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.

Separating experience reporting from content recall in dream reports across sleep stages

Peiyan Zhao

peiyan.jason@gmail.com

Brunel University London https://orcid.org/0009-0001-7172-9419

Article

Keywords: dream reports, white dreams, REM sleep, NREM sleep, measurement, data harmonisation

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11519432/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.

Abstract

Contentful dream reports are less frequent after non-rapid eye movement (NREM) than rapid eye movement (REM) awakening. This gap could reflect fewer reports of any experience or less recalled content within those reports. Distinguishing them requires retaining white dreams, experiences reported without recalled content. We audited the questions and codes behind all 20 datasets in the Dream EEG and Mentation database. Fourteen supported comparisons of any reported experience, ten supported content recall within experience reports, and one mapping merged white dreams with negative reports. A Bayesian hierarchical model decomposed the gap within six datasets supporting both comparisons at both stages. Averaged equally across datasets, contentful reports were 21.4 percentage points less frequent after NREM awakening. Most of this primary gap came from fewer reports of any experience: 16.6 points (95% credible interval 9.6 to 22.9), compared with 4.9 (− 4.1 to 12.3) for less recalled content. The experience-report contribution remained positive across weighting, sample and model checks. Content contributions depended on the datasets and weights averaged; both contributions were clearly positive in a four-dataset subset with at least ten records per stage. Uncertain-answer coding shifted the balance between negative reports and white dreams. Contrasting contentful with negative reports combined experience reporting with content recall. Theories must explain why any experience is reported less often after NREM awakening. Neural classifiers targeting any reported experience should retain white dreams in the positive class and assess content recall separately.

1 Introduction

Dream content is reported less often after awakening from non-rapid eye movement (NREM) than rapid eye movement (REM) sleep (Dement & Kleitman, 1957; Nielsen, 2000). Yet someone with no content to describe may still have the impression of having experienced something. The REM–NREM gap in contentful reports can therefore arise in two ways: NREM awakenings may yield fewer reports of any experience, or more of the reported experiences may lack recalled content.

Three familiar categories distinguish these outcomes (Fig. 1a): experience with recalled content (E), experience without recalled content (EWR, often called a white dream), and the category conventionally labelled no experience (NE; Siclari et al., 2013, 2017). We use experience report for E or EWR, meaning recalled content or an impression that an experience occurred. NE is the protocol's negative report. Depending on the dataset, it can mean no memory or impression, uncertainty, or an explicit denial of experience (Methods 2.3). These categories describe reports after awakening, which can reflect memory as well as the preceding experience (Windt et al., 2016).

The Dream EEG and Mentation (DREAM) database makes this distinction consequential for current neural research. It pools reports and pre-awakening electroencephalography (EEG) across laboratories to support comparisons of conscious and unconscious sleep, defines NE through absent recall and absent impression of experience, and acknowledges that negative reports may reflect failed recall (Wong et al., 2025a). Treating measures with the same name as measuring the same construct is the jingle fallacy (Flake & Fried, 2020). An original No Content category combining inability to remember with nothing having occurred, for example, was mapped to NE in DREAM (Zhang & Wamsley, 2019; Section 3.1). We examined each dataset's interview and coding to establish whether an impression without content could be retained and what a negative answer meant (Waterman et al., 1993; Windt, 2013; Demšar & Windt, 2024; Fortier et al., 2017).

Earlier work established the value of separating these outcomes. Elce et al. (2025) fitted two mixed models for morning reports: any reported experience versus a negative report, and recalled content versus white dreams within experience reports. Ataei et al. (2026), in a preprint, linked natural nocturnal awakenings to these two morning outcomes. Stucky (2025) reviewed three report categories and analysed participant-level stage proportions in DREAM version 4, describing a roughly even division of the NREM loss of contentful reports between white dreams and negative reports. DREAM's original three-outcome model used study and participant intercepts, with each stage coefficient held constant across studies (Wong et al., 2025a). These studies leave open whether pooled labels preserve the same report distinctions and how much each report probability contributes to the stage gap across protocols.

The distinction also bears on theories of NREM mentation. Early awakening studies established mental activity outside REM sleep (Foulkes, 1962). One-generator accounts propose shared processes for REM and NREM mentation, two-generator accounts propose distinct processes, and covert REM offers another account of NREM experiences (Nielsen, 2000). Each connects experience to the later report through memory and elicitation. Losing details while retaining an impression can yield a white dream. Fewer experiences, loss of the impression, or a stricter reporting criterion can yield a negative report. The stage difference in both report outcomes gives these accounts a shared set of observations to explain (Fig. 1).

We asked which contrasts DREAM's shared labels preserve and how much of the REM–NREM gap in contentful reports reflects fewer reports of any experience rather than less content recalled within those reports. Earlier white-dream research suggested less recalled content among NREM experience reports (Siclari et al., 2017; Fazekas et al., 2019), but we had no prespecified expectation about the relative contributions. We audited all 20 datasets and decomposed the gap within six eligible datasets before averaging. We then examined protocol differences, uncertain-answer coding and the consequences of omitting white dreams for neural comparisons.

a, A conceptual partition of final report codes. E denotes recalled content, EWR an experience without recalled content, and NE the protocol's negative report. The probabilities a and c concern any reported experience and recalled content within experience reports. b, The questions answered by common binary contrasts. c, Experience occurrence, memory and response criteria can affect the observed probabilities. Each protocol defines its own interview order and reference period; Methods 2.1 gives the full glossary and Supplementary Table A1 formalises reporting assumptions.

2 Methods

2.1 Report categories and the decomposition

Using the deposited codes (Table 1), we compared E + EWR with NE for experience reporting and E with EWR for content recall within experience reports (Fig. 1). Datasets that merge E and EWR contribute only to the experience-report contrast.

Table 1
Definitions used in the analysis

Term	Definition
E	Experience with recalled content
EWR	Reported experience without recalled content (white dream)
NE	The protocol's negative category, including absent recall or an explicit denial of experience and, in three datasets, uncertainty or perceived wakefulness
Experience-report probability a	P(E or EWR) among classified records
Conditional content probability c	P(E | E or EWR) among experience reports
Contentful-report gap	REM minus NREM probability of E, in percentage points (pp)
Experience reporting and content contributions	The two additive contributions to that gap, defined in the equations below
Stage association	REM-versus-NREM odds ratio (OR) for experience reporting or conditional content
Record and equal dataset weights	Weights proportional to retained records, or identical for each included dataset

The definitions distinguish report categories, stage associations and averaging targets. A contribution is an additive term in the contentful-report gap; the two terms are defined below. Negative-report meanings differ across protocols (Table 2).

For stage s, write a_s for the experience-report probability and c_s for the conditional content probability. Then

$$\mathrm{P}\left(\mathrm{E} \mid \mathrm{s}\right) = a_s c_s$$

$$\mathrm{P}\left(\mathrm{EWR} \mid \mathrm{s}\right) = a_s(1 - c_s)$$

$$\mathrm{P}\left(\mathrm{NE} \mid \mathrm{s}\right) = 1 - a_s$$

E versus all other reports measures the product of experience-report probability and conditional content probability. EWR/(E + EWR) equals one minus conditional content probability. EWR/(EWR + NE) uses a different denominator and depends on both probabilities.

To express their contributions on the same absolute scale, we used the standard symmetric decomposition of a difference between two products (Kitagawa, 1955; Das Gupta, 1993, Section 2.2). We also calculated both one-sided decompositions by weighting the experience-report difference by NREM or REM conditional content (Supplementary Table D40). Using R and N to denote the two stages,

$$\Delta_E = a_R c_R - a_N c_N = D_A + D_C$$

$$D_A = \frac{(a_R - a_N)(c_R + c_N)}{2}$$

$$D_C = \frac{(c_R - c_N)(a_R + a_N)}{2}$$

D_A is the experience-report contribution and D_C the content contribution. Each multiplies one stage difference by the average of the other probability. The contributions may oppose one another but sum exactly to the gap. Supplementary Section A1 illustrates the arithmetic. Supplementary Sections A2–A3 give the measurement assumptions and content-availability selection bounds.

Each contribution depends on baseline report rates as well as on stage associations. Identical odds ratios can divide the gap differently in different datasets (Supplementary Section A1). Memory and response criteria can affect both probabilities.

2.2 DREAM data and records

2.2.1 Sources and record selection

We used all 20 datasets in DREAM version 9 (Wong et al., 2025b), first accessed on 30 July 2026 and rechecked on 9 October 2026. Versioned inputs and hashes accompany the code. This secondary analysis collected no new participant data. Dataset denotes a deposited data unit, protocol its study procedures, and source the contributing study or documentation. Supplementary Material B documents all 20 datasets.

The six modelled datasets differed in setting and awakening procedure (Table 2). LODE collected home morning reports, TWC USA used morning laboratory naps, and Sleep Talking used morning diaries after laboratory nights. SCAN sampled frequent dream recallers during laboratory nights, DATA1 used serial laboratory awakenings, and Aamodt evening used serial alarms during evening sleep.

We retained the highest-amendment row for each dataset–case identity, reducing 3,191 export rows to 2,669 record identities. Analysis fields agreed across all 261 repeated identities. Later amendments added 59 identities. Record identities are the analysis units. Metadata do not establish a one-to-one link with physiological awakenings. Excluding combined NE/EWR, Unknown, missing-stage and Wake-labelled records left 2,280 classified sleep records (Supplementary Figure B1). The six-dataset sample comprised 178 REM and 588 NREM records, including seven REM white dreams from six participant clusters.

Table 2
Observed report categories and study designs

Dataset	Design	Elicitation route	Meaning of NE	REM E/EWR/NE	NREM E/EWR/NE
LODE Set 3 Elce et al. (2023)	Home; spontaneous final-morning reports	Report-based coding	No memory or impression	46/3/14 n = 63	72/11/44 n = 127 [7/118/2]
TWC USA Set 4 Konkoly et al. (2021)	Morning lab naps; lucidity cues; mixed awakenings	Report-based coding	No recall or perceived failure to sleep	18/1/1 n = 20	8/0/4 n = 12 [6/6/0]
SCAN Set 8 Eichenlaub et al. (2018)	Lab night; frequent recallers; induced early N3/later REM	Report-based coding	Nothing remembered; impression coded EWR	44/0/2 n = 46	19/7/13 n = 39 [0/0/39]
DATA1 Set 16 Wong et al. (2020)	Lab; induced early-night serial awakenings	Structured question	Dreamless; uncertain answers excluded	37/1/4 n = 42	70/62/112 n = 244 [0/75/169]
Sleep Talking Set 10 DREAM contributor description (Set 10)	Lab night; final-morning diary	Report-based coding	No recall or uncertainty without a dream feeling	2/1/2 n = 5	9/0/8 n = 17 [1/16/0]
Aamodt evening Set 17 Aamodt et al. (2023)	Lab evening; serial alarms; rose 3 h early	Structured question	Explicit no experience, distinct from EWR	1/1/0 n = 2	108/26/15 n = 149 [10/95/44]
All six	—	—	—	148/7/23 n = 178	286/106/196 n = 588

Counts are E/EWR/NE in the retained common sample; n is the stage total. Brackets give NREM N1/N2/N3 counts. LODE, TWC USA, SCAN and DATA1 form the four-dataset sensitivity analysis. Elicitation route is independent of NE meaning. Detailed participant characteristics, awakening procedures, report timing and source evidence are in Supplementary Sections B1–B2.

