Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature: DocuSigned by: Marta Migó 81BF97DCB3644D9...

Marta Migó
Name

4/5/2023 | 10:07 AM EDT
Date

Title A Dynamic Model of Spontaneous Stuck Thought

Author Marta Migó

Degree Master of Arts

Program Psychology

Approved by the Committee

DocuSigned by: Michael Treadway E5E6836C7DCE4A1...	Michael Treadway *Advisor*
DocuSigned by: 4A8807E27D2542D...	Philip Kragel *Committee Member*
DocuSigned by: Alexandra Cohen 005A8C6CC7DC4CC...	Alexandra Cohen *Committee Member*
DocuSigned by: 79E8E140FC9841E...	Jessica Cooper *Committee Member*

Accepted by the Laney Graduate School

DocuSigned by: Kimberly Jacob Arriola 699065B1043A4FB... *5/2/2023 | 1:01 PM EDT*	Kimberly Jacob Arriola, PhD, MPH *Dean, James T. Laney School of Graduate Studies*

A Dynamic Model of Spontaneous Stuck Thought

By

Marta Migó

B.A./B.F.A., New York University

Advisor: Michael Treadway, PhD

An abstract of a thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Master of Arts in Psychology

2023

Abstract

A Dynamic Model of Spontaneous Stuck Thought

By Marta Migó

While worry and rumination are highly prevalent features of internalizing disorders, their cognitive mechanisms remain unclear. Moreover, uncovering these mechanisms has proved challenging experimentally, as worried and ruminative thoughts often occur in the absence of measurable behaviors. To date, free-association paradigms and semantic word associations have shown some promise for uncovering mechanisms of worry and rumination, though this work remains in early stages. Here, we analyze word-association data using a dynamic attractor-state modelling that conceptualizes repetitive negative thinking as a phenomenon of spontaneously navigating a multidimensional semantic space while in the presence of a strong maladaptive attractor space. The previously validated Free Association of Semantics Task (FAST) was used to collect word-associations from two samples of participants: 79 online and 65 in-person. This task required users to submit single word responses to prompts based on the psychoanalytic procedure of "free association". Submitted words were first embedded using a pre-trained GloVe model, and the resulting multi-dimensional semantic space was reduced to 7 dimensions, explaining 90% of the data variance. Using dynamic attractor-state modelling, data simulations were conducted and compared to the collected data using cosine similarity. An optimizer helped find the best fitting parameters, which were later clustered using unsupervised k-means. Consistently, one of three resulting clusters yielded higher levels of perceived pathology, lower levels of enjoyment, and revealed a pattern of thought inflexibility and repetition, which resembled worried or ruminative thinking. Indeed, individuals whose data mostly clustered into that cluster also self-reported higher measures of rumination and worry. Our results support a conceptual model of repetitive negative thinking derived from attractor-state dynamics, thereby providing promise to a novel method for measuring rumination and worry severity.

A Dynamic Model of Spontaneous Stuck Thought

By

Marta Migó

B.A./B.F.A., New York University

Advisor: Michael Treadway, PhD

A thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Master of Arts in Psychology

2023

Table of Contents

1. Introduction

Mind-wandering thoughts are task-unrelated, stimulus-independent thoughts that spontaneously pop into our minds as we go about our daily life. These mind-wandering trains of thought freely flow between ideas in any and all topics and valences, without staying on any one concept for too long (Christoff et al., 2016). When these thoughts become frequent, repetitive, and constrained to the negative valence, individuals are no longer said to be mind-wandering, and instead are said to have begun ruminating or worrying (Nolen-Hoeksema et al., 2008). Although most everybody ruminates and worries from time to time, individuals who excessively engage in these repetitive negative thinking (RNT) patterns are at a higher risk for the development and maintenance of internalizing psychopathologies, such as depression, anxiety, and post-traumatic stress disorder (Hoyer et al., 2009; Nolen-Hoeksema, 2000; Spinhoven et al., 2015). Crucially, there is a surge of interest in moving away from the categorical study of mental illnesses (Compton et al., 2007; Kessler et al., 1996) and towards the better understanding of transdiagnostic symptoms and dimensional risk factors (Cuthbert, 2014; Ringwald et al., 2021). Given the vulnerability that worry and rumination confers individuals, RNT is a promising transdiagnostic process that may confer risk for internalizing psychopathology.

Scientist and philosophers alike have longed theorized about spontaneous thought. However, formal assessments and measurement tools for RNT were not developed until the late 1980s. Self-report scales like the Ruminative Response Scale (RRS) (Nolen-Hoeksema & Morrow, 1991) and the Penn State Worry Questionnaire (PSWQ) (Meyer et al., 1990) filled a very important gap in the literature when they were first published and, to this day, they remain widely-used measures of rumination and worry. These self-report instruments are often used to gage how frequent and problematic participants believe their spontaneous negative cognitions to be, and to explore how they perceive their abilities to engage or disengage from these repetitive thoughts. However, there is a general agreement in the field of mind-wandering and RNT that these self-report tests suffer from a number of limitations. First, they heavily rely on retrospective report, which exhibit a number of flaws (for a review, see Conner & Barrett, 2012). Specifically, studies comparing ecological momentary assessment data and retrospective reports have shown significant discrepancies in how people remember how difficult quitting smoking was (Shiffman et al., 1997), how stable their mood was throughout a month (Solhan et al., 2009), and how successful they were at coping with distress and regulating emotion (Stone et al., 1998). Retrospective report is also affected by the peak-end effect – the phenomenon in which individuals tend to best remember the most arousing moment of an event, and how the event ended (Kahneman et al., 1993). This phenomenon makes it so that individuals are likely to forget about most of an event, which effects subsequent reporting. Crucially, not only is retrospective report provenly flawed, but spontaneous thoughts like mind-wandering, worry, and rumination lack meta-awareness, which makes retrospective reflection upon those trains of thought increasingly challenging (Smallwood & Schooler, 2015). A second limitation of these self-report measures is their unclear construct validity. It is also doubtful that all items of these scales are equally related to constructs rumination and worry. For example, the RRS has been shown to tap into several additional constructs, like reflection and depression (Treynor, 2003). Finally, these questionnaires are unable to generate hypotheses about possible mechanisms underlying thought dynamics that give rise to RNT patterns. Taken together, it does not seem feasible to fully understand the severity of an individual's worry and rumination by solely relying on retrospective self-report.

In seeking to overcome these limitations, several scientists in the last 15 years have turned towards the use of behavioral tasks and computational algorithms to expand on what these limited self-report measures can tell us. For instance, a mechanistic study successfully predicted when individuals would become distracted during a task, as task performance would decrease, and what they would think about. This model was based on the believe that individuals' neural resources are in constant competition between attending to the task at hand and mind-wondering/ruminating (Taatgen et al., 2021). Because mind-wandering and rumination are memory-retrieval processes, easier-to-activate memory chunks (for instance, highly arousing, negative, ruminative memories) were more likely to be spontaneously recalled, leading to decreased task performance (M. K. van Vugt et al., 2018). Another study recently used a free recall task and showed participants several lists of words to investigate what words, and in what order, these individuals would freely remember (Gupta et al., 2022). Using the Adaptive Control of Thought-

Rational model, they showed that people with high self-reported perseverative-thinking scores were less likely to be affected by primacy and recency effects, generally less accurate in their recalls, but more likely to freely recall negative words, and to continue to recall negative words following the first recall. Relatedly, the Free Association of Semantics Task (FAST) – a free word association paradigm—, was recently developed and modelled using Markov chains to show that individuals with high RRS scores are more likely to spontaneously transition into negative thoughts, and more likely to stay within the negative valence than individuals with low RRS scores (Andrews-Hanna et al., 2022). Together, these tasks and computational models have been developed to better describe and understand the nature of RNT. However, no study to date has attempted to develop a model to measure the severity of a worry or a rumination, with the ultimate goal of serving as a diagnostic tool. The present paper aims to use spontaneous word-association data and attractor dynamic modelling to estimate the severity of an individual's RNT.

We conceptualized mind-wandering as the free navigation of concepts within a semantic pool or the spontaneous and unconstrained association of ideas within a multidimensional semantic space. By contrast, we thought of worry and rumination as the attempt to navigate this same semantic space while in the presence of a maladaptive attractor space, specifically located in a negatively-valenced semantic space. This strong attractor space would lead individuals who to repeatedly return to negative topic following a variety of prompts. An "attractor space" is a concept from dynamical systems theory that represents a stable state that a dynamical system will tend to gravitate toward (O'Reilly et al., 2012). Dynamic systems modelling and attractor space dynamics, which were originally developed in the field of physics, have previously been applied in psychology to study excitatory and inhibitory neural feedback loops. This work has successfully modelled neural activity and behavior related to top-down imagery, top-down ambiguity resolution, and pattern cognition (O'Reilly et al., 2012). These models have also recently been used, for the first time, in the context of spontaneous thought dynamics, which has proven an important step forward (Amir & Bernstein, 2023). This novel study proposed the Attention-to-Thought (A2T) model – a dynamic systems model of internally-directed cognition that computationally defines momentary states and predicts differential temporal trajectories of internal attention, working memory, and emotion using experimental simulations. Importantly, in the same issue of *Psychological Inquiry* (February 2023), this first approach was highly praised for its novel conceptualization and promising computational approach, but also heavily criticized. The A2T model appears incapable of capturing detailed behavioral patterns quantitatively, and counts on excessive levels of flexibility and redundancy, which likely inflate its general explanatory power (M. van Vugt & Jamalabadi, 2023). Hence, while dynamic modelling shows real potential as a means to elucidate the mechanisms and patterns of internally-directed cognition, this tool has not yet been extensively explored and can benefit from further research.

