Mapping OCD Symptom Triggers with Large Language Models

Dorothée Bentz[1] and Dirk U. Wulff[2,1]
[1]University of Basel
[2]Max Planck Institute for Human Development

Abstract

Recent advancements in natural language processing (NLP) and large language models (LLMs) offer new avenues for exploring previously under-researched areas in mental health. Their capacity to automatically and meaningfully analyze large-scale text data makes them particularly valuable for studying highly individualized clinical phenomena, such as triggers of obsessive-compulsive symptoms (OCS), where pattern identification is often challenging. To address this gap, we asked 1,495 individuals to identify their most common triggers, rate their intensity, and report the severity of their contamination-related OCS. Using LLM-based embeddings, we generated a map of key trigger categories, revealing their diversity across ecological domains and varying degrees of semantic similarity. Monte Carlo simulations further showed that individuals frequently reported semantically similar trigger pairs that differed in intensity. These findings are clinically significant, providing a foundation for a more fine-grained understanding of OCS treatment mechanisms and paving the way for novel therapeutic approaches.

Main

Rapid advances in automatic text data processing, particularly in natural language processing (NLP) and large language models (LLM), have the potential to bridge knowledge gaps in understanding mental disorders. NLP and LLM excel at analyzing large volumes of text data in a meaningful and automated manner (Chandran et al., 2019), and the growing availability of open-source LLMs further bolsters such research by enhancing transparency and reproducibility, which are crucial when investigating sensitive topics in mental health (Binz et al., 2025; Hussain et al., 2024; Wulff et al., 2024). This allows large populations to be surveyed on clinically relevant topics using open text formats, with their responses being subsequently efficiently analyzed (Feuerriegel et al., 2025; Hussain et al., 2024). This approach is particularly useful for studying clinical phenomena that are both widespread and highly idiosyncratic, such as triggers of compulsive behaviors and obsessive thoughts related to contamination fears.

Contamination fears are a central concern for approximately half of the 2 to 3 percent of individuals diagnosed with obsessive-compulsive disorder (OCD) over their lifetime, often leading to excessive and maladaptive cleaning rituals, commonly referred to as contamination-related OCD (C-OCD) (APA, 2013; Karno et al., 1988; Rasmussen & Eisen, 1992). In addition, obsessive-compulsive symptoms related to contamination (C-OCS), which are prevalent in the general population, share key characteristics with clinically significant C-OCD and are best conceptualized dimensionally (Abramowitz et al., 2014; Adam et al., 2012; Angst et al., 2004; Haslam et al., 2005; Mataix-Cols et al., 2003; Mataix-Cols et al., 2005).

Individuals with contamination-related obsessive-compulsive symptoms (C-OCS) typically respond to a range of distinct triggers (also called cues), such as specific objects (e.g., toilets or garbage bins) or situations (e.g., shaking hands or using public transportation) (APA, 2013). However, comprehensive studies that systematically examine the nature and prevalence of these triggers are lacking. The existing knowledge comes largely from anecdotal reports from patients and OCD experts (Cullen et al., 2021; Mataix-Cols et al., 2009; Rachman, 2004; Simon et al., 2012; Sousa et al., 2024), as well as from symptom provocation studies in individuals with OCS, which have assessed triggers based on theoretically derived dimensions such as fear or disgust (Cullen et al., 2021; Mataix-Cols et al., 2009; Simon et al., 2012; Sousa et al., 2024). Consequently, general data on the frequency of specific triggers in both clinical and non-clinical populations remain scarce, as well as more detailed information on demographic patterns, such as whether certain triggers are more common in one gender or age group.

