Mapping Cognition Across Lab and Daily Life Using Experience Sampling

By

Louis Chitiz

A thesis submitted to the Graduate Program in Psychology (Cognitive Neuroscience)
in conformity with the requirements for the
Degree of Master of Science

Queen's University
Kingston, Ontario, Canada
September, 2024

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.

Co-Authorship

This study was conducted at Queen's University in the THinC Lab by Louis Chitiz, under the supervision of Dr. Jonathan Smallwood. Louis Chitiz, Raven Wallace, Dr. Brontë Mckeown, Ian Goodall-Halliwell, and Bridget Mulholland all contributed to study design, with additional consultations from Dr. Robert Leech, Dr. Giulia Poerio, Raven Wallace, Dr. Brontë Mckeown, and Ian Goodall-Halliwell. The analysis used previously published data collected by Bridget Mulholland, Dr. Brontë Mckeown, Dr. Nerissa Siu-Ping Ho, and Dr. Delali Konu. Software design was completed by Louis Chitiz, with consultation from Ian Goodall-Halliwell and Dr. Brontë Mckeown. Formal analysis was completed by Louis Chitiz. Figure development was completed by Louis Chitiz. The final manuscript was written by Louis Chitiz, with writing and editing contributions from Dr. Jonathan Smallwood, Raven Wallace, Dr. Giulia Poerio, Dr. Elizabeth Jefferies, Dr. Arno Klein, and Dr. Robert Leech.

Acknowledgments

I would first like to thank my family (Mom, Dad, Ruth, and Sophie) for being so supportive of me through this whole process. None of this, not anything, would have been possible without your steadfast support and love. I love you guys.

Next, I would like to thank my supervisor Dr. Jonathan Smallwood. Beyond just being an invaluable guide through this process and a voice of reason and wisdom to me as a student, you have been good to me as a person, and seen me through my highs and lows with unwavering support, empathy and understanding. For that I'm really grateful to you Jonny, thank you.

Last but not least I would of course like to thank my lab-mates over the course of the past two years Raven Wallace, Bridget Mulholland, Ian Goodall-Halliwell, Dr. Samyogita Hardikar, and Dr. Brontë Mckeown. You were all an essential part of this for me, but more importantly it was a pleasure making friends with all of you. I look forward to more collaborations and good times to come with the squad!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

fMRI	**Functional Magnetic Resonance Imaging**
mDES	**Multidimensional Experience-Sampling**
PCA	**Principal Components Analysis**
PC	**Principal Component**
RHom.	**R-Homologue**
TCC	**Tucker's Congruence Coefficient**

Chapter 1: Introduction

A central goal of psychological science 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., 2023). Ecologically valid methods such as field studies that examine phenomena 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., 2021; Kingstone et al., 2003; Matusz et al., 2019; Neisser, 1980; Orne, 1979; Osborne-Crowley, 2020). As such 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, partly because there are few tools that researchers can use to record cognition in both situations (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.

The 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 lab and during brain imaging; for a review see Smallwood 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). Recent studies have successfully used mDES in both the lab (Iwata et al., 2024; 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).

Since mDES can be easily administered across 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. To identify thought patterns common across lab and daily life that might form this shared space, we leveraged the established compatibility of mDES with pattern analysis approaches (e.g., Principal Components Analysis (PCA), hierarchical clustering, chain P-factor analysis, etc.) (Smallwood et al., 2021). More specifically, we used PCA, which has the strongest precedent in prior experience-sampling research, to capture groupings of thought characteristics (i.e., "thought patterns"), which account for large chunks of variance in subjects' reported thought (Smallwood et al., 2021). Our analysis used data collected 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 would constitute an important step in the generation of mechanistic theories of human behaviour that are grounded in the real world.

Chapter 2: Literature Review

2.1 – Early Experience-Sampling Approaches for Connecting Lab and Daily Life

While the use of experience-sampling to measure cognition is not novel, even in the real-world, there are surprisingly few formal analyses of how it can be used to compare lab and daily-life (Prescott, Csikszentmihalyi, and Graef, 1976; Csikszentmihalyi, Larson, and Prescott, 1977; Kane et al., 2007). Early approaches from Kane and colleagues examined the relationship between lab-based cognitive measures and “feral cognition”, which they probed using Personal Digital Assistants that they provided to subjects (Kane et al., 2007, 2017; McVay, Kane, and Kwapil, 2009). These studies relied on a unidimensional experience-sampling approach focused on measuring the occurrence of mind-wandering (operationalized as task-unrelated thought) in the lab and daily-life (Kane et al., 2007, 2017; McVay, Kane, and Kwapil, 2009). The sampling approach they adopted consisted of an initial self-assessment of whether a subject’s thoughts pertained to their external task and, if they did not, an additional 4 questions to further characterize their task-unrelated state (e.g., were their thoughts deliberately task-unrelated). Afterwards, whether or not the subject reported being on task, they would answer a set of 18 further Likert-type questions to describe their surrounding internal and external context (e.g., their emotional state, sobriety, their interest in/the nature of the task they were doing) (Kane et al., 2007; McVay, Kane, and Kwapil, 2009).

Rather than seeking to characterize cognitive states, these studies focused more intently on how internal/external context in daily-life as well as task-performance on lab measures predicted the occurrence of task-unrelated thought. While finding several predictors of task-unrelated thought within daily-life, these studies importantly found that performance on lab-based measures of cognition associated with mind-wandering rate in the real-world. Some of these lab-to-real-world relationships were direct, for example subjects who reported more task-unrelated thoughts during a sustained attention task in the lab (Sustained Attention to Response task; Robertson et al., 1997) reported more frequent task-unrelated thought in daily life (McVay, Kane, and Kwapil, 2009). Others were more indirect, individuals who performed more strongly on working memory tasks (SPAN tasks, where individuals held a string of letters in memory while performing distracting processing tasks between memory intake trials) seemed to exert better control over their task-focus in daily-life, remaining more on-task as they tried harder to concentrate and as their tasks became more difficult than those with a lower working memory capacity (Kane et al., 2007). Interestingly, these studies found that lab-based task performance also associated with thought contents reported during daily-life, such that subjects who reported more task-unrelated thoughts during a sustained attention task tended to also report more worrying in the real-world (McVay, Kane, and Kwapil, 2009).

While these earlier studies established that cognition as measured in laboratory settings likely has some predictive bearing on the cognitive states people inhabit in the real world, their results often failed to replicate in later research (Kane et al., 2017). A more recent study with a larger sample size found that subjects' frequency of task-unrelated thought during a variety of executive-control tasks (a sustained-attention task, a Stroop task, 2 variations on the flanker task, and a 2-back task) was unrelated to their rate of task-unrelated thought in daily life (Kane et al., 2017). In addition, participants' working memory capacities still predicted their ability to suppress or allow mind-wandering as they tried harder or less hard to concentrate, but no longer related to the association between task-unrelated thought and whether a task was challenging (Kane et al., 2017). Interestingly, this study identified that other indicators of stronger executive functioning such as lower rates of task-unrelated thought and stronger performance on attentional control tasks also predicted a greater negative association between reported concentration and wandering off-task (Kane et al., 2017). In other words, subjects whose concentration efforts better suppressed mind-wandering in daily life tended to also report fewer off-task thoughts and also exhibit greater attentional control during lab tasks (Kane et al., 2017). With regards to the association between task-performance and thought content in daily-life, while Kane et al. (2017) failed to replicate the association between task-unrelated thought in the lab and worrying in daily-life, they found that individuals who performed more poorly on attention restraint tasks (e.g., ignoring a stimulus on one side of a screen to make a response based on a less salient target on the other side of the screen) and reported more task-unrelated thought tended to describe their thoughts as being less controllable and more "racing" in daily life. Altogether, earlier experience-sampling approaches to comparing lab and daily-life found several indicators that cognitive performance and reported thought in the lab related to reported thought in daily-life, but generally focused on only a limited array of cognitive states (i.e., task-unrelated thought) and often yielded findings which did not replicate in larger samples (Kane et al., 2007, 2017; McVay, Kane, and Kwapil, 2009).