2.2.2 Participant identifiers

We clustered by dataset–participant identity because 51 participant identifier strings recurred across datasets and cross-study linkage was unavailable. The deduplicated export contained 508 such participant clusters; the People export identifies contributors rather than participants.

2.2.3 Sleep stage and awakening procedures

We used the deposited Last sleep stage as supplied by each source and combined N1, N2 and N3 as NREM. Four of the 20 deposited descriptions named a scoring standard: three explicitly named the American Academy of Sleep Medicine (AASM), and one cited Rechtschaffen and Kales (1968) together with Iber et al. (2007). Supplementary Material B documents experimenter-induced, spontaneous and mixed awakenings and the period each report was intended to describe. Table 2 gives the retained category counts.

LODE (Set 3; Elce et al., 2023) requested experience immediately before final awakening and excluded E/EWR reports referring only to earlier periods. E/EWR reports were retained only within one hour of awakening. This limit did not apply to NE. Human reinspection followed automated headband staging. The source lists eleven final-epoch corrections (Supplementary Section B1).

2.3 Documentary audit

The experience-report contrast required E or EWR to remain distinguishable from NE. Conditional content additionally required separate E and EWR codes and a documented question, instruction or assignment rule capable of producing EWR. A documented route remained eligible with zero observed EWR records. We recorded contrast eligibility, the elicitation route and the meaning of NE separately.

We began with each dataset's highest-amendment Experimental description and consulted linked methods and supplied native files. Native checks covered Sets 1, 3, 5, 12, 13, 16, 18 and 20. Sources, study links and unresolved mappings are documented in Supplementary Material B.

Among the datasets supporting the experience-report contrast, we distinguished three elicitation routes: a structured question separating no experience from white dreams; coding from reported memory or impressions without such a question; or unresolved. A structured question may use a fixed follow-up, conditional clarification or a three-way choice. The NE column of Table 3 records whether negative answers include uncertainty or perceived wakefulness. Table 3 separately flags content-based coding that lost the EWR boundary and samples selected for dream reports. Table 3 and Supplementary Table B4 also record what the initial question targets and how answers became final codes.

Each category contrast was eligible, ineligible or indeterminate according to whether its boundary was documented, lost through coding or selection, or insufficiently specified. We recorded question wording, follow-up, reference period, coding role and negative-report meaning, together with the content threshold. Five modelled datasets accept any recalled mental content, whereas Sleep Talking does not specify its threshold (Supplementary Material B). A category can reflect the participant's memory or impression even when a researcher assigns its final code. We recorded this basis separately from whether the interview used a structured question.

2.4 Statistical analyses

We fitted one hierarchical categorical model to the six eligible datasets (Sections 2.4.1–2.4.2). We predicted each dataset's report probabilities at both stages and decomposed its gap before averaging (Section 2.4.3). The primary target averages the six datasets' contributions equally. Record-weighted summaries and the four-dataset subset with ≥ 10 records per stage are secondary targets (Table 4).

2.4.1 Common sample and report likelihood

The joint model used the six datasets in Table 2 (766 records, 112 clusters). The joint decomposition requires both category boundaries and both stages. Datasets supporting only experience reporting enter the ten-dataset auxiliary model (Section 2.5.3). We also fitted the four-dataset subset (593 records, 75 clusters), selected post hoc by requiring ≥ 10 records per stage to reduce extreme stage extrapolation. This subset retained 171 REM and 422 NREM records, including five REM white dreams from four clusters.

For record i, a_i and c_i denote experience reporting and conditional content probabilities. One categorical likelihood assigns E, EWR and NE probabilities $a_i c_i$, $a_i(1 - c_i)$ and $1 - a_i$, respectively. Both category contrasts use the same sample.

2.4.2 Dataset variation and participant effects

Both reporting logits included dataset-specific intercepts and stage slopes and participant random intercepts nested within datasets. Stage was zero for REM and one for NREM. Dataset slopes followed normal distributions. Dataset intercepts had Normal(0, 2.5²) priors, slope means Normal(0, 1.5²), slope standard deviations Half-Normal(1), and participant standard deviations Half-Normal(1.5). These weakly informative priors regularise stage effects and participant variation. An LKJ(2) prior allowed the two participant intercepts to correlate. Supplementary Sections D1 and D12 give the hierarchy and correlation estimates. Records were conditionally independent given the predictors and effects, and stage was observational.

All analyses used Python 3.12.14. Models were fitted with the No-U-Turn sampler in PyMC 6.3.2 (Abril-Pla et al., 2023), using four chains with 2,000 warm-up and 2,000 retained draws (samples from the fitted parameter distribution) per chain (4,000 each for the participant-slope model). Diagnostics were calculated with ArviZ 1.3.0 (Kumar et al., 2019). All fits met R-hat < 1.01, bulk and tail effective sample sizes ≥ 400, and no divergent transitions (Supplementary Section D3.2).

2.4.3 Standardisation and additive contributions

Within each posterior draw, we predicted every retained record at both stages, keeping its dataset and participant effects. Predicting an unobserved stage within a participant or dataset requires extrapolation. We averaged E, EWR and NE probabilities within datasets over the same records at both stages. Experience reporting is E + EWR, and conditional content is E/(E + EWR). Records with a higher predicted probability of an experience report therefore contribute more to the conditional-content average.

Equal protocol weights target the finite mean of the six selected protocols (Seaman et al., 2014), giving each protocol equal influence despite differences in record counts. Sleep Talking and Aamodt evening together receive one third of the weight and rely heavily on model-based predictions for REM. Record weights target the retained record mixture, of which DATA1 and Aamodt evening supply 37% and 20%. Refits use their retained records and renormalised weights.

We report posterior means and 95% equal-tailed credible intervals (CrIs), which contain 95% of the posterior probability for each estimated quantity. The experience-report share, D_A/Δ_E, is summarised by its median and CrI. Shares can fall outside 0–100% when contributions have opposite signs and become unstable when the gap is near zero. We emphasise absolute contributions.

Hierarchical mean log-odds ratios are summarised by posterior means, and their odds ratios by posterior medians. Hierarchical log-odds associations describe a typical protocol, whereas record-standardised absolute contributions use probabilities averaged over the same retained records within each protocol. Their intervals can differ in apparent certainty because they describe different targets.

2.5 Sensitivity and secondary analyses

2.5.1 Sensitivity analyses and descriptive comparisons

These checks addressed dependence on the sample and averaging target, model assumptions, and the timing and sequence of reports. We refitted the six-dataset hierarchy after omitting each dataset in turn, retaining its likelihood and priors. Other checks restricted NREM to N2, added timing, report history or participant stage slopes, or retained only participants observed at both stages (Supplementary Sections D5–D6 and D11). One sample restriction excluded TWC USA and Sleep Talking. A supplementary six-dataset refit compared the existing record-standardised target with integration over the model's correlated participant-effect distribution, using 40-point Gauss–Hermite quadrature per dimension. A separate model constrained content stage slopes to be common. We compared its conditional white-dream count predictions with the varying-slope fit (Supplementary Section D12.4). For each model, the posterior predictive P value is the fraction of replications whose discrepancy exceeds the observed discrepancy. These checks reuse most records, so the range of posterior means describes sensitivity to analysis choices.

We also refitted the six-dataset model with Half-Normal slope-SD scales of 0.5 and 2, in addition to the primary scale of 1. For new-protocol predictions, we integrated the normal distribution defined by each saved draw's hierarchical mean and slope SD, inverted the averaged cumulative distribution at .025, .5 and .975, and exponentiated these quantiles (Supplementary Table D42).

We grouped protocols as experimenter-induced (SCAN, DATA1 and Aamodt evening), final-morning (LODE and Sleep Talking), or mixed (TWC USA). Within each posterior draw, we averaged dataset contributions equally within each group (Supplementary Table D33).

Figure 2 displays raw category proportions for all six modelled datasets, including sparse cells. Descriptive white-dream comparisons used dataset–stage cells with at least 20 records. Gap-share comparisons required at least 20 records at each stage. These rules did not apply to model inclusion or Fig. 2. Supplementary Table B2 includes every content-eligible dataset.

For DREAM, we obtained raw intervals by resampling participants with all their records kept together (a participant-cluster bootstrap). Published aggregate counts used multinomial bootstraps. Supplementary Tables D19–D20 compare observed and model-based contributions and category changes for all six datasets and report non-estimable sparse-cell resamples. Experiment 2 of Siclari et al. (2017) supplies a descriptive three-category benchmark.

Secondary common-slope logistic models, participant-stratified analyses and dataset-cluster uncertainty checks are detailed in Supplementary Material C (McNeish, 2026; MacKinnon et al., 2025; Webb, 2023). They target different estimands from the joint hierarchy.

2.5.2 Category changes and reporting assumptions

For the category comparison, we calculated the NREM-minus-REM increase in NE minus the corresponding increase in EWR within each posterior draw, then averaged equally across datasets (Supplementary Table D16). This difference is denoted G in the supplementary tables. When the contentful-report gap is positive, $G > 0$ corresponds to an NE share above one half. For NE's share of the decline, we distinguish the ratio of mean category changes from the median and CrI of draw-wise ratios (Supplementary Table D41). These summaries describe redistribution among categories, separately from the two product-decomposition contributions.

An exploratory, unadjusted DATA1 check compared N2 and N3 conditional white-dream proportions using a participant-cluster percentile bootstrap (Supplementary Table C6). Bias-corrected and accelerated (BCa) participant-cluster bootstrap intervals for uncertain-answer rates use 100,000 paired-stage resamples and delete-participant jackknife acceleration (Efron, 1987; Supplementary Table D39).

To address selection into experience reports (Section 3.4.3; Hernán et al., 2004), we calculated partial-identification bounds on the stage difference in content available at collection (Manski, 2003), under the reporting assumptions in Supplementary Section A3. Availability includes encoding and retrieval up to the report.

For the category-boundary scenarios, each dataset, stage and draw moved a specified fraction of NE probability to EWR while holding E fixed. REM and NREM fractions independently ranged from zero to one in 0.05 steps (441 scenarios per target). We recomputed contributions within datasets before applying the same fixed weights. This deterministic transformation holds fitted participant effects and partial pooling fixed. Reassigning observations in a refit could alter both. We examined the NE-minus-EWR increase, the experience-report contribution and the difference between the two contributions separately (Supplementary Section D9).