However, a comprehensive empirical understanding of the various triggers of C-OCD and their semantic relationships is crucial to improving treatment outcomes. First, exposure

Dorothée Bentz https://orcid.org/0000-0003-0291-2356 Dirk U. Wulff https://orcid.org/0000-0002-4008-8022

Correspondence concerning this article should be addressed to Dorothée Bentz, Clinical Psychology and Translational Psychotherapy Research, Faculty of Psychology, University of Basel, Missionsstrasse 62A, 4055 Basel, Switzerland. E-mail: dorothee.bentz@unibas.ch and Dirk U. Wulff, Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin

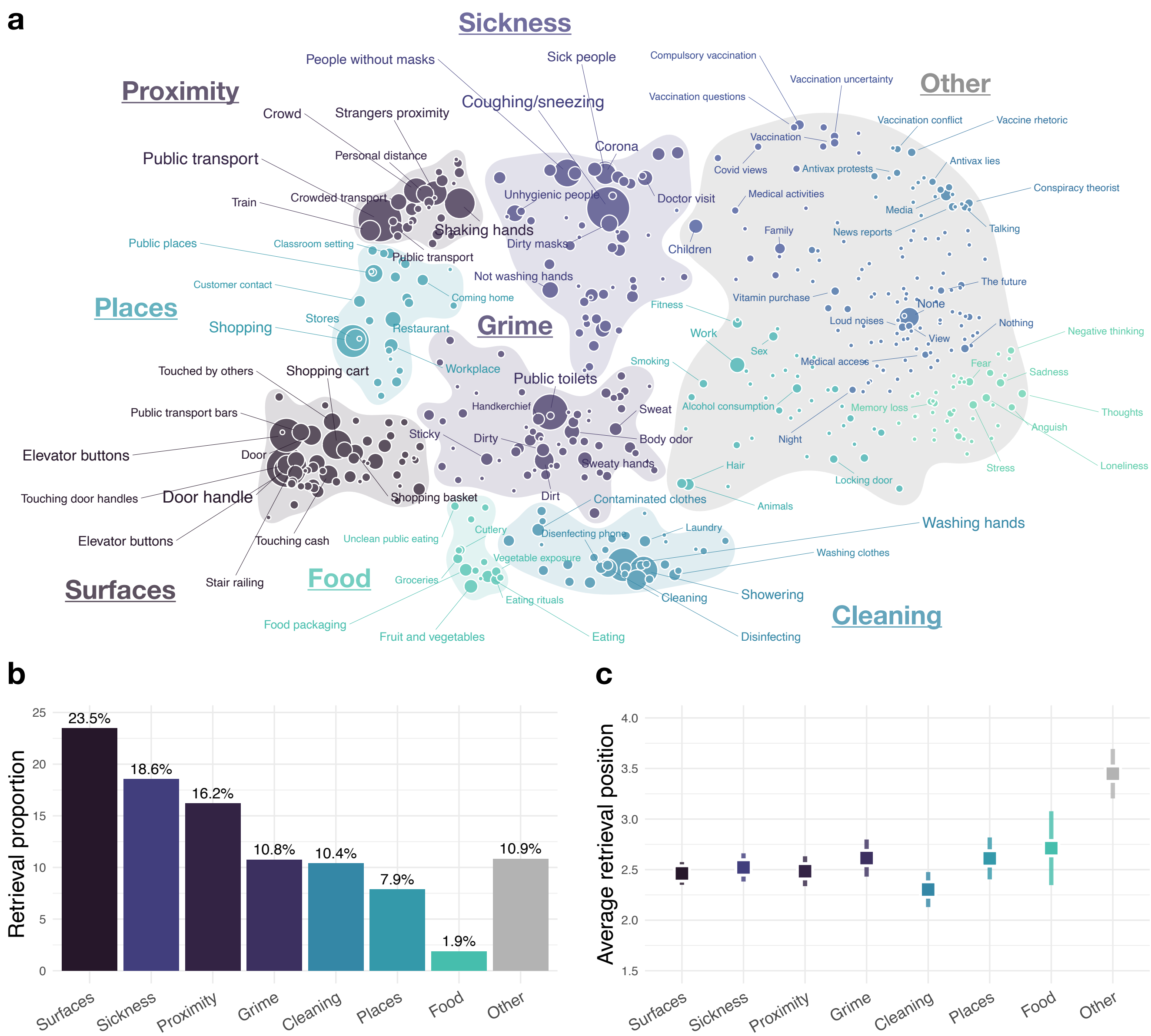

Figure 1

Mapping the C-OCS triggers. Panel A shows the C-OCS trigger map, generated using the PaCMAP dimensionality reduction algorithm. We identified twelve C-OCS trigger categories, each represented by a different color, including seven prominent categories identified as valid C-OCS triggers. Labels were assigned to best represent the semantic content of each trigger cluster. Panel B and C show, respectively, the proportion and average position of participant responses belonging to each cluster.

