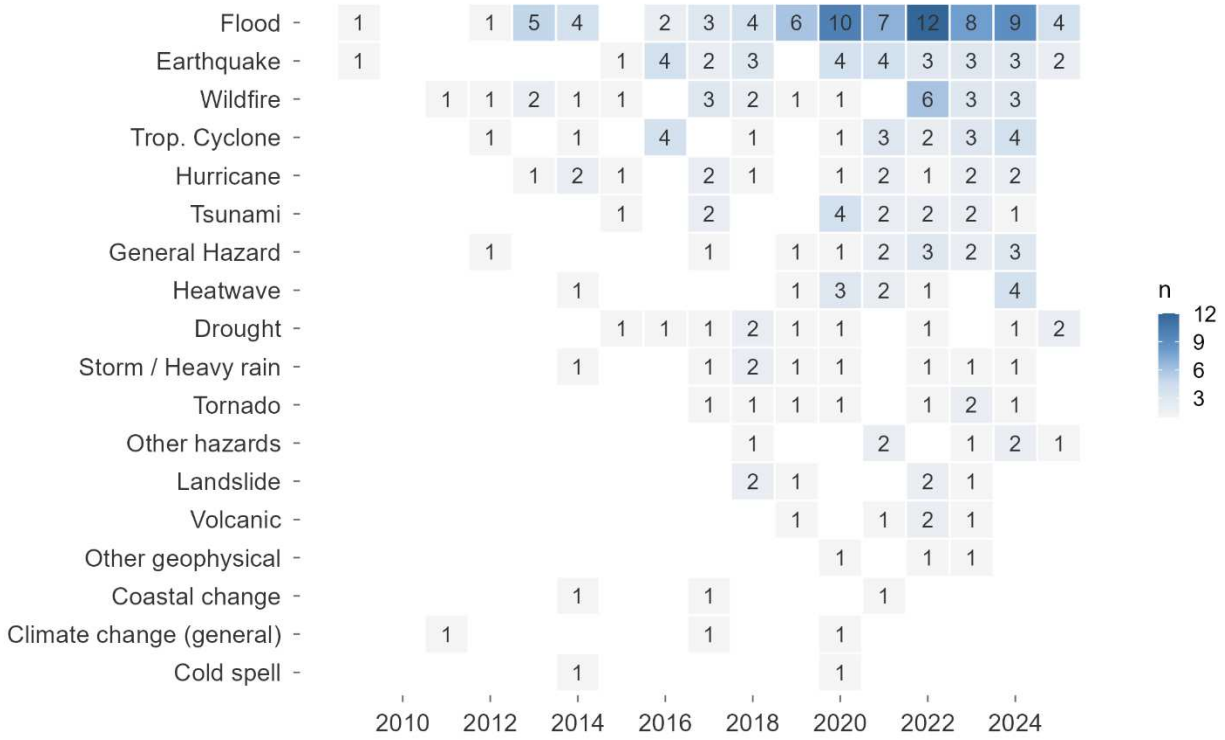
Hazard distribution across the corpus. Number of studies addressing each harmonised hazard category, with single-hazard and multi-hazard studies distinguished. Multi-hazard studies contribute to every hazard they cover, so the bars sum to more than 222.

