Article:

Chitiz, Louis, Mckeown, Bronte, Mulholland, Bridget et al. (11 more authors) (2025) Mapping cognition across lab and daily life using Experience-Sampling. Consciousness and cognition. 103853. ISSN 1053-8100

https://doi.org/10.1016/j.concog.2025.103853

Consciousness and Cognition 131 (2025) 103853

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/yccog

Full Length Article

Mapping cognition across lab and daily life using Experience-Sampling

Louis Chitiz [a,*], Bronte Mckeown [a], Bridget Mulholland [a], Raven Wallace [a], Ian Goodall-Halliwell [b], Nerissa Siu Ping-Ho [c], Delali Konu [d], Giulia L. Poerio [e], Jeffrey Wammes [a], Michael Milham [f], Arno Klein [f], Elizabeth Jefferies [g], Robert Leech [h], Jonathan Smallwood [a]

[a] *Department of Psychology, Queen's University, Kingston, Ontario, Canada*
[b] *Centre for Neuroscience Studies, Queen's University, Kingston, Ontario, Canada*
[c] *School of Psychology, University of Plymouth, Plymouth, UK*
[d] *Department of Psychology, Durham University, Durham, UK*
[e] *School of Psychology, University of Sussex, Falmer, UK*
[f] *Child Mind Institute, New York, NY, USA*
[g] *Department of Psychology, University of York, York, UK*
[h] *Department of Neuroimaging, King's College, London, UK*

ARTICLE INFO

Keywords:
Ecological validity
Experience-sampling
Ecological momentary assessment
Spontaneous thought
Cognitive-neuroscience
Principal component analysis

ABSTRACT

The goal of psychological research is to understand behaviour in daily life. Although lab studies provide the control necessary to identify cognitive mechanisms behind behaviour, how these controlled situations generalise to activities in daily life remains unclear. Experience-sampling provides useful descriptions of cognition in the lab and real world and the current study examined how thought patterns generated by multidimensional experience-sampling (mDES) generalise across both contexts. We combined data from five published studies to generate a common 'thought-space' using data from the lab and daily life. This space represented data from both lab and daily life in an unbiased manner and grouped lab tasks and daily life activities with similar features (e.g., working in daily life was similar to working memory in the lab). Our study establishes mDES can map cognition from lab and daily life within a common space, allowing for more ecologically valid descriptions of cognition and behaviour.

1. Introduction

A central goal of psychological science and cognitive neuroscience is to establish how cognition and behaviour operate in the real world (Cialdini, 2009; Kingstone & Smilek, 2008; Matusz et al., 2019; Nastase et al., 2020; Orne, 1962; Pooja et al., 2024). Ecologically valid methods such as field studies that examine thoughts and actions in daily life are useful for providing insights into real world behaviour, but lack the control needed to understand underlying mechanisms (Cialdini, 2009; Fischhoff, 1996; Matusz et al., 2019). Lab-based studies, on the other hand, provide the experimental control needed to identify specific mechanisms behind discrete behaviours, but often at the cost of being difficult to generalise to daily life situations (Broadbent, 1971; Fischhoff, 1996; Kihlstrom et al.,

* Corresponding author.
E-mail address: 17lssc@queensu.ca (L. Chitiz).

https://doi.org/10.1016/j.concog.2025.103853
Received 18 July 2024; Received in revised form 9 March 2025; Accepted 24 March 2025
Available online 10 April 2025

2021; Kingstone et al., 2003; Matusz et al., 2019; Neisser, 1980; Orne, 1979; Osborne-Crowley, 2020). In principle, therefore, lab and field studies in combination produce more compelling accounts of psychological phenomena than either on their own (Lin et al., 2021). However, we currently lack empirical tools that enable translation between data captured in ecologically valid and controlled settings (Bromham et al., 2016; Finn et al., 2023; Lin et al., 2021; Matusz et al., 2019). The ability to integrate data from the lab into daily life (and vice versa), therefore, is an important next step in developing more comprehensive theoretical perspectives on aspects of psychological functioning including mental health, creativity and productivity.

Our current study examines whether experience-sampling, which leverages people's capacity for introspection to characterise mental states at different moments, can be used as a tool to identify a common space within which similarities and differences in cognition between laboratory and real world contexts can be assessed (for prior examples of the utility of experience-sampling mapping thoughts in the lab and daily life see Kane et al., 2007; Kane et al., 2017; Ho et al., 2020). We used multidimensional experience-sampling (mDES), an established method in which individuals rate their experiences along a set of dimensions (e.g., the level of detail in a person's thoughts or their relationship to an external task). mDES provides a flexible method for sampling cognition (it can be administered in daily life via smartphones, in the behavioral lab and during brain imaging; for a review see Smallwood, Turnbull et al., 2021) and affords information about mental states that are tied to the current task or activity as well as to covert mental states (see for example research on mind-wandering; Smallwood and Schooler, 2015). Studies have successfully used mDES in both the lab (Konishi et al., 2017; Konu et al., 2021; Simola et al., 2023), the scanner (Konu et al., 2020; Karapangiotidis et al., 2020; Mckeown et al., 2023) and in daily life (Mckeown et al., 2021; Mulholland et al., 2023; Turnbull et al., 2021). Recently, we established that the representations of ongoing thought that are generated by mDES are sufficiently precise that they can accurately estimate patterns of brain activity in both tasks (McKeown et al., 2025) and in videos (Wallace et al., 2025).

Since mDES can be easily administered in both the lab and daily life, we sought to examine whether this tool can be harnessed to integrate data across contexts, forming a 'shared space' within which researchers can directly compare the patterns of thought participants report in the lab and in daily life. A shared "thought-space" that aligned patterns of thought in the field to those collected in the laboratory would be invaluable for psychology since it would allow estimations of cognition derived from mechanistic investigations in the lab with those collected in daily life. For example, an early study examining this question by Ho and colleagues (2020) was able to identify that laboratory patterns of off-task thought tended to over emphasise social features compared to daily life. In support of this conclusion social features to off-task thought were associated with greater BOLD responses to faces in regions of ventral visual cortex that are assumed to play a key role in face perception (Kanwisher, 1997).

The goal of our analysis was to build on these prior studies to produce a formal shared space that accommodates patterns of thought generated in the laboratory and those that are generated in the field. Accordingly our analysis used data from 5 datasets collected from just under 400 participants in a variety of daily life and lab contexts to examine the capacity of mDES to produce this 'shared space' by addressing three specific questions: (i) Does mDES produce similar thought patterns across lab and daily life contexts? (ii) Can these thought patterns be used to create a 'shared space', which captures the mental states that characterise each of its constituent contexts? (iii) To test the validity of this shared space, does it organise contexts in a conceptually coherent manner? To the extent that mDES can generate a non-biased shared space that integrates cognition from the lab and field would constitute an important step in the generation of mechanistic theories of human behaviour that are grounded in the real world.

2. Method

2.1. Participants

To examine how well mDES functions as a means of integrating cognition across contexts, we used mDES data from a variety of lab- and daily-life-contexts. The total sample comprised 370 participants (280 females, 82 males, 3 non-binary, 4 unspecified; mean(age) = 21.57yrs) yielding a total 8459 experience-sampling probes. See Table 1 for a summary of the sample/probe characteristics for each context. See the relevant publications referenced in the table for more in-depth descriptions of each sample.

We derived our sample from five previously published datasets. Konu et al. (2020) ("fMRI") had a sample of 62 undergraduate students (41 females, 21 males; mean(age) = 23.29 ± 4.53yrs) in the UK complete mDES probes in fMRI while they conducted a simple go/no-go task. Ho et al. (2020) ("UK") administered 5 mDES probes per day for 7 days to a sample of 78 undergraduate students (57

Table 1
Environment/sample characteristics for the five analysed datasets.

Environment	Publication	*N*	Probes
UK/Lab/fMRI	Konu et al. (2020)	62	1456
UK/Daily life/Pre-COVID	Ho et al. (2020)	78	1996
UK/Lab/Tasks	Konu et al. (2021)	70	2302
UK/Daily life/COVID	Mckeown et al. (2021)	59	1257
Canada/Daily life/Post-COVID	Mulholland et al. (2023)	101	1458

Table 1. Our study used mDES data sampled from five datasets encompassing a variety of daily life and lab contexts. 'N' = Total number of participants in each study; 'Probes' = Total number of probes in each study.

females, 21 males; mean(age) = 19.64 ± 1.62yrs) in the UK using a smartphone app during daily life. In Konu et al. (2021) ("Lab"), 70 UK undergraduate students (60 females, 10 males; mean(age) = 20.60 ± 2.10yrs) completed mDES probes while performing a battery of 13 different cognitive tasks (e.g., visual/verbal semantics tasks, self/other reference tasks, working memory tasks, TV/audiobook tasks, etc.). Mckeown et al. (2021) ("UK – COVID") notified 59 participants recruited online in the UK (40 females, 17 males, 1 binary, 1 unspecified; mean(age) = 24.22 ± 4.07yrs) to complete mDES probes in daily life during the COVID lockdown 5 times per day for 7 days. Finally, Mulholland et al. (2023) ("Canada"), *after* the COVID lockdown, administered mDES probes 8 times per day for 7 days to 101 undergraduate students in Canada (82 females, 13 males, 2 non-binary, 3 unspecified; mean(age) = 24.22 ± 4.07yrs).