Using dynamic and attractor state modelling and word-association data, the present paper aims to model thought trajectories. Our approach enables us to find the semantic locations that individuals may revisit during trains of thought, as well as determine the presence of problematic attractor spaces. Our analyses show how our output measures cluster into ruminative/worried patterns and are predictive of perseverative-thinking self-report measures, and we argue for the increased validity and reliability of our behavioral and computational measures as diagnostic tools (Zorowitz & Niv, 2023).

2. Materials & Methods

2.1. Participants

An *a priori* power analysis was conducted using G*Power version 3.1.9.7 (Faul et al., 2007) to determine the minimum sample size required to test the study hypotheses. Results indicated the required sample size to achieve 90% power for detecting a medium effect, at a significance criterion of $\propto = 0.05$, was N = 59 for a 4-level multiple linear regression model. Accordingly, two samples were collected. First, we collected data from a sample of freshmen and sophomore undergraduate students from Emory University (N = 72). After data cleaning, 65 participants from this sample were included in the final analysis (43 female). These participants were recruited through the Sona System, an online tool to manage university participant pools, and received a course credit in exchange for their time (about 1 hour). A second community sample was collected online using the online participant recruitment platform

Prolific (N = 100). After data cleaning, 79 participants (M_{age} = 37.1 years, SD_{age} = 12.5; 35 female) from this sample were included in the analysis. Prolific participants received $12/hour as compensation for their time. All data collection, storage, and usage took place in accordance with and under the approval of Emory University's Institutional Review Board. For a breakdown of our participants' demographics, see *Table 1*. The samples did not differ in average RRS, PSWQ, PSS, DARS, PHQ, or GAD scores. However, the in-person sample did report significantly or trend-level higher measures of positive mood (In-Person: M = 32.9, SD = 1.12; Online: M = 28.4, SD = 1.23, $t(142)$ = -2.65, p = 0.0098), negative mood (In-Person: M = 23.2, SD = 1.29; Online: M = 20.0, SD = 1.29, $t(142)$ = -1.68, p = 0.097), and dysfunctional attitudes (In-Person: M = 21.7, SD = 0.90; Online: M = 18.2, SD = 0.79, $t(142)$ = -2.95, p = 0.0043) than the online group.

2.2. Task

Participants began by completing an adapted written version of the FAST (Andrews-Hanna et al., 2022). This paradigm consisted of displaying an initial seed word on the screen, and then allowing participants to type in a related word. Upon submission of the first word, participants continued submitting related words until they reached 10 word-submissions for each seed word. Notably, each word had to be related to the immediately prior word, not the seed word. For instance, when presented with the seed word "spouse", a participant initially thought of "together", then "apart", and then "alone". While "alone" is more immediately related to "apart" it is no longer as immediately related to "spouse". If a participant failed to submit a new word within the time allotted (4.5 s), the previous word automatically re-submitted. Participants were instructed not to submit proper nouns, names of places or brands, numbers, or acronyms.

Importantly, participants were given only 4.5 s to type a word. This time interval was chosen after an online pilot dataset (N = 42), collected on Prolific, revealed that participants took an average of 3.01 seconds (SD = 1.24) to come up with and type in a word. A standard deviation was added to that average, and rounded up, to account for slower participants. However, crucially, more time was not granted to prevent individuals from self-editing their answers and encouraging them to submit their first thought as an answer.

Subjects completed two practice sequences before moving on to the actual task. A total of 21 seed words were presented, and 210 words were manually (by the participant) or automatically (by the program after 4.5 s) submitted. According to the Affective Norms for English Words (ANEW) database (Bradley & Lang, 2010), a third of the seed words were negative in valence, another third were neutral, and the final third were positive. Seed words were presented in a random order.

After completing the FAST, participants were presented with all the unique words that they submitted (no duplicates were shown), and asked to provide two ratings on each submission. For each word, participants were asked "how much do you enjoy thinking about [word]?" (Enjoyment rating) and "do you believe you think about [word] more than you should in your daily life?" (Perceived Pathology rating). Participants' answers on Enjoyment and Perceived Pathology ranged from "Not at all" (0) to "Very much" (100).

Once the task portion of the study was completed, participants filled out eight self-report measures: the RRS (containing three subscales: Brooding, Reflection, and Depression), the PSWQ, the Perceived Stress Scale (PSS), the Dysfunctional Attitudes Scale (DAS), the Generalized Anxiety Disorder-7 (GAD-7), the Patient Health Questionnaire-9 (PHQ-9), the Positive and Negative Affect Schedule (PANAS), and the Dimensional Anhedonia Rating Scale (DARS). These tools were chosen in order the test the validity, specificity, and sensitivity of our model outputs. Specifically, while our model outputs should be most predictive of the RRS and the PSWQ, they should also be moderately predictive of related constructs, measured by tools such as the PSS, the DAS, the GAD-7, the PHQ-9, and the PANAS. In contrast, our model outcomes should be least predictive of the DARS, a measure of anhedonia, as these phenotypes tend to behaviorally cluster together but have not been linked at the cognitive level (Rutherford et al., 2023). The study took a total of 50-60 minutes to complete and was coded in JsPsych and ran locally (in person sample) or remotely, on Pavlovia (online sample).

2.3. Data Preprocessing

Data was manually checked for spelling mistakes and other errors. Submissions that contained proper nouns, names of places and brands, numbers, or acronyms, were treated as "late word submissions" and substituted with the previous word in the sequence. Two-word expressions or phrasal verbs were substituted by one-word synonyms. After data cleaning, participants who had failed to submit more than 1/3 of words within the 4.5s time limit (meaning, submitted less than 138 unique words) were excluded from the analysis. 7 in-person participants and 21 online participants were removed from the sample, as mentioned above, due to not meeting minimal data quantity requirements.

The cleaned data was then embedded using a pre-trained GloVe model (Pennington et al., 2014), trained with 2 billion tweets/27 billion tokens, which yielded 25 dimensional vectors for each submitted word. These vectors were numerical representations of the word location in a multidimensional semantic space. Principal component analysis was then used to reduce the number of dimensions to 7, explaining 90% of the variance in the data, in order to alleviate computational requirements of subsequent analyses.

2.4. Modelling

2.4.1. Dynamic Modelling

Dynamic modelling was used to describe the 21 thought trajectories that each participant engaged in. Our general formula – $f(t) = \hat{k} - (\hat{k} - k_0) \cdot e^{-a \cdot t}$— allowed us to estimate, given a starting point (seed word), k_0, the location towards which individuals were converging, $\hat{k}$, and the rate of approach towards that location, a. Importantly, the interpretation of a is dependent on the total distance traveled during the word-chain. The further away from an attractor space that a ruminator/worrier begins the word-chain at, the quicker they will travel towards that strong attractor space, yielding high values of a. However, if a ruminator/worrier begins and ends a word chain in the same general location, that behavior will yield small values of a, and will also be an indication of "stuckness". Mind-wanderers ought to travel large distances while not being influenced by a strong attractor space: small values of a. Our model additionally generated the locations in the semantic space where individuals found themselves at each of the 10 time-points, or 10 word-submissions. Our general function was turned into a system of differential equations for further editing and testing. In the end, four systems of differential equations were tested using Euler's method. 500 data simulations for each word-chain were conducted, and the generated data was compared to the collected data using cosine similarity. The Nelder-Mead optimization method, a gradient-free method for well-conditioned, high dimensional data, was used to find the parameter values that best explained the collected data, which were the ones that yielded the highest cosine similarity (closest to 1). The four systems of differential equations that we tested were:

1. $$\begin{cases} f'(t) = a \cdot (\hat{k} - k_0) \cdot e^{-a \cdot t} \\ f(0) = k_0 \end{cases}$$

2. $$\begin{cases} f'(t) = a \cdot (\hat{k} - k_0) \cdot e^{-b \cdot t} \\ f(0) = k_0 \end{cases}$$

3. $$\begin{cases} f'(t) = v \cdot \breve{k} + a \cdot (\hat{k} - k_0) \cdot e^{-a \cdot t} \\ f(0) = k_0 \end{cases}$$

4. $$\begin{cases} f'(t) = v \cdot \breve{k} + a \cdot (\hat{k} - k_0) \cdot e^{-b \cdot t} \\ f(0) = k_0 \end{cases}$$

In some of the variations above, we broke down the rate of approach, *a*, into two variables, *a* and *b*. Those differential equations still described a converging trajectory towards a semantic location, but in the following way: $f(x) = k_0 + \frac{a}{b} \cdot (\hat{k} - k_0) - \frac{a}{b} \cdot (\hat{k} - k_0) \cdot e^{-b \cdot t}$. Unfortunately, we soon realized that several combinations of *a* and $\hat{k}$ were yielding the same resulting trajectory towards $k_{final} = \lim_{t \to \infty} f(t) = \left(1 - \frac{a}{b}\right) \cdot k_0 + \frac{a}{b} \cdot \hat{k}$, approaching at rate *b*, which made the model overly difficult to interpret and unusable for the purposes of this study. However, this model was crucial to inform our choice of parameter ranges in subsequent models, as it revealed that individuals were sometimes trending towards very far away location, further than we originally expected. In some of the other variations above, we further included a term aimed at capturing noise: at every time point, we generated a random direction in the 7d space, $\breve{k}$, towards which an individual may have been pulled at a strength of *v*. While this measure of noise did help improve fit for some participants and trials, after further examination, it became clear that it was improving fit in a highly heterogeneous fashion, making model interpretations challenging. Accordingly, we used our simplest model for further testing, although, we acknowledge that including a parameter that more consistently captures noise will be a necessary next step in future work:

$$\begin{cases} f'(t) = a \cdot \left(\hat{k} - k_0\right) \cdot e^{-a \cdot t} \\ f(0) = k_0 \end{cases}$$

Final parameter ranges were chosen rationally after inspecting the collected and initially simulated data. After extracting word embedding from all of our word submissions, and inspecting the coordinates that participants were trending towards in the initial round of data simulations, parameter ranges for $\hat{k}$ were set to -2000 to 2000, encompassing all data-points and most initial k_{final}s. Our values of a needed to be positive and non-zero to ensure a convergence trajectory. Accordingly, we set the ranges of this parameter to (0, 40], after taking into account our device's computational capacity and the fact that we could only perform 500 simulations per trial, given our tight timeline. A wider range of parameters would have been too sparsely explored, given the simulation constraints.