2.5.3 Auxiliary models and uncertain answers

The ten-dataset auxiliary model used 1,383 classified records from ten experience-report-eligible REM/NREM datasets, respecting separate or merged EWR labels. Over the same retained records and weights, we predicted the E-versus-NE and E + EWR-versus-NE stage gaps. The effect of omitting EWR is the first gap minus the second, averaged equally over the seven sources storing EWR separately (Supplementary Tables D26–D27 and D36).

Further checks changed record weights, participant weights, the slope prior, or the sample to datasets with at least five participants observed at each stage. That sample refit includes five sources with separate EWR labels. The evening and morning Aamodt datasets contribute only two and one REM participant clusters, and Children dreaming and Young adults have no participants observed at both stages. Aamodt morning's eligibility for the broad conditional-content contrast remains indeterminate. This exploratory comparison retains its native labels (Supplementary Sections D10 and D12).

For uncertain answers, we added DATA1's 34 and Multiple awakenings' 21 joint NE/EWR records to the ten-dataset auxiliary model and refitted the model under four assignments. Uncertain answers at each stage were assigned wholly to NE or wholly to EWR. Multiple awakenings contributes only its observed combined-positive-versus-negative likelihood, so no E/EWR split was created. The resolved records, priors and hierarchy were unchanged. Each fit standardised over its expanded record mixture. Supplementary Section D11 retains the raw arithmetic and Supplementary Section D12.6 gives model results.

2.6 Analysis chronology and verification records

All analyses were exploratory and unregistered. The audit began before the analysis results were known. Eligibility rules were formalised after frequency inspection, later reassessments used the results then available, and the reporting hierarchy was chosen after the analyses had been run (Supplementary Section D3.1).

The author verified sources, code, numerical results and key inferences. Re-running the analysis package in a separate computing environment reproduced the reported outputs. Supplementary Sections D3.1 and E summarise the analysis chronology and verification.

3 Results

We first establish which report contrasts the datasets support. We then ask whether experience reporting contributes to the stage gap, whether it contributes more than content recall, and whether negative reports increase more than white dreams. Finally, we examine what changes when white dreams are omitted.

3.1 Shared labels support different comparisons

Of DREAM's 20 datasets, 14 preserved a comparison of any reported experience with NE, and 10 also preserved recalled content versus EWR within experience reports (Table 3). Among the 14, five documented a structured awakening question separating no experience from white dreams, six coded reported memory or impressions, and three left that response route unresolved. Three broadened NE to include uncertainty or perceived wakefulness.

Six datasets supported both contrasts and contained REM and NREM records (Table 2). DATA1 and Aamodt evening used structured awakening questions, whereas the other four coded reported memory or impressions. NE also included perceived wakefulness in TWC USA and uncertainty in Sleep Talking, which did not specify the report reference period.

Zhang and Wamsley (2019) show how a shared label can lose a distinction. Their original No Content category combined inability to remember with nothing having occurred. Two deposited examples described experience without recalled content but carried NE codes (Supplementary Table B6). Mapping No Content to NE erased the white-dream boundary. We excluded this dataset from both contrasts.

Table 3
Documentary support for report contrasts across DREAM

Dataset and set	Elicitation route	Meaning of NE	Initial question target	Experience report	Content	Broadened NE
REM Turku (Set 6)	Structured question	No dream experience after clarification	Last image, then whole dream	Eligible	Eligible	No
Siclari serial (Set 13)	Structured question	Explicit no-experience alternative	Most recent experience	Eligible	Eligible	No
DATA1 (Set 16)	Structured question	Dreamless; uncertain answers excluded	Dream experience; defined as every image or thought	Eligible	Eligible	No
Aamodt evening (Set 17)	Structured question	Explicit no experience, distinct from EWR	Last thing through the mind; any mental activity	Eligible	Eligible	No
MEG Kyushu (Set 20)	Structured question	No experience after recall follow-up	Dreaming; defined as any mental experience	Eligible	Eligible	No
LODE (Set 3)	Report-based coding	No memory or impression	Everything through the mind before awakening	Eligible	Eligible	No
SCAN (Set 8)	Report-based coding	Nothing remembered; impression coded EWR	What was going through the mind	Eligible	Eligible	No
Young adults (Set 11)	Report-based coding	Absent recall under a feeling-of-dreaming instruction	Feeling of having dreamed	Eligible	Ineligible	No
TWC USA (Set 4)	Report-based coding	No recall or perceived failure to sleep	Initial wording not reported	Eligible	Eligible	Yes
Children dreaming (Set 7)	Report-based coding	No recall or uncertainty without a dream feeling	Feeling of dreaming; initial wording unreported	Eligible	Ineligible	Yes
Sleep Talking (Set 10)	Report-based coding	No recall or uncertainty without a	Feeling of dreaming; initial	Eligible	Eligible	Yes

Dataset and set	Elicitation route	Meaning of NE	Initial question target	Experience report	Content	Broadened NE
		dream feeling	wording unreported			
Oudiette N1 (Set 2)	Route unresolved	Nothing through the mind	What went through the mind	Eligible	Eligible	No
Donders (Set 5)	Route unresolved	Not documented	Not documented	Indeterminate	Indeterminate	—
Multiple awakenings (Set 9)	Route unresolved	Negative experience judgment; combined answers excluded	Dream experience; defined as any distinct mental activity	Eligible	Ineligible	No
Older adults (Set 12)	Route unresolved	REC/NREC mapping unresolved	All remembered mental activity	Indeterminate	Indeterminate	—
Aamodt morning (Set 18)	Route unresolved	No experience; exact EWR route unresolved	Last thing through the mind	Eligible	Indeterminate	No
Zhang–Wamsley (Set 1)	Content-based coding	Absence of described content; includes EWR-like reports	Any mental activity; recall prompt	Ineligible	Ineligible	—
UFRN (Set 14)	Content-based coding	No remembered dream	Whether a dream is remembered	Ineligible	Ineligible	—
Kumral audiobook (Set 19)	Content-based coding	No recalled mental content	What was going through the mind; recall prompts	Ineligible	Ineligible	—
Motor tDCS (Set 15)	Dream-selected sample	General negatives not represented	Everything through the mind, then bodily experience	Ineligible	Ineligible	—

Structured question denotes a fixed follow-up, conditional clarification or three-way choice distinguishing no experience from white dreams. Report-based coding assigns labels from reported memory or impressions. Initial-question targets retain each protocol's broad definitions. In the final column, Yes marks broadened NE definitions that include uncertainty or perceived wakefulness; No marks the other experience-report-eligible datasets, and dashes mark ineligible or indeterminate experience boundaries. Eligibility concerns the report

contrast; the NE column separately records the meaning of a negative answer. Source evidence is in Supplementary Tables B1 and B4.

3.2 Fewer experience reports account for most of the primary gap

Observed E, EWR and NE proportions before modelling. Labels give records (n) and participants (k) at each stage; the same person may contribute to both. Hatched, paler bars have fewer than 10 records. Segments below 6% are unlabelled; exact counts are in Methods 2.2.1.

With equal weights across the six protocols, the gap in contentful reports was 21.4 percentage points (pp). Fewer reports of any experience contributed 16.6 pp (95% CrI 9.6 to 22.9), and less recalled content within experience reports contributed 4.9 pp (− 4.1 to 12.3; Table 4).

The experience-report contribution exceeded the content contribution by 11.7 pp (95% CrI 1.6 to 22.1) and had a median share of 76% (53 to 136%; Supplementary Tables D14 and D23). For a positive total gap, shares above 100% mean that a negative content contribution offsets part of the experience-report contribution. It remained larger on average under both one-sided decompositions, but not clearly so when NREM conditional content supplied the weight (Supplementary Table D40). Its share was also not clearly above one half when protocols were weighted by record counts (Supplementary Table D14).

Table 4
Samples and contributions to the REM–NREM contentful-report gap

Analysis target	Records / clusters	REM EWR records / clusters	Total gap (pp)	Experience-report contribution (pp)	Content contribution (pp)
Six protocols Equal weights	766 / 112	7 / 6	21.4 (10.2 to 31.6)	16.6 (9.6 to 22.9)	4.9 (− 4.1 to 12.3)
≥ 10 records per stage Four protocols; equal weights	593 / 75	5 / 4	32.9 (24.7 to 41.3)	20.2 (13.7 to 27.1)	12.7 (7.6 to 18.6)
Six protocols Record weights	766 / 112	7 / 6	26.8 (16.4 to 35.7)	16.7 (11.7 to 21.3)	10.2 (1.1 to 17.2)
Participants at both stages Equal protocol weights	583 / 61	6 / 5	26.5 (14.0 to 37.5)	17.2 (10.1 to 24.5)	9.3 (− 1.2 to 18.5)
Participants at both stages Record weights	583 / 61	6 / 5	35.9 (27.8 to 43.4)	19.8 (14.3 to 25.2)	16.1 (10.3 to 21.3)
Integrated participant effects Equal protocol weights	766 / 112	7 / 6	20.9 (9.4 to 31.2)	16.3 (9.4 to 22.9)	4.6 (− 4.8 to 12.3)
Exclude TWC USA and Sleep Talking Equal protocol weights	712 / 81	5 / 4	27.3 (16.0 to 36.0)	15.2 (9.3 to 20.3)	12.1 (1.5 to 18.8)

Posterior means and 95% CrIs in pp. Contributions sum to the gap before rounding. Records / clusters gives retained records and dataset–participant clusters; REM EWR gives white-dream records / contributing clusters. The primary target averages the six protocols equally. The ≥ 10-per-stage subset contains LODE, TWC USA, SCAN and DATA1. Both-stage rows retain only participants observed at REM and NREM. Integration averages over the model's participant distribution. The final row excludes TWC USA's broadened NE category and Sleep Talking's unspecified report reference period. These checks change the sample or averaging target; full specifications are in Supplementary Tables D30 and D34.

3.3 Experience-report contributions persist across analysis choices

The pooled experience-report contribution ranged from 11.5 to 20.6 pp across weighting schemes, sample restrictions and model specifications, and every 95% CrI excluded zero. For a typical protocol, experience

reporting was more likely after REM than NREM awakening (log odds ratio 1.52, 95% CrI 0.48 to 2.42; Table 4; Supplementary Tables D5 and D8–D9).

The four datasets with at least ten REM records each had higher observed experience-report probabilities after REM than NREM awakening (Fig. 2). Sleep Talking and Aamodt evening agreed in sign but had only five and two REM records. All six modelled means were positive (Fig. 3), as was the pooled contribution in every leave-one-dataset-out refit. The pooled experience-report contribution was also positive when we averaged experimenter-induced, final-morning and mixed awakening procedures separately (Supplementary Table D33).