As well as sampling experience, the daily life sampling studies collected data regarding participants' current social environments (e.g., 'alone', 'interacting with others', etc.; Ho et al., 2020, Mckeown et al., 2021, Mulholland et al., 2023). Further, Mckeown et al. (2021) and Mulholland et al. (2023) collected data regarding what participants were doing (e.g., doing homework, social media, etc.), their virtual social environments (e.g., passively receiving texts, on a call, etc.), and physical locations (e.g., "inside at home", "outside in a town/city", etc.). See the relevant publications in Table 1 for a summary of sampled contexts.

Although each study used a similar multidimensional approach to sample experience, the specific set of questions employed in each study varied across studies. To harmonise datasets we first identified the set of questions that are common to each study. This resulted in a set of nine dimensions common across studies. One dimension, 'people', which measures how much thought content involves other people, was divided into two dimensions in two datasets (UK, UK-COVID) to separate thought contents relating to others close and not close to the participant (Ho et al., 2020; Mckeown et al., 2021). As both scores pertain to thinking about another person, we used the higher of the two scores to represent an individual's 'people' score (i.e., how much their thoughts concerned "other people", independent of those people's identities), which strongly correlated with the mean of the two questions, $r(4973) = 0.92$, $p < 0.001$. While this recoded item tended to organize probes in a manner consistent with its analogues in other datasets (see Results), the impact of more content-specific probes on participant responses remains unclear. In future, refining question wording across studies could enhance mDES' ability to capture cognitive variation across contexts (see Discussion). To standardise the combined data, all scores were re-coded into a continuous 1–5 scale using Min-Max normalization (Jain et al., 2005), and then z-scored. See Table 2 for the set of dimensions included in the analysis. See Table S1 for the questions used to measure each dimension in each dataset.

2.2. Statistical analysis

2.2.1. Component reproducibility analyses

The goal of our analysis was to examine how different lab and daily life contexts were comparable when combined into a shared "thought space". The analysis focused on three main questions: 1) the reproducibility of thought patterns across environments, 2) the degree to which a set of common thought patterns represent those associated with each environment, and 3) the reliability of context mappings between the lab and daily life. As is common in mDES studies (see Smallwood, Turnbull et al., 2021), our analysis involves the decomposition of the set of common questions using Principal Components Analysis (PCA). These analyses were conducted using the ThoughtSpace package in Python 3.12.0 (Centrum voor Wiskunde en Informatica,1995; https://github.com/Bronte-Mckeown/ThoughtSpace/releases/tag/v1.0.1). The ThoughtSpace package uses scikit-learn (Pedregosa et al., 2011) for PCA computations and factor analyzer (https://pypi.org/project/factor-analyzer/) for component rotation.

In order to assess the utility of mDES in generating a viable common space we measured how well the components generated using one dataset were reproduced by other datasets (e.g., how well do the components from one sampling environment reproduce the components produced from other environments). As different datasets often produced similar components in a different order (based on the variance accounted for) we based our estimates of component reproducibility on a given component's similarity to its most similar equivalent in another solution (i.e., its "homologue"). We measured component reproducibility using two metrics: 1) loading similarity (i.e. how similar are the features in solutions from mDES when the same method of decomposition is applied to each dataset) and 2) component score similarity (i.e., whether similar thought patterns from different datasets organize each other's moments of experience in a similar way).

Table 2
Common dimensions of thought measured across contexts.

Dimension	Question (from Mulholland et al., 2023)	Scale (1 … 5)
Focus/Task	*My thoughts were focused on an external task or activity:*	Not at all … Completely
Future	*My thoughts involved future events:*	Not at all … Completely
Past	*My thoughts involved past events:*	Not at all … Completely
People*	*My thoughts involved other people:*	Not at all … Completely
Self	*My thoughts involved myself:*	Not at all … Completely
Problem	*I was thinking about solutions to problems (or goals):*	Not at all … Completely
Detailed	*My thoughts were detailed and specific:*	Not at all … Completely
Deliberate	*My thoughts were:*	Spontaneous … Deliberate
Emotion	*The emotion of my thoughts was:*	Negative … Positive

Table 2. Nine common mDES items sampled in each of the studies. Question wording marginally differed across the studies for the 9 dimensions (See Supplementary material for the full set of questions). All items were coded to score from 1 to 5. *In the UK daily life datasets,'People' was divided into thoughts about people 'close' or 'not close' to the participant. In our analysis, the higher score of the two was used as the overall 'People' score.

We calculated component loading similarity using Tucker's Congruence Coefficient (TCC; Lovik et al., 2020; Tucker, 1951), where a score $\geq$ 0.85 indicates fair similarity and $\geq$ 0.95 indicates exact similarity (Lorenzo-Seva & Ten-Berge, 2006). We computed component score similarity between homologous components based on the correlation between the component scores they each generate for the same set of data (RHom.; Mulholland et al., 2023). As a correlation $\geq|.80|$ is often used to indicate redundancy in predictors in regression analysis, we used a cutoff of 0.80 to indicate reproduction of components (i.e., variables that *should* be the same appearing redundant) (Berry & Feldman, 1985).

To compare the solutions in each dataset to a common template which is constructed in a similar manner (i.e., contains the same number of dimensions), we generated an overall PCA solution using the full set of data, and the number of components extracted in this analysis was used for comparison between samples and to aggregated data. We determined the number of components to extract for this overall solution based on examination of the scree plot in tandem with parallel analysis (Franklin et al., 1995; Horn, 1965). Finally, we used information regarding the maximally reliable solution to guide component extraction (Everett, 1983). To test the robustness of the determined solution in each dataset, a bootstrapped split-half reliability was conducted for each sample as well as for the overall dataset (bootstrap resampling halves of data and assessing component similarity; Mulholland et al., 2023).

2.2.2. *Sample-to-Sample reproducibility*

Our first objective was to assess how well the thought patterns generated by the decomposition of mDES in different samples naturally generalise to each other. To quantify this, we conducted a 'direct-projection' analysis, in which we compared the components associated with each environment. In each comparison, we generated principal components for each sample. We then compared each pairing of environments, selecting one to be a 'reference' solution, which we Varimax rotated, and the other to be a 'comparate' solution, which we procrustean rotated to the reference solution to maximally align the two sets of components (Gower, 1975). After computing loading similarity, we estimated component scores for the full dataset based on each set of components and computed score similarity among homologous pairs across the two selected datasets (see Fig. 1 for a visualisation of this process). We bootstrapped these comparisons by dividing each sample into 5 equal folds and comparing each combination of folds in the reference set to each

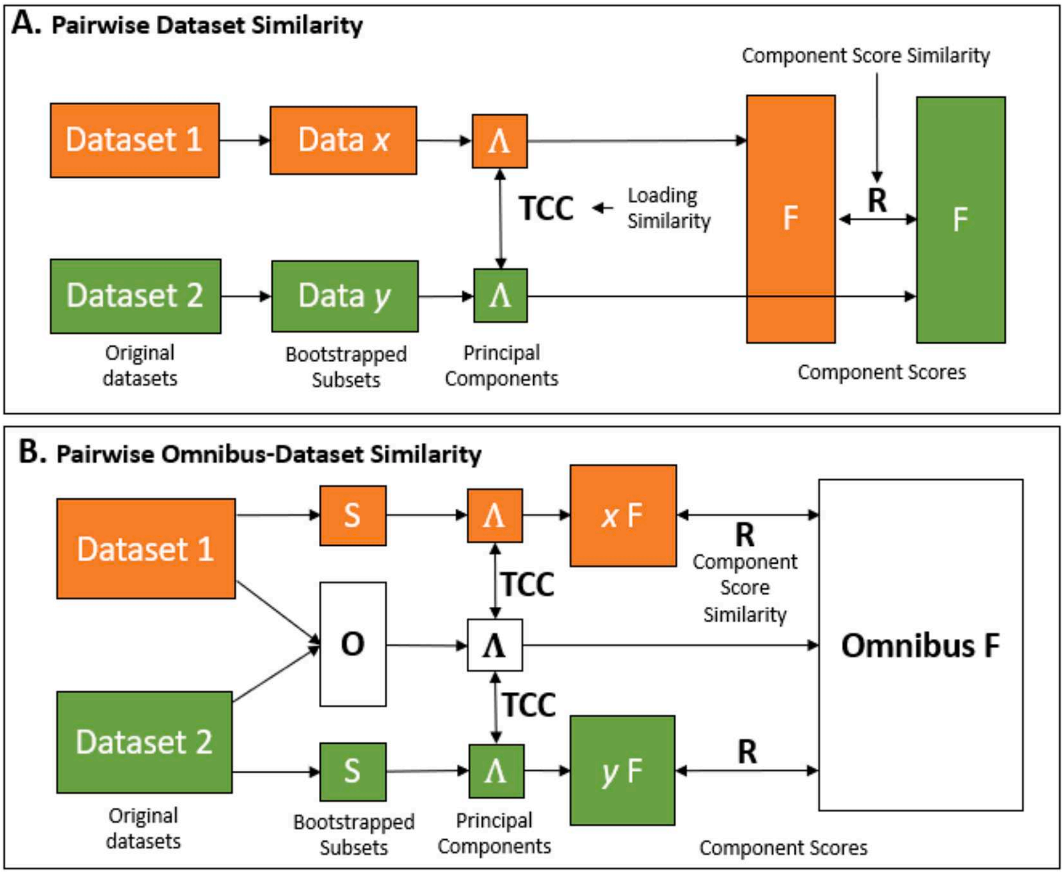


Fig. 1. Complementary ways of assessing component reproducibility across datasets. Panel A shows a visualisation of the 'direct-projection' method used to assess whether environments independently produced similar thought patterns with mDES. Λ = Principal components, F = Component scores. Panel B shows a visualisation of the 'omnibus-sample' method used to determine whether the 'common' thought patterns of the shared space reproduced the components seen in each data set individually. O = 'Omnibus', S = 'sample', F = Component scores, TCC = Tucker's Congruence Coefficient.

combination of folds in the comparison set.