2.2 – Multidimensional Experience-Sampling Approaches

In contrast to studies that use experience-sampling to catch researcher-defined states such as mind-wandering, other researchers have taken an experience-sampling approach to characterizing states of ongoing thought more broadly (Martinon et al., 2019; Smallwood et al., 2021). Rather than focus on one aspect of internal experience, these studies use a technique called "multidimensional experience-sampling" (mDES) in which subjects score their most recent thoughts according to a series of characteristics (e.g., my thoughts were deliberate, about the past, in the form of words, etc.) on Likert scales (Smallwood et al., 2021). To reduce the relatively high dimensional datasets associated with mDES into more manageable data structures, researchers often make use of dimensionality reduction techniques such as Principal Components Analysis (PCA) (Martinon et al., 2019). The resultant components represent "thought patterns" comprising groupings of thought characteristics that tend to covary (Smallwood et al., 2021). In response to the definitional controversy associated with many researcher-defined states such as mind-wandering, many in the field argue that defining states in a more data-driven manner such as is done with mDES could allow for a more objective, consistent approach to studying ongoing thought (Seli et al., 2018; Smallwood et al., 2021).

As discussed in the introduction, mDES' flexibility and access to the full range of cognitive states (i.e., both externally and internally oriented states) lends the method potential for comparing cognition across lab and daily life. However, despite this potential, there is little literature using mDES in this regard. In the studies we could find, researchers administering mDES in both lab and daily-life typically probed subjects over the course of some cognitive task in a laboratory setting and then probed them via a smartphone app in daily life (Linz et al., 2019; Ho et al., 2020). In two connected studies, Engert et al. (2014) administered 7-item mDES probes to a sample of 99 subjects while they completed a choice reaction time task and a working memory task in a lab setting and then Linz et al. (2019) conducted mDES using the same questionnaire in a largely overlapping sample of 87 subjects over two days of daily-life. First, decomposing their daily life probes using PCA, the researchers identified three components: a future self-relevant component, a past-oriented social-cognition component, and a component related to thought valence. Afterwards, to examine the correspondence in reported thought between lab and daily life, Linz et al. (2019) correlated within-subject averages for each of the 7 items in the questionnaire across lab and daily life probes. The researchers found that reported thought between lab and daily life related differently depending on the dimension of thought in question, such that subjects only rated the degree to which their thoughts were negative, future-oriented, about others, and about oneself in a significantly similar manner between lab and daily-life (Engert et al., 2014; Linz et al., 2019). In addition to reported thought, both studies measured salivary cortisol as an indicator of subject stress levels throughout data collection (Engert et al., 2014; Linz et al., 2019). Interestingly, while cortisol levels associated with aspects of reported thought in the lab (negative thoughts were linked to higher cortisol and future- and socially-oriented thoughts were linked to decreased cortisol), there was no association observed between them in daily-life. Of note, while they used pattern analysis approaches to identify patterns in ongoing thought in daily life and the lab separately, they used different approaches in each environment and conducted no analysis to examine how the underlying structure of the dimensions of ongoing thought that they measured compared across the two sampling environments (Engert et al., 2014; Linz et al., 2019).

Extending the research of Linz et al. (2019), Ho et al. (2020) administered a 13-item mDES questionnaire to a sample of 78 subjects while doing a variation of the n-back task (0- vs 1-back task) in fMRI and then over 7 days of daily life. Unlike the analysis of Linz et al. (2019), the researchers in Ho et al. (2020) compared mDES across lab and daily life by comparing the similarity of their PCA solutions rather than correlating the two based on individual questionnaire items. The authors decided to extract 4 components for lab and daily-life, which they labelled based on their loadings: 1) *Vivid, Evolving & Detailed*, 2) *Off-Task & Self-Relevant*, 3) *Deliberate & Task-Relevant Emotion*, and 4) *Modality*. Importantly, like Linz et al. (2019), Ho et al. (2020) identified that the degree of similarity in ongoing thought between lab and daily-life differed depending on the central items in question. More specifically, components

1, 2, and 4 were all significantly correlated across lab and daily-life, but *Deliberate & Task-Relevant Emotion* did not significantly correlate across the sampling environments (Ho et al., 2020). Following up this analysis by qualitatively comparing components separately extracted from lab and daily life probes, Ho et al. (2020) determined that in a lab-setting, thoughts described as 'off-task' tended to be much more associated with other people than off-task thought in daily life. In addition to comparing reported thought across lab and daily life, the researchers also examined how internal experience in lab and daily life related to measures of cortical thickness as well as neural activation while viewing clear or scrambled faces and scenes. The researchers found some brain-thought associations were consistent across lab and daily life, greater thickness of the bilateral parahippocampus related to more detailed thought across both sampling environments. In other cases, how brain activation related to ongoing thought in the lab shed less insight into how thought and neural activation might relate in the real word. For example, greater activation of a cluster around the fusiform gyrus during facial processing associated with greater frequency of off-task thoughts in the lab, which tended to be more self-relevant, while the same activation did not predict the more socially-oriented off-task thought observed in daily life (Ho et al., 2020). This illustrates the capacity of mDES to highlight the possible differences that off-task thought can have when it occurs in different contexts (e.g., in the lab vs daily life).

In summary, the mDES approach, with its unassuming approach to defining states of internal experience in a data-driven manner, demonstrates significant potential as one method for comparing cognition across lab and daily life. However, there is little literature assessing its potential for doing so. What research has been done has indicated that ongoing experience is largely consistent across lab and daily life with some notable differences in terms of the actual experiential content itself as well as how that experience relates to physiological markers of stress (i.e., salivary cortisol) and lab-based patterns of neural activation (e.g., off-task thought and face-related fusiform activation). Still, despite the contributions of these studies, certain questions remain: how effectively do patterns of reported thought generalize across different samples? Could representing lab and daily life using a single set of components as was done in Ho et al. (2020) bias the patterns observed to be more representative of one environment over the other? Further, neither Linz et al. (2019) nor Ho et al. (2020) considered the context surrounding the probes administered in daily life (i.e., where is the subject? Are they interacting with anyone? what are they doing? etc.). Would associated patterns of reported thought organize these contexts in a coherent manner? Could mDES further be used to assess, say, how well lab-based tasks match the internal experience of the everyday situations they are intended to emulate?

2.3 – The Current Study

The present study extends this prior work by analyzing mDES in a manner not yet attempted in experience-sampling research. Rather than systematically administering probes within the same sample across lab and daily life and comparing ongoing thought within-subject, we sought to test whether mDES probes form a generalizable enough component-structure to be synthesized across samples. If so, this would allow researchers to take a more data-driven approach to establishing how contexts across the lab and daily life might align and diverge in terms of their associated psychological states. This quantified system for determining similarity and difference between contexts based on internal experience would afford researchers the capacity to develop psychological theories with a stronger consideration for cognitive functioning in the real world. Using previously published data collected in a range of sampling environments (fMRI, the behavioural lab, and daily life), we tested: 1) how mDES' component-structure translated across different sampling environments, 2) whether a single set of components could represent all of the datasets in an unbiased manner, and 3) if that single set of components organized the surrounding contexts for each probe in a coherent fashion (e.g., do probes while doing in homework in Canada versus the UK associate with consistent patterns of reported thought). Should this more meta-analytic approach of analyzing mDES across separate studies produce reliable and interpretable results, the study would be a promising indication that mDES could allow for quantifiable comparisons of cognitive state across lab and daily life hitherto unexamined in cognitive research.

Chapter 3: Materials and Methods

3.1 – Participants and Dataset Characteristics

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) = 21.57 years) 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.

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

Note. 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.

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 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 Appendix A for a summary of sampled contexts.

3.2 – Common Dimensions of Experience and Combination of Datasets

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, $r(4973) = .92, p < .001$. To standardise the combined data, all scores were re-coded into a continuous 1-5 scale and then *z*-scored. See Table 2 for the set of dimensions included in the analysis. See Appendix A for the questions used to measure each dimension in each dataset.

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

Note. 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.