Experience reporting also differed within NREM sleep. In secondary logistic models with dataset intercepts and a shared stage association, it was lower in N3 than N2 (OR 0.51, 95% confidence interval 0.30 to 0.89). The corresponding conditional-content estimate was uncertain (OR 0.93, 0.52 to 1.67; Supplementary Table C3; Methods 2.5).

Filled circles show dataset means and 95% CrIs from the six-dataset model; open circles show observed contributions and participant-cluster bootstrap intervals. Diamonds show equal-protocol summaries from the six-dataset fit and the ≥ 10-per-stage four-dataset fit. The lower group shows complete leave-one-dataset-out refits with equal weights. Figures 3 and 4 use identical horizontal scales. Raw estimates use the observed records at each stage; model estimates standardise both stages to the same records. Bootstrap intervals use estimable resamples. Sources: Supplementary Tables D9 and D19.

3.4 Content contributions depend on the protocols averaged

3.4.1 Protocol differences and averaging choices

SCAN and DATA1 had the clearest positive content estimates, with experience reporting and content recall each contributing roughly half of the gap (Fig. 4). The symmetric decomposition gave median experience-report shares of 53.9% in SCAN and 47.5% in DATA1, compared with 77.0% in LODE (Supplementary Table D14).

Pooled mean content contributions ranged from − 0.3 to 16.3 pp across the same checks. Omitting DATA1 reduced the equal-protocol content contribution to − 0.3 pp (95% CrI − 11.1 to 8.3). Excluding TWC USA and Sleep Talking (Section 3.1) instead increased it to 12.1 pp (1.5 to 18.8), retaining LODE, SCAN, DATA1 and Aamodt evening.

Dataset estimates, pooled targets and complete leave-one-dataset-out refits follow Fig. 3, on the same horizontal scale. Filled circles show model means and 95% CrIs; open circles show observed contributions and participant-cluster bootstrap intervals; diamonds show the six-protocol and ≥ 10-per-stage four-protocol summaries. Sources: Supplementary Tables D9 and D19.

Protocols differed in how strongly content recall depended on stage. Simulated white-dream counts matched the observed counts better when this association varied across protocols than when it was shared (posterior predictive P = 0.362 versus 0.025; Supplementary Section D12.4). For a typical protocol, the stage association was uncertain (log odds ratio 0.55, 95% CrI − 1.14 to 1.98). The magnitude of protocol variation was also poorly determined. The between-protocol standard deviation had a posterior mean of 1.65 under the primary prior and ranged from 0.9 to 2.5 under narrower or wider priors (Supplementary Tables D5 and D17).

3.4.2 Sleep depth, awakening procedure, timing and report history

Sleep depth, awakening procedure, elicitation route and time of night varied together across protocols. The protocols with the clearest content estimates sampled NREM early in the night, mostly in N3. All 39 SCAN NREM records and 69% of DATA1's were N3, whereas LODE and Sleep Talking collected morning reports mainly from N2 and TWC USA used morning naps. Restricting the comparison to N2 also removes SCAN, combining a change of sleep depth with a change of protocols (Supplementary Table D8).

When we grouped protocols by awakening procedure, experimenter-induced awakenings had a mean content contribution of 10.3 pp (95% CrI − 4.3 to 21.0). Means were near zero for final-morning reports (− 0.8 pp) and TWC USA's mixed procedure (− 0.2 pp). All three 95% intervals included zero (Supplementary Table D33).

DATA1 showed no clear increase in white dreams among experience reports in N3 compared with N2. The N3-minus-N2 difference was − 8.1 pp, with a 95% interval of − 25.0 to 5.1 from resampling participants (Supplementary Table C6). Adjusting DATA1 for night, awakening order and elapsed time left its mean content contribution unchanged at 25.2 pp. Adding report history changed the pooled division little and improved predictions for consecutive reports from the same participant and night (Supplementary Tables D4 and D7–D8). These checks leave the separate roles of depth, procedure and timing unresolved.

3.4.3 Content availability among experiences

Content is observed only after an experience report, so REM and NREM comparisons may select different subsets of experiences. We used bounds to assess whether differential reporting alone could explain the conditional-content gap, assuming that experience reports refer to actual experiences and that available content always produces an E report (Supplementary Section A3). Under these assumptions, DATA1's lower bound for the REM–NREM difference in content available at collection was 23.1 pp (95% CrI 8.8 to 36.7). Across the pooled targets, the lower bound's interval was above zero only in the record-weighted four-dataset summary (9.7 pp; 1.4 to 17.5; Supplementary Table A2). DATA1 contributed nearly half the records in that target. Reweighting the same fit without DATA1 gave − 2.2 pp (95% CrI − 10.8 to 6.6; Supplementary Table D22).

3.5 Sparse cells and coding shape the balance of negative reports and white dreams

We next asked how the loss of contentful reports was distributed between NE and EWR. With the deposited codes, NE increases accounted for about 92% of the primary loss. This category share differs from the 76% experience-report contribution share in Section 3.2. An NE increase equals the full fall in experience-report probability, whereas the decomposition multiplies that fall by average conditional-content probability. It therefore assigns 16.6 pp to experience reporting, compared with the 19.6 pp NE increase (Methods 2.1; Table 4; Supplementary Table D15).

The negative-report share was 70–78% under the other six- and four-dataset weighting targets. Calculated within each posterior draw rather than from mean changes, the primary share had a median of 90% (95% CrI 64 to 166%; Supplementary Table D41). With a positive contentful-report gap, this category share can exceed 100%

when white-dream reports decline. The mean increase in NE exceeded the increase in white dreams by 17.8 pp (6.0 to 30.2; Supplementary Tables D15–D16).

Sparse REM samples helped shape this balance. Sleep Talking and Aamodt evening predicted fewer white dreams after NREM awakening, whereas DATA1 and SCAN predicted more (Supplementary Table D20). Together the two sparse protocols supplied about two-thirds of the white-dream probability in the primary REM summary. The median negative-report share of the NREM fall in contentful reports was 62.1% in SCAN and 64.5% in DATA1, compared with 85.5% in LODE (Supplementary Table D41).

Two datasets recorded uncertain answers jointly as NE/EWR. Their observed rates were higher after NREM than REM awakening: a difference of 9.6 pp in DATA1 (95% bootstrap confidence interval − 0.1 to 14.5) and 6.0 pp in Multiple awakenings (3.1 to 10.1). Stage-specific rates, counts, participant numbers and resampling details are in Supplementary Table D39 and Methods 2.5.

Coding uncertain answers shifted DATA1's NREM loss between negative reports and white dreams. We refitted the ten-dataset auxiliary model, which includes sources that merged E and EWR, with uncertain answers assigned wholly to NE or wholly to EWR at each stage (four assignments; Methods 2.5). With resolved answers only, DATA1's NE increase exceeded its EWR increase by 12.0 pp (95% CrI − 4.1 to 28.9). The difference rose to 21.4 pp (4.8 to 38.0) when all uncertain answers were NE and fell to 0.0 pp (− 15.2 to 16.8) when all were EWR (Fig. 5). DATA1's experience-report contribution stayed positive in all four assignments; its content contribution varied with the coding (Supplementary Table D37).

a–b, DATA1 estimates from the ten-dataset auxiliary model, first for resolved answers and then after adding uncertain answers under four REM/NREM assignments. Labels give the uncertain-answer code at REM / NREM. c–d, Six-protocol equal-weight estimates along specified transfers of NE probability to EWR: NREM only (grey dashed) or the same fraction at both stages (purple). The top row refits the model after adding recorded answers; the bottom row transforms fitted probabilities. Points and lines show posterior means, with 95% CrIs as intervals or shading. Sources: Supplementary Tables D23 and D37.

Hypothetically reassigning 10% of NREM negative-report probability to white dreams made both the difference between contributions and the NE-minus-EWR increase uncertain. The experience-report contribution then exceeded the content contribution by 5.1 pp (95% CrI − 4.4 to 15.2), and the NE increase exceeded the EWR increase by 11.0 pp (− 0.3 to 23.0; Supplementary Table D23). Reassigning half at both stages still left an experience-report contribution of 7.0 pp (4.1 to 9.5; Fig. 5).

3.6 Omitting white dreams changes the estimated stage difference

Omitting white dreams changes the contrast algebraically. At each stage, the odds of E versus NE equal the odds of any experience report multiplied by the conditional content probability (Fig. 1a). The REM–NREM odds ratio combines the experience-report odds ratio with the REM-to-NREM ratio of conditional content probabilities. It is larger when conditional content is higher after REM. The two targets coincide where E and EWR are coded together.

The REM–NREM probability gap was 2.0–7.6 pp larger after omitting EWR than for the experience-report contrast across the auxiliary checks (posterior means). The increase was 2.7 pp (95% CrI – 1.0 to 4.8) when the seven sources with separate EWR codes were averaged equally (Supplementary Table D36; see Supplementary Section D13.5 for the odds identity and its relation to probability gaps).

4 Discussion

We found that shared dream-report labels can conceal different measurements. Across eligible protocols, fewer reports of any experience accounted for most of the primary REM–NREM gap and remained a positive contribution across analysis choices. The content contribution depended more on the protocols averaged, while uncertain-answer coding changed the balance between negative reports and white dreams. These findings inform accounts of NREM mentation (Section 4.1) and the targets used in neural classification (Section 4.3).

4.1 What the report gap requires theories to explain

The main theoretical question is why any experience, including a bare impression, is reported less often after NREM awakening. One-generator, two-generator and covert-REM accounts can address this through differences in experience occurrence, encoding, retrieval or reporting criteria (Nielsen, 2000). Each account must connect sleep processes to both report probabilities. Physiology within NREM can test covert-REM proposals, while length-controlled report comparisons address the richness of experiences whose content was recalled (Antrobus, 1983; Foulkes & Schmidt, 1983).

We separated the gap into experience-report and content contributions. Distinguishing the processes behind them, namely experience occurrence, memory and response criteria (Fig. 1c), requires interventions within a stage (Section 4.4) or the reporting assumptions in Supplementary Table A1. Under the reporting assumptions in Section 3.4.3, differing tendencies to report an experience without available content could not alone explain lower content recall after NREM awakening in DATA1.

White dreams have been interpreted as weakly differentiated experience, forgetting, or a metacognitive feeling of having dreamed (Fazekas et al., 2019; Nemeth, 2023; Alcaraz Sánchez, 2026). Sparse REM cells and uncertain-answer coding changed the estimated balance between white dreams and negative reports, so theories based on white-dream rates also depend on how this boundary is measured.