To compare the associations between datasets (i.e., whether some samples yielded more similar thought-patterns than others). We conducted Fisher's r-to-z tests comparing the bootstrapped component-score similarity between pairs of solutions, estimating standard error using their 95 % confidence intervals (Altman & Bland, 2011; Fisher, 1921). We used Cohen's q as an estimate of effect size, interpreting an effect of $q < 0.1$ as indicative of no effect, an effect of $0.1 \leq q < 0.3$ as a small effect, $0.3 \leq q < 0.5$ as a medium effect, and $q \geq 0.5$ as a large effect (Cohen, 1988).

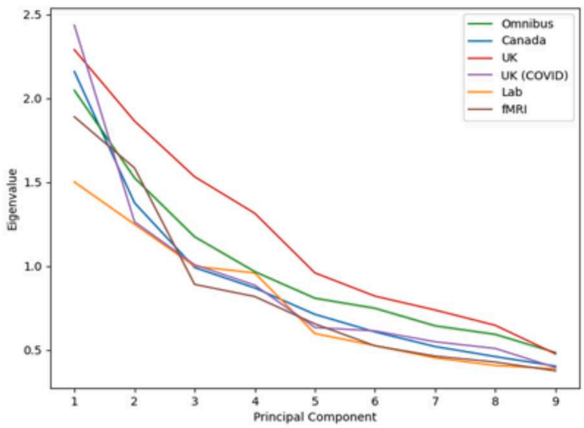
2.2.3. Omnibus-to-Sample reproducibility

Our second objective was to assess whether aggregated data could form a 'shared space' that maintained relationships observed in each environment. This analysis helps quantify whether data that is combined from each situation produces a set of 'common' components that reproduced those generated by individual contexts. To examine this, we repeatedly randomly assigned equal subsets of each sample's data to be used an 'omnibus' solution, where it would be combined with probes from all other environments to generate 'common' thought patterns and assigned its remaining probes to a 'sample' solution, which represented the components of the given sample. Each iteration yielded a set of components for each sample and an 'omnibus' set of components generated from the aggregated data. We Varimax rotated the omnibus components and procrustean rotated each sample's components to those of the omnibus. We then computed loading similarity and component score similarity for homologous pairs using the same projection process as above (see Fig. 1 for a visualisation of this analysis).

To obtain more specific results, we conducted a 'by-component' variation on this analysis, where we held one omnibus solution stable and examined component reproducibility on a component-by-component level relative to each collected sample. As split-half solutions often produced equivalent, but inverted, components, we decided to use only component-score similarity to assess by-component performance.

2.2.4. Shared-Space context mappings

Beyond examining the reliability with which mDES generates thought patterns across lab and daily life contexts, our final goal was to explore the theoretical potential of this 'shared space' by examining whether it maps contexts one might expect to be similar (e.g., Working in Canada versus the UK) in consistent ways. Our next analysis examined whether mDES can group behavioural tasks in the laboratory with real-world contexts in which those tasks' behaviours might be applied (e.g., watching movies in the lab and watching TV in daily life). Collected data included information regarding task/activity contexts (e.g., homework, n-back task, etc.), physical social environments (e.g., alone, hanging out with friends, etc.), virtual social environments (e.g., on a video call, passively viewing text notifications, etc.), and locations (e.g., at home, in the scanner, etc.). For the sake of space, we only report results for task/activity



Dataset	Split-Half	95% CI
Canada	.99	[.98, 1]
UK	.99	[.99, 1]
UK (COVID)	.97	[.94, 1]
fMRI	.97	[.93, 1]
Lab	.99(6)	[.99, 1]
Overall	.99(6)	[.99, 1]

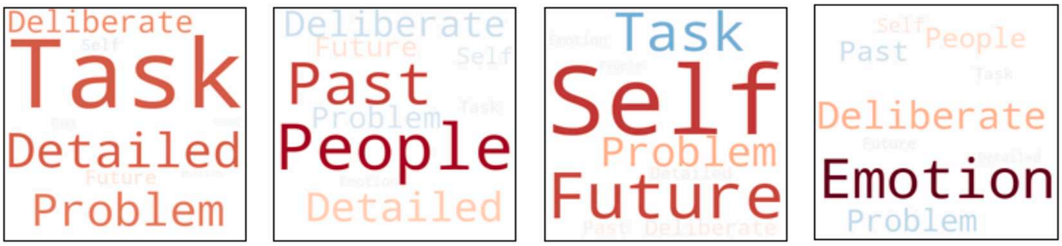


Fig. 2. 4-PC solution performance across datasets. (Top left) Scree plot for each dataset and the full combined set ("Omnibus"). (Top right) Split-half reliability for 4-component solutions for each dataset and the combined set. (Bottom) Word-clouds visualising the loading structures of the 'thought patterns' for the overall 4-PC solution. Word size indicates loading strength (i.e., larger = stronger loading), word colour indicates loading directionality (i.e., warmer = more positive loading, cooler = more negative loading).

contexts in our main analysis. See Supplementary materials for results focused on other contextual variables.

To quantify the reliability of context mappings we used permutation testing to calculate which contexts on average generated component scores greater than might be produced by chance. As the sample size of probes differed based on each task/activity, we used a stratified permutation approach, where we matched permutation sample-size to the sample-size of the context focus. In assessing significance, we used the Bonferroni-Holm method to adjust for inflation of type-I error (Aickin & Gensler, 1996; Holm, 1979). We then examined among significant context-component associations whether the same thought patterns characterised theoretically similar contexts in the shared space.

3. Results

3.1. Initial PCA results

To establish a consistent number of thought patterns to extract throughout the analysis, we conducted an initial PCA based on the full set of combined data. Based on examination of the Scree plot and parallel analysis, it was unclear as to whether the solution should include 3 or 4 principal components (PC). Subsequent examination of component reliability found that the 4-PC solution produced more reproducible components (see Fig. S2-5 for the 3-PC results). The 4 components accounted for approximately 63.5 % of the total variance in the full dataset, and accounted for a consistent proportion of variance across each dataset (See Fig. 2, Table S3).

Further, the 4-component solution demonstrated excellent split-half reliability across the total dataset, as well as for the 4-PC solution in each dataset (See Fig. 2). The specific loadings for each component are represented as word clouds in Fig. 2 in which the size of the item describes its importance (bigger = more important) and the colour indicates the polarity (items with a similar colour behave in a similar way). The specific loadings from which these clouds are presented in Table 3. Based on their loadings and names for similar components in earlier studies, we termed the 4 components: 1) “Detailed Task-Focus”, 2) “Episodic Social-Cognition”, 3) “Future Problem-solving”, and 4) “Positive Engagement”. Each has precedents in prior research, including studies beyond the datasets analyzed here (e.g., Karapanagiotidis et al., 2020; Ruby et al., 2013; Turnbull, Wang, Murphy et al., 2019), so we aligned our naming with established conventions.

3.2. Does mDES produce similar thought patterns across lab and daily life contexts?

After establishing an underlying component structure, we examined whether the application of PCA to different datasets produces components that are similar to those seen in other datasets. We generated 4-PC solutions for each sample and compared them based on their loading similarity and the correlation of their component scores.

According to loading similarity, all comparisons but one met the TCC $\geq$ 0.85 cutoff for fair similarity according to Tucker’s Congruence Coefficient (See Fig. 3; Lavek et al., 2020). The only pair that failed to meet this threshold was the components generated based on data from the UK during the COVID lockdown compared to PCs generated from fMRI data sampled during a go/no-go task (TCC = 0.83). While this did not significantly differ from most of the other comparisons, the 95 % confidence intervals (CI) for TCC between the UK pre-lockdown both with the lab and with Canada after the lockdown excluded the upper limit for loading similarity between the UK during COVID and fMRI (See Fig. 3).