3.3 – Statistical Analysis

3.3.1 – Component Reproducibility Analyses

The goal of our analysis was to examine how different lab and daily life contexts were comparable in a shared cognitive 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 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 (https://github.com/Bronte-Mckeown/ThoughtSpace/releases/tag/v1.0.1) in Python 3.12.0 (Centrum voor Wiskunde en Informatica,1995). Specifically, component-similarity analyses used the RHom module of the package. 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 of 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).

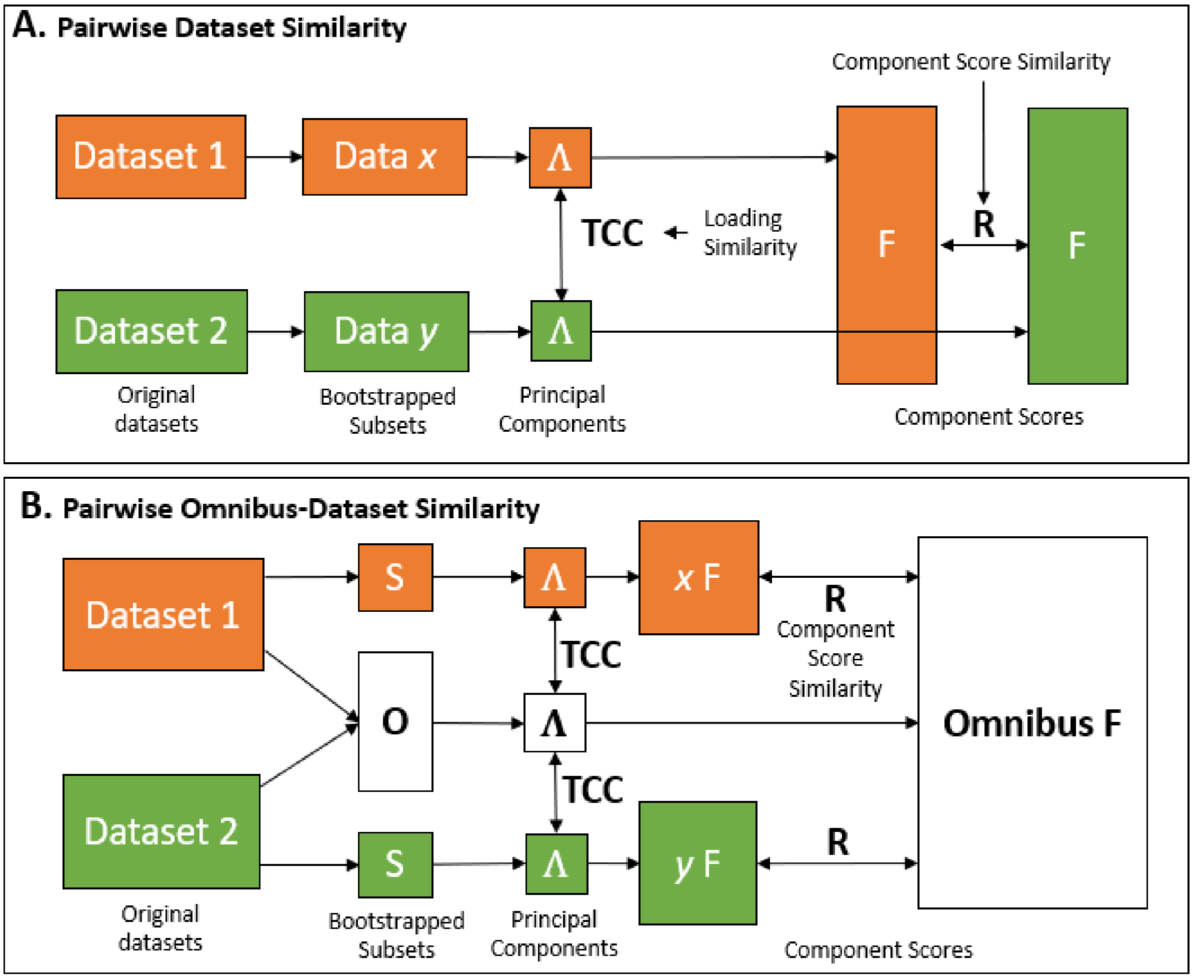
We calculated component loading similarity using Tucker's Congruence Coefficient (TCC; Lovik et al., 2020; Tucker, 1951), where a score $\geq .85$ indicates fair similarity and $\geq .95$ indicates exact similarity (Lorenzo-Seva & Ten-Berge, 2006). We computed component score similarity between homologous 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 cut-off of .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).

3.3.2 – Sample-to-Sample Reproducibility

Our first objective was to assess how well the thought patterns generated by the decomposition of mDES in different settings 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 combination of folds in the comparison set.

Figure 1.



Note. **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.

3.3.3 – 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 followed this up by conducting 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.

3.3.4 Shared-Space Context Mappings

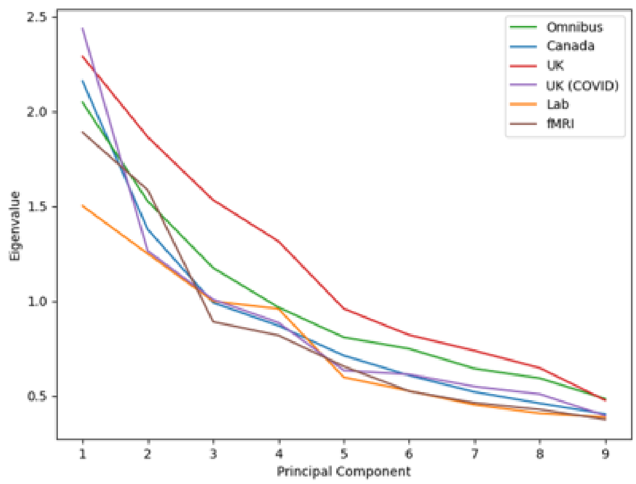
Beyond examining the reliability with which mDES generates thought patterns across lab and daily life contexts, our final goal was to explore the theoretical potential of this ‘shared space' by examining whether it maps contexts one might expect to be similar (e.g., Working in Canada versus the UK) in consistent ways. We further 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.) (See Appendix A for the full list of activities/tasks sampled). We only report results for task/activity contexts in our main analysis. See Appendix B 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.

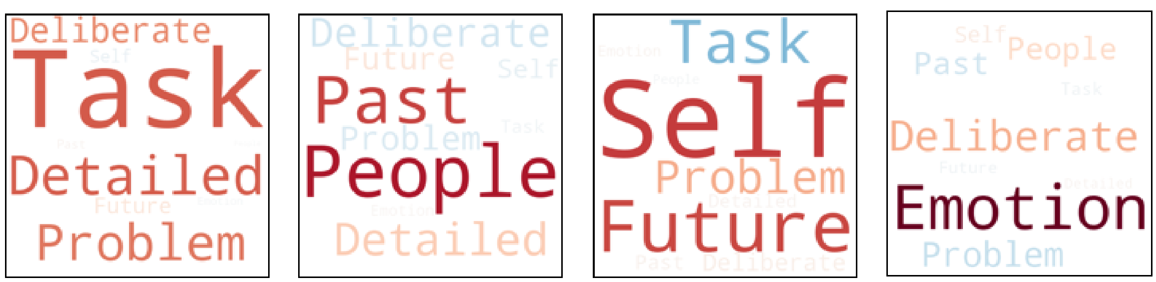
Chapter 4: Results

4.1 – Initial PCA Results

Figure 2.



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]



Note. **4-PC solution performance across datasets.** (Top left) Scree plot for each dataset as well as 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 word = stronger loading), word colour indicates loading directionality (i.e., warmer colours = more positive loading, cooler = more negative loading).

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 supplementary materials for the 3-PC results). The 4 components accounted for approximately 63.5% of the total variance in the dataset, and accounted for a consistent proportion of variance across each dataset (See Fig. 2, Table S1). 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). In our paper we focus on the 4-component solution analyses and include separate results for the 3-PC solution in the supplementary materials.

Table 3. *PCA loadings for each component.*

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

Note. **Loadings for each component on the omnibus PCA**. Displayed loadings are Varimax rotated.