2.4.2. Kernel Density Estimation

Kernel density estimation is a non-parametric tool used to estimate the density distribution of a dataset. After we used cross-validation to find the best bandwidth for our subsequent estimations, we used Python's sklearn.neighbors package to first train our model on each individual's dataset, and then find the approximate density at each individual's 21 $\hat{k}$s. This method allowed us to investigate whether people were travelling towards locations that they had previously visited during the task. This was important to capture the repetitive and topic-revisiting nature of rumination and worry.

2.4.3. Unsupervised Clustering

Once best-fitting parameter values were obtained, unsupervised K-Means clustering was conducted. This unsupervised method was used to cluster to data into 1-21 different groups, and find the best fitting number of clusters, 3, using the knee method (Satopa et al., n.d.). Various combinations of variables yielded similar clustering results, but inertias –a measure of how far data points are from their cluster centroid— were used to compare models. The least complex model, the one using exclusively the parameter a to conduct the clustering, yielded the smallest inertia and was used for subsequent analyses.

2.5. Statistical Analyses

Statistical analyses were conducted to explore 1. the relationship between the collected rating scales and self-report measures, and 2. the parameter clusters that most resembled ruminative/worried behavior and predicted self-report measures.

2.5.1. Relationship Between Rating Scales and Self-Report Measures

Based on previous literature, we first looked to replicate findings suggesting that individuals with high self-report rumination, worry, depression, and anxiety scores typically freely associate and freely recall, on average, more negative concepts than individuals with low scores in these measures (Andrews-Hanna et al., 2022; Gupta et al., 2022). Accordingly, we ran simple linear regressions to see how rating-scales of average Perceived Pathology and Enjoyment predicted clinical self-report measures. Next, we estimated the Euclidian distances between each individual's 21 $\hat{k}$'s (or the locations towards which participants were trending during each word-chain) and their submitted words. Using that information, we computed a weighted average, where values of Perceived Pathology and Enjoyment associated with words that appeared closer in space to $\hat{k}$ were weighted more heavily than the rating values of those words that appeared far from $\hat{k}$ in the semantic space. Regression analyses were again conducted to see if individuals with high rumination, worry, depression, and anxiety scores would also trend towards less enjoyable and more problematic semantic spaces. Crucially, these regression analyses controlled for differences in minimum distance between estimated $\hat{k}$'s and submitted words.

2.5.2. Clustering of Model Parameter and Prediction of Self-Report Measures

Next, we used the aforementioned clusters to investigate thinking patterns across individuals. We used one-way ANOVAs and post-hoc Tukey Honestly Significant Difference tests to compare the average measures of a, Enjoyment, Perceived Pathology, and kernel density estimates across clusters.

After inspection, clusters that were believed to most closely resemble ruminative/worried behavior were submitted to further analyses. Specifically, we computed the proportion of trials that each participant spent in the "ruminative/worried" cluster. Those proportions were compared to clinical self-report measures using linear regression.

3. Results

3.1. Rating-Scales

In the online sample, regression analyses revealed that individuals' brooding, depression, anxiety, perceived stress, and negative mood scores negatively predicted average Enjoyment ratings at weak-to-moderate effect sizes (see *Table 2*), meaning that individuals with high perseverative thinking and internalizing disorder scores tended to bring up concepts that they did not enjoy thinking about. Positive mood scores positively and weakly predicted Enjoyment ratings. Rumination, depressed cognitions, and worry scores negatively predicted average Enjoyment ratings at weak-to-moderate effect sizes, while the associations between reflection, dysfunctional attitudes, and anhedonia scores and Enjoyment remained negative, but much weaker. Crucially, while the direction of these relationships did replicate in our in-person sample, effect sizes did not replicate in any case. In our in-person sample, Enjoyment was only moderately negatively correlated with measures of dysfunctional attitudes and anhedonia, and moderately positively correlated with positive mood. For a full break-down, see *Table 2*.

In the online sample, identical regression analyses revealed that individuals' perceived stress and negative mood scores positively predicted average Perceived Pathology ratings (see *Table 3*). In other words, participants with high stress and negative mood scores tended to bring up concepts that they perceived to be, on average, significantly more problematic (meaning, they believed they thought about those concepts more than they should in their day-to-day life) than individuals with lower self-report scores. This positive association between worry and depression scores, and average Perceived Pathology measures was of a medium effect size, while the effect size of all other regressions were small. Importantly, the directions of these relationships did replicate in our in-person sample, but effect sizes again did not, see *Table 3*.

When investigating the Enjoyment and Perceived Pathology scores associated with the coordinates towards which individuals were traveling in their word-chains, regression analyses in only the online sample showed that individuals with high perseverative-thinking and internalizing disorders scores tended to travel and/or get stuck in locations that they perceived to be less enjoyable and more problematic to think about in their day-to-day life than individuals who scored low in these measures. For a full break down of these results, see *Table 4* and *Table 5*.

3.1. Model Clustering

3.1.1. Overview

Across samples, the fit of our dynamic model was estimated using cosine similarity. Cosine similarity values range from -1 (suggesting that our data simulations are completely dissimilar and opposite from our actual data) to 1 (suggesting a perfect replication of the data). In our online sample, the average cosine similarity across all participants and trials was 0.45 (minimum: 0.15, maximum: 0.89; *Figure 1*). In our in-person sample, the average cosine similarity was 0.47 (minimum; 0.18, maximum: 0.87; *Figure 2*). Our models estimated the location towards which participants were trending during their word chains, as well as the rate of approach towards that location, *a*. As explained in the Methods section, the rate of approach *a* was submitted to unsupervised k-means clustering, which consistently yielded 3 best-fitting clusters in both samples, see *Figure 3*.

The absolute average values of semantic distance traveled during a word-chain, and average cosine similarities consistently and accurately replicated within clusters, across samples. Only relative values of kernel density estimates, and average measures of Enjoyment and Perceived Pathology replicated across clusters and samples. In other words, while the absolute values of certain measures did not replicate across samples, we vastly replicated patterns. For example, our online group showed significantly greater kernel estimates than our in-person data in all clusters (possibly due to differences in sample size). However, kernel density estimates were largest in cluster 1 and smallest in clusters 1 and 2, across samples. For a visual comparison, see *Figure 4*, and for a detailed statistical comparison, see *Table 4*. Across samples, the largest proportion of trials made it into cluster 1 (online sample: 48.2%, in-person sample: 50.0%), followed by cluster 2 (online sample: 39.7%, in-person sample: 38.6%), and finally cluster 3 (online sample: 12.1%, in-person sample: 11.4%).

3.1.2. Cluster Descriptions

Cluster 1was, on average, the best-fitting cluster (Online: M = 0.45, SD = 0.09, In-Person: M = 0.49, SD = 0.12). This unsupervised cluster yielded the smallest values of a across samples, and resulted in low average measures of Enjoyment, significantly below-average semantic distances traveled per word-chain, and high kernel density estimates and average measures of Perceived Pathology. Cluster 2 was, on average, the worst-fitting cluster (Online: M = 0.43, SD = 0.09, In-Person: M = 0.43, SD = 0.10). This cluster averaged the largest values of a (while even the largest values of a remained relatively small), and resulted in high average measures of Enjoyment and semantic distance traveled, but low average measures of Perceived Pathology and kernel density estimates. Finally, cluster 3 yielded average values of a, and resulted in small kernel density estimates and semantic distance traveled per word-chain, but large average values of Enjoyment. Average values of Perceived Pathology did not replicate across samples, as the online sample reported low average values of Perceived Pathology in cluster 3, and the in-person sample reported high average values of this same measure. Next, we will discuss our interpretations of these findings.

3.1.3. Cluster Interpretations

After visually inspecting the clusters, we determined that cluster 1 most closely resembled a pattern that reflected rumination and worry. Not only did the trials in this cluster tend to revisit previously explored topics (kernel density estimates were high), but the word-chains in this cluster also appeared to travel short distances in a slow fashion (small rate of approach, a). As mentioned in our Methods section, small values of a can indicate “stuckness” or thought inflexibility when individuals travel short distances in their word-chains. In those cases, individuals ended their word-chains at a similar location as their starting points (seed words) leading them to move slowly around a small semantic space. This pattern indicated higher levels of overall “stuckness” from beginning to end. Furthermore, this cluster also revealed an average propensity to stay in and travel towards less enjoyable topics and topics that are also perceived to be problematic in day-to-day life. Cluster 2 described thought trajectories in which individuals traveled big semantic distances at a faster rate. Trials in this cluster did not tend to revisit previous topics or trend towards problematic or unenjoyable semantic locations. Accordingly, this cluster resembled most our *a priori* views of adaptive, flexible mind-wandering. Finally, cluster 3 was most challenging to interpret due to its less consistent replication across samples. This cluster revealed low levels of revisiting patterns, but average rates of approach while travelling small semantic distances, which could indicate higher levels of "stuckness". Importantly, most $\hat{k}$s in this cluster were located in enjoyable locations across samples, and, in the online sample, most $\hat{k}$s were also found in semantic locations that were not perceived to be problematic in day-to-day life. However, the in-person sample reported that most of those $\hat{k}$s were indeed located in problematic semantic spaces. Taken together, while this overall cluster did not fully resemble an RNT, it did reveal some level of "stuckness" that was perceived as pathological in one of the two groups.