Metacognition could also lower experience reporting. A weaker feeling of having dreamed after NREM awakening would reduce reports without any change in the preceding experience. The higher NREM rate of uncertain answers in Multiple awakenings offers a clue to this possibility. Randomising how experience is elicited after awakening would test for a reporting pathway, while keeping stage and preceding physiology comparable. Such an intervention cannot alter the preceding experience, so a shift in reports or confidence would show that elicitation affects reporting. Unchanged reports would leave occurrence and memory explanations open.

4.2 Relation to earlier reports

Both Siclari et al. (2017, experiment 2) and Stucky (2025) described a roughly even division of the NREM loss between negative reports and white dreams. SCAN and DATA1 are closer to this pattern than the pooled DREAM

summary. Equal weights allow sparse protocols to influence that summary as much as protocols with more observations. Weighting and coding choices help explain why a pooled DREAM summary can differ from an individual study.

We weighted the change in experience reporting by mean conditional content within each dataset (Section 3.5), so nearly equal category changes need not imply equal contributions. In Siclari's experiment 2, the experience-report contribution accounts for about 35% of the gap, leaving about two-thirds for content recall. The negative-report increase accounts for 54%; multiplying this by mean REM and NREM conditional content of 0.64 gives the 35% contribution share (Supplementary Tables D19 and D41). Reconstructing counts from Stephan et al.'s (2021) published percentages under the assumptions in Supplementary Material F also gave a larger content contribution. Results from Kumar et al. (2025), who used dream-focused labels, varied by night and NREM substage.

Elce et al. (2025) found that vulnerability to interference predicted white dreams relative to contentful morning reports. That outcome corresponds to the conditional-content contrast here and is compatible with a role for memory in its contribution. Their E + EWR outcome, described as an estimate of generated dreams, corresponds to our experience-report contrast. Ataei et al. (2026), in a preprint, likewise found different associations of nocturnal awakenings with morning experience reporting and content recall. We identified which of these contrasts each protocol supports and then quantified how the two report probabilities form the stage gap, allowing baseline rates and stage associations to differ across protocols.

Sleep-depth results likewise depend on the outcome. In the present analyses, experience reporting was lower in N3 than N2, whereas DATA1 showed no clear N3–N2 difference in the white-dream proportion among experience reports. Possible explanations for lower experience reporting include fewer experiences, sleep inertia affecting the retained impression, and a stricter reporting criterion after deep-sleep awakening. In DREAM's original model, contentful reports became less likely from N1 to N3 relative to both negative reports and white dreams. The balance between white dreams and negative reports showed no clear depth trend (Wong et al., 2025a). Stucky (2025) found no significant N2–N3 difference. These studies compared different report categories and protocol mixtures, so their depth patterns need not agree.

4.3 Using report definitions in neural comparisons

We linked each usable report contrast to its source questions, answer paths and final codes (Supplementary Section B4). Other databases can use the same audit to check whether shared labels support comparable measurements, addressing the jingle fallacy identified by Flake and Fried (2020). A machine-readable crosswalk gives dataset-specific guidance for EEG analyses across all 20 datasets (Supplementary Section B5).

In DREAM's initial EEG classification analysis (Wong et al., 2025a, Fig. 5), E versus NE omits white dreams where they are separately coded. Re-running eligible classifiers with EWR in the positive class would target any reported experience. A separate E-versus-EWR classifier would target recalled content within experience reports. Three-category models retain all outcomes.

Within-laboratory EEG studies link the difference between white dreams and negative reports to low-frequency and slow-wave features. They link the difference between contentful reports and white dreams to high-

frequency and frontal slow-wave features (Siclari et al., 2017, 2018). With matched cohorts, preprocessing and validation, classifiers can be trained once for E versus NE and once for E + EWR versus NE.

Features that primarily track content recall should contribute less to held-out prediction in the second target, whereas features that track any experience should keep their predictive value. Comparing held-out contributions avoids confounding changes in raw coefficients with correlations among features.

4.4 Studies that can distinguish experience, memory and reporting

Within a sleep stage, randomising question wording, pre-interview practice or post-awakening interference could test how elicitation and retrieval affect each report probability. Siclari's experiment 2 included two weeks of home questionnaire practice, which might help participants retain or report a bare impression. A shift from NE to EWR with little change in E would locate the effect at the boundary altered by uncertain-answer coding. Case-linked confidence in experience and content would support response-confidence analyses (Demšar & Windt, 2024). Separating sensitivity from criterion also needs an accuracy reference or explicit latent-state assumptions (Maniscalco & Lau, 2012).

Sleep inertia and interference after awakening may affect access to content or the impression of experience (Trotti, 2017; Vallat et al., 2019, 2020). Studies can test these possibilities by varying awakening procedure within a protocol at comparable stage and timing, or by randomising report delay. Recording delay, elapsed sleep and sleep inertia alongside both report probabilities would help distinguish content-specific changes from a general fall in experience reporting.

Richer response options can retain reports of minimal experiences before harmonisation. Using Michalak et al.'s (2026) N2 categories, contentful experiences would map to E, both rich and simple white dreams to EWR, and explicit no-experience responses to NE. Sense of presence should retain a minimal-experience subcode on the positive side. Its E/EWR assignment depends on whether recalled presence or time passing meets the declared content threshold. Checklist 1 records the core information needed to preserve these distinctions.

Prospective multisite studies are needed to estimate how the experience-report association varies between protocols. The six protocols supplied little information about this variation, so its estimated distribution still largely reflected the prior assumptions (Supplementary Table D17). For a new protocol, the 95% predictive interval for the odds ratio was about 1.2–19 with the narrowest prior, 0.5–32 with the primary prior and 0.2–51 with the widest (Supplementary Table D42).

Future studies should preregister both report contrasts and the dataset and participant structure, and neural analyses should separate local spectral power, microevents and global stage probabilities. Studies will need many REM experience reports to obtain enough white dreams for these comparisons. In dataset–stage cells with at least 20 records, white dreams comprised 0–6.1% of REM experience reports (Supplementary Table B2), while Siclari's experiment 2 reached 19.1% (Supplementary Section D8). As an illustration, comparing 6% with 13% requires about 275 independent experience reports per stage for 80% power at two-sided $\alpha = .05$. At 6%, this would yield about 17 REM white dreams, versus five in the four-dataset subset. Multisite planning should use protocol-specific rates and simulate repeated observations and dataset variation (Supplementary Section D8).

Checklist 1 Core information for post-awakening interviews

Core items

1. Exact initial and follow-up wording, definitions and requested reference period.
2. Initial answer, each follow-up answer and final code, including uncertainty.
3. Meaning of NE and separate rules for EWR and any white-dream subtypes.
4. Counts by stage and category, including excluded or unclassifiable answers.
5. Awakening method, reporting delay, stage-scoring procedure and scorer blinding. For automated staging, include final-epoch agreement matrices.

Additional measures

Separate confidence in experience and recalled content (Demšar & Windt, 2024).

Minimal-experience probes and rich/simple white-dream subtypes.

Sleep onset, time of night, awakening order and preceding awakenings (Picard-Deland et al., 2023).

4.5 Limitations

Protocols differed in elicitation and negative-report meaning, and two had broader or less specified negative reports (Section 3.1). LODE applied a reporting-delay restriction only to E/EWR (Methods 2.2.3), which can raise retained NE proportions and alter stage contrasts in either direction.

A single human adjudicator made all audit judgments, some with results known (Methods 2.6). Native checks covered eight datasets, and documentation remained uneven.

Rare REM white dreams weaken dataset-specific precision and bootstrap calibration (Table 4; Methods 2.2.1). DATA1 supplied most EWR observations, and each dataset recording uncertainty had just one uncertain REM answer. Cross-dataset participant linkage was unavailable, so clustering could not account for any participants shared across datasets (Methods 2.2.2).

Only 61 of 112 clusters were observed at both stages, so predictions for unobserved stages rely on the assumed distributions of dataset and participant effects (Methods 2.4.2). These predictions supplied 91.4% of Aamodt evening's REM target and 54.5% of Sleep Talking's (Supplementary Table D32). Retaining only participants observed at both stages, or averaging over the model's participant distribution, still gave experience-report contributions of 17.2 and 16.3 pp (Table 4).

Timing and report-history adjustment may leave residual within-night dependence, and the ten-dataset auxiliary models omitted those terms (Section 3.4.2). We compared external studies descriptively because report definitions, denominators and participant dependence were not consistently comparable across sources (Section 4.2; Supplementary Material F).

Stage scoring varied across sources, and LODE supplied about one-quarter of the records from automated headband staging with correction of clear final-epoch errors. Stage errors can blur contrasts or, when related to report type, bias them. The requested report window can also extend beyond the final scored epoch.

Participants, stage depth, timing and elicitation varied together. SCAN selected frequent dream recallers, TWC USA used lucidity cues during morning naps, and Aamodt evening followed rising three hours earlier than usual. In DATA1, serial NREM awakenings produced selective REM pressure and occasional sleep-onset REM intrusions (Wong et al., 2020). This sampling differs from ordinary later-night REM. The conclusions concern the included protocols and their report procedures.

4.6 Conclusion

Shared dream-report labels should be checked against the questions that produced them before pooling. In the six eligible DREAM datasets, fewer reports of any experience accounted for 16.6 of the 21.4 percentage points in the REM–NREM gap, whereas the content contribution depended on the protocols averaged. Neural studies of any reported experience should keep white dreams in the positive class. Within-stage interventions on elicitation or retrieval can test why experience is reported less often after NREM awakening.

5 Data availability

DREAM version 9 is available at https://doi.org/10.26180/22133105.v9 (Wong et al., 2025b). The analysis archive listed in Code availability contains versioned inputs and documentary evidence. Its audit records link category-contrast eligibility, negative-report meaning, reference periods and evidence locators to Supplementary Tables B1, B2 and B4. Supplementary Materials A–D give detailed methods and results. Supplementary Material E summarises verification and reproduction.

6 Code availability

Analysis archive: https://doi.org/10.5281/zenodo.23664363. The accompanying package contains custom analysis code, model specifications, saved posterior draws, figure scripts and the reproduction entry point reproduce_all.py. It distinguishes deterministic reproduction from refitting and records their software environments.

Declarations

CRediT authorship contribution statement. Peiyan Zhao: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, Writing – review and editing, Project administration.

Competing interests. The author declares no competing interests.

Funding. This research received no specific grant from a funding agency in the public, commercial or not-for-profit sectors.

Declaration of generative AI and AI-assisted technologies. The author is grateful to AI models. After originating the research question, the author discussed it with Claude (Anthropic) and narrowed it to the present report-measurement problem. Throughout the research, the author used ChatGPT (OpenAI) and Claude (Anthropic) to assist with code editing, language polishing, methodological critique, internal consistency and computational checks. Neither system is an author or an independent human adjudicator. The author is responsible for the evidence, audit judgments, analyses, interpretation and final submitted text.