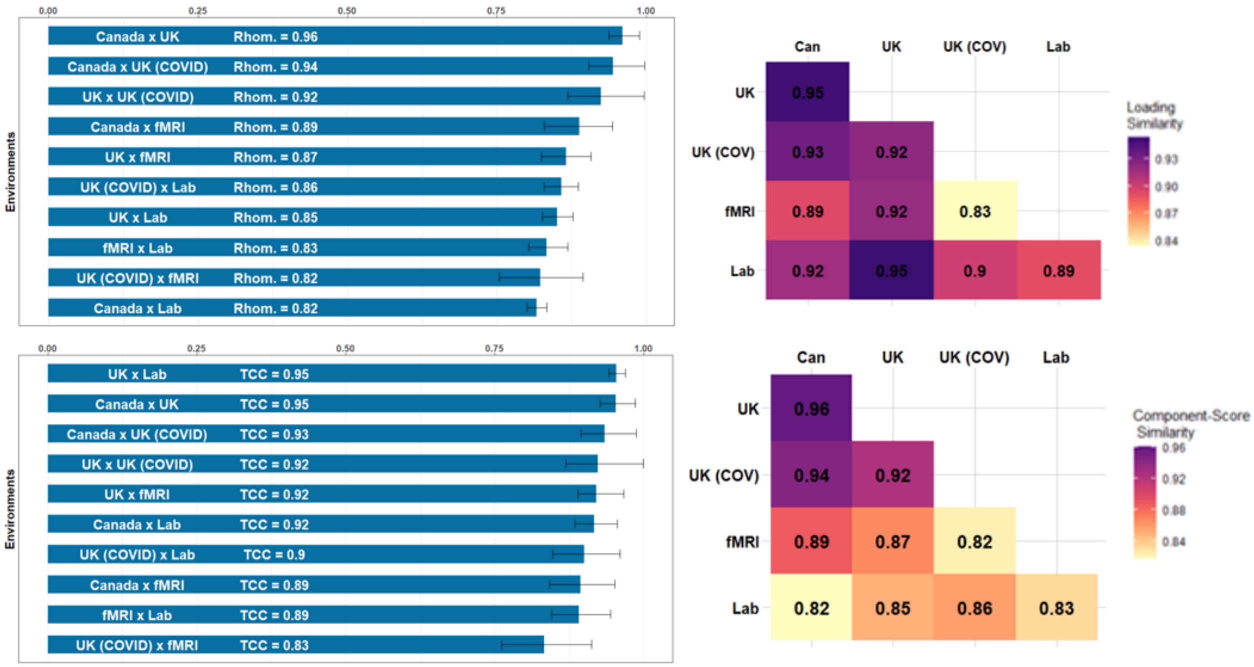
The specific loadings for each component are represented as word clouds in Figure 2 in which the size of the item describes its importance (bigger = more important) and the colour indicates the polarity (items with a similar colour behave in a similar way). The specific loadings from which these word clouds are derived are presented in Table 3. Based on their loadings, we termed the 4 components: 1) "*Detailed Task-Focus*", 2) "*Episodic Social-Cognition*", 3) "*Future Problem-solving*", and 4) "*Positive Engagement*". There are precedents for all the components observed, even among literature outside of the datasets included in this analysis (Karapanagiotidis et al., 2020; Ruby et al., 2013; Turnbull et al., 2019).

4.2 – Do Different Contexts Produce Similar Thought Patterns Using mDES?

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 $\geq$ .85 cut-off for fair similarity according to Tucker's Congruence Coefficient (See Fig. 3; Lavek et al., 2020). The only pair to not 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 = .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).

Figure 3.



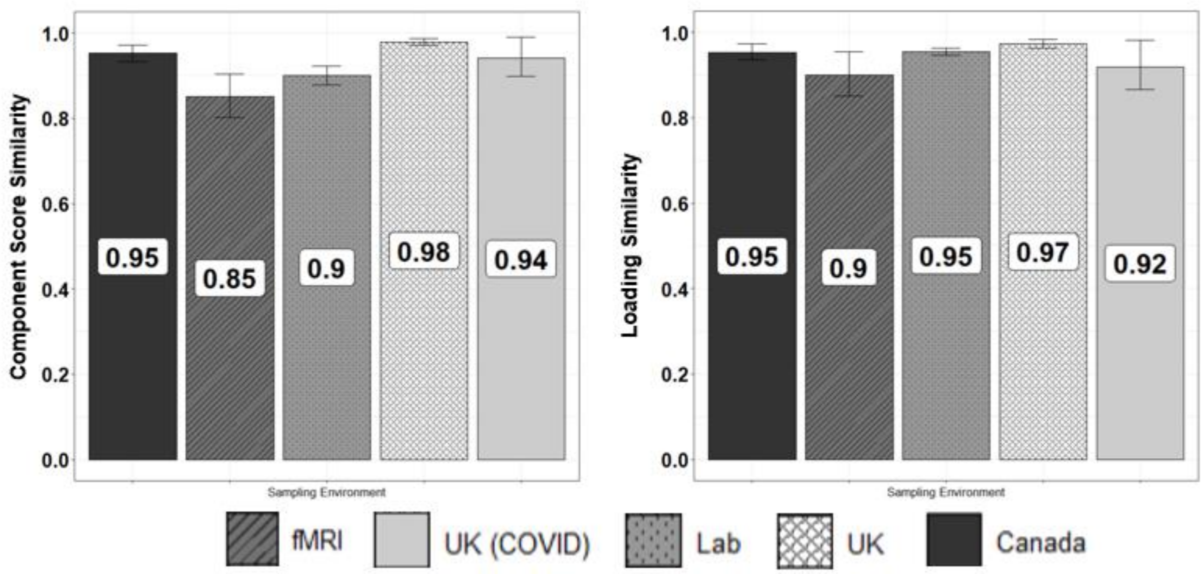
Note. **Results of the sample-to-sample 'thought pattern' reproducibility analysis.** (Top left) Loading similarity with 95% CIs for each sample-sample combination. (Top right) Heatmap for sample-sample loading similarity. (Bottom left) Component-score similarity with 95% CIs for each sample-sample combination.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

Next, we compared the datasets based on the similarity of the components they produced. All pairings met our RHom. >|.80| cut-off 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 >|.90|, which is conventionally used in split-half reliability within the same dataset, (See Fig. 3; Everett, 1983). 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).

4.3 – Can a Shared Space Comprise Thought Patterns Representative of Aggregated Environments?

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.

Figure 4.



Note. **Results of the omnibus-to-sample 'thought pattern' reproducibility analysis**. Bar graphs represent M±95% CIs for omnibus to each sample for: (right) loading similarity and (left) component score similarity.

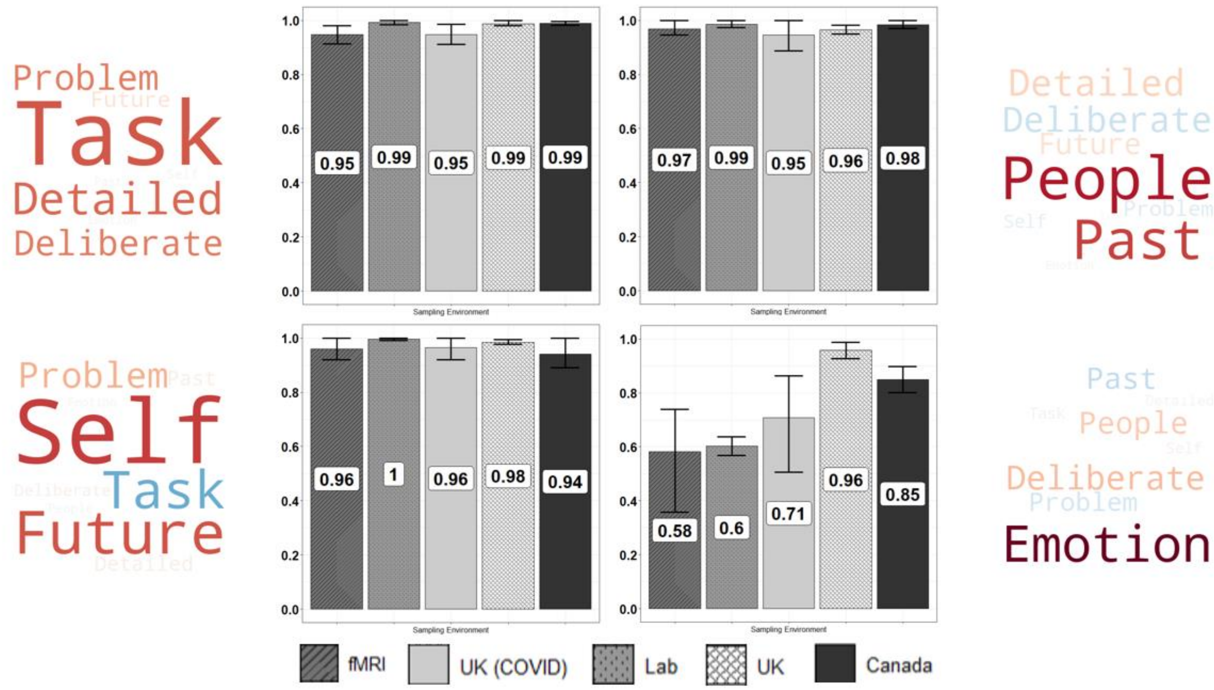
According to both component score and loading similarity, the shared space successfully represented each of the constituent contexts, surpassing both thresholds for reasonable similarity (TCC $\geq$ .85, RHom. $\geq$ |.80|) on all 5 contexts. It further met more conservative thresholds for excellent similarity (TCC $\geq$ .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).