with response prevention (ERP), the first-line psychotherapeutic treatment for OCD, involves systematic exposure to these triggers (for Mental Health (UK). Leicester (UK): British Psychological Society (UK), 2006). Second, ERP faces a key challenge common to all exposure-based therapies: Success with one trigger does not always translate into success with other triggers (Jacoby & Abramowitz, 2016; Kodzaga et al., 2025). Third, the associative learning processes that are described as key mechanisms for symptomatology and the effectiveness of ERP treatment for OCS (Pittig et al., 2016) are highly dependent on the characteristics of the triggers (Kodzaga et al., 2025). In fear learning, for example, the knowledge of associations between one stimulus and another is transferred depending on their perceived similarity (Hall, 1996; Shepard, 1987, 2004), while generalization in extinction occurs only to a limited extent or is absent (Kodzaga et al., 2025; Vervliet et al., 2010; Vervliet et al., 2005).

In the context of OCD, few studies have explored the generalization of triggers (Cooper & Dunsmoor, 2021). Particularly at the conceptual level (Dunsmoor et al., 2011; Vervoort et al., 2014) and the semantic relational level (Boyle et al., 2016) studies are lacking. Yet, conditioning paradigms that assess generalization along these conceptual or semantic dimensions and not only the perceptual dimension may best capture the often complex and abstract nature of triggers for obsessive-compulsive symptoms (OCS) (Cooper & Dunsmoor, 2021). A key reason for the scarcity of research in this area might be the limited understanding of the semantic relationships of the C-OCS triggers.

Taken together, advancing research on extinction generalization - and ultimately improving the treatment of OCS - requires a comprehensive understanding of the nature of OCS triggers and their semantic relationships. This study aims to address this gap by leveraging large language models (LLMs) to map C-OCS triggers using a large dataset of free text responses from individuals in the general population who experience OCS. Specifically, we examine the frequency and content of various triggers, categorize them into distinct categories of triggers of C-OCS, and analyze their distribution between individuals with varying severity of symptoms, as well as across different age and gender groups. In addition, we investigate the relationship between trigger intensity and semantic similarity and explore whether less triggering, yet semantically related, stimuli can substitute for highly triggering ones. This approach lays the foundation for future research on associative learning in OCS, with the potential to inform novel intervention strategies.

Results

Our results are structured as follows. First, we present a map of C-OCS trigger categories derived from LLM embeddings and subsequent cluster analysis, addressing our aim to identify and categorize these triggers. Second, we examine these categories in terms of retrieval order, and how their prominence relates to symptom severity, age, and gender, to document key individual differences as outlined in our study objectives. Third, we analyze how semantic similarity influences trigger co-occurrences within individuals and variations in trigger intensity, exploring a potential pathway to enhance understanding of mechanisms relevant to treatment. Our analysis includes 1,213 individuals with a score greater than zero on the OCI-R washing/contamination dimension, out of 1,399 participants who reported C-OCS triggers. In total, they identified 3,508 triggers.

Mapping C-OCS triggers

To create a map of trigger categories, we first embedded the triggers using a fine-tuned embedding model to understand the semantic organization of C-OCS triggers (see huggingface.co/dwulff/mpnet-cocs). Based on this model, we then grouped C-OCS triggers into semantically cohesive groups and projected the average embedding of these groups into a two-dimensional map using PaCMAP. Finally, we partitioned the map into twelve clusters using hierarchical clustering, creating highly interpretable trigger categories. We manually selected the number of clusters based on the interpretability of the cluster solution.

In 1A, the C-OCS trigger categories are represented by different colored points and shaded areas on the map. Analyzing the twelve clusters, we identified seven as containing cohesive groups of C-OCS triggers. The remaining five clusters contained groups of "Other" responses that, in some cases, may include C-OCS triggers, but mostly reflect associations with the topic (e.g., emotions such as "Loneliness" or "Stress"), current social topics during data collection (e.g., "Vaccination" or "Antivax lies"), or none responses (e.g., "None" or "Nothing"). Although this may risk losing some valid C-OCS triggers, we decided not to include these five clusters in the analysis. We see it as a strength of our LLM-based approach that we can exclude invalid responses, which are common in free-text responses and would otherwise distort our analysis. These invalid responses were a small minority of responses (10.9%; see 1B) and were retrieved considerably later than the valid responses (see 1C), supporting the claim that they are affected by more associative processes.