3.1.4. Cluster Testing

To test the hypothesis that cluster 1 resembled ruminative and worried behavior most, we calculated the proportion of trials that every individual had clustered into cluster 1. When regressing those trial proportions to self-report measures of worry (PSWQ) and rumination (RRS), we found a positive correlation of a weak-to-moderate effect size between the proportions of cluster 1 trials and RRS scores in the online sample, and a very weak, slightly negative relation between proportion of cluster 1 trials and PSWQ scores in the in-person sample (RRS: β = 1.12, $t(63)$ = 2.30, p < .05*, PSWQ: β = -0.04, $t(63)$ = -0.11, p = .91). In the online sample, we found two positive relations of small effect size between these variables (RRS: β = 0.98, $t(77)$ = 1.59, p = .12, PSWQ: β = 0.47, $t(77)$ = 1.19, p = .24). These results mostly suggested that cluster 1 indeed captured some variance in ruminative behavior. For a visual representation of these regressions, see *Figure 5*. Next, we investigated cluster 3. Although we did not expect cluster 3 to positively predict RNT in the online sample, we did expect it to predict thought "stuckness" in the in-person sample, given the high average Perceived Pathology ratings that resulted from that cluster in the in-person sample. Our online sample revealed negative relations of small effect sizes between the proportions of cluster 3 trials and our self-reported measures of rumination and worry (RRS: β = -2.64, $t(77)$ = -2.26, p < .05*, PSWQ: β = -1.32, $t(77)$ = -1.73, p = .08+), while our in-person sample yielded the expected positive relationship of small effect size between the proportions of cluster 3 trials and worry measures, and a very weak negative relationship between the proportions of cluster 3 trials and rumination measures (RRS: β = 0.29, $t(63)$ = 0.28, p = .78, PSWQ: β = 1.39, $t(63)$ = 1.98, p < .05*). For a visual representation of these regressions, see *Figure 6*. For completion, we also investigated how the mind-wandering-looking cluster, cluster 2, related to self-reported measures of worry and rumination. As expected, all of those relationships were negative and of weak or weak-to-moderate effect size, meaning that the more trials that individuals spent in cluster 2, the less they tended to worry or ruminate (Online: RRS: β = -0.27, $t(77)$ = -0.39, p = .70, PSWQ: β = -0.06, $t(77)$ = -0.15, p = .88, In-Person: RRS: β = -1.52, $t(63)$ = -2.78, p < .01**, PSWQ: β = -0.39, $t(63)$ = -0.99, p = .97). For a visual representation of these regressions, see *Figure 7*.

4. Discussion

Our project aimed to study ruminative/worried thought trajectories using dynamic attractor-state modelling. Our modelling allowed us to investigate the location towards which individuals were travelling during multiple word-chains, total semantic distance travelled, topic-revisiting patterns, and strength of attractor spaces. First, simple regression analyses mostly replicated previous literature showing that individuals with high self-report measures of depression and anxiety will spontaneously recall negative-valenced words and will continue to freely associate negatively-valenced words significantly more than healthy individuals (Andrews-Hanna et al., 2022). While our modelling approach is the first of its kind to try and capture RNT severity and it cannot easily be compared to prior studies, resulting parameter clusters exhibited a strong degree of agreement across two independent samples and consistently revealed one cluster that more closely resembled RNT behavior as described in previous literature (Christoff et al., 2016; Nolen-Hoeksema et al., 2008). Moreover, this cluster was consistently related to higher self-report measures of rumination, and higher self-report measures of worry in one of the samples, which is promising. Importantly, two of these relationships were non-significant, and our regression failed to positively predict self-report measures of worry in our second sample. This could be due to 1. flaws in our model, which will be discussed in the next section, 2. the inherent limitations of self-report measures, namely retrospective report errors and construct validity problems (Conner & Barrett, 2012; Treynor, 2003), and/or 3. the trait-like nature of the questionnaires that we selected. Both measures of perseverative thinking that we collected, the RRS and the PSWQ, are questionnaires that exclusively rely on retrospective report and are aimed at understanding trait-like thinking patterns that individuals believe they experience in day-to-day life. None of our selected RNT questionnaires examined the emotional and cognitive state of our participants at the time of testing, nor did they provide objective (biological, behavioral, etc.) measures of trait worry or rumination. Our model was built to capture ruminative and worried-like patterns of thinking at the time of testing. It could have been the case that our community-sample participants were not actively experiencing ruminative or worried episodes when they partook in our study, even if they sometimes did believe themselves to be ruminators or worries in day-to-day life. Similarly, we may have captured some ruminative and worried-like thinking patterns in individuals who do not believe themselves to often engage with this type of maladaptive thinking. Because of that, while we were expecting our models to capture some of the variance in our self-report measures, we are not surprised to observe only weak-to-moderate associations with our self-report.

Interestingly, our model allowed for two types of stuck thinking: thinking that started, ended, and moved slowly within a small problematic semantic space, and thinking that quickly traveled long distances towards a problematic semantic space. Cluster 1 exclusively resembled the former stuck-thinking pattern that we proposed. This finding is intriguing as it also speaks to prior research showing that RNT is related to decreased creativity (Davis & Nolen-Hoeksema, 2000; Folkman & Lazarus, 1980, 1986). Our results show that individuals who self-reported higher ruminative and worried scores were also individuals that often did not travel far in the semantic space, hence showing a pattern of word-associations that was not as imaginative, flexible, or creative in nature.

Together, we believe this paradigm and model show great potential for future work. We are particularly encouraged by having replicated our clustering results across samples, discovering a robust ruminative/worried-like cluster and a mind-wandering-like cluster, according to prior literature, and finding individual differences in thinking patterns and strategies that captures some of the variance found in clinical self-report measures.

4.1. Limitations and Next Steps

Our model should be considered in light of several limitations, which also reveal necessary future steps. First, and most importantly, our model fits were lower than desired. This likely was caused by two primary factors: restricted parameter ranges, and insufficient number of data simulations. As explained in the Methods section, due to the limited computational power of our devices and the restricted timeline of this study, data simulations and parameter ranges had to be set to a smaller interval than we realize would have been ideal. Most noticeably, the parameter ranges that were provided to estimate $\hat{k}$ were likely problematically narrow. In the near future, we plan on moving our analyses to a new analysis machine to become able to widen our parameter ranges and significantly increase the number of data simulations that we conduct. Also, with the goal to increase our model fits, we realize the importance of developing and including a parameter into our model that will capture some of the noise in our data and will help better fit those individuals who were truly engaging in mind-wandering-like thinking patterns. Our current model was a RNT model, aimed at detecting attractor spaces and their strength. Accordingly, our model often failed to capture non-RNT-like thinking or thinking that was not affected by an attractor space and freely traveled the semantic space. Further editing our model will be necessary as we aim to robustly and consistently differentiate ruminative/worried clusters from mind-wandering ones.

On a second note, because of the novelty of our model, we were not able to check previous literature to help guide our choice of community sample size. As a result, we are unsure that we counted on large enough samples, and fear that we may have been underpowered to capture robust ruminative and worried thinking patterns in our non-clinical samples. Further evidence of our potentially underpowered sample size may be the fact that we did not fully replicate previous findings: while our online sample (N = 79) did yield weak-to-moderate relationships between RRS, PSWQ, PHQ, PSS, DAS, PANAS, and GAD, and Enjoyment and Perceived Pathology ratings, our in-person sample (*N* = 65) often revealed very weak relationships of equal directionality. On a similar note, we also question the quality of the rating scales data that we collected. Because we interviewed most in-person participants after they completed the study, we know that most participants reported feeling tired by the end of the study and finding the ratings section tedious and excessively long. This reality may have negatively impacted the quality of our data. Consequently, we acknowledge a need to shorten the task significantly by either collect less rating scales or finding a new way of collecting data on enjoyment and perceived pathology measures.

A combination of the aforementioned limitations also makes us uncertain about the Enjoyment and Perceived Pathology ratings that we computed for our $\hat{k}$s. Often in word-chains that yielded poor model fits, our $\hat{k}$s were in very far-away locations from our participants' collected data. This reality raises the question of whether the Perceived Pathology and Enjoyment ratings that we computed for those locations are accurate. We believe that finding ways of improving our parameter fits and choosing different tools to collect Enjoyment and Perceived Pathology ratings may help this limitation. Finally, in upcoming data collections, we plan on adding a psychometric measure of English fluency. Although all participants attested to being fluent English speakers, and all data collection happened in the United States, it will be beneficial to control for English fluency in future analyses.