4.4 – 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.

Figure 5



Note. **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.

4.5 – Does the Shared-Space Map Contexts in Reliable Ways?

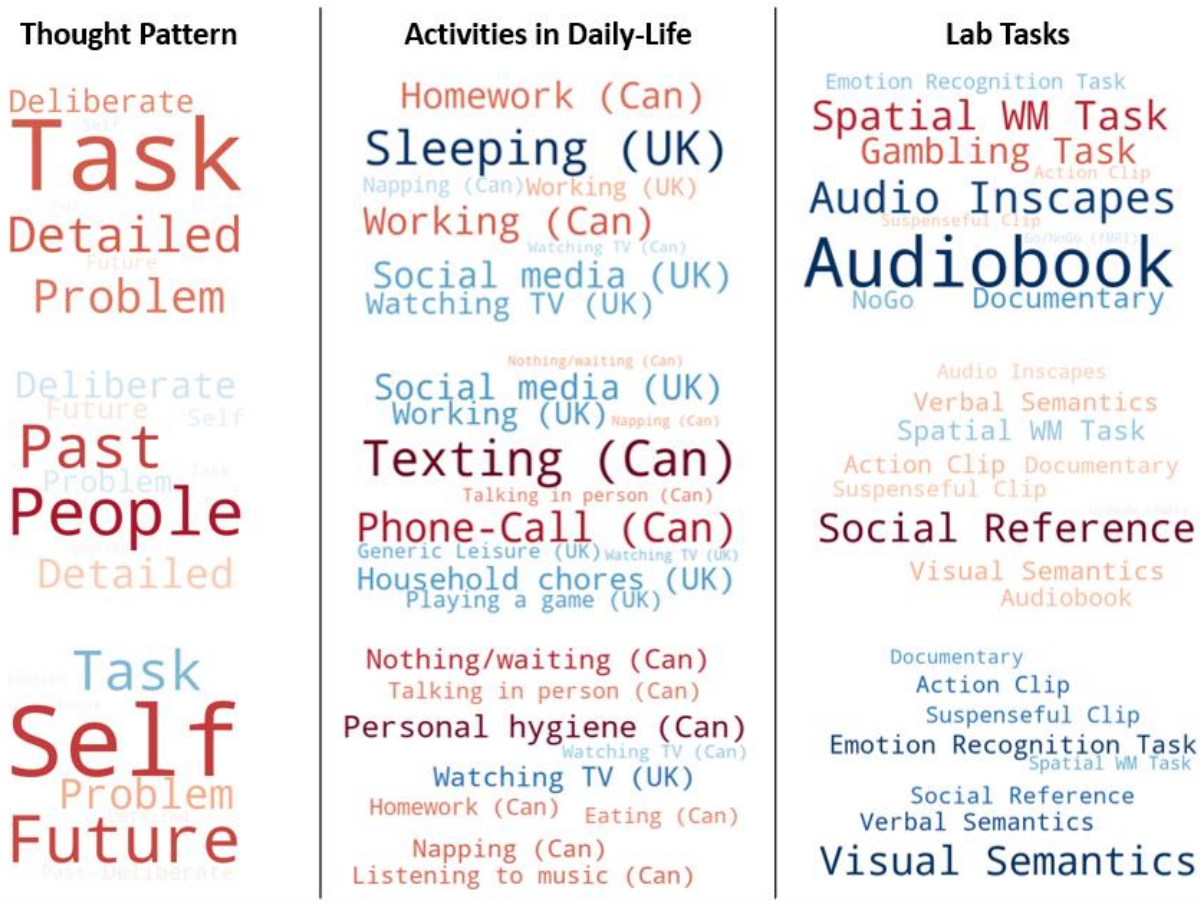
So far, our analysis has established that mDES can produce a similar set of dimensions when it is applied across different datasets and highlighted that 3 of the 4 components yield highly reproducible dimensions across 5 datasets. Our next analysis explores how these common dimensions organise tasks 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 dataset). See Appendix A for the activities selected from in each 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 permutation testing (using Holm-Bonferroni to adjust for inflation of type-I error) to determine if tasks/activities on average scored significantly higher on a given component than might occur by chance. See Figure 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 Figure 6 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 (Hasson et al., 2008), social-cognition reference tasks in the lab and virtual/in-person conversation in daily life (de Caso et al., 2017). See Table 4 for a summary of the top five strongest significant associations and dissociations with each thought pattern.

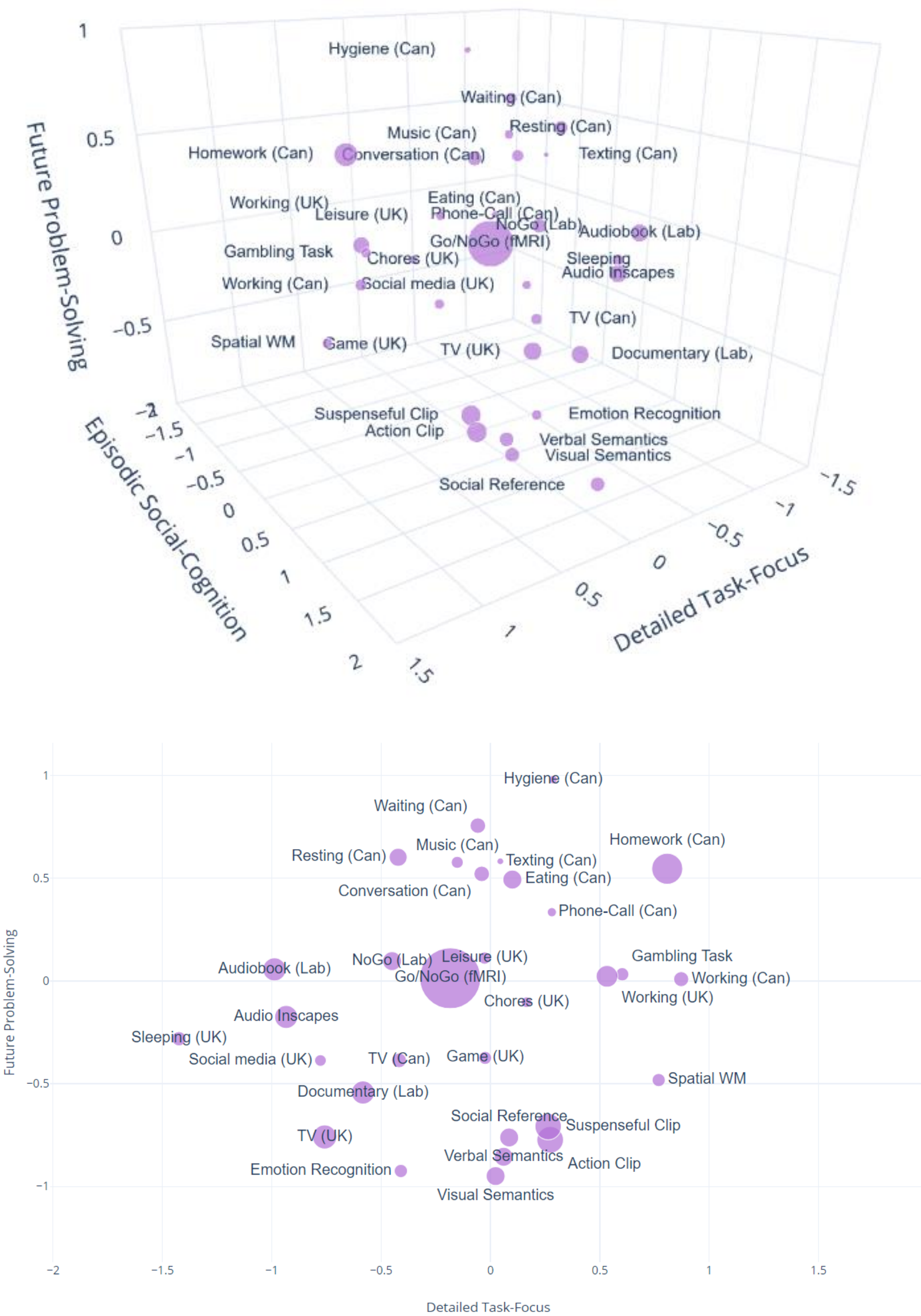
We conducted an additional exploratory analysis examining context mappings on the level of social environment (e.g., interacting with someone in-person/virtually, being alone, etc.), which was recorded across all 5 contexts. Due to its more exploratory nature, we detail this analysis and its results 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.