Considering valid responses, the largest C-OCS trigger cluster, labeled *Surfaces*, accounts for 23.5% of responses and contains triggers such as "Door handle," "Elevator buttons," and "Shopping cart." The second-largest cluster, labeled *Sickness* (18.5% of responses), contains triggers such as "Coughing/Sneezing," "People without masks," and "Corona." The third-largest cluster, labeled *Proximity* (16.2%), contains triggers such as "Public transport," "Shaking hands," and "Crowd." The fourth-largest cluster, labeled *Grime* (10.8%), contains triggers such as "Public toilets," "Dirt," and "Body odor." The fifth-largest cluster, labeled *Cleaning* (10.4%), contains triggers such as "Washing hands," "Showering," and "Cleaning." The sixth-largest cluster, labeled *Places* (7.9%), contains triggers such as "Shopping," "Stores," and "Restaurant." Finally, the seventh-largest cluster, labeled *Food* (7.9%), contains triggers such as "Fruit and vegetables," "Food packaging," and "Eating."

Taken together, our LLM-driven approach successfully mapped the diverse landscape of C-OCS triggers, revealing their semantic organization and providing a foundational, data-driven understanding of the key categories that provoke C-OCS, which has previously been reliant on anecdotal reports.

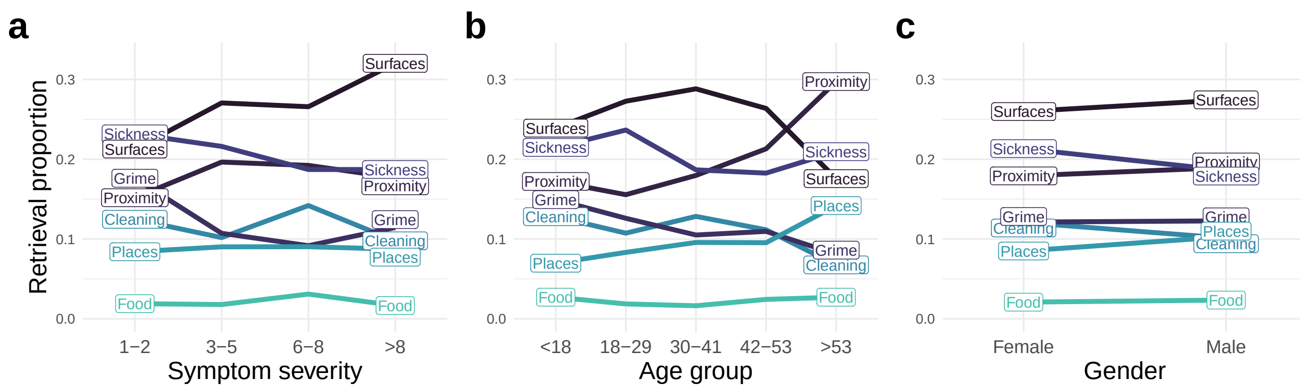

Figure 2

C-OCS trigger moderators. The figure shows the relative frequency of triggers from the 7 trigger categories as a function of C-OCS symptom severity (a), age (b), and gender (c).

Individual differences in C-OCS trigger importance

Our second goal was to understand individual differences in C-OCS trigger importance. We analyzed the relative response frequencies by symptom strength, age, and gender. Symptom strength was determined based on the three items measuring the OCI-R washing/contamination dimension. Figure 2 shows the result for the seven valid categories.

Overall, the importance of categories was relatively stable across the three individual differences variables, as indicated by rank-correlations across levels of $r = .91$ (symptom strength), $r = .87$ (age group), and $r = .93$ (gender). However, some notable deviations emerged. Analyzing the responses from each cluster as a function of all three individual variables simultaneously, we found that *Surfaces* was more prominent among high compared to low-symptom individuals, while the opposite was true for *Sickness* and *Grime*. Furthermore, *Proximity* and *Places* were more prominent among older as compared to younger adults, whereas the opposite was true for *Grime*. No significant differences were observed for gender.