Next, we compared the datasets based on the similarity of the components they produced. All pairings met our Rhom. $\geq|.80|$ cutoff for notable similarity, indicating that based on component- score similarity all environments produced at least reasonably similar thought patterns independently of one another. However, datasets collected in daily life met a more conservative threshold of $\geq|.90|$, which is conventionally used in split-half reliability within the same dataset, (See Fig. 3; Everett, 1983). Fisher’s equivalence tests further confirmed this separation, showing that thought patterns in daily life outside the COVID-19 lockdown were significantly more correlated with large effect sizes than those between daily life and the lab or between lab environments themselves (See Fig. S6, Table S4 for the results of this analysis). In other words, patterns of thoughts sampled in daily life environments were more similar to each other than they were to a basic lab environment and were more similar to each other than the lab was to fMRI (See Fig. 3).

Table 3
Loadings for each component on the omnibus PCA. Displayed loadings are Varimax rotated.

Dimension	Detailed Task-Focus	Episodic Social-Cognition	Future Problem-Solving	Positive Engagement
Task	0.54	−0.03	−0.40	−0.04
Future	0.11	0.11	0.59	−0.04
Past	0.03	0.60	0.04	−0.18
Self	−0.07	−0.07	0.62	0.07
People	0.00	0.72	−0.01	0.17
Problem	0.47	−0.13	0.33	−0.21
Deliberate	0.43	−0.17	0.05	0.32
Detailed	0.54	0.23	0.03	0.03
Emotion	−0.02	0.02	0.02	0.88

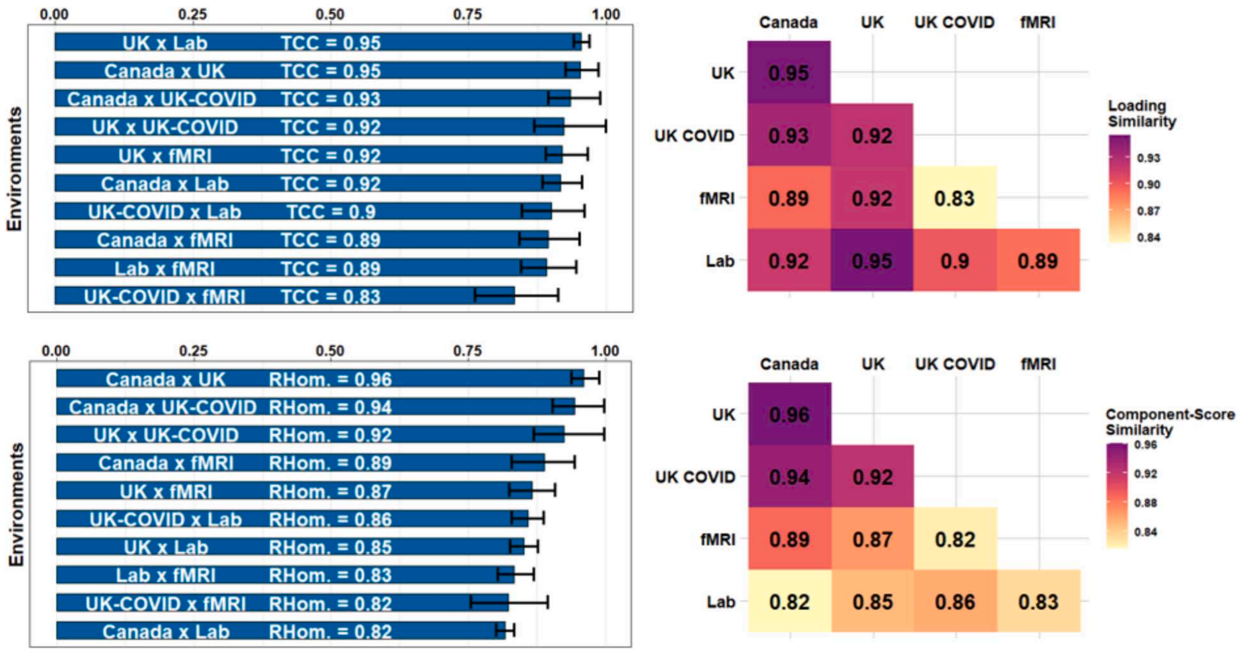


Fig. 3. Results of the sample-to-sample 'thought pattern' reproducibility analysis. (Top left) Loading similarity with 95% CIs for each sample-sample combination. (Top right) Heatmap for sample-sample loading similarity. (Bottom left) Component-score similarity with 95% CIs for each sample-sample combination. (Bottom right) Heatmap for sample-sample component-score similarity. "Can" = "Canada", "UK (COV)" = "UK – COVID".

3.3. Can these thought patterns form a space that captures the mental states of constituent contexts?

Having assessed the similarity between the overall set of four dimensions generated by PCA across datasets, our next analysis examined whether the thought patterns from one situation reproduced those derived from each dataset individually.

According to both component score and loading similarity, the shared space successfully represented each of the constituent contexts, surpassing both thresholds for reasonable similarity (TCC $\geq$ 0.85, Rhom. $\geq |.80|$) for all 5 datasets. It further met more conservative thresholds for excellent similarity (TCC $\geq$ 0.95, Rhom. $\geq |.90|$) for all 3 daily life contexts based on loading similarity as well as the basic-lab environment based on component score similarity (See Fig. 4).

3.3.1. Pattern-dependent stability in the shared space

Leveraging the excellent split-half reliability of each of the 4-PC solutions, we used stable omnibus and sample halves to examine shared-space performance on a by-component level. This revealed a component-specific effect such that the first three thought patterns all replicated with excellent similarity across all five contexts, but the 4th component ('*Positive Engagement*') performed worse. *Positive Engagement* achieved moderate similarity with the two daily life environments outside of the lockdown but fell just short of moderate similarity with daily life during the COVID lockdown and failed to do so with the two lab environments (See Fig. 5). As typical real-world living comprises significantly more self-directed behaviour (i.e., people tend to choose to engage in activities they enjoy), this sheds some insight as to the experiential differences between the lab and daily life as well as between daily life during and outside of the COVID-19 lockdown (Rubio-Tomás et al., 2022).

Altogether, the shared-space effectively reproduced the first 3 thought patterns produced by each context (*Detailed Task-Focus*, *Episodic Social-Cognition*, *Future Problem-Solving*), but better represented daily life contexts on the 4th component.

3.4. Does the shared space organise contexts in a conceptually coherent manner?

So far, our analysis has established that mDES can produce a similar set of dimensions when it is applied across different datasets and highlighted that 3 of the 4 components yield highly reproducible dimensions across 5 datasets. Our next analysis explores how these common dimensions organise tasks in the lab and activities in daily life, with the aim of understanding if they group situations in both contexts that are conceptually similar. Our data sampled from a set of 61 different task-settings across 4 of the 5 contexts (Note: no activity data was recorded for the UK pre-COVID lockdown dataset).

To assess the tasks/activities characterised by each sufficiently stable common thought pattern, we estimated average component scores for each of the 61 tasks/activities on each of the 3 stable components. To classify a task/activity as being 'characterized' by a certain pattern of thinking, we used a stratified permutation testing approach, assessing whether the average component score of the probes linked to a given task/activity was significantly higher than might occur by chance (i.e., random samples of probes the same size

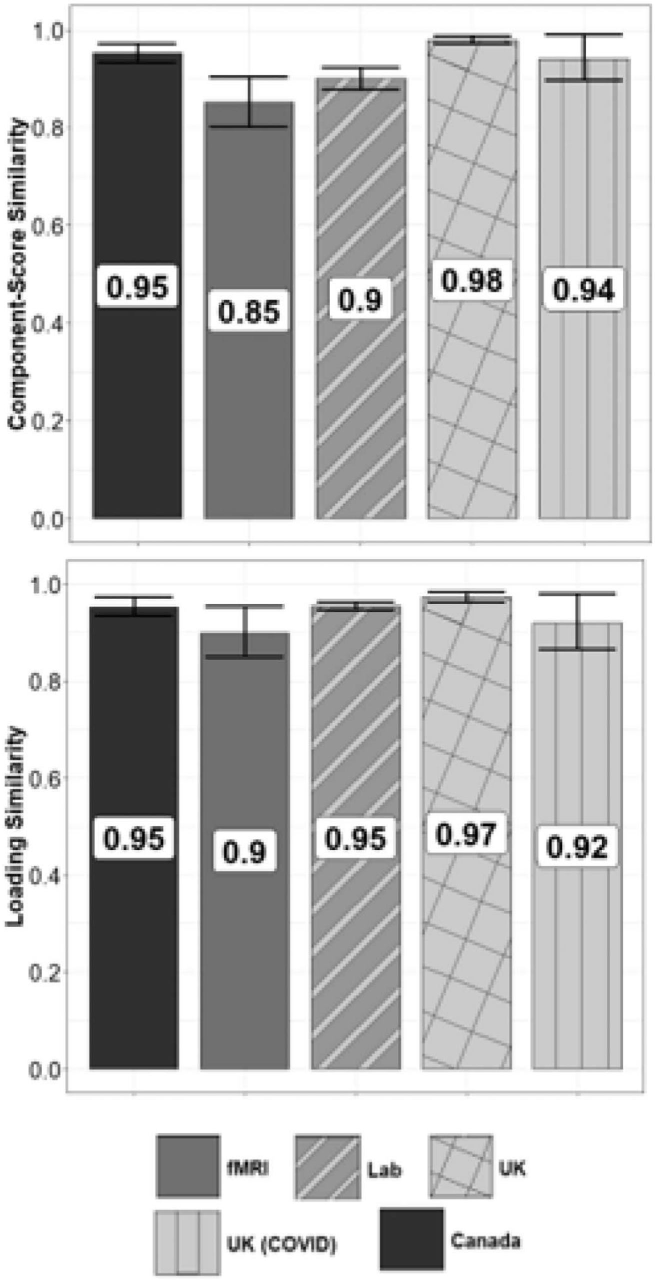


Fig. 4. Results of the omnibus-to-sample reproducibility analysis. Bar graphs represent M ± 95 % Cls for omnibus to each sample for: (Top) component-score similarity and (Bottom) loading similarity.

as the number of probes associated with the activity). We used the Holm-Bonferroni method to adjust for familywise error rate across permutation tests for each daily life activity and lab task. See Fig. 6 for word clouds summarising the context mappings for each thought pattern that survived correction for multiple comparisons.