5. Tables

Sample 1			
		N	**%**
Gender	Cis-Man	22	33.8
	Cis-Woman	43	66.2
Race	White	33	50.8
	Black	9	13.8
	Asian	23	35.4
Ethnicity	Latino/Hispanic	9	13.8
	Non-Latino/Hispanic	38	86.2
Sample 2			
		N	**%**
Gender	Cis-Man	38	48.1
	Cis-Woman	35	44.3
	Trans-Man	1	1.3
	Other	5	6.3
Race	White	61	77.2
	Black	7	8.9
	Asian	5	6.3
	Native America	3	3.8
	Other/Unknown	3	3.8
Ethnicity	Latino/Hispanic	3	3.8
	Non-Latino/Hispanic	76	96.2
Age		**M**	**SD**
		37.1	12.5

Table 1: Sample 1 and Sample 2 Participant Demographics

Predictor: Average Enjoyment					
Outcome Variable	**Coefficient**	R^2	**F-statistic**	***p*-value**	**Sample**
RRS	-0.60	0.046	3.73	0.057+	Online
	-0.09	0.001	0.05	0.816	In-Person
Brooding	-0.17	0.065	5.31	0.024*	Online
	-0.07	0.009	0.55	0.462	In-Person
Reflection	-0.05	0.006	0.46	0.499	Online
	0.06	0.005	0.35	0.558	In-Person
Depression	-0.37	0.048	3.88	0.053+	Online
	-0.09	0.002	0.13	0.723	In-Person
PSWQ	-0.37	0.044	3.51	0.065+	Online
	-0.15	0.005	0.30	0.587	In-Person
PHQ-9	-0.24	0.051	4.13	0.046*	Online
	-0.25	0.033	2.15	0.148	In-Person
GAD	-0.24	0.055	4.50	0.037*	Online
	-0.13	0.011	0.71	0.404	In-Person
PSS	-0.62	0.127	11.20	0.001**	Online
	-0.29	0.027	1.74	0.192	In-Person
DAS	-0.17	0.031	2.48	0.119	Online
	-0.36	0.105	7.41	0.008**	In-Person
PANAS Neg.	-0.45	0.092	7.76	0.007**	Online
	-0.23	0.017	1.11	0.296	In-Person
PANAS Pos.	0.38	0.058	4.77	0.032*	Online
	0.62	0.116	8.24	0.006**	In-Person
DARS	-0.35	0.034	2.70	0.104	Online
	-0.85	0.125	9.02	0.004**	In-Person

**Note*: +$p < 0.1$, *$p < 0.05$, **$p < 0.01$

Table 2: Average "Perceived Enjoyment" Predicts Clinical Measures; Both Samples

Predictor: Average Perceived Pathology					
Outcome Variable	**Coefficient**	R^2	**F-statistic**	***p*-value**	**Sample**
RRS	0.21	0.006	0.47	0.494	Online
	0.34	0.018	1.16	0.286	In-Person
Brooding	0.10	0.026	2.08	0.153	Online
	0.003	0.000	0.001	0.973	In-Person
Reflection	-0.03	0.002	0.14	0.706	Online
	0.07	0.013	0.80	0.374	In-Person
Depression	0.13	0.007	0.50	0.480	Online
	0.24	0.025	1.58	0.213	In-Person
PSWQ	0.34	0.039	3.16	0.080+	Online
	-0.03	0.000	0.02	0.878	In-Person
PHQ-9	0.19	0.037	2.93	0.091+	Online
	0.12	0.012	0.76	0.387	In-Person
GAD	0.16	0.028	2.21	0.141	Online
	0.01	0.000	0.009	0.927	In-Person
PSS	0.38	0.051	4.17	0.045*	Online
	0.06	0.002	0.10	0.753	In-Person
DAS	0.12	0.016	1.28	0.262	Online
	-0.008	0.000	0.005	0.943	In-Person
PANAS Neg.	0.48	0.115	9.98	0.002**	Online
	0.06	0.002	0.12	0.726	In-Person
PANAS Pos.	-0.01	0.000	0.001	0.973	Online
	0.08	0.003	0.18	0.675	In-Person
DARS	-0.16	0.008	0.63	0.428	Online
	0.12	0.004	0.24	0.624	In-Person

**Note*: $+p < 0.1$, $*p < 0.05$, $**p < 0.01$

Table 3: Average "Perceived Pathology" Predicts Clinical Measures; Both Samples

Predictor: Enjoyment *at* $\hat{k}$					
Outcome Variable	**Coefficient**	**Std Err**	**t-statistic**	***p*-value**	**Sample**
RRS	-0.58	0.313	-1.86	0.067+	Online
	-0.02	0.399	-0.06	0.954	In-Person
Brooding	-0.17	0.076	-2.20	0.031*	Online
	-0.05	0.100	-0.53	0.600	In-Person
Reflection	-0.05	0.079	-0.69	0.495	Online
	0.07	0.096	0.78	0.437	In-Person
Depression	-0.36	0.191	-1.89	0.062+	Online
	-0.06	0.244	-0.24	0.812	In-Person
PSWQ	-0.36	0.202	-1.79	0.078+	Online
	-0.13	0.274	-0.46	0.648	In-Person
PHQ-9	-0.24	0.120	-2.00	0.049*	Online
	-0.23	0.168	-1.37	0.177	In-Person
GAD	-0.22	0.111	-2.03	0.046*	Online
	-0.11	0.150	-0.76	0.449	In-Person
PSS	-0.61	0.188	-3.24	0.002**	Online
	-0.27	0.220	-1.21	0.230	In-Person
DAS	-0.17	0.111	-1.56	0.123	Online
	-0.33	0.132	-2.53	0.014*	In-Person
PANAS Neg.	-0.44	0.164	-2.68	0.009**	Online
	-0.19	0.212	-0.90	0.374	In-Person
PANAS Pos.	0.34	0.171	2.00	0.049*	Online
	0.64	0.211	3.02	0.004**	In-Person
DARS	-0.35	0.211	-1.65	0.103	Online
	-0.83	0.280	-2.97	0.004**	In-Person

**Note*: $+p < 0.1$, $*p < 0.05$, $**p < 0.01$

Table 4: "Perceived Enjoyment" at $\hat{k}$ Predicts Clinical Measures; Both Samples

Predictor: Perceived Pathology *at* $\hat{k}$					
Outcome Variable	**Coefficient**	**Std Err**	**t-statistic**	***p*-value**	**Sample**
RRS	0.23	0.297	0.79	0.434	Online
	0.34	0.316	1.074	0.287	In-Person
Brooding	0.11	0.072	1.49	0.139	Online
	0.003	0.080	0.04	0.969	In-Person
Reflection	-0.01	0.074	-0.20	0.841	Online
	0.07	0.076	0.93	0.356	In-Person
Depression	0.14	0.181	0.78	0.437	Online
	0.24	0.193	1.24	0.220	In-Person
PSWQ	0.35	0.188	1.85	0.069+	Online
	-0.007	0.219	-0.03	0.973	In-Person
PHQ-9	0.19	0.112	1.72	0.090+	Online
	0.12	0.135	0.89	0.376	In-Person
GAD	0.16	0.104	1.54	0.127	Online
	0.003	0.120	0.02	0.981	In-Person
PSS	0.39	0.181	2.17	0.033*	Online
	0.06	0.178	0.32	0.747	In-Person
DAS	0.12	0.104	1.14	0.257	Online
	-0.02	0.111	-0.14	0.886	In-Person
PANAS Neg.	0.49	0.149	3.26	0.002**	Online
	0.04	0.171	0.25	0.801	In-Person
PANAS Pos.	0.004	0.163	0.02	0.981	Online
	0.07	0.181	0.40	0.692	In-Person
DARS	-0.17	0.199	-0.84	0.404	Online
	0.13	0.239	0.53	0.598	In-Person

**Note*: +$p < 0.1$, *$p < 0.05$, **$p < 0.01$

Table 5: "Perceived Pathology" at $\hat{k}$ Predicts Clinical Measures; Both Samples

			Perceived Pathology		Enjoyment		Cosine Similarity		Traveled Distance	
One-way ANOVA			**F-statistic**	***p***	**F-statistic**	***p***	**F-statistic**	***p***	**F-statistic**	***p***
		Online	0.39	.68	0.57	.56	26.9	3.3E-12***	308.7	1.2E-114***
		In-Pers	3.29	.04*	1.43	.24	35.6	8.2E-16***	258.8	5.5E-96 ***
Post-hoc Tukey Tests										
	Cluster	**Cluster**	**meandiff**	***p*-adj**	**meandiff**	***p*-adj**	**meandiff**	***p*-adj**	**meandiff**	***p*-adj**
Online	1	2	-0.19	.81	0.29	.57	-0.041	.0***	1352.15	.0***
In-Pers.			-0.79	.03*	0.28	.50	-0.053	.0***	693.71	.0***
Online	1	3	0.35	.71	-0.28	.79	0.018	.07+	1352.15	.0***
In-Pers.			-0.096	.98	0.58	.28	-0.021	.07+	-1400.87	.0***
Online	2	3	0.17	.93	0.0045	.99	-0.023	.02*	2073.65	.0***
In-Pers.			0.69	.34	0.30	.72	0.032	.004**	-2094.58	.0***

			Kernel density		RRS		PSWQ	
One-way ANOVA			**F-statistic**	***p***	**F-statistic**	***p***	**F-statistic**	***p***
		Online	3.29	.04*	2.58	.08+	2.09	.12
		In-Pers	2.45	.09+	6.64	.0001***	2.44	.09+
Post-hoc Tukey Tests								
	Cluster	**Cluster**	**meandiff**	***p*-adj**	**meandiff**	***p*-adj**	**meandiff**	***p*-adj**
Online	1	2	-0.0	.04*	-0.83	.53	-0.38	.72
In-Pers.			-0.0	.09+	-2.79	.0008***	-0.39	.74
Online	1	3	0.0	.25	2.56	.07+	1.51	.10
In-Pers.			-0.0	.37	-1.34	.49	1.45	.17
Online	2	3	-0.0	1	1.73	.30	1.13	.29
In-Pers.			0.0	1	1.45	.45	1.84	.07+

**Note*: +p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table 5: Unsupervised Clusters Statistics: We conducted an initial one-way ANOVA to test for differences in average measures of Perceived Pathology, Enjoyment, cosine similarity, traveled distance per word-chain, kernel density estimates, RRS, and PSWQ across clusters (see top 2 rows). Next, we conducted *post-hoc* Tukey Honestly Significant Difference tests to improve our understanding of exactly how the clusters differed (see bottom 6 rows). Note that none of the variables in this table were used in the unsupervised clustering process.