These findings suggest that while the core C-OCS trigger categories identified are broadly relevant, their prominence can exhibit subtle yet potentially meaningful variations related to symptom severity and age. This nuanced understanding contributes to a more detailed picture of how C-OCS triggers manifest across different segments of the population experiencing these symptoms.

Evaluating the potential of trigger substitution

Our third goal was to evaluate the potential of trigger substitution as a means to improve therapeutic approaches within the ERP framework based on semantic similarity. We reasoned that C-OCS triggers that frequently co-occur among the triggers of the same person, while receiving different strength scores, represent attractive targets for substitution in ERP. We further reasoned that semantic similarity, as captured by our LLM, may potentially present a proxy for identifying substitution candidates.

To analyze the potential for trigger substitution, we first developed a simulation-based approach to quantify the co-occurrence propensity of triggers. We simulated ten million synthetic respondents under the assumption of trigger independence and then calculated the propensity by z-standardizing the empirical co-occurrence frequencies using the synthetic ones. This approach corrects for the effect of trigger frequency on co-occurrences, producing a cleaner measure of trigger co-occurrence Goñi et al., 2011; Wulff et al., 2022.

Figure 3A shows the relationship between the trigger co-occurrence propensity for each trigger pair with a minimum co-occurrence frequency of five and the corresponding strength delta, calculated as the absolute differences between the trigger strength scores of the trigger pair. The labels show pairs that are high in either frequency, co-occurrence propensity, or strength delta. The two variables show a mild positive relationship ($r = .20$), with higher co-occurrences going hand-in-hand with higher strength deltas. This pattern implies that there exist trigger pairs that frequently co-occur but do not elicit C-OCS with the same strength. For instance, "Showering" co-occurred frequently with both "Cleaning" and "Washing hands" but elicited substantially smaller strength ratings. Such trigger pairs could represent effective candidates for substitution in ERP.

To evaluate the potential for using semantic similarity as a proxy for trigger substitution, we analyzed its connection to co-occurrence propensity and strength delta. As can be seen in Figure 3B-C, semantic similarity was positively related to co-occurrence propensity ($r = .35$, $p < .001$) and negatively related to strength delta ($r = -.07$, $p = .445$), implying that high-similarity triggers are more frequently named by the same person and are more similar in strength. However, cru-