Extended Data Fig. 6



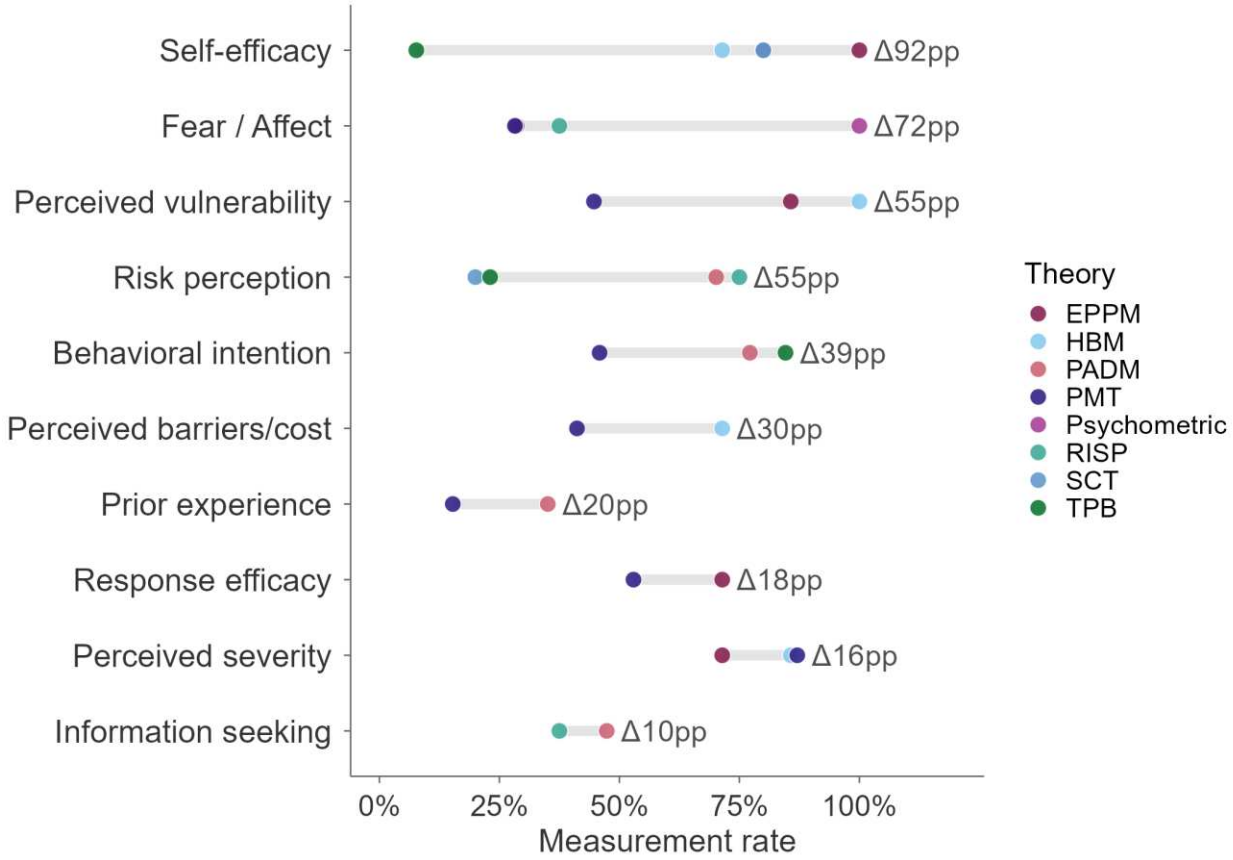
Construct operationalisation for the four core theories not shown in Figure 3. As Figure 3, for Cultural Theory of Risk, the Health Belief Model, Expected Utility Theory and Social Cognitive Theory. Points show the proportion of applications measuring each construct with Wilson 95% confidence intervals; filled circles denote core constructs and open diamonds extended constructs.