6. Figures

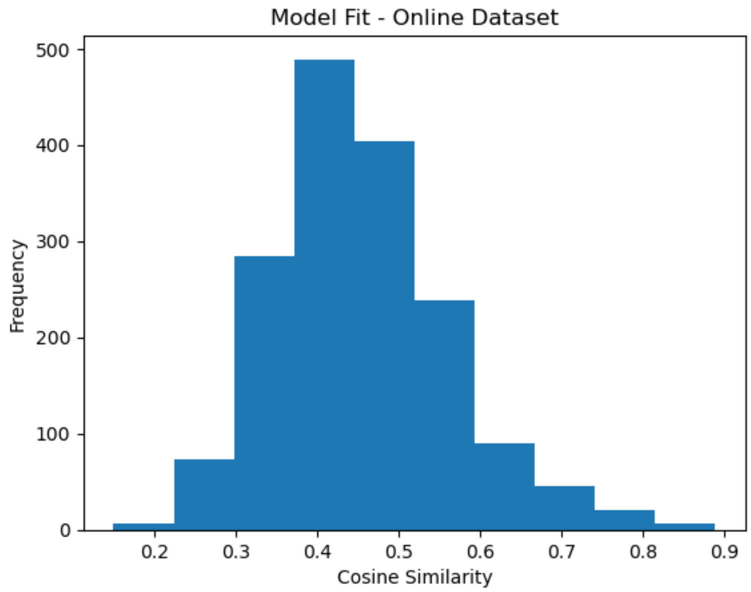


Figure 1: Model Fit (Cosine Similarities) Distribution of Online Sample

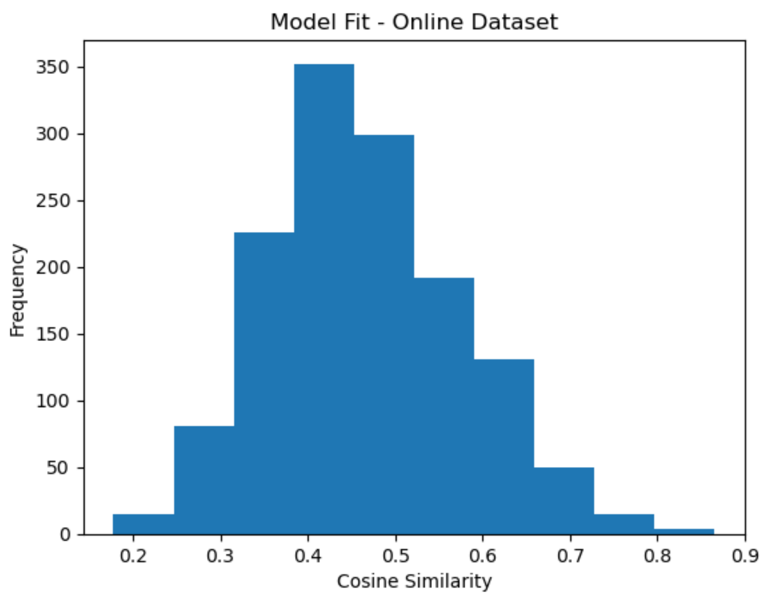


Figure 2: Model Fit (Cosine Similarities) Distribution of In-Person Sample

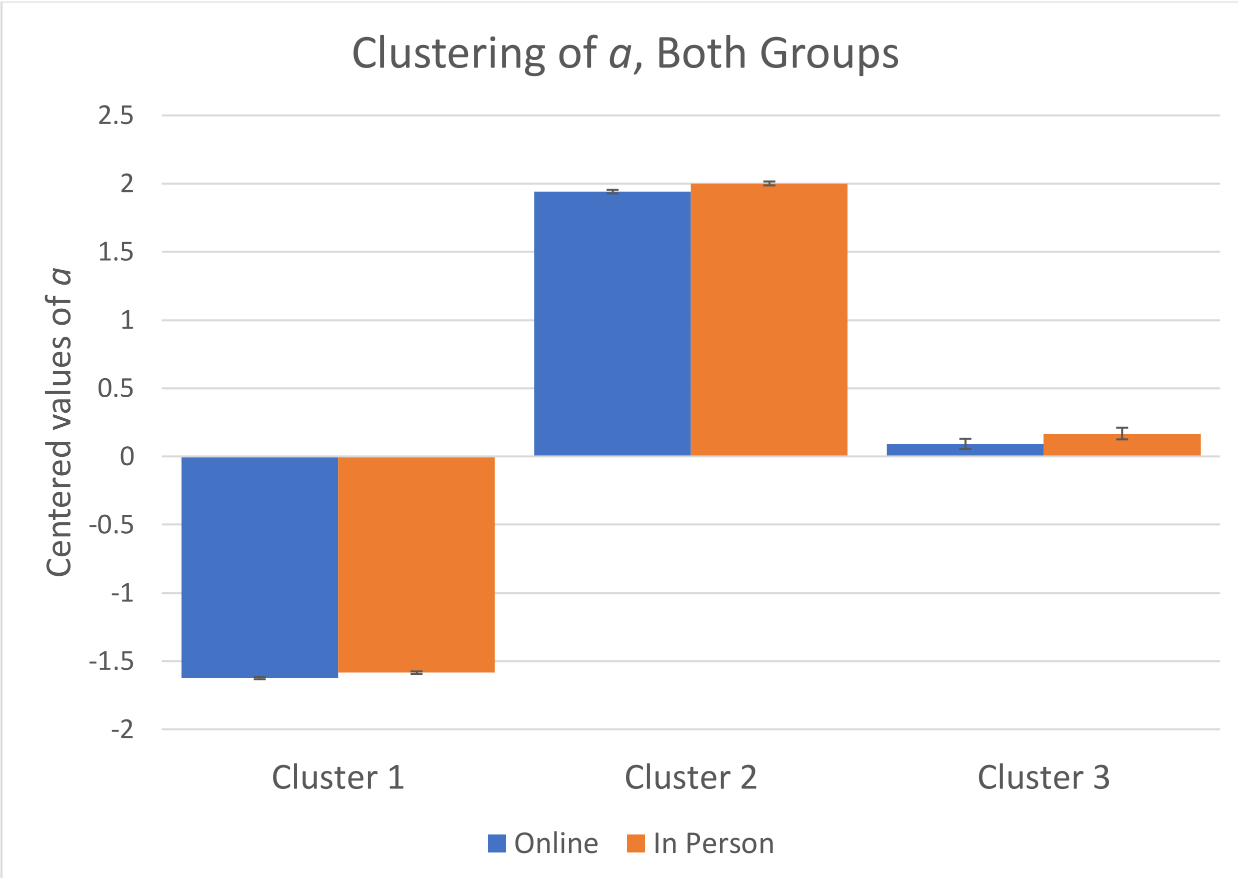
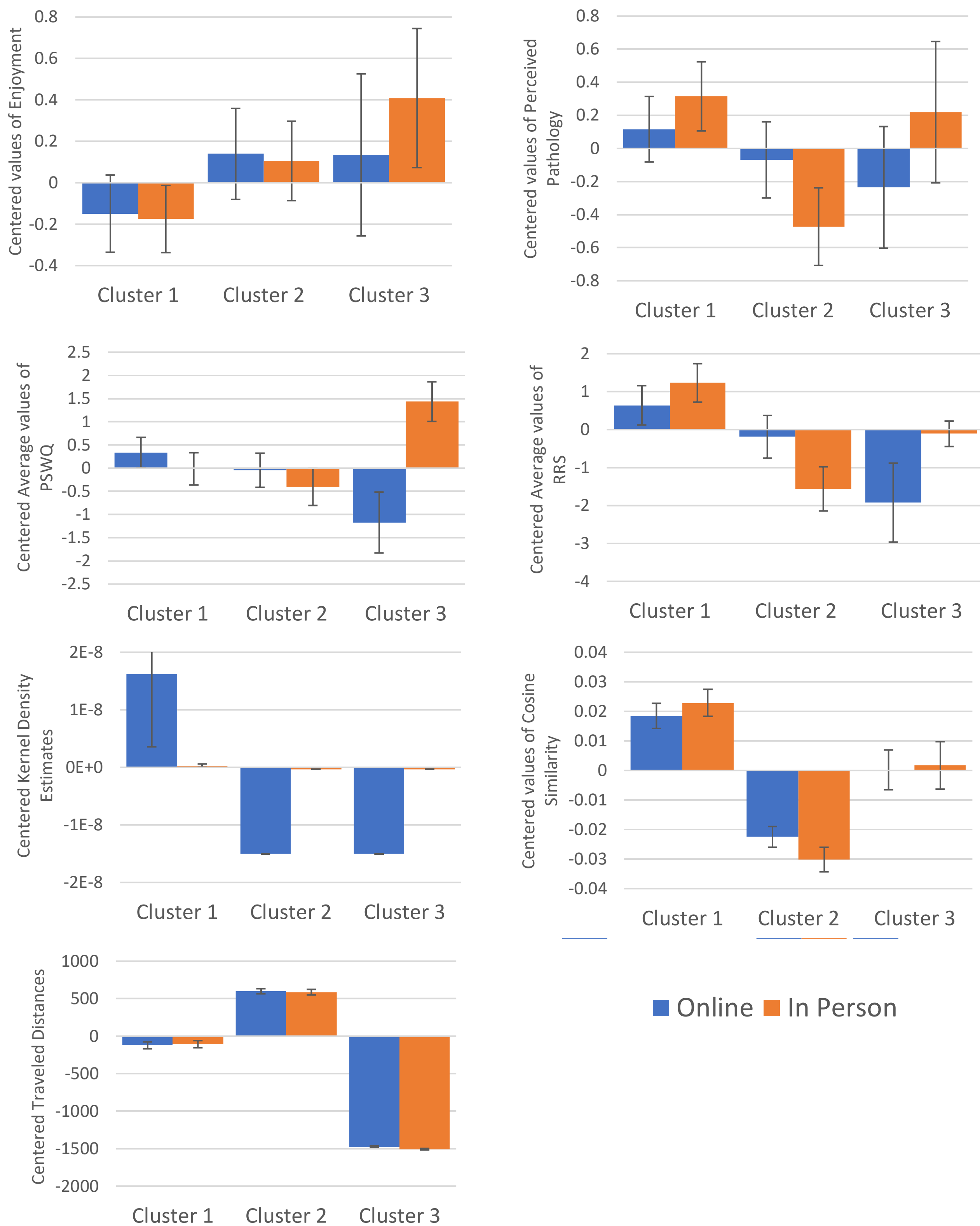
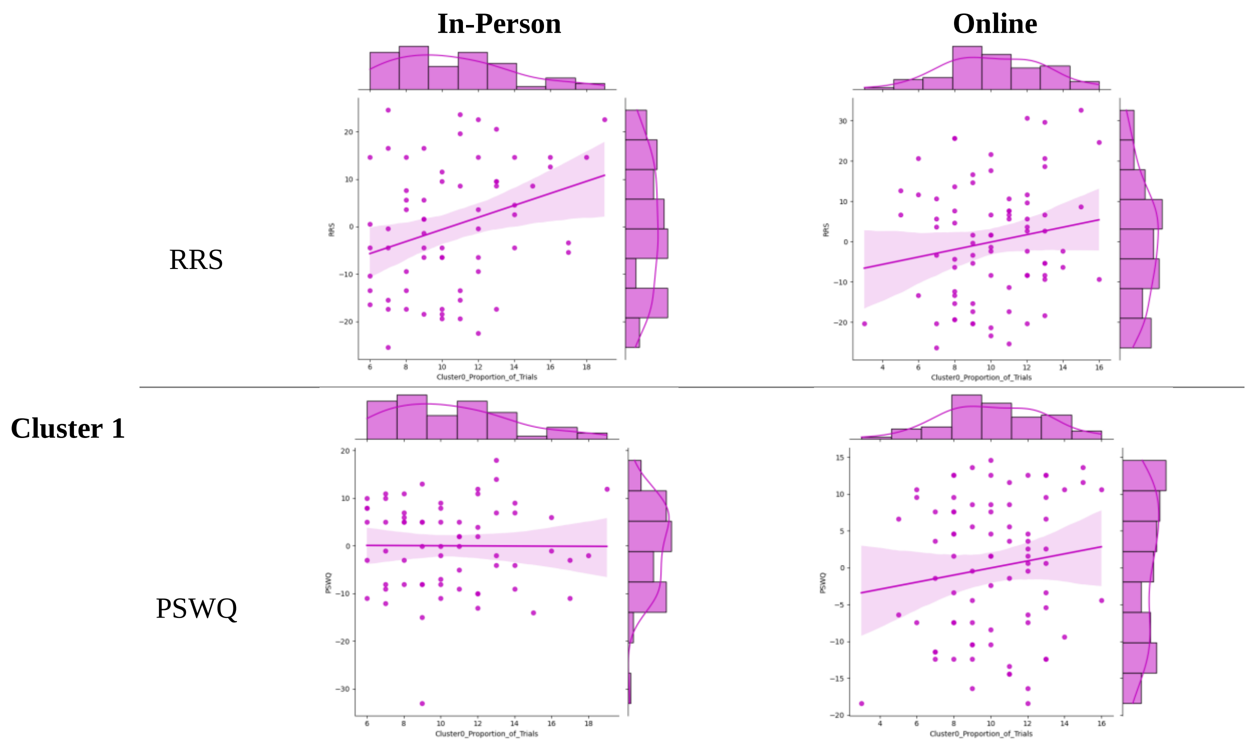