From the grouping of activities and tasks shown in Figs. 6 and 7 two noteworthy observations can be drawn about the shared space. First, it reliably maps many similar daily life contexts sampled in separate real-world samples (e.g., watching TV in the UK and Canada, working in the UK and Canada, etc.). Second, it often groups daily life task-contexts together with the lab-tasks intended to target their conceptually relevant cognitive processes. Examples include homework/working in daily life with a spatial working memory task in the lab (Ackerman, 2005; Cowan, 2017), watching TV in daily life and movie-viewing tasks in the lab (Note they all fall to the same end of *Future Problem-Solving*; Hasson et al., 2008), social-cognition reference tasks in the lab and virtual/in-person conversation in daily life (de Caso et al., 2017, Murphy et al., 2019). See Table 4 for a summary of the five strongest significant associations with each thought pattern.

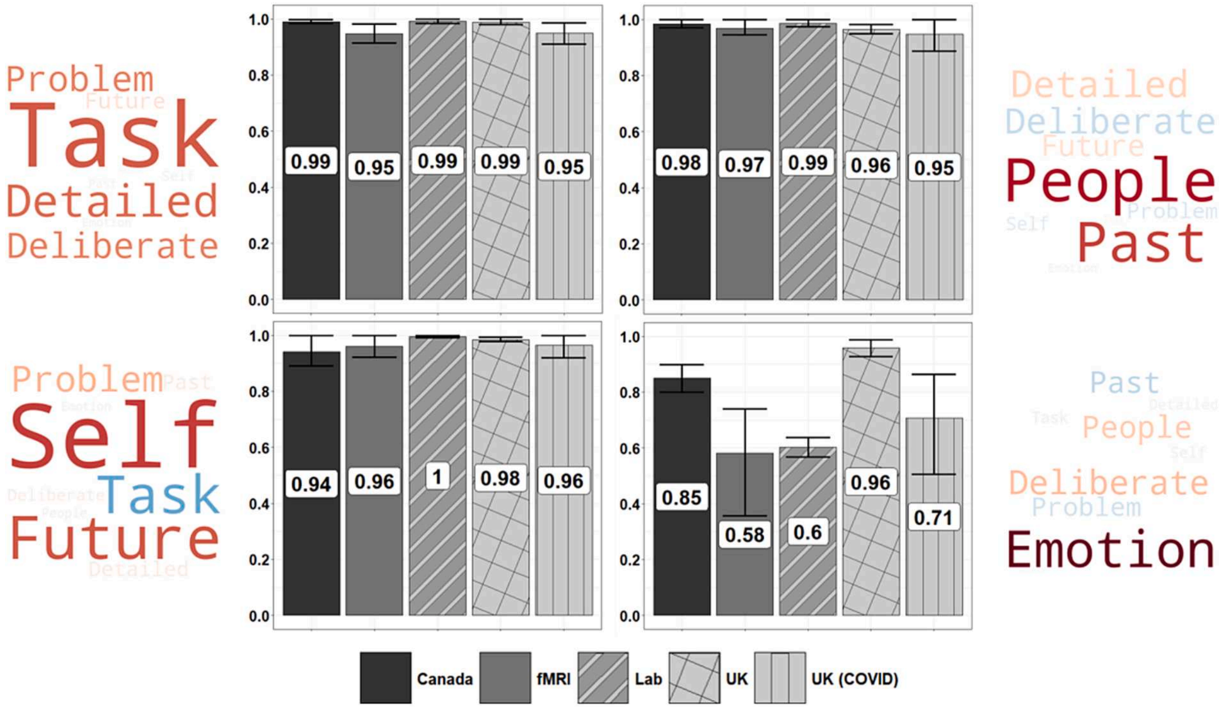


Fig. 5. 'Omnibus-to-sample' component reproducibility, broken down by component. Bar graphs represent M ± 95 % CIs for component score similarity. As certain components were reproduced but inverted in some split-half solutions, loading similarity provided a less representative measure of component similarity.

We conducted an additional exploratory analysis examining context mappings on the level of social environment (e.g., interacting with someone in-person/virtually, being alone, etc.), which was recorded across all 5 contexts. Due to its more exploratory nature, we detail this analysis and its results in our supplementary materials. Overall, the shared space appeared to consistently cluster social-cognitive states observed in the UK before COVID together with those sampled in Canada after COVID, and clustered social-states sampled during the lockdown with those measured in the lab.

4. Discussion

Psychological scientists often propose mechanistic theories that utilise laboratory studies to explain human cognition and behaviour as observed in real-world situations from field studies. While the combination of lab and field methods have complementary strengths and weaknesses for theory generation, there are relatively few ways to conceptualise how observations of cognition derived from one research setting relate to observations made in the other. In the current study, we examined whether multidimensional experience-sampling (mDES), a low-cost, flexible method for quantifying ongoing thought patterns, can produce sufficiently consistent dimensions across sampling contexts to produce a shared 'thought space', which integrates and enables direct comparison of cognition across the lab and real-world. To do so, we compared dimensions of thought generated by the application of PCA to a set of nine mDES questions measured in nearly 400 participants from 5 published datasets collected across a variety of both daily life (UK & Canada; pre-, during, and post-lockdown) and lab settings (during behavioural tasks, in fMRI).

In our first analysis, different sampling contexts produced reasonably consistent dimensions. While all contexts met thresholds for reasonable component score similarity, daily life sampling demonstrated significantly higher stability than the two lab contexts. The stability of cognition as measured by mDES in the real-world data is important, and counterintuitive, because sampling behaviour outside of the lab is typically conceptualised as producing noisier, less consistent data (Fischoff, 1996; Goodwin & Goodwin, 2016; Nastase et al., 2020). One possible reason why dimensions of thought produced by mDES from daily life may be more stable than those seen in the lab is that in daily life people engage in activities that have greater motivation (e.g., hobbies) or greater incentives (e.g., working for money) than in lab-based studies (Ho et al., 2020; Smallwood, Turnbull et al., 2021). Indeed, among lab-tasks, unstimulating tasks such as the Go/No-Go task tend to show the widest within-subject variability in thought content (Konu et al., 2021). Consistent with this view, our analysis of a component-dependent effect found that while the first three 'common' components successfully represented those produced by each environment, the fourth component (*Positive Engagement*) better represented daily life environments than lab-based ones (see Fig. 5). Conversely, in a recent study we found that a similar go-no go task was dominated by two patterns – one highlighting patterns of detailed task focus (and loading on brain regions in frontoparietal cortex linked to concentrated focus and known as the multiple-demand network; Duncan, 2010) and a second component highlighting a distracted