Extended Data Fig. 7



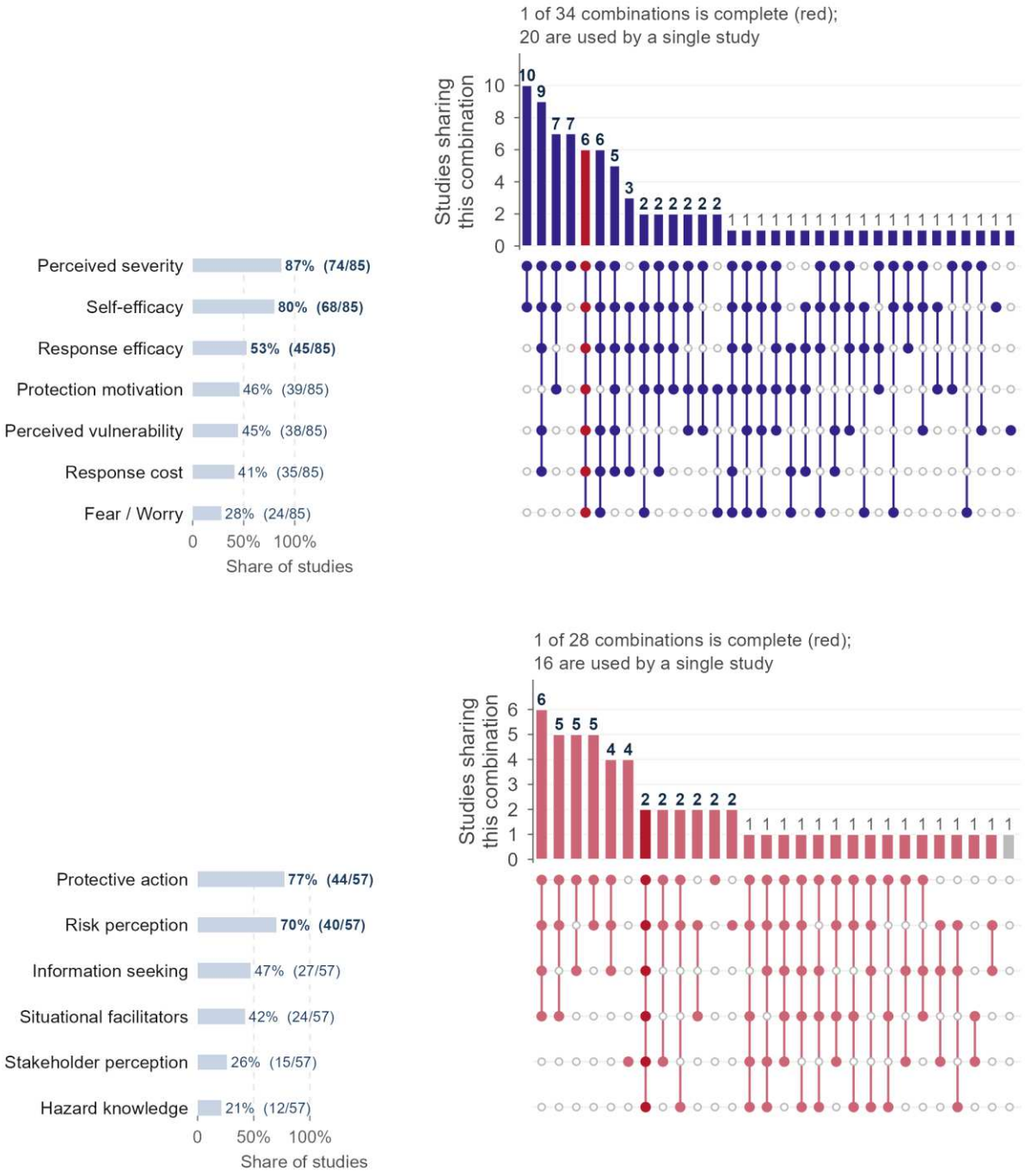
Hazard coverage by publication year. Number of studies per hazard category and publication year, showing when each hazard entered the theory-guided literature and how coverage has developed.