Figure 3: Unsupervised Clusters Based on *a*: Summary of *a*



**Note: Centered cluster averages are displayed*

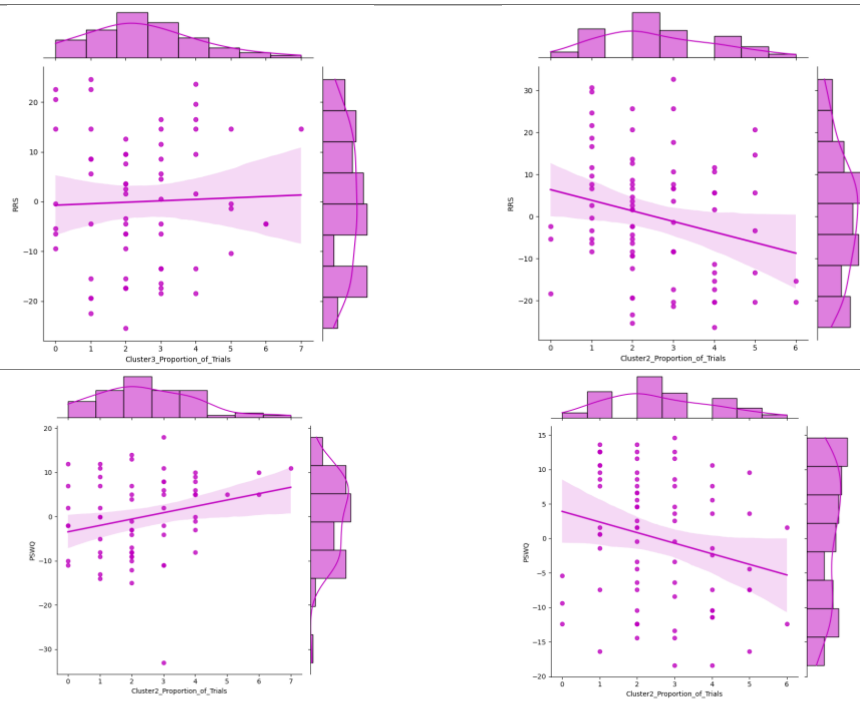
Figure 4: Unsupervised Clusters Based on *a*: Summary of Variables that were not Used for Clustering



**Note: Centered RRS and PSWQ scores are displayed*

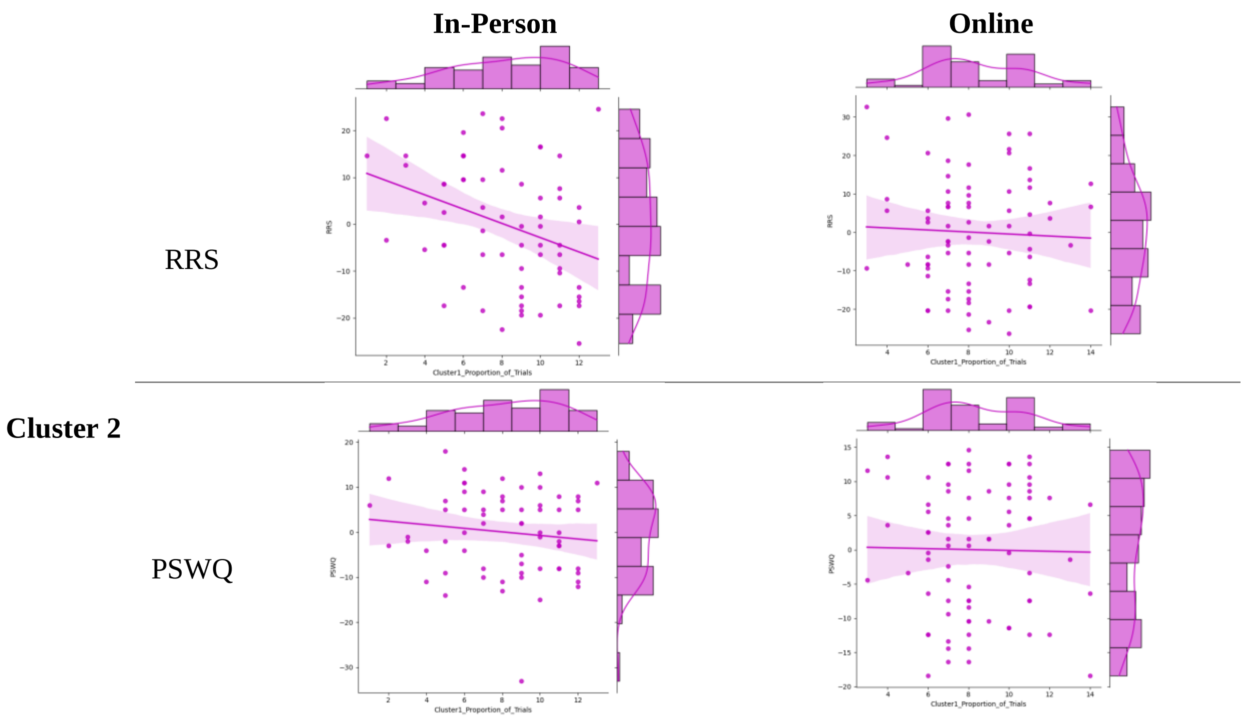
Figure 5: Relation between Proportion of Trials in Cluster 1 and self-report measures of rumination (RRS) and worry (PSWQ)

	In-Person	Online
Cluster 3 RRS		
PSWQ		



**Note: Centered RRS and PSWQ scores are displayed*

Figure 6: Relation between Proportion of Trials in Cluster 3 and self-report measures of rumination (RRS) and worry (PSWQ)



**Note: Centered RRS and PSWQ scores are displayed*

Figure 7: Relation between Proportion of Trials in Cluster 2 and self-report measures of rumination (RRS) and worry (PSWQ)

7. Citation Diversity Statement

Work from several fields of science has revealed that there is a bias in citation practices, as the research of women and racial/ethnic minority scholars is often under-cited relative to the number of such papers in the field (Bertolero et al., 2020; Dworkin et al., 2020). Here we worked to choose references that would reflect the diversity in our field. In an attempt to increase transparency and bring visibility to this problem, we report the gender and racial/ethnic breakdown of our references. Regarding gender, our references contain 20% woman(first)/woman(last), 20% man/woman, 22.9% woman/man, and 37.1% man/man. Regrading race/ethnicity, our references contain 8.6% author of color (first)/author of color(last), 5.7% white author/author of color, 5.7% author of color/white author, and 80% white author/white author. We acknowledge that further work needs to be done to promote and support equitable practices in science and academia.