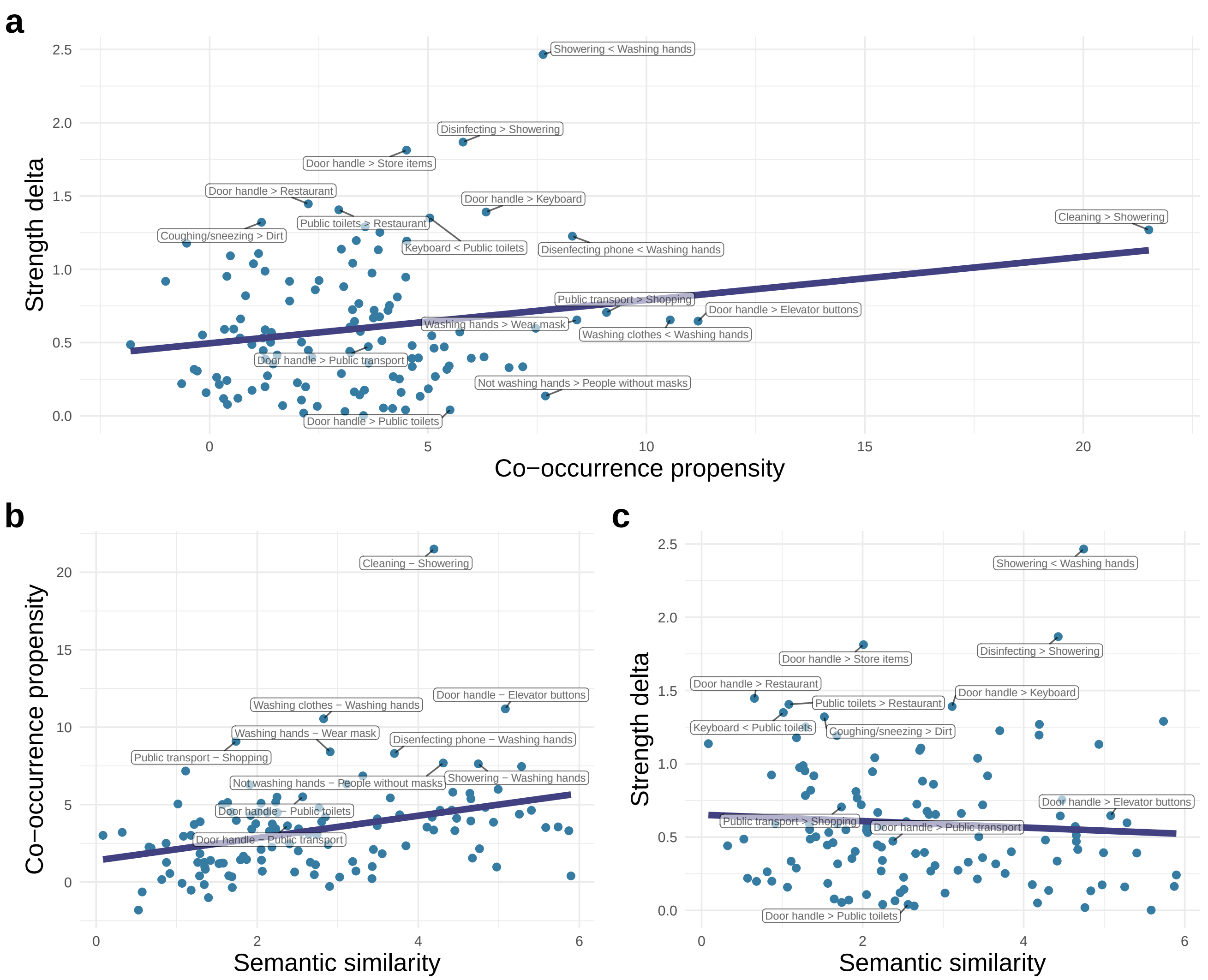

Figure 3

Potential for C-OCS substitution. The figure shows the relationships between semantic trigger similarity, co-occurrence propensity, and strength delta. Panel a shows the relationship between co-occurrence propensity and trigger strength delta (difference in trigger strength among pairs). Panels b and c show the relationship of semantic similarity to co-occurrence propensity and strength delta, respectively.

cially, only the former connection is substantially and significantly different from zero. This suggests that semantic similarity represents a useful proxy to identify triggers that are conceptually related and thus likely to co-occur, even if they do not always elicit the same intensity of C-OCS.

All in all, these analyses demonstrate that semantically similar C-OCS triggers often co-occur within an individual's reported triggers, yet can significantly differ in their reported intensity. This finding highlights the potential of using semantic similarity as an empirical guide to select and sequence stimuli in exposure therapy, thereby paving the way for more nuanced and potentially more effective treatment strategies aimed at improving extinction generalization for OCS.

Discussion

Leveraging a large language model (LLM) to analyze free-text responses about triggers of contamination-related obsessive-compulsive symptoms (C-OCS), our study demonstrates that these triggers—often considered highly idiosyncratic—can be meaningfully grouped into distinct categories. Specifically, our LLM-driven mapping and clustering identified seven primary C-OCS trigger categories: *Surfaces*, *Sickness*, *Proximity*, *Grime*, *Cleaning*, *Places*, and

Food. These categories span various semantic domains and are linked to different ecological contexts of individuals. Unlike previous research that often relied on predefined stimuli rated for intensity (Mataix-Cols et al., 2009), our approach gathered free-text responses from participants about their personal triggers, which were then rated for intensity. Our findings extend descriptive work such as that of Rachman et al. (Rachman, 2004), who identified heterogeneous contamination triggers and grouped them broadly. While some of our empirically identified categories show conceptual overlap with theirs (e.g., *Sickness* with diseases/germs, *Grime* with dirt/pollution), our study uniquely quantifies their prevalence and semantic organization based on a large dataset. The valid trigger categories reported the most frequently in our sample were *Surfaces* (23.5%), followed by *Sickness* (18.5%) and *Proximity* (16.2%), highlighting the key areas of concern for individuals with C-OCS.