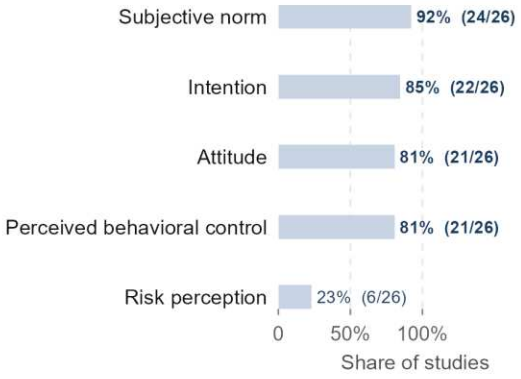
Extended Data Fig. 8

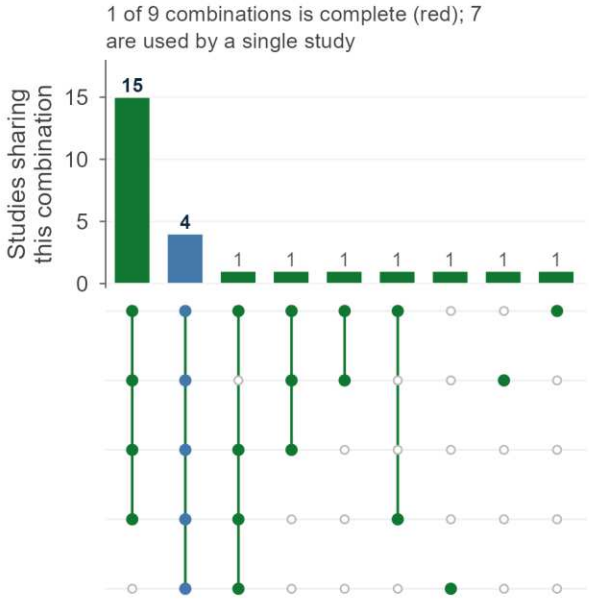


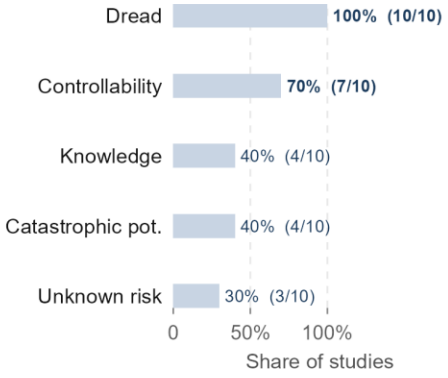
Cross-theory comparison of shared constructs. Measurement rates for constructs that appear in more than one core theory, plotted by theory. Differences in the rate at which the same construct is measured indicate that theoretical framing, rather than the construct itself, shapes what is operationalised.

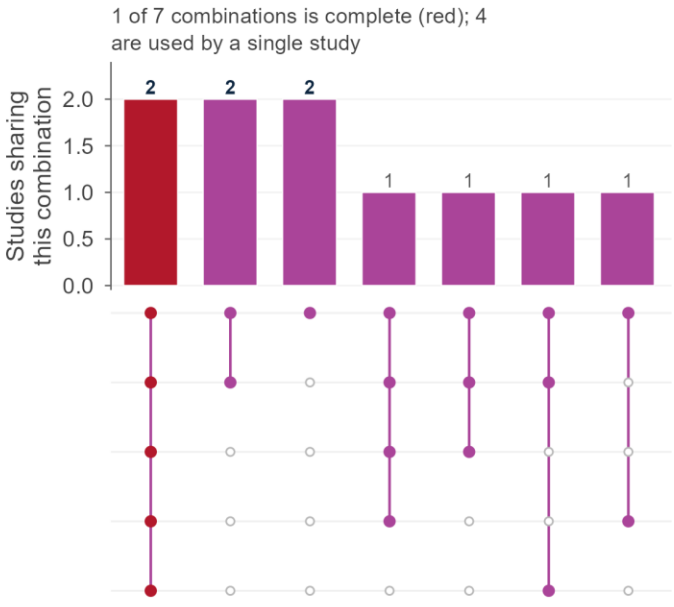
Extended Data Fig. 9




Subjective norm
92% (24/26)
Intention
85% (22/26)
Attitude
81% (21/26)
Perceived behavioral control
81% (21/26)
Risk perception
23% (6/26)
0
50%
100%
Share of studies