Ethics. This secondary analysis used openly released, de-identified DREAM records. It involved no participant contact or access to identifiable participant information, and no additional ethical review was sought. Contributing studies obtained their own approvals and confirmed participant consent for sharing and secondary use. The DREAM project was approved by the Monash University Human Research Ethics Committee (Project ID 31269).

References

1. Aamodt, A., Sevenius Nilsen, A., Markhus, R., Kusztor, A., HasanzadehMoghadam, F., Kauppi, N., Thürer, B., Storm, J. F., & Juel, B. E. (2023). EEG Lempel-Ziv complexity varies with sleep stage, but does not seem to track dream experience. Frontiers in Human Neuroscience, 16, 987714. https://doi.org/10.3389/fnhum.2022.987714
2. Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C. C., Martin, O. A., Osthege, M., Vieira, R., Wiecki, T., & Zinkov, R. (2023). PyMC: A modern, and comprehensive probabilistic programming framework in Python. PeerJ Computer Science, 9, e1516. https://doi.org/10.7717/peerj-cs.1516
3. Alcaraz Sánchez, A. (2026). Did I have a dream last night? White dreaming as metacognitive feelings. Mind & Language, 41(4), 450–468. https://doi.org/10.1111/mila.70027
4. Antrobus, J. (1983). REM and NREM sleep reports: Comparison of word frequencies by cognitive classes. Psychophysiology, 20(5), 562–568. https://doi.org/10.1111/j.1469-8986.1983.tb03015.x
5. Ataei, S., Jafarzadeh Esfahani, M., Axmacher, N., Dresler, M., & Schoch, S. F. (2026). How awakenings shape dream recall: A multilevel study (Version 1) [Preprint]. bioRxiv. https://doi.org/10.64898/2026.05.20.722678
6. Das Gupta, P. (1993). Standardization and decomposition of rates: A user's manual (Current Population Reports, P23-186). U.S. Bureau of the Census. https://www2.census.gov/library/publications/1993/demographics/p23-186.pdf
7. Dement, W., & Kleitman, N. (1957). The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. Journal of Experimental Psychology, 53(5), 339–346. https://doi.org/10.1037/h0048189
8. Demšar, E., & Windt, J. (2024). Studying dream experience through dream reports: Points of contact between dream research and first-person methods in consciousness science. In D. Gregory & K. Michaelian (Eds.), Dreaming and Memory: Philosophical Issues (pp. 85–117). Springer. https://doi.org/10.1007/978-3-031-68204-9_5
9. Efron, B. (1987). Better bootstrap confidence intervals. Journal of the American Statistical Association, 82(397), 171–185. https://doi.org/10.1080/01621459.1987.10478410
10. Eichenlaub, J.-B., van Rijn, E., Gaskell, M. G., Lewis, P. A., Maby, E., Malinowski, J. E., Walker, M. P., Boy, F., & Blagrove, M. (2018). Incorporation of recent waking-life experiences in dreams correlates with frontal theta activity in REM sleep. Social Cognitive and Affective Neuroscience, 13(6), 637–647. https://doi.org/10.1093/scan/nsy041
11. Elce, V., Bergamo, D., Avvenuti, G., Bellesi, M., & Bernardi, G. (2023). LODE [Data set]. figshare. https://doi.org/10.6084/m9.figshare.22147085

12. Elce, V., Bergamo, D., Bontempi, G., Pedreschi, B., Bellesi, M., Handjaras, G., & Bernardi, G. (2025). The individual determinants of morning dream recall. Communications Psychology, 3, 25. https://doi.org/10.1038/s44271-025-00191-z
13. Fazekas, P., Nemeth, G., & Overgaard, M. (2019). White dreams are made of colours: What studying contentless dreams can teach about the neural basis of dreaming and conscious experiences. Sleep Medicine Reviews, 43, 84–91. https://doi.org/10.1016/j.smrv.2018.10.005
14. Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. Advances in Methods and Practices in Psychological Science, 3(4), 456–465. https://doi.org/10.1177/2515245920952393
15. Fortier, I., Raina, P., Van den Heuvel, E. R., Griffith, L. E., Craig, C., Saliba, M., Doiron, D., Stolk, R. P., Knoppers, B. M., Ferretti, V., Granda, P., & Burton, P. R. (2017). Maelstrom Research guidelines for rigorous retrospective data harmonization. International Journal of Epidemiology, 46(1), 103–115. https://doi.org/10.1093/ije/dyw075
16. Foulkes, D., & Schmidt, M. (1983). Temporal sequence and unit composition in dream reports from different stages of sleep. Sleep, 6(3), 265–280. https://doi.org/10.1093/sleep/6.3.265
17. Foulkes, W. D. (1962). Dream reports from different stages of sleep. Journal of Abnormal and Social Psychology, 65(1), 14–25. https://doi.org/10.1037/h0040431
18. Hernán, M. A., Hernández-Díaz, S., & Robins, J. M. (2004). A structural approach to selection bias. Epidemiology, 15(5), 615–625. https://doi.org/10.1097/01.ede.0000135174.63482.43
19. Iber, C., Ancoli-Israel, S., Chesson, A. L., Jr., & Quan, S. F. (2007). The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications. American Academy of Sleep Medicine.
20. Kitagawa, E. M. (1955). Components of a difference between two rates. Journal of the American Statistical Association, 50(272), 1168–1194. https://doi.org/10.1080/01621459.1955.10501299
21. Konkoly, K. R., Appel, K., Chabani, E., Mangiaruga, A., Gott, J., Mallett, R., Caughran, B., Witkowski, S., Whitmore, N. W., Mazurek, C. Y., Berent, J. B., Weber, F. D., Türker, B., Leu-Semenescu, S., Maranci, J.-B., Pipa, G., Arnulf, I., Oudiette, D., Dresler, M., & Paller, K. A. (2021). Real-time dialogue between experimenters and dreamers during REM sleep. Current Biology, 31(7), 1417–1427.e6. https://doi.org/10.1016/j.cub.2021.01.026
22. Kumar, G., Naaz, S., Jabin, N., Sasidharan, A., Nagendra, R. P., Yadav, R., & Kutty, B. M. (2025). Neurophysiological features of dream recall and the phenomenology of dreams: Auditory stimulation impacts dream experiences. Consciousness and Cognition, 132, 103869. https://doi.org/10.1016/j.concog.2025.103869
23. Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. (2019). ArviZ a unified library for exploratory analysis of Bayesian models in Python. Journal of Open Source Software, 4(33), 1143. https://doi.org/10.21105/joss.01143
24. MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. D. (2025). Cluster-robust jackknife and bootstrap inference for logistic regression models (Version 2) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2406.00650
25. Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Consciousness and Cognition, 21(1), 422–430.

26. Manski, C. F. (2003). Partial identification of probability distributions. Springer.
27. McNeish, D. (2026). A practical guide to selecting and blending approaches for clustered data: Clustered errors, multilevel models, and fixed-effect models. Psychological Methods, 31(2), 225–251. https://doi.org/10.1037/met0000620
28. Michalak, A., Marzoli, D., Pietrogiacomi, F., Bergamo, D., Elce, V., Pedreschi, B., Mosca, G., Navari, A., Emdin, M., Ricciardi, E., Handjaras, G., & Bernardi, G. (2026). Immersive NREM2 dreaming preserves subjective sleep depth against declining sleep pressure. PLOS Biology, 24(3), e3003683. https://doi.org/10.1371/journal.pbio.3003683
29. Nemeth, G. (2023). The route to recall a dream: theoretical considerations and methodological implications. Psychological Research, 87, 964–987. https://doi.org/10.1007/s00426-022-01722-7
30. Nielsen, T. A. (2000). A review of mentation in REM and NREM sleep: “Covert” REM sleep as a possible reconciliation of two opposing models. Behavioral and Brain Sciences, 23(6), 851–866. https://doi.org/10.1017/S0140525X0000399X
31. Picard-Deland, C., Konkoly, K., Raider, R., Paller, K. A., Nielsen, T., Pigeon, W. R., & Carr, M. (2023). The memory sources of dreams: Serial awakenings across sleep stages and time of night. Sleep, 46(4), zsac292. https://doi.org/10.1093/sleep/zsac292
32. Rechtschaffen, A., & Kales, A. (Eds.). (1968). A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. U.S. Government Printing Office.
33. Seaman, S., Pavlou, M., & Copas, A. (2014). Review of methods for handling confounding by cluster and informative cluster size in clustered data. Statistics in Medicine, 33(30), 5371–5387. https://doi.org/10.1002/sim.6277
34. Siclari, F., Baird, B., Perogamvros, L., Bernardi, G., LaRocque, J. J., Riedner, B., Boly, M., Postle, B. R., & Tononi, G. (2017). The neural correlates of dreaming. Nature Neuroscience, 20(6), 872–878. https://doi.org/10.1038/nn.4545
35. Siclari, F., Bernardi, G., Cataldi, J., & Tononi, G. (2018). Dreaming in NREM sleep: A high-density EEG study of slow waves and spindles. Journal of Neuroscience, 38(43), 9175–9185. https://doi.org/10.1523/JNEUROSCI.0855-18.2018
36. Siclari, F., LaRocque, J. J., Postle, B. R., & Tononi, G. (2013). Assessing sleep consciousness within subjects using a serial awakening paradigm. Frontiers in Psychology, 4, 542. https://doi.org/10.3389/fpsyg.2013.00542
37. Stephan, A. M., Lecci, S., Cataldi, J., & Siclari, F. (2021). Conscious experiences and high-density EEG patterns predicting subjective sleep depth. Current Biology, 31, 5487–5500. https://doi.org/10.1016/j.cub.2021.10.012
38. Stucky, B. (2025). We are the sensors of consciousness! A review and analysis on how awakenings during sleep influence dream recall. Nature and Science of Sleep, 17, 709–729. https://doi.org/10.2147/NSS.S506461
39. Trotti, L. M. (2017). Waking up is the hardest thing I do all day: Sleep inertia and sleep drunkenness. Sleep Medicine Reviews, 35, 76–84. https://doi.org/10.1016/j.smrv.2016.08.005
40. Vallat, R., Meunier, D., Nicolas, A., & Ruby, P. (2019). Hard to wake up? The cerebral correlates of sleep inertia assessed using combined behavioral, EEG and fMRI measures. NeuroImage, 184, 266–278.