8. References

Amir, I., & Bernstein, A. (2022). Dynamics of Internal Attention and Internally-Directed Cognition: The Attention-to-Thoughts (A2T) Model. *Psychological Inquiry*, *33*(4), 239–260. https://doi.org/10.1080/1047840X.2022.2141000

Andrews-Hanna, J. R., Woo, C.-W., Wilcox, R., Eisenbarth, H., Kim, B., Han, J., Losin, E. A. R., & Wager, T. D. (2022). The conceptual building blocks of everyday thought: Tracking the emergence and dynamics of ruminative and nonruminative thinking. *Journal of Experimental Psychology: General*, *151*(3), 628–642. https://doi.org/10.1037/xge0001096

Bertolero, M. A., Dworkin, J. D., David, S. U., Lloreda, C. L., Srivastava, P., Stiso, J., Zhou, D., Dzirasa, K., Fair, D. A., Kaczkurkin, A. N., Marlin, B. J., Shohamy, D., Uddin, L. Q., Zurn, P., & Bassett, D. S. (2020). *Racial and ethnic imbalance in neuroscience reference lists and intersections with gender* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.10.12.336230

Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: A dynamic framework. *Nature Reviews Neuroscience*, *17*(11), 718–731. https://doi.org/10.1038/nrn.2016.113

Compton, W. M., Thomas, Y. F., Stinson, F. S., & Grant, B. F. (2007). Prevalence, Correlates, Disability, and Comorbidity of DSM-IV Drug Abuse and Dependence in the United States: Results From the National Epidemiologic Survey on Alcohol and Related Conditions. *Archives of General Psychiatry*, *64*(5), 566. https://doi.org/10.1001/archpsyc.64.5.566

Conner, T. S., & Barrett, L. F. (2012). Trends in Ambulatory Self-Report: The Role of Momentary Experience in Psychosomatic Medicine. *Psychosomatic Medicine*, *74*(4), 327–337. https://doi.org/10.1097/PSY.0b013e3182546f18

Cuthbert, B. N. (2014). The RDoC framework: Facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology: Forum - The Research Domain Criteria Project. *World Psychiatry*, *13*(1), 28–35. https://doi.org/10.1002/wps.20087

Davis, R. N., & Nolen-Hoeksema, S. (2000). Cognitive Inflexibility Among Ruminators and Nonruminators. *Cognitive Therapy and Research*, *24*(6), 699–711. https://doi.org/10.1023/A:1005591412406

Dworkin, J. D., Linn, K. A., Teich, E. G., Zurn, P., Shinohara, R. T., & Bassett, D. S. (2020). The extent and drivers of gender imbalance in neuroscience reference lists. *Nature Neuroscience*, *23*(8), 918–926. https://doi.org/10.1038/s41593-020-0658-y

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gupta, A., Kaiser, C., Everaert, J., van Vugt, M. K., & Roy, P. P. (2022). Modeling effects of rumination on free recall using ACT-R.

Folkman, S., & Lazarus, R. S. (1980). An Analysis of Coping in a Middle-Aged Community Sample. *Journal of Health and Social Behavior*, *21*(3), 219. https://doi.org/10.2307/2136617

Folkman, S., & Lazarus, R. S. (1986). Stress processes and depressive symptomatology. *Journal of Abnormal Psychology*, *95*(2), 107–113. https://doi.org/10.1037/0021-843X.95.2.107

Hoyer, J., Gloster, A. T., & Herzberg, P. Y. (2009). Is worry different from rumination? Yes, it is more predictive of psychopathology! [Text/html]. *GMS Psycho-Social-Medicine; 6:Doc06; ISSN 1860-5214*. https://doi.org/10.3205/PSM000062

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When More Pain Is Preferred to Less: Adding a Better End. *Psychological Science*, *4*(6), 401–405. https://doi.org/10.1111/j.1467-9280.1993.tb00589.x

Kessler, R. C., Nelson, C. B., McGonagle, K. A., Liu, J., Swartz, M., & Blazer, D. G. (1996). Comorbidity of DSM–III–R Major Depressive Disorder in the General Population: Results from the US National Comorbidity Survey. *British Journal of Psychiatry*, *168*(S30), 17–30. https://doi.org/10.1192/S0007125000298371

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour Research and Therapy*, *28*(6), 487–495. https://doi.org/10.1016/0005-7967(90)90135-6

Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology*, *109*(3), 504–511. https://doi.org/10.1037/0021-843X.109.3.504

Nolen-Hoeksema, S., & Morrow, J. (1991). A prospective study of depression and posttraumatic stress symptoms after a natural disaster: The 1989 Loma Prieta earthquake. *Journal of Personality and Social Psychology*, *61*(1), 115–121. https://doi.org/10.1037/0022-3514.61.1.115

Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking Rumination. *Perspectives on Psychological Science*, *3*(5), 400–424. https://doi.org/10.1111/j.1745-6924.2008.00088.x

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., & Contributors. (2012). *Computational Cognitive Neuroscience* (4th Edition (2020)). Wiki Book. https://CompCogNeuro.org

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. http://www.aclweb.org/anthology/D14-1162

Ringwald, W. R., Forbes, M. K., & Wright, A. G. C. (2021). Meta-analysis of structural evidence for the Hierarchical Taxonomy of Psychopathology (HiTOP) model. *Psychological Medicine*, 1–14. https://doi.org/10.1017/S0033291721001902

Rutherford, A. V., McDougle, S. D., & Joormann, J. (2023). "Don't [ruminate], be happy": A cognitive perspective linking depression and anhedonia. *Clinical Psychology Review*, *101*, 102255. https://doi.org/10.1016/j.cpr.2023.102255

Satopa, V., Albrecht, J., Irwin, D., & Raghavan, B. (n.d.). *Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior*.

Shiffman, S., Hufford, M., Hickcox, M., Paty, J. A., Gnys, M., & Kassel, J. D. (1997). Remember that? A comparison of real-time versus retrospective recall of smoking lapses. *Journal of Consulting and Clinical Psychology*, *65*(2), 292–300. https://doi.org/10.1037/0022-006x.65.2.292.a

Smallwood, J., & Schooler, J. W. (2015). The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. *Annual Review of Psychology*, *66*(1), 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Solhan, M. B., Trull, T. J., Jahng, S., & Wood, P. K. (2009). Clinical assessment of affective instability: Comparing EMA indices, questionnaire reports, and retrospective recall. *Psychological Assessment*, *21*(3), 425–436. https://doi.org/10.1037/a0016869

Stone, A. A., Schwartz, J. E., Neale, J. M., Shiffman, S., Marco, C. A., Hickcox, M., ... & Cruise, L. J. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of personality and social psychology*, *74*(6), 1670.

Spinhoven, P., Penninx, B. W., Krempeniou, A., van Hemert, A. M., & Elzinga, B. (2015). Trait rumination predicts onset of Post-Traumatic Stress Disorder through trauma-related cognitive appraisals: A 4-year longitudinal study. *Behaviour Research and Therapy*, *71*, 101–109. https://doi.org/10.1016/j.brat.2015.06.004

Taatgen, N. A., van Vugt, M. K., Daamen, J., Katidioti, I., Huijser, S., & Borst, J. P. (2021). The resource-availability model of distraction and mind-wandering. *Cognitive Systems Research*, *68*, 84–104. https://doi.org/10.1016/j.cogsys.2021.03.001

Treynor, W. (2003). Rumination Reconsidered: A Psychometric Analysis. *Cognitive Therapy and Research*, *27*(3), 247–259. https://doi.org/10.1023/A:1023910315561

van Vugt, M., & Jamalabadi, H. (2022). Too Much Flexibility in a Dynamical Model of Repetitive Negative Thinking? *Psychological Inquiry*, *33*(4), 276–279. https://doi.org/10.1080/1047840X.2022.2149195 van Vugt, M. K., van der Velde, M., & ESM-MERGE Investigators. (2018). How Does Rumination Impact Cognition? A First Mechanistic Model. *Topics in Cognitive Science*, *10*(1), 175–191. https://doi.org/10.1111/tops.12318

Zorowitz, S., & Niv, Y. (2023). Improving the reliability of cognitive task measures: A narrative review. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, S2451902223000423. https://doi.org/10.1016/j.bpsc.2023.02.004