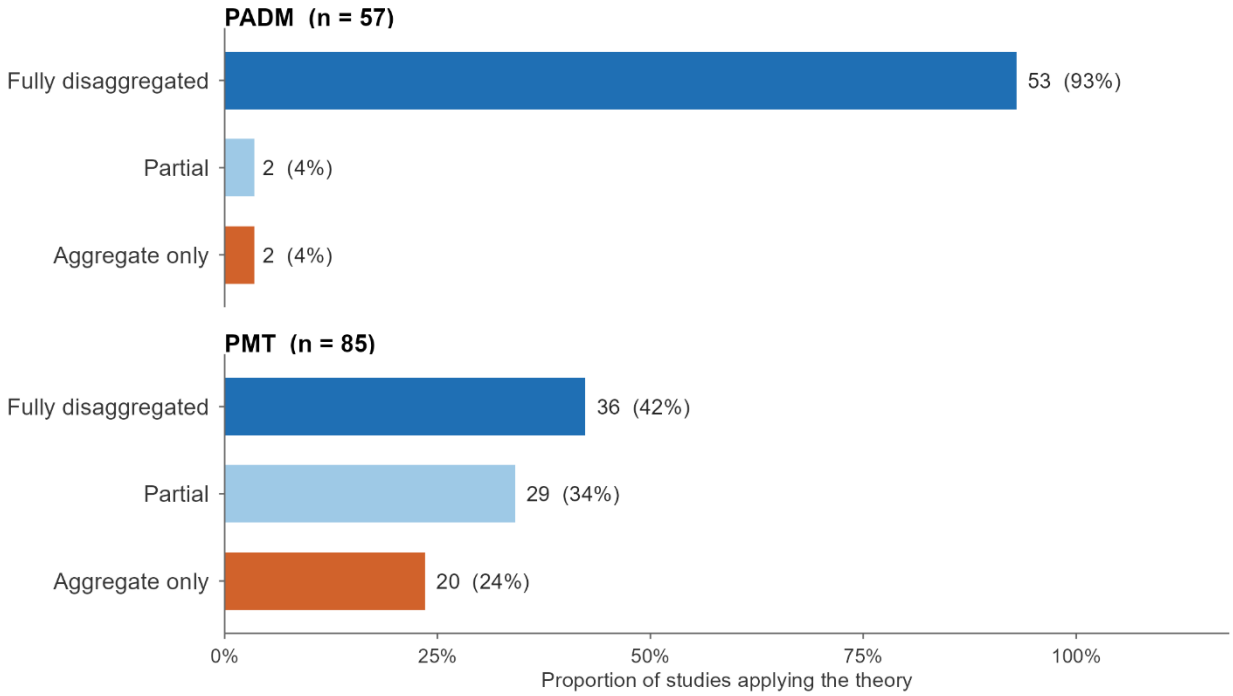

1 of 9 combinations is complete (red); 7 are used by a single study
Studies sharing this combination
15
10
5
0
15
4
1
1
1
1
1
1
1


Dread
100% (10/10)
Controllability
70% (7/10)
Knowledge
40% (4/10)
Catastrophic pot.
40% (4/10)
Unknown risk
30% (3/10)
0
50%
100%
Share of studies


1 of 7 combinations is complete (red); 4 are used by a single study
Studies sharing this combination
2.0
1.5
1.0
0.5
0.0
2
2
2
1
1
1
1

Fragmented operationalisation of the four most-applied theories. UpSet plots for a, PMT (n = 85 applications with codable construct descriptions); b, PADM (n = 57); c, TPB (n = 26); and d, the Psychometric Paradigm (n = 10). In each panel every column is a unique combination of the theory's core constructs; the top bar gives the number of studies sharing that combination, sorted most to least frequent, and the dot matrix shows which constructs are measured (filled circles) or omitted (open circles). The complete combination, where every core construct is measured, is highlighted. The left bar gives the overall proportion of studies measuring each construct.

Extended Data Fig. 10



Aggregate versus disaggregated operationalisation of PMT and PADM. Share of applications classified as fully disaggregated (all named sub-constructs), partially disaggregated, or aggregate-only (reporting bundle labels such as threat appraisal or coping appraisal without naming the component constructs).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformationSubmitted.docx
- releaseddatasetpaperlevel.csv
- S2searchterms.csv
- READMEreleaseddataset.md
- releaseddatasettheoryapplicationlevel.csv
- CodingInstruction.docx
- nccsubmissionfigurespublished.txt