Figure 6.



Note. **Task/Activity-Context associations with patterns of thought in the shared space.** The word clouds in the leftmost column represent the first three common components generated from the aggregated data, where word size indicates loading strength (larger = higher absolute loading) and colour indicates loading directionality (warmer = positive, cooler = negative). In the task-clouds on the middle and 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).

Figure 7.



Note. **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.

The top 5 highest/lowest scoring tasks/activities on each of the three generalizable components.

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***

Note. Five strongest significant associations and dissociations and their average component scores for each thought-pattern.

* $p < .01$, ** $p < .0008$ (Corrected for 61 comparisons), *** $p < .00027$ (Corrected for 183 comparisons).

Chapter 5: Discussion

5.1 – Summary of Results

Psychological scientists often propose mechanistic theories that typically 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 cognitive states, can produce sufficiently consistent dimensions across sampling contexts to produce a 'shared 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 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 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 Figure 5).

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. 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.

5.2 – 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 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 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.

5.3 – 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, largely female (76%) undergraduate student sample, and it is not clear how these relate to patterns of thought identified using mDES in other populations (e.g., older, male, non-WEIRD, etc.) (Turnbull et al., 2021). Given the cost-effective nature of mDES it would be relatively easy to build a more inclusive shared space in the future and certain types of conclusions (such as what thought patterns are the most representative in daily life) should wait until such data is available. Second, 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. Additionally, our generation of a shared space relies on dimension reduction techniques, showing a high degree of similarity between Principal Components analysis 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. 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 (Turnbull et al., 2019; Karapanagiotidis et al., 2020; Konu et al., 2020) and can discriminate traits within normal population (Turnbull et al., 2020; Vatansever et al., 2018). 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.

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.

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.

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.

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

Csikszentmihalyi, M., Larson, R., & Prescott, S. (1977). The ecology of adolescent activity and experience. *Journal of Youth and Adolescence*, *6*(3), 281-294.

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.

Engert, V., Smallwood, J., & Singer, T. (2014). Mind your thoughts: Associations between self-generated thoughts and stress-induced and baseline levels of cortisol and alpha -amylase. *Biological psychology*, *103*, 283-291.

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.

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.

Gonzalez-Castillo, J., Kam, J. W., Hoy, C. W., & Bandettini, P. A. (2021). How to interpret resting-state fMRI: ask your participants. *Journal of Neuroscience*, *41*(6), 1130-1141.

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*, 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.

Iwata, T., Yanagisawa, T., Ikegaya, Y., Smallwood, J., Fukuma, R., Oshino, S., Tani, N., Khoo, H. M., & Kishima, H. (2024). Hippocampal sharp-wave ripples correlate with periods of naturally occurring self-generated thoughts in humans. *Nature Communications*, *15*(1).

Jonas, E., & Kording, K. P. (2017). Could a neuroscientist understand a microprocessor? *PLoS computational biology*, *13*(1), e1005268.

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 acrosslaboratory 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*, 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*, 103139.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. *Neuron*, *93*(3), 480-490.

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.

Linz, R., Pauly, R., Smallwood, J., & Engert, V. (2019). Mind-wandering content differentially translates from lab to daily life and relates to subjective stress experience. *Psychological Research*, *85*(2), 649-659.

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.

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.

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), 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*, 103530.

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

Pooja, R., Ghosh, P., & Sreekumar, V. (2023). Towards an ecologically valid naturalistic cognitive neuroscience of episodic memory and event cognition: A literature review and synthesis. *PsyArXiv*.

Prescott, S., Csikszentmihalyi, M., & Graef, R. (1976). *Environmental effects on cognitive and affective states: The experiential sampling approach*. University of Chicago.

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), e77554.

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).

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.

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* (PRS-984). (Personnel Research Section Reports, Issue. Armed Services Technical Information Agency.

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., 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.

Vatansever, D., Bozhilova, N. S., Asherson, P., & Smallwood, J. (2019). The devil is in the detail: exploring the intrinsic neural mechanisms that link attention deficit/hyperactivity disorder symptomatology to ongoing cognition. *Psychological medicine*, *49*(7), 1185 -1194.

Vatansever, D., Bozhilova, N. S., Asherson, P., & Smallwood, J. (2019). The devil is in the detail: exploring the intrinsic neural mechanisms that link attention-deficit/hyperactivity disorder symptomatology to ongoing cognition. *Psychological medicine*, *49*(7), 1185-1194.

Vatansever, D., Karapanagiotidis, T., Margulies, D. S., Jefferies, E., & Smallwood, J. (2020). Distinct patterns of thought mediate the link between brain functional connectomes and well-being. *Network neuroscience*, *4*(3), 637-657.

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.

Appendix A: Supplementary Materials

Supplementary Table 1.

Original mDES probes from each dataset.

Item	fMRI/Lab	UK/UK (COVID)	Canada	Scale low	Scale high
Task	**fMRI:** *My thoughts were focused on the task I was performing:* **Lab:** *My thoughts were focused on the task:*	*My thoughts were related to my current activity and/or external environment:*	*My thoughts were focused on an external task or activity:*	Not at all	Completely
Future	*My thoughts involved future events:*	*My thoughts were about the future:*	*My thoughts involved future events:*	Not at all	Completely
Past	*My thoughts involved past events:*	*My thoughts were about the past:*	*My thoughts involved past events:*	Not at all	Completely
Self	*My thoughts involved myself:*	*My thoughts involved myself:*	*My thoughts involved myself:*	Not at all	Completely
Person	*My thoughts involved other people:*	**Close-others:** *My thoughts involved other people close to me:* **Distant-others:** *My thoughts involved other people NOT close to me:*	*My thoughts involved other people:*	Not at all	Completely
Emotion	*The emotion of my thoughts was:*	*My thoughts were…*	*The emotion of my thoughts was:*	Negative	Positive
Detailed	*My thoughts were detailed and specific:*	*My thoughts were detailed and specific:*	*My thoughts were detailed and specific:*	Not at all	Completely
Deliberate	*My thoughts were:*	*My thoughts were…*	*My thoughts were:*	Spontaneous	Deliberate

Item	fMRI/Lab	UK/UK (COVID)	Canada	Scale low	Scale high
Problem	*I was thinking about solutions to problems (or goals):*	**Pre-lockdown:** *To what extent are your thoughts focused on solving a problem?* **Lockdown:** *I was thinking about solutions to problems (or goals):*	*I was thinking about solutions to problems (or goals):*	Not at all	Completely

Note. Experience-sampling questions for each dataset. For the 'Person' item in the UK, the higher score was taken between the distant- and close-others items since in either case the subject could be said to be thinking about another person. This was also strongly correlated with the average of the two items, $r(4973) = .92$, $p < .001$. Additionally, the scale ends for the 'Deliberate' item in the UK datasets used 'Completely Spontaneous' and 'Completely Deliberate' and for the 'Emotion' item used 'Very Negative' and 'Very Positive'.