Examining individual differences revealed a nuanced picture regarding the prominence of these trigger categories. The influence of the severity of the OCS symptoms, for example, showed that while the overall importance of the cluster rank was relatively stable, supporting a dimensional approach to OCS (Abramowitz et al., 2014; García-Soriano et al., 2011), some significant deviations emerged. *Surfaces* were more prominent for individuals with higher symptom severity, whereas *Sickness* and *Grime* were more prominent for those with lower severity, suggesting that the salience of certain trigger categories may shift with symptom intensity. Age-related variations were also notable: triggers related to *Proximity* and *Places* were more prominent among older adults, while *Grime* was more so among younger adults. These age-based differences might reflect genuine developmental shifts in C-OCS trigger prominence or could be partially influenced by the data collection period during the COVID-19 pandemic, which heightened concerns about public spaces and proximity, particularly for older individuals (Andrighetto et al., 2024). In contrast to some existing literature suggesting gender differences in C-OCS categories (Bogetto et al., 1999; Labad et al., 2008), our analysis did not reveal significant differences in the prominence of the identified trigger categories between genders. Together, these findings underscore that while core trigger categories are broadly relevant, their specific importance can vary across different segments of the population experiencing C-OCS, offering potential avenues for tailoring therapeutic interventions.

A clinically significant finding is that semantically similar triggers, which frequently co-occur within an individual's reported triggers, can differ notably in their perceived intensity. Our analysis showed a positive relationship between semantic similarity and co-occurrence propensity, but no substantial relationship between semantic similarity and the difference in trigger strength (strength delta). This suggests that individuals often report multiple semantically related triggers, but these triggers do not necessarily elicit the same level of distress. This insight has direct implications for exposure with response prevention (ERP): if future research confirms a robust generalization of extinction learning between semantically similar stimuli, as suggested by studies on associative learning in OCD (Cooper & Dunsmoor, 2021), then lower intensity triggers could potentially be used during ERP to facilitate generalization to related, higher intensity triggers. Such an approach would potentially enhance treatment tolerability and reduce dropout rates, as exposure to highly intense triggers is a known barrier to ERP engagement and completion (Ong et al., 2016; Öst et al., 2015; Rosa-Alcázar et al., 2008).

The methodological approach employed in this study, particularly its reliance on open LLM technologies, holds considerable promise for broader applications. The use of fine-tuned models available through open-source platforms reduces barriers to entry and encourages the adaptation of these powerful tools for various clinical research questions (Hussain et al., 2024; Wulff et al., 2024). The use of LLMs to systematically analyze and map idiosyncratic, free-text descriptions of symptom triggers is not limited to C-OCS (Aeschbach et al., 2025). This data-driven technique could be readily adapted to explore the trigger landscapes of other OCD presentations, such as checking compulsions (e.g., identifying common fears or behavior excesses), or symmetry and ordering concerns. Beyond OCD, this methodology could be invaluable in understanding triggers and expressions of symptoms in other mental health conditions such as PTSD (characterizing trauma-related signals), anxiety disorders (mapping phobic stimuli or content of worries), or even depressive disorders (identifying patterns in automatic negative thoughts). Furthermore, in general medicine, analyzing patient-reported outcomes or qualitative descriptions of symptoms through LLMs could help in understanding complex, multifaceted conditions where individual experiences vary widely, such as chronic pain or autoimmune diseases, thereby fostering a more nuanced, patient-centered approach to clinical research and practice.

In terms of limitations, study participants were not a fully representative sample of the Swiss population, having initially been recruited for a study on stress and behavioral changes during the COVID-19 pandemic. This might introduce a selection bias towards individuals more attuned to mental health issues or experiencing higher stress. Data collection during the pandemic may have also influenced the prominence of certain contamination triggers, particularly those related to infectious diseases (e.g., the *Sickness* cluster). Although participants were asked to report triggers beyond official hygiene measures, the prevailing context could have heightened awareness of such triggers. This aligns with observed increases in contamination-related compulsive symptoms during that period (Otte et al., 2025). Therefore, while these findings are highly relevant, replication in diverse representative samples and in a post-pandemic context would strengthen their generalizability.