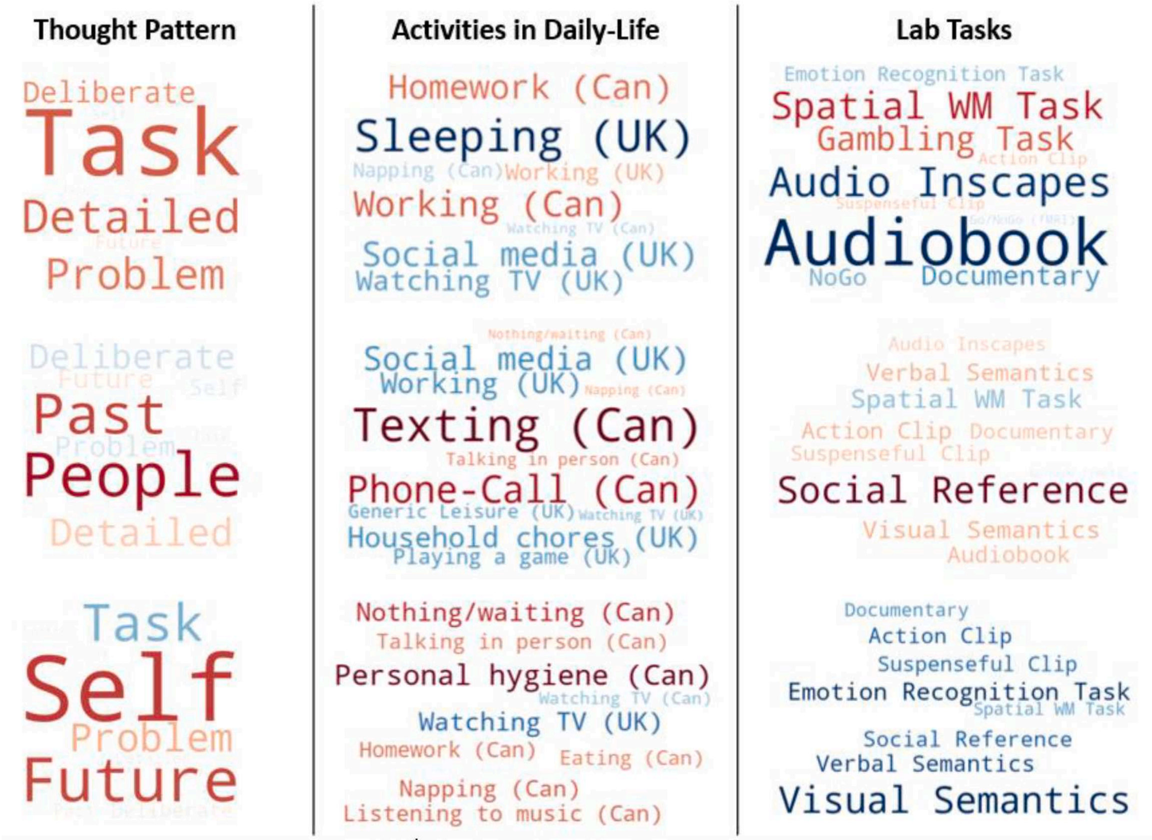


Fig. 6. Task/Activity-Context associations with patterns of thought in the shared space. In the task-clouds on the right, word size indicates size of average score for that activity on that component (larger = higher absolute mean score), and word colour indicates the direction of the relationship (warmer colour = positive score, cooler = negative).

state (which highlighted regions of the default mode network that are important when people engage in thought patterns linked to memory; Smallwood, Bernhardt et al., 2021). Importantly prior studies have shown that the transition from task focused states to self-generated states depends on a transition from the dominance of the multiple demand system to one dominated by the default mode network (Turnbull et. al., 2020). It seems possible that variablility in thought content in certain lab situations reflects competition between these two large scale systems which emerges in lab contexts because they often involve relatively unengaging tasks (in comparison to more engaging activities like watching TV or interacting with other people that are both important in daily life).

As well as characterising the different contexts in a similar manner, the shared space also helped organise the situations in which mDES was administered both within the lab, within daily life and importantly did so in a conceptually sensible way across both contexts (i.e., showing construct validity). For example, theoretically similar tasks in the lab (e.g., the spatial working memory task and the Cambridge Gambling task) were grouped together in one area of the space, decision making tasks (e.g., social reference, visual/verbal semantics) were grouped in another region, and media-viewing tasks in an area of their own. The shared space also grouped daily life activities in similar ways. For example, task-focussed states like homework and working were grouped together in both Canada and the UK, while task-unrelated states such as pre-/post-sleeping (UK) and napping (Canada) were grouped together, while activities entailing interactions with a device, such as watching TV, playing video games, and engaging with social media, also grouped together. Perhaps most notably, however, the shared space grouped daily life activities with lab tasks that have relatively similar cognitive features. For example, the shared space grouped tasks that target forms of executive-functioning (e.g., spatial working memory) with working and doing homework (Ackerman, 2005; Cowan, 2017), tasks that involve appraisals of other people (e.g., social reference and semantic tasks) were grouped with social interaction (e.g., texting, calling, conversation) (de Caso et al., 2017), and film-viewing in both the lab and daily life were grouped together (Hasson et al., 2008). In this way, our shared space provides some of the first evidence that cognitive similarities between states as they are targeted in the lab and their real-world manifestations can be quantified in a comparable manner.

It is worth noting that while the shared space generally grouped contexts together interpretably, this was not always the case. For example, *Future Problem-Solving* positively characterised personal hygiene care, doing nothing/waiting, napping/resting, listening to music, and Homework. At the level of conjecture, these activities do all share moments of relatively automated engagement, where individuals may think more passively. Although it remains a matter of speculation regarding what these disparate contexts share that

Fig. 7. The location of tasks and activities in the shared space. (Top) A three-dimensional representation of the space. (Bottom) Two-dimensional distribution of tasks and activities according to Detailed Task-Focus and Future Problem-Solving. 'Can' = Canada, 'UK' = UK – COVID.

Table 4

Five strongest significant associations and dissociations and their average component scores for each thought-pattern. * $p < 0.01$, ** $p < 0.0008$ (Corrected for 61 comparisons), *** $p < 0.00027$ (Corrected for 183 comparisons).

Detailed Task-Focus		*Episodic Social-Cognition*		*Future Self-Relevant*	
Task/Activity-Context	**Mean**	**Task/Activity-Context**	**Mean**	**Task/Activity-Context**	**Mean**
5 Strongest Associations		*5 Strongest Associations*		*5 Strongest Associations*	
Working (Canada)	0.87***	Other Reference (Lab)	1.73***	Personal hygiene care (Canada)	0.98***
Homework (Canada)	0.81***	Texting by phone (Canada)	0.98**	Nothing or waiting (Canada)	0.88***
Spatial Working Memory Task (Lab)	0.77***	Talking on the phone (Canada)	0.97***	Napping or resting (Canada)	0.60***
Gambling Task (Lab)	0.60***	Visual Semantics Task (Lab)	0.62***	Listening to music (Canada)	0.58***
Working (UK – COVID)	0.53***	Verbal Semantics Task (Lab)	0.60***	Homework (Canada)	0.55***
5 Strongest Dissociations		*5 Strongest Dissociations*		*5 Strongest Dissociations*	
Sleeping (UK – COVID)	−2.11***	Spatial Working Memory Task (Lab)	−0.68***	Visual Semantics Task (Lab)	−0.95***
Audiobook (Lab)	−0.99***	Generic Leisure (UK – COVID)	−0.59**	Emotion Recognition Task (Lab)	−0.93***
Audio Inscapes (Lab)	−0.93***	Social media (UK – COVID)	−0.57***	Verbal Semantics Task (Lab)	−0.86***
Social media (UK – COVID)	−0.78***	Working (UK – COVID)	−0.57***	Action Clip (Lab)	−0.77***
Watching TV (UK – COVID)	−0.76***	Household Chores (UK – COVID)	−0.56*	Other Reference (Lab)	−0.76***

underpins these common cognitive features, it is worth noting that this thought pattern is supressed by lab tasks that require the processing of complex semantic information (e.g. videos or semantic judgement task). This underpins one of the important advances that our shared “thought-space” provides to psychological science since it is possible that the features of daily life activities that encourage this form of cognition can be understood by exploring laboratory studies designed to test how semantic knowledge impacts on patterns of thought (see Wang et al. (2020) and Poerio et al. (2017) for prior examples of such studies).

4.1. Experience sampling can connect the lab to the field

The observation that mDES can match daily life activities with laboratory tasks has important implications for future theory development. First, in sub-clinical populations mDES can track symptoms of disorders such as autism (Strawson et al., in press; Turnbull et al., 2020) and depression (Konu et al., 2021). Consequently, it may be possible to evaluate the mDES fingerprints of different clinical populations in daily life and examine how these patterns of thoughts relate to detailed cognitive measurements recorded in the laboratory (e.g., Andrews-Hanna et al., 2020; Happe et al., 2017; Osborne-Crowley & McDonald, 2016). This would help advance mechanistic accounts of psychiatric disturbance and other features of mental health. Second, since mDES can easily be measured in daily life using smartphones, it enables studies to sample cognition within a broad range of cultural contexts, and across a broader range of participants than would normally be possible in standard laboratory environments. Consequently, it will be possible to build a more inclusive model of psychological research that extends beyond the predominantly WEIRD demographics that characterise undergraduate student populations (Gurven, 2018; Harrington & Gelfland, 2014). Third, it will allow the extension of lab models of cognition in real world environments such as courtrooms, conversations or in-person lectures (Goodman & Hahn, 1987; Rose, 2017) to examples of these behaviours as they occur naturally within the field. Finally, as mDES has been shown to stably relate to patterns of neural activation via fMRI (e.g., Ho et al., 2020; Karapanagiotidis et al., 2020; Konu et al., 2020; Sormaz et al., 2018; Turnbull, Wang, Murphy et al., 2019; Wang et al., 2018) and EEG (e.g., Simola et al., 2023), it may even be possible to offer insight into the brain activation associated with cognitive states in the context within which they naturally occur.

4.2. Open questions for future research

Our study highlights the capacity for mDES to capture comparable patterns in the lab and daily life, and to highlight conceptual groupings between lab tasks and activities in daily life. Nonetheless, important open questions remain unanswered. For example, our study used a young, predominantly female (~76 %) white undergraduate sample, and it remains unclear how well our findings may generalize among other populations (e.g., older, male, non-WEIRD, etc.) (Turnbull et al., 2021). The overrepresentation of white, English-speaking participants remains a pervasive issue in psychological research, potentially biasing researchers’ conclusions (Blasi et al., 2022; Roberts and Mortenson, 2023; Thalmayer et al., 2021). mDES offers a cost-effective and flexible tool for improving inclusivity in future studies, but broader conclusions about thought patterns in daily life should await more representative data.