https://doi.org/10.1016/j.neuroimage.2018.09.033

41. Vallat, R., Nicolas, A., & Ruby, P. (2020). Brain functional connectivity upon awakening from sleep predicts interindividual differences in dream recall frequency. Sleep, 43(12), zsaa116. https://doi.org/10.1093/sleep/zsaa116
42. Waterman, D., Elton, M., & Kenemans, J. L. (1993). Methodological issues affecting the collection of dreams. Journal of Sleep Research, 2(1), 8–12. https://doi.org/10.1111/j.1365-2869.1993.tb00053.x
43. Webb, M. D. (2023). Reworking wild bootstrap-based inference for clustered errors. Canadian Journal of Economics/Revue canadienne d'économique, 56(3), 839–858. https://doi.org/10.1111/caje.12661
44. Windt, J. M. (2013). Reporting dream experience: Why (not) to be skeptical about dream reports. Frontiers in Human Neuroscience, 7, 708. https://doi.org/10.3389/fnhum.2013.00708
45. Windt, J. M., Nielsen, T., & Thompson, E. (2016). Does consciousness disappear in dreamless sleep? Trends in Cognitive Sciences, 20(12), 871–882. https://doi.org/10.1016/j.tics.2016.09.006
46. Wong, W., Andrillon, T., Decat, N., Herzog, R., Noreika, V., Valli, K., Windt, J., & Tsuchiya, N. (2025b). The DREAM database (Version 9) [Data set]. Monash University. https://doi.org/10.26180/22133105.v9
47. Wong, W., Herzog, R., Andrade, K. C., Andrillon, T., de Araujo, D. B., Arnulf, I., Ataei, S., Avvenuti, G., Baird, B., Bellesi, M., Bergamo, D., Bernardi, G., Blagrove, M., Decat, N., Demirel, Ç., Dresler, M., Eichenlaub, J.-B., Elce, V., Gais, S., … Tsuchiya, N. (2025a). A dream EEG and mentation database. Nature Communications, 16, 7495. https://doi.org/10.1038/s41467-025-61945-1
48. Wong, W., Noreika, V., Móró, L., Revonsuo, A., Windt, J., Valli, K., & Tsuchiya, N. (2020). The Dream Catcher experiment: Blinded analyses failed to detect markers of dreaming consciousness in EEG spectral power. Neuroscience of Consciousness, 2020(1), niaa006. https://doi.org/10.1093/nc/niaa006
49. Zhang, J., & Wamsley, E. J. (2019). EEG predictors of dreaming outside of REM sleep. Psychophysiology, 56(7), e13368. https://doi.org/10.1111/psyp.13368