Supplementary Table 2.

Daily life activities and lab tasks engaged in at the time of probing.

Sampling Environment	Question	Task(s)/Activities
fMRI	NA	Go/No-Go Task
Lab	NA	Go/No-Go Task Self-Reference Task Social-Reference Task Spatial Working Memory Task Cambridge Gambling Task Set Shift Task Suspenseful Movie Clip Action Movie Clip Visual Semantics Task Verbal Semantics Task Emotion Recognition Task Documentary Clip Audiobook Excerpt Audio Inscapes
UK (COVID)	"*What were you DOING just before taking this survey?* *If you were doing more than one thing, please select your PRIMARY activity…*"	Caring for an adult(s) Childcare Cooking Eating and/or drinking Exercising Getting ready for bed Getting ready for the day Household chores Arts and crafts Gardening Listening to music Listening to radio/podcast Leisure: other Playing a game Reading for pleasure Watching TV/film Reading/listening to/watching the news Shopping Sleeping

Sampling Environment	Question	Task/Activity
		Social media
		Talking/conversation (in person)
		Talking/conversation (virtually)
		Working
		Other
Canada	"*What were you doing?*"	Eating
		Homework
		Household chores
		Listening to music
		Napping or resting
		Nothing or waiting
		Personal exercise
		Personal hygiene care
		Physical leisure or sports
		Reading
		Shopping
		Talking in person
		Talking on the phone
		Texting by phone
		Travelling or commuting
		Using a computer or an electronic device
		Walking the dog
		Working (paid or volunteer)
		Other activity

Note. Activities/Tasks engaged with at the time of probing. During daily life studies, this was determined via a probe where subjects selected their primary activity from a checklist (the options of which are provided in the rightmost column).

Appendix B: Supplementary Analyses

4-Component Solution Extra Figures

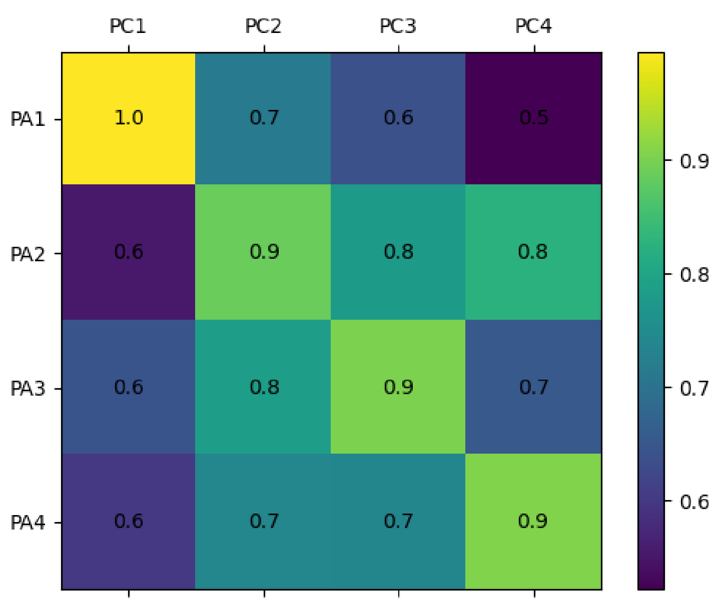
Supplementary Table 3.

Explained Variance of the Omnibus Components for Each Sampling Environment.

Dataset	Explained Variance (%)				
	Detailed Task-Focus	*Episodic Social-Cognition*	*Future Problem-Solving*	*Positive Engagement*	**Total**
Canada	21.75%	13.78%	19.52%	8.84%	63.90%
UK	17.10%	13.84%	16.10%	19.30%	66.34%
UK - COVID	23.12%	14.65%	16.51%	8.42%	62.71%
Lab	16.10%	16.22%	18.22%	11.56%	62.10%
fMRI	19.84%	16.55%	16.60%	7.40%	60.40%

Supplementary Figure 1.

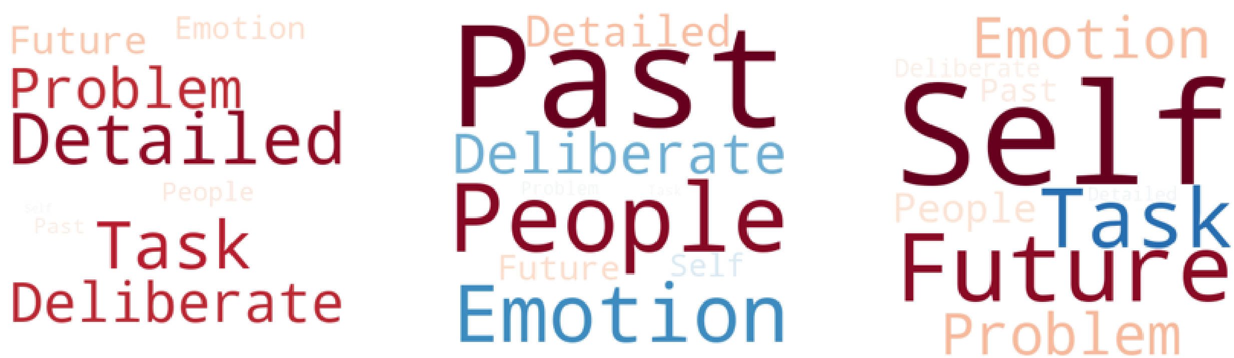
Congruence Between the 4 Component Solution in PCA and the 4 Factor Solution in EFA



3-Component Solution Results

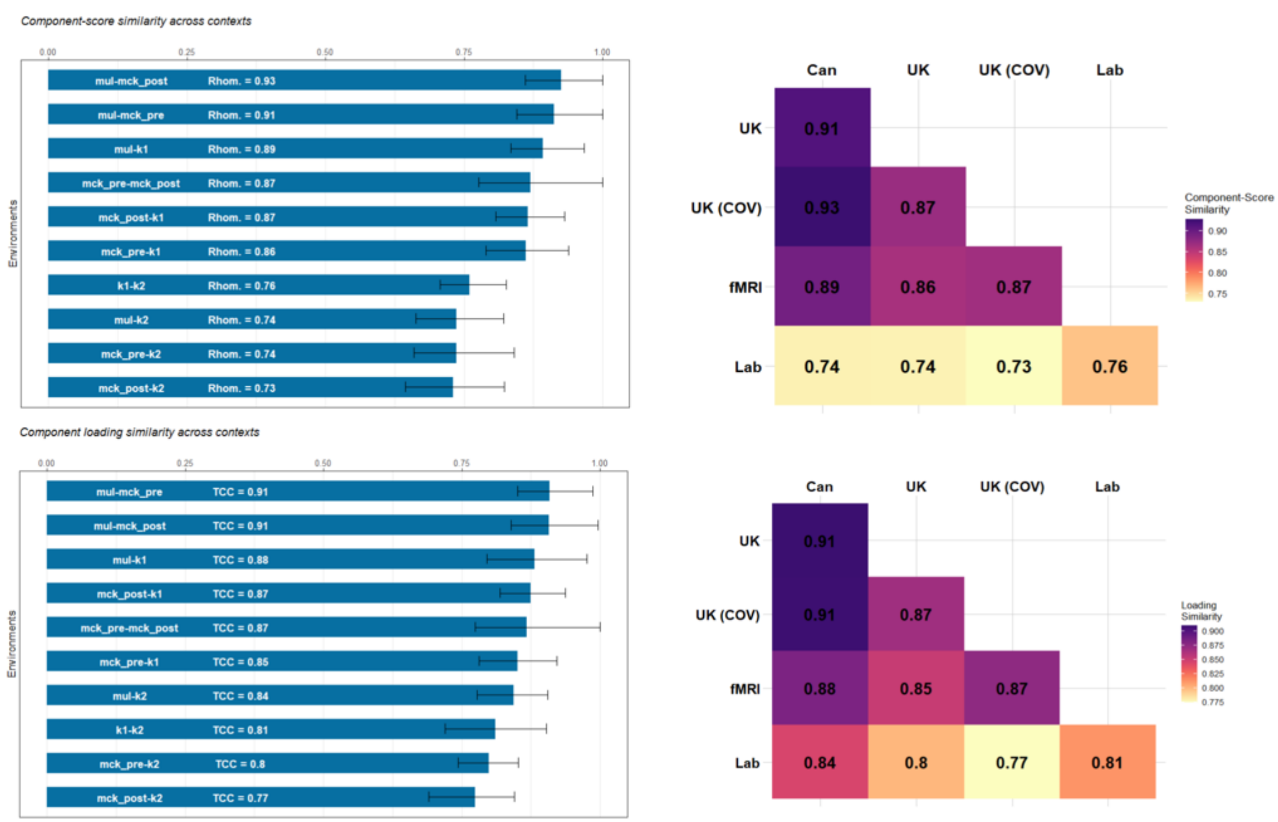
Supplementary Figure 2.