While psychological science often suffers from samples being too homogeneous, the opposite is the case in terms of methodology (e.g., Sanbonmatsu et al., 2021; Seli et al., 2018). While the dimensions extracted from the datasets used in this analysis formed consistent components across lab and daily life, each parent study slightly differed in their overall questionnaire and probing schedule (see Methods, Table S1, S2). In one sense the consistency of component structure we observed in spite of these differences speaks to a strength of mDES as a tool for harmonizing data from disparate sources. Still, as touched on above, further examination is warranted as to how reported thought may change as participants are probed to describe their state in greater detail, more frequently, and at different times of the day (Mulholland et al., 2023; Wallace et al., 2025). An intriguing next step would be to adopt a longitudinal approach by repeatedly administering mDES across diverse settings over an extended period. This would allow us to model the effects of context on cognition and compare how psychological state fluctuations throughout the day and year differ across environments.

Third, although the questions used in these studies can generalise thought patterns from the lab and daily life, other experience-sampling question sets may capture this common space even more efficiently or reveal dimensions that better distinguish contexts in the lab and real world. Additionally, our generation of a shared space relies on dimension reduction techniques, showing a high degree of similarity between PCA and Exploratory Factor Analysis (see Supplementary Fig. S1). Nonetheless there may be more sophisticated approaches for forming a shared space that can account for other factors (such as ongoing activities or trait variance of the participants) that could be employed in the future. For example, while PCA is useful for modelling linear relationships in data, dimensionality reduction techniques such as t-distributed stochastic neighbour embedding (t-SNE; van der Maaten & Hinton, 2008) or uniform manifold approximation and projection (UMAP; McInnes et al., 2020) may more effectively capture non-linear patterns in ongoing thought. Ultimately, such an approach, combined with machine learning optimization, could reveal the best set of questions for explaining human cognition as it emerges in both the lab and daily life (see Lorenz et al., 2018).

Finally, it is worth noting that the key limitation in the use of experience sampling to map cognition is its reliance on introspection. Although introspection is known to have limits, mDES has been shown to be sensitive to objective indicators of cognition such as pupillometry (Konishi et al., 2017), electroencephalography (Simola et al., 2023) and functional magnetic resonance imaging (Karapanagiotidis et al., 2020; Konu et al., 2020Turnbull, Wang, Schooler et al., 2019) and can discriminate traits within normal population (Turnbull et al., 2020). There are likely to be features of the organisation of cognition that techniques like mDES are unable to describe efficiently, however, as the current work shows it does have the capacity to generate common patterns that show construct validity in their capacity to distinguish activities with similar features. It is possible that such sensitivity emerges because introspection allows individuals to distinguish between the sorts of cognitive operations that occur in different circumstances rather than because they have insight into the underlying processes which support the states themselves. Fortunately, it will be possible in the future to explore this possibility by capitalising on the associations between approaches like mDES and objective indicators of cognition to explore which features of mDES map onto which underlying features of cognition.

5. Transparency Statement

5.1. General disclosures

Conflicts of interest: All authors declare no conflicts of interest. Funding: This research is supported by award to Dr. Jonathan Smallwood and Dr. Jeffery D. Wammes from the Government of Canada's New Frontiers in Research Fund (NFRF) [grant ID NFRF-2021–00183]. Artificial intelligence: No artificial intelligence assisted technologies were used in this research or the creation of this article. Ethics: This research complies with the Declaration of Helsinki (2023), aside from the requirement to preregister human subjects research, and received approval from a local ethics board (GREB IRB Number: IRB00003062). Computational reproducibility: The code used in this analysis is available on the ThoughtSpace Github linked in the article. The relevant datasets used in this article are available in their source publications.

5.2. Study 1

Preregistration: No aspects of the study were preregistered. Materials: Study materials are presented in the article and any idiosyncratic features are published in the relevant papers from which the datasets came. Data: the dataset used in the present analysis is publicly available on Mendeley Data (https://doi.org/10.17632/7wjm45zmyw.1). Analysis scripts: all analysis scripts are publicly available on the ThoughtSpace Github linked in the article.

CRediT authorship contribution statement

Louis Chitiz: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Bronte Mckeown:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Bridget Mulholland:** Writing – review & editing, Resources, Methodology, Formal analysis, Data curation. **Raven Wallace:** Writing – review & editing, Resources, Methodology, Data curation. **Ian Goodall-Halliwell:** Writing – review & editing, Visualization, Software, Resources, Methodology, Formal analysis, Conceptualization. **Nerissa Siu Ping-Ho:** Writing – review & editing, Resources, Data curation. **Delali Konu:** Writing – review & editing, Resources, Data curation. **Giulia L. Poerio:** Writing – review & editing, Conceptualization. **Jeffrey Wammes:** Writing – review & editing, Funding acquisition, Conceptualization. **Michael Milham:** Writing – review & editing. **Arno Klein:** Writing – review & editing, Software. **Elizabeth Jefferies:** Writing – review & editing. **Robert Leech:** Writing – review & editing, Formal analysis. **Jonathan Smallwood:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2025.103853.

Data availability

The data is available on Mendeley Data (link in manuscript)

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*(1), 30.

Aickin, M., & Gensler, H. (1996). Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm methods. *American Journal of Public Health, 86*(5), 726–728.

Altman, D. G., & Bland, J. M. (2011). How to obtain the confidence interval from a P value. *BMJ : British Medical Journal, 343*, Article d2090.

Andrews-Hanna, J. R., Christoff, K., & O'Connor, M. F. (2020). Dynamic regulation of internal experience. Neuroscience of enduring change: implications for psychotherapy, 89.

Berry, W. D., & Feldman, S. (1985). *Multiple regression in practice, No. 50*. Sage.

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences, 26*(12), 1153–1170.

Broadbent, D. E. (1971). *Decision and stress*. London: Academic Press.

Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature, 534*(7609), 684–687.

Centrum voor Wiskunde en Informatica (1995). Python tutorial. Technical Report CS R9526. Amsterdam: van Rossum, G.

Cialdini, R. B. (2009). We have to break up. *Perspectives on Psychological Science, 4*(1), 5–6.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* ((2nd ed.).). L. Erlbaum Associates.

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic bulletin & review, 24*, 1158–1170.

de Caso, I., Karapanagiotidis, T., Aggius-Vella, E., Konishi, M., Margulies, D. S., Jefferies, E., & Smallwood, J. (2017). Knowing me, knowing you: Resting-state functional connectivity of ventromedial prefrontal cortex dissociates memory related to self from a familiar other. *Brain and Cognition, 113*, 65–75.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences, 14*(4), 172–179.

Everett, J. E. (1983). Factor comparability as a means of determining the number of factors and their rotation. *Multivariate Behavioral Research, 18*(2), 197–218.

Finn, E. S., Poldrack, R. A., & Shine, J. M. (2023). Functional neuroimaging as a catalyst for integrated neuroscience. *Nature, 623*(7986), 263–273.

Fischhoff, B. (1996). The real world: What good is it? *Organizational Behavior and Human Decision Processes, 65*(3), 232–248.

Fisher, R. A. (1921). On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample. *Metron, 1*, 3–32. Article 14.

Franklin, S. B., Gibson, D. J., Robertson, P. A., Pohlmann, J. T., & Fralish, J. S. (1995). Parallel analysis: A method for determining significant principal components. *Journal of Vegetation Science, 6*(1), 99–106.

Goodwin, K. A., & Goodwin, C. J. (2016). *Research in psychology: Methods and design*. John Wiley & Sons.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika, 40*, 33–51.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The neuroscience of film. *Projections, 2*(1), 1–26.

Ho, N. S. P., Poerio, G., Konu, D., Turnbull, A., Sormaz, M., Leech, R., & Smallwood, J. (2020). Facing up to the wandering mind: Patterns of off-task laboratory thought are associated with stronger neural recruitment of right fusiform cortex while processing facial stimuli. *Neuroimage, 214*, Article 116765.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65–70.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185.

Jain, A., Nandakumar, K., & Ross, A. (2005). Score normalization in multimodal biometric systems. *Pattern recognition, 38*(12), 2270–2285.

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science, 18*(7), 614–621.

Kane, M. J., Gross, G. M., Chun, C. A., Smeekens, B. A., Meier, M. E., Silvia, P. J., & Kwapil, T. R. (2017). For whom the mind wanders, and when, varies across laboratory and daily life settings. *Psychological Science, 28*(9), 1271–1289.

Karapanagiotidis, T., Vidaurre, D., Quinn, A. J., Vatansever, D., Poerio, G. L., Turnbull, A., & Smallwood, J. (2020). The psychological correlates of distinct neural states occurring during wakeful rest. *Scientific Reports, 10*(1), 21121.

Kihlstrom, J. F. (2021). Ecological validity and "ecological validity". *Perspectives on Psychological Science, 16*(2), 466–471.

Kingstone, A., Smilek, D., Ristic, J., Kelland Friesen, C., & Eastwood, J. D. (2003). Attention, researchers! It is time to take a look at the real world. *Current Directions in Psychological Science, 12*(5), 176–180.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology, 99*(3), 317–340.