Figures

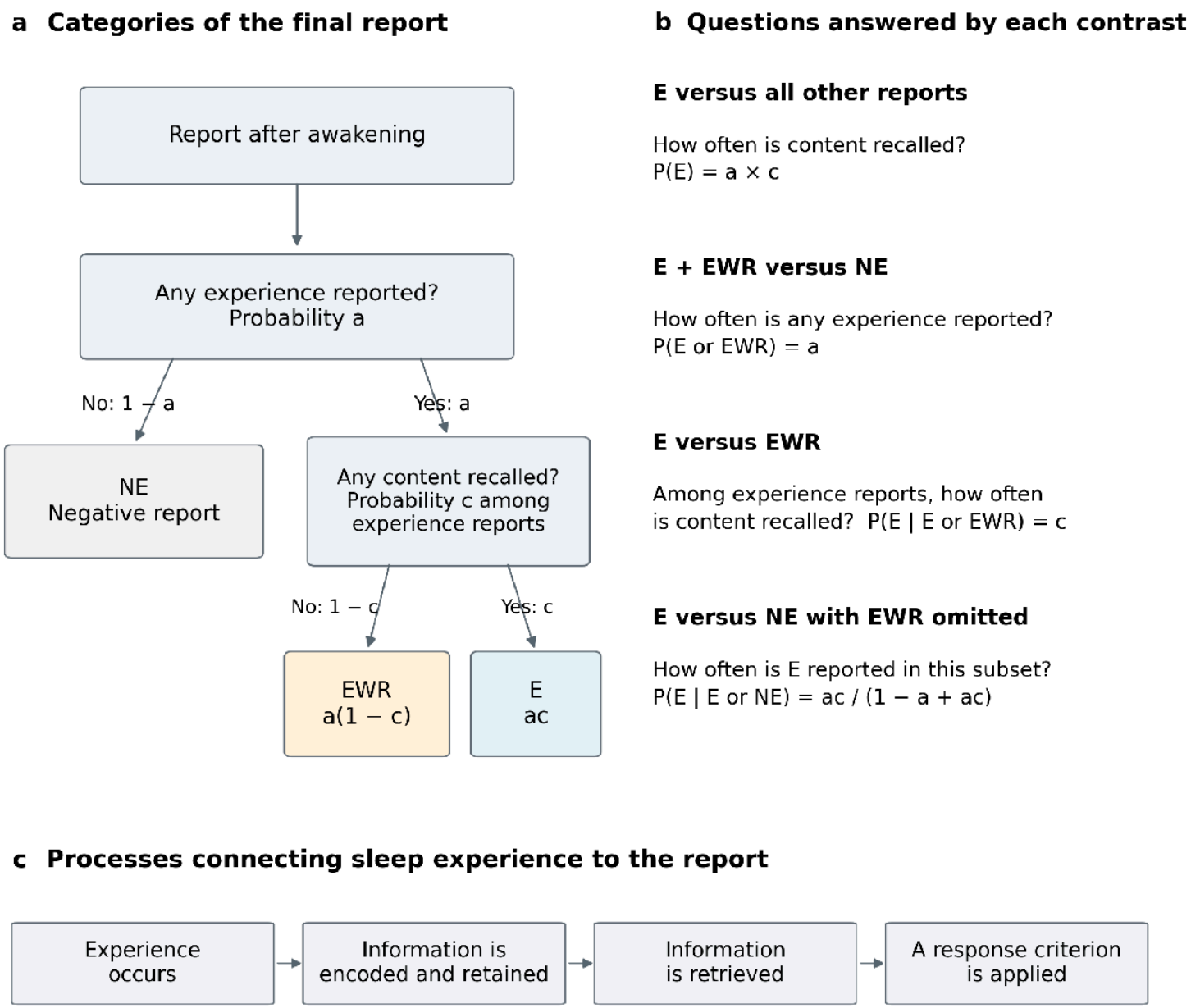


Figure 1

Report contrasts and the processes connecting experience to a report

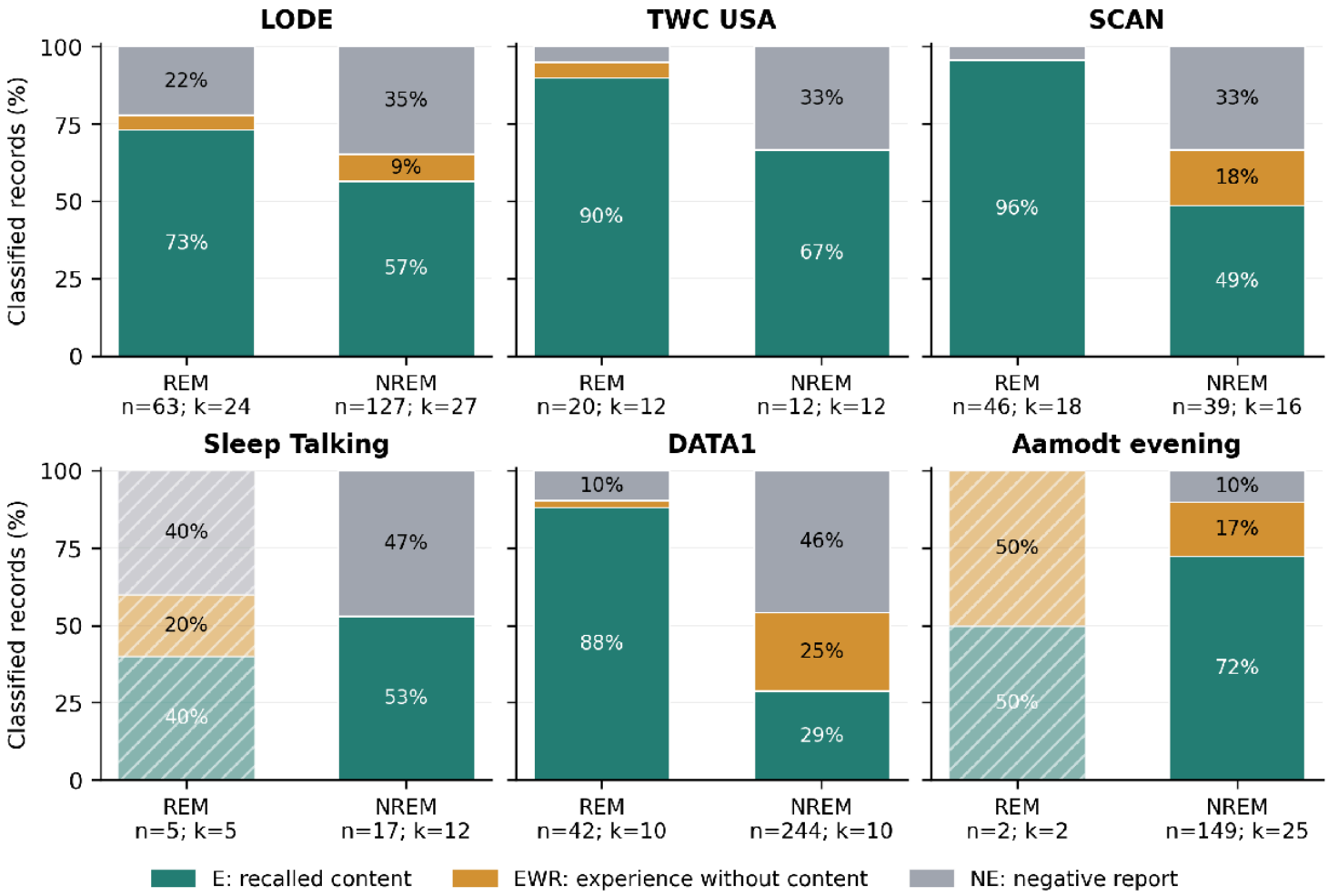


Figure 2

Observed report categories by dataset and stage

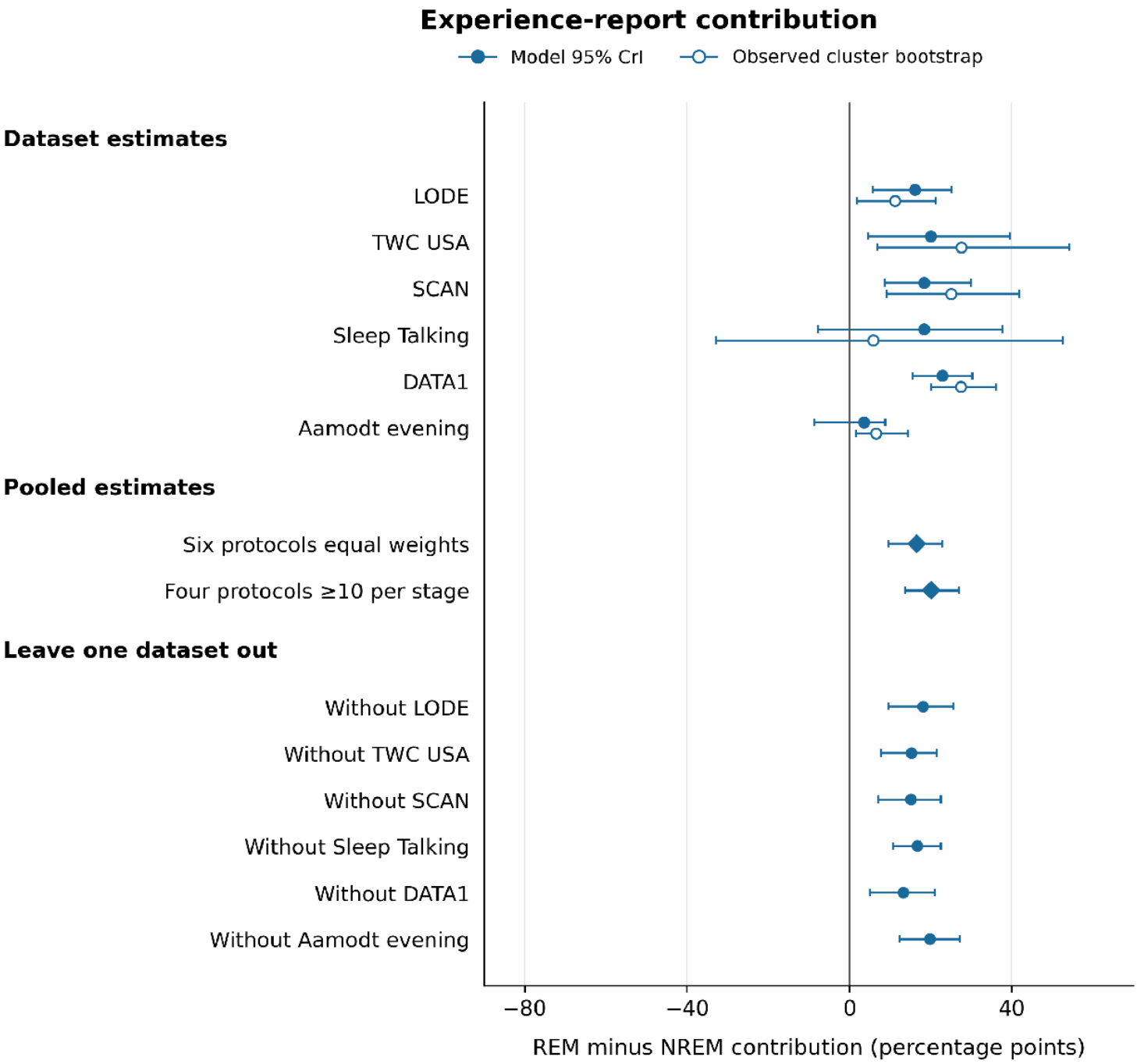


Figure 3

Experience-report contributions by dataset and analysis

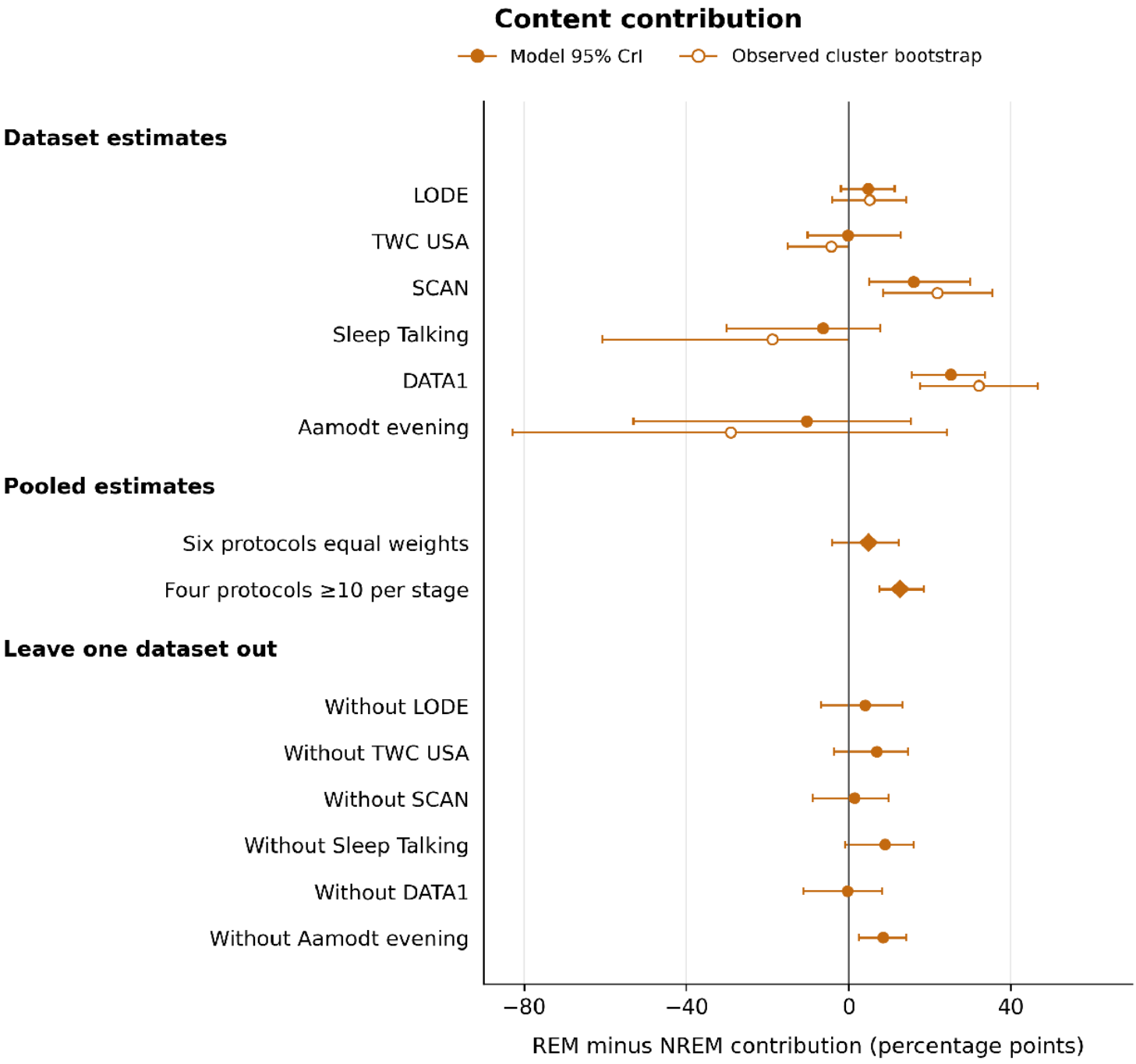


Figure 4

Content contributions by dataset and analysis

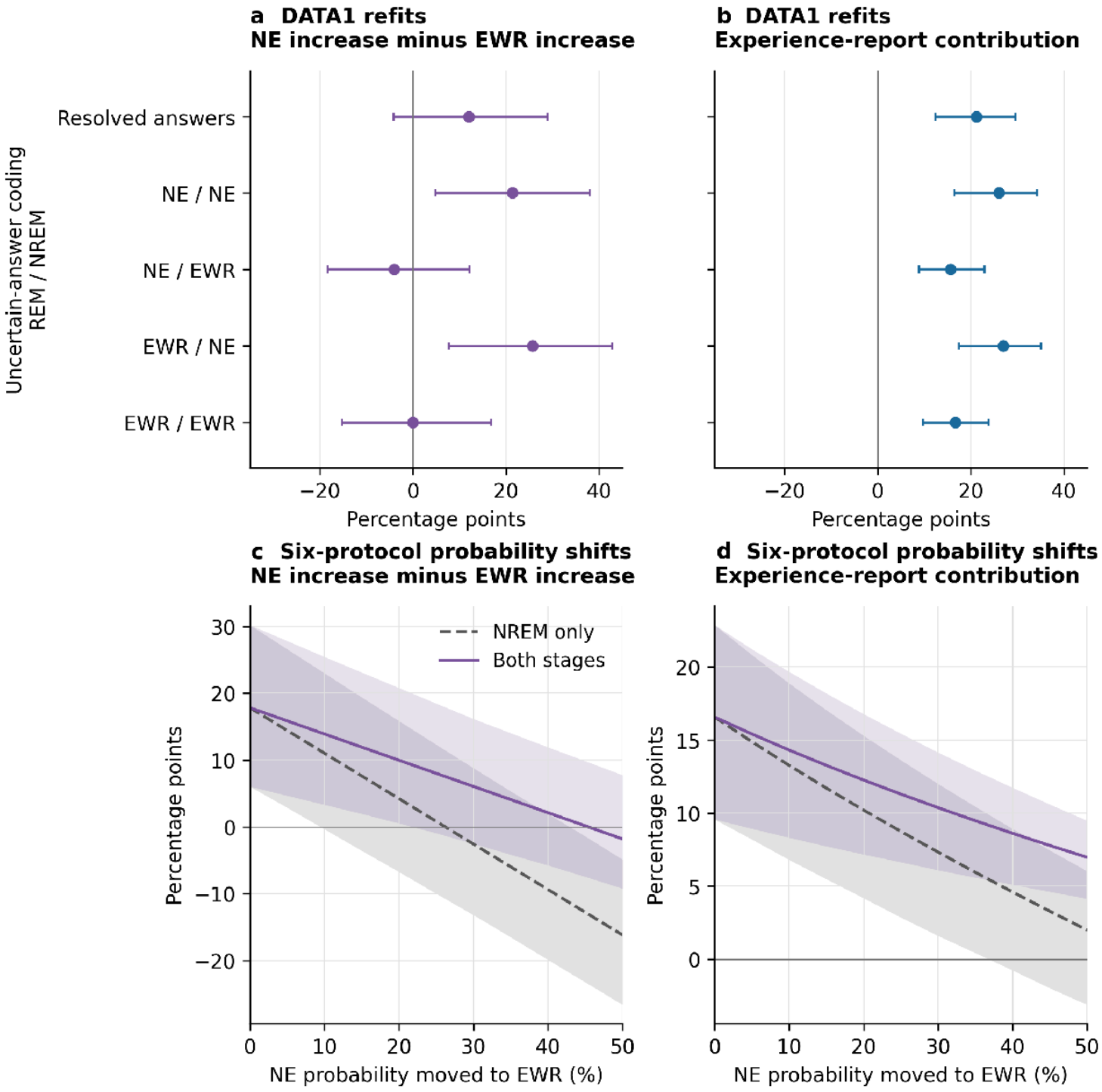


Figure 5

Recorded uncertainty and hypothetical category shifts

Supplementary Files

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

- XXXSupplementaryInformationCommunicationsPsychology.pdf