Wordclouds for the 3-Component Solution



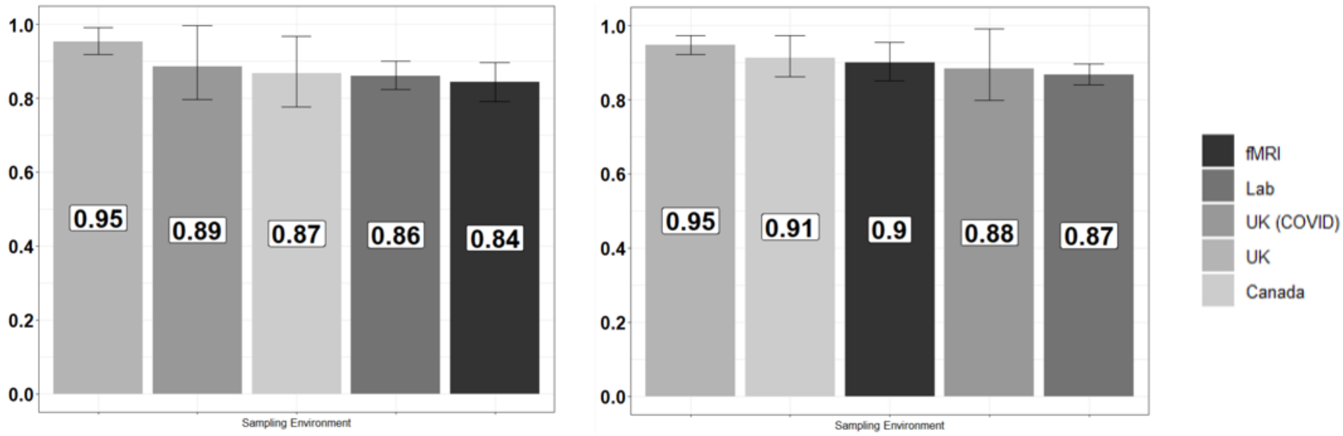
Supplementary Figure 3.

Sample-to-Sample Reproducibility for the 3-Component Solution



Supplementary Figure 4.

Omnibus-to-Sample Reproducibility for the 3-Component Solution



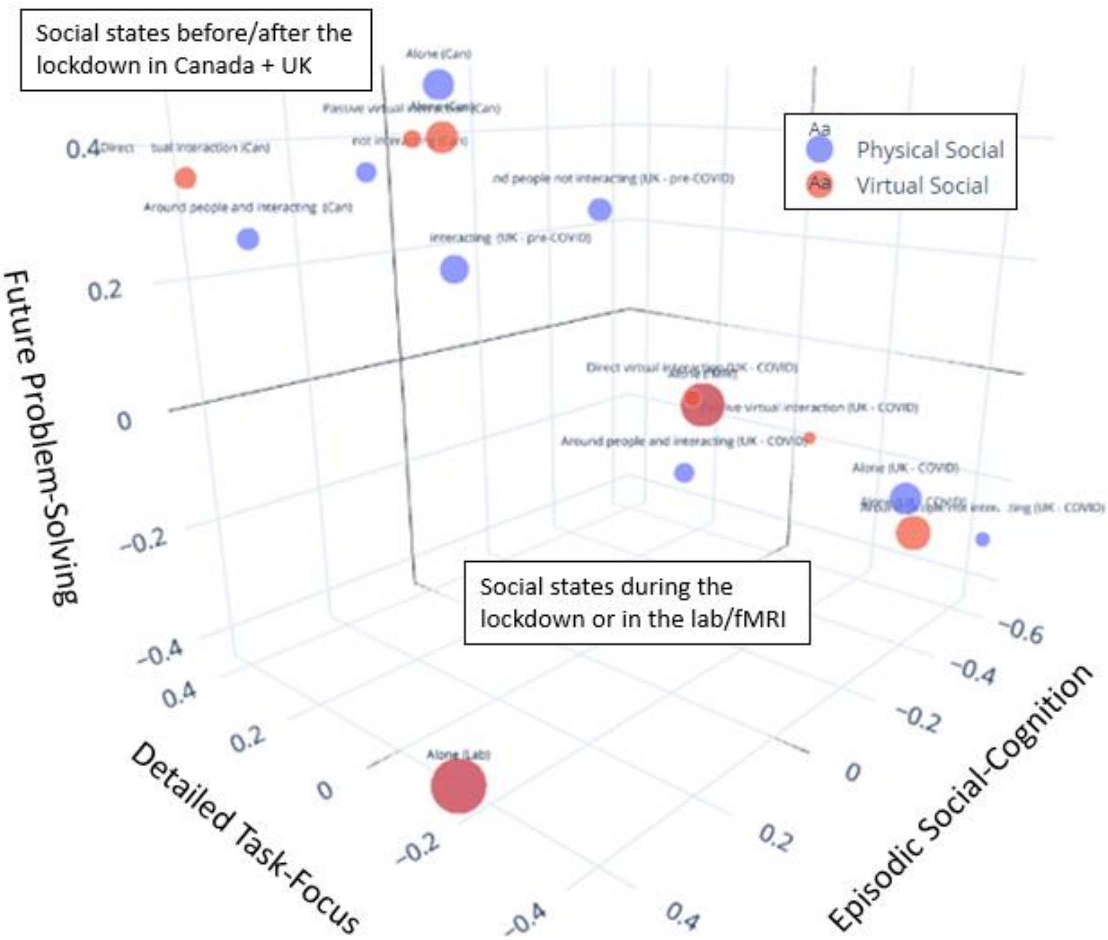
Supplementary Figure 5.

By-Component Omnibus-to-Sample Reproducibility for 3 Components

Social-Environment Analysis

Supplementary Figure 6.

Social states in the shared-space.



Note. **The shared space separates typical daily life social states from those sampled during the lockdown or in the lab.** To assess the social environments characterized by each stable thought pattern, we computed average component-scores on each stable component. We then used permutation testing with Bonferroni-Holm correction to establish whether average scores were significantly greater than occurred by chance. Interestingly, both *Detailed Task-Focus* and *Future Problem-Solving* significantly dissociated states sampled in typical daily life (before and after the COVID-19 lockdown) from states sampled either in daily life during the lockdown and those measured in the lab.

Using the 4-Component solution, we also examined how the shared space distinguished social environments. Both the *Detailed Task-Focus* and *Future Problem-Solving* patterns tended to positively characterise mental-states sampled during typical daily life pre-COVID in the UK and post-lockdown in Canada and inversely characterise states sampled from the lab/fMRI together with those sampled during daily life in the UK *during* the lockdown (See Fig. S6). This distribution appears to reflect how the lifestyle consequences of the COVID-19 lockdown (e.g., reduced physical activity & social interaction, limited self-directed behaviour, spending most time in one place, etc.) caused individuals' daily-experiences to in many ways align with experiential features characteristic of the lab setting (Ho et al., 2020; Rubio-Tomas et al., 2022).