Konishi, M., Brown, K., Battaglini, L., & Smallwood, J. (2017). When attention wanders: Pupillometric signatures of fluctuations in external attention. *Cognition, 168*, 16–26.

Konu, D., Turnbull, A., Karapanagiotidis, T., Wang, H. T., Brown, L. R., Jefferies, E., & Smallwood, J. (2020). A role for the ventromedial prefrontal cortex in self-generated episodic social cognition. *Neuroimage, 218*, Article 116977.

Konu, D., Mckeown, B., Turnbull, A., Ho, N. S. P., Karapanagiotidis, T., Vanderwal, T., & Smallwood, J. (2021). Exploring patterns of ongoing thought under naturalistic and conventional task-based conditions. *Consciousness and Cognition, 93*, Article 103139.

Lin, H., Werner, K. M., & Inzlicht, M. (2021). Promises and perils of experimentation: The mutual-internal-validity problem. *Perspectives on Psychological Science, 16*(4), 854–863.

Lorenz, R., Violante, I. R., Monti, R. P., Montana, G., Hampshire, A., & Leech, R. (2018). Dissociating frontoparietal brain networks with neuroadaptive Bayesian optimization. *Nature Communications, 9*(1), 1227.

Lorenzo-Seva, U., & Ten Berge, J. M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*(2), 57–64.

Lovik, A., Nassiri, V., Verbeke, G., & Molenberghs, G. (2020). A modified Tucker's congruence coefficient for factor matching. *Methodology, 16*(1), 59–74.

van der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research, 9*(11), 2579–2605.

Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2019). Are we ready for real-world neuroscience? *Journal of Cognitive Neuroscience, 31*(3), 327–338.

McInnes, L., Healy, J., & Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction (arXiv: 1802.03426v3). Tutte Institue for Mathematics and Computing.

McKeown, B., Goodall-Halliwell, I., Wallace, R., Chitiz, L., Mulholland, B., Karapanagiotidis, T., Hardikar, S., Strawson, W., Turnbull, A., Vanderwal, T., Ho, N., Xu, H.-T., Wang, T., Milham, M., Wang, X., Zhang, M., Gonzalez Alam, T. R., Vos De Wael, R., Bernhardt, B., & Smallwood, J. (2025). Self-reports map the landscape of task states derived from brain imaging. *Communications Psychology, 3*(1).

Mckeown, B., Poerio, G. L., Strawson, W. H., Martinon, L. M., Riby, L. M., Jefferies, E., & Smallwood, J. (2021). The impact of social isolation and changes in work patterns on ongoing thought during the first COVID-19 lockdown in the United Kingdom. *Proceedings of the National Academy of Sciences, 118*(40), Article e2102565118.

Mckeown, B., Strawson, W. H., Zhang, M., Turnbull, A., Konu, D., Karapanagiotidis, T., & Smallwood, J. (2023). Experience sampling reveals the role that covert goal states play in task-relevant behavior. *Scientific Reports, 13*(1), 21710.

Mulholland, B., Goodall-Halliwell, I., Wallace, R., Chitiz, L., Mckeown, B., Rastan, A., & Smallwood, J. (2023). Patterns of ongoing thought in the real world. *Consciousness and Cognition, 114*, Article 103530.

Murphy, C., Poerio, G., Sormaz, M., Wang, H.-T., Vatansever, D., Allen, M., Margulies, D. S., Jefferies, E., & Smallwood, J. (2019). Hello, is that me you are looking for? A re-examination of the role of the DMN in social and self relevant aspects of off task thought. *PLOS ONE, 14*(11), Article e0216182.

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: Rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage, 222*, Article 117254.

Neisser, U. (1980). On" Social Knowing". *Personality and Social Psychology Bulletin, 6*(4), 601–605.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*(11), 776–783.

Orne, M. T. (1979). Commentary on "Citation Classic - On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications.". *Current Contents/Social & Behavioral Sciences, 13*, 56.

Osborne-Crowley, K. (2020). Social cognition in the real world: Reconnecting the study of social cognition with social reality. *Review of General Psychology, 24*(2), 144–158.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, É. (2011). Scikit-learn: machine learning in Python. *the Journal of machine Learning research, 12*, 2825–2830.

Poerio, G. L., Sormaz, M., Wang, H.-T., Margulies, D., Jefferies, E., & Smallwood, J. (2017). The role of the default mode network in component processes underlying the wandering mind. *Social Cognitive and Affective Neuroscience, 12*(7), 1047–1062. https://doi.org/10.1093/scan/nsx041

Pooja, R., Ghosh, P., & Sreekumar, V. (2024). Towards an ecologically valid naturalistic cognitive neuroscience of memory and event cognition. *Neuropsychologia.,, 203*, Article 108970.

Roberts, S. O., & Mortenson, E. (2023). Challenging the White = Neutral Framework in Psychology. *Perspectives on Psychological Science, 18*(3), 597–606.

Rubio-Tomás, T., Skouroliakou, M., & Ntountaniotis, D. (2022). Lockdown due to COVID-19 and its consequences on diet, physical activity, lifestyle, and other aspects of daily life worldwide: A narrative review. *International Journal of Environmental Research and Public Health, 19*(11), 6832.

Ruby, F. J., Smallwood, J., Engen, H., & Singer, T. (2013). How self-generated thought shapes mood—the relation between mind-wandering and mood depends on the socio-temporal content of thoughts. *PloS one, 8*(10), Article e77554.

Sanbonmatsu, D. M., Cooley, E. H., & Butner, J. E. (2021). The Impact of Complexity on Methods and Findings in Psychological Science. *Frontiers in psychology, 11*, Article 580111.

Seli, P., Kane, M. J., Smallwood, J., Schacter, D. L., Maillet, D., Schooler, J. W., & Smilek, D. (2018). Mind-Wandering as a Natural Kind: A Family-Resemblances View. *Trends in Cognitive Sciences, 22*(6), 479–490.

Simola, J., Silander, T., Harju, M., Lahti, O., Makkonen, E., Pätsi, L. M., & Smallwood, J. (2023). Context independent reductions in external processing during self generated episodic social cognition. *Cortex, 159*, 39–53.

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology, 66*, 487–518.

Smallwood, J., Turnbull, A., Wang, H. T., Ho, N. S., Poerio, G. L., Karapanagiotidis, T., & Jefferies, E. (2021). The neural correlates of ongoing conscious thought. *Iscience, 24*(3).

Smallwood, J., Bernhardt, B. C., Leech, R., Bzdok, D., Jefferies, E., & Margulies, D. S. (2021). The default mode network in cognition: A topographical perspective. *Nature Reviews Neuroscience, 22*(8), 503–513.

Sormaz, M., Murphy, C., Wang, H. T., Hymers, M., Karapanagiotidis, T., Poerio, G., & Smallwood, J. (2018). Default mode network can support the level of detail in experience during active task states. *Proceedings of the National Academy of Sciences, 115*(37), 9318–9323.

Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist, 76*(1), 116–129.

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* (Vol. 984). Princeton, NJ: Educational Testing Service.

Turnbull, A., Wang, H. T., Murphy, C., Ho, N. S. P., Wang, X., Sormaz, M., & Smallwood, J. (2019). Left dorsolateral prefrontal cortex supports context dependent prioritisation of off-task thought. *Nature Communications, 10*(1), 3816.

Turnbull, A., Wang, H. T., Schooler, J. W., Jefferies, E., Margulies, D. S., & Smallwood, J. (2019). The ebb and flow of attention: Between-subject variation in intrinsic connectivity and cognition associated with the dynamics of ongoing experience. *Neuroimage, 185*, 286–299.

Turnbull, A., Garfinkel, S. N., Ho, N. S., Critchley, H. D., Bernhardt, B. C., Jefferies, E., & Smallwood, J. (2020). Word up–Experiential and neurocognitive evidence for associations between autistic symptomology and a preference for thinking in the form of words. *Cortex, 128*, 88–106.

Wallace, R. S., McKeown, B., Goodall-Halliwell, I., Chitiz, L., Forest, P., Karapanagiotidis, T., Mulholland, B., Turnbull, A., Vanderwal, T., Hardikar, S., Gonzalez Alam, T. R., Bernhardt, B. C., Wang, H.-T., Strawson, W., Xu, M., Milham, T., Margulies, D. S., Poerio, G. L., Jefferies, E., & Smallwood, J. (2025). Mapping patterns of thought onto brain activity during movie-watching. *eLife, 13*.

Wang, H. T., Bzdok, D., Margulies, D., Craddock, C., Milham, M., Jefferies, E., & Smallwood, J. (2018). Patterns of thought: Population variation in the associations between large-scale network organisation and self-reported experiences at rest. *Neuroimage, 176*, 518–527.

Wang, H.-T., Ho, N. S. P., Bzdok, D., Bernhardt, B. C., Margulies, D. S., Jefferies, E., & Smallwood, J. (2020). Neurocognitive patterns dissociating semantic processing from executive control are linked to more detailed off-task mental time travel. *Scientific Reports, 10*(1).