All in all, our LLM-based approach offers a more fine-grained and data-driven understanding of the semantic landscape of C-OCS triggers. This provides crucial insights that can inform the refinement of treatment strategies, particularly ERP. By moving beyond predefined or purely anecdotal categorizations to a comprehensive semantic mapping, we extend current knowledge and enable further research into the generalization of extinction. This is in line with a recent systematic review on extinction generalization following exposure in anxiety disorders, which emphasizes that clinical progress depends on a more nuanced understanding of stimulus complexity and the factors driving generalization (Kodzaga et al., 2025). Understanding these semantic relationships is vital for advancing research into the associative learning mechanisms that underpin successful OCS treatment and for guiding the development of more personalized and potentially more effective therapeutic interventions.

Conclusions

We have demonstrated that recent advances in natural language processing, specifically large language models, can make a significant contribution to psychopathological research, particularly in understanding obsessive-compulsive symptoms. Our analysis revealed that C-OCS triggers, often perceived as highly idiosyncratic, can be systematically mapped and grouped into meaningful semantic categories based on individuals' free-text descriptions. These findings not only provide a richer, empirically grounded taxonomy of C-OCS triggers but also enable further investigation into their role in associative learning experiments. Ultimately, this data-driven approach to understanding symptom triggers paves the way for enhancing the efficacy and personalization of OCD treatment, and offers a versatile methodology for exploring similar phenomena across a range of psychological and medical conditions.

Methods

Design, Setting, and Participants

The study utilized an anonymous online survey to investigate OCS triggers related to washing and contamination symptoms. Participation was offered as a follow-up to the Swiss Corona Stress Study (Survey 4) (Freytag et al., 2022), which was conducted between November 16 and 28, 2021, to assess the COVID-19 pandemic's impact on mental well-being in Switzerland. Participants for the Swiss Corona Stress Study were recruited from all Swiss regions via media releases from the University of Basel, local newspapers, radio interviews, and social media. Inclusion criteria for the Swiss Corona Stress Study were Swiss residency, age over 14 years, and no participation in previous iterations of the study (Surveys 1-3) (de Quervain et al., 2020a, 2020b). As the survey was anonymous, formal ethics approval was not deemed necessary. All participants provided written informed consent prior to participation and received no monetary compensation. From the 11,167 individuals in the Swiss Corona Stress Study (Survey 4) (Freytag et al., 2022), 3,615 voluntarily proceeded to a survey on compulsions and obsessions. Of these, 1,213 participants (80.6

Procedure

The anonymous online survey was structured in two parts. The first part comprised the Swiss Corona Stress Study (Survey 4) (Freytag et al., 2022), and the second was an optional, independent survey focusing on compulsions, obsessions, and triggers within the washing/contamination dimension of OCS. The first part took approximately 15 minutes to complete, and the second part took about 5 minutes. Participants accessed the survey via the website www.coronastress.ch, available in German, French, and Italian. The survey was implemented using SoSci Survey software (Leiner, 2019), which recorded the participation date but not IP addresses or timestamps. Completion required responses to all items.

Following general study information and informed consent, the Swiss Corona Stress Study collected sociodemographic data (gender, age, nationality, education, profession, self-declared psychiatric conditions) and information on stress and behavioral changes during the pandemic. Upon completing the first part, participants were invited to continue with the survey on compulsions and obsessions. These were defined as: "compulsions and obsessions can be activities or thoughts that you repeat or think about over and over again, e.g., washing very often, checking things or counting, even if you do not want to or if they seem excessive or irrational to you."

First, current obsessive-compulsive symptoms in the washing/contamination dimension were assessed using three items from the Obsessive-Compulsive Inventory-Revised (OCI-R). Second, participants were asked to list their "most common triggers for intrusive and distressing thoughts about germs and contamination and/or recurring behaviors or rituals to prevent contamination (e.g. washing, cleaning, showering)" in a free-text format and to indicate the intensity of distress caused by each trigger. At the study's conclusion, participants received automated stress management recommendations based on their responses in the initial survey part (as part of the Swiss Corona Stress Study, Survey 4). Additionally, general information on OCS, treatment options, and contact details for professional support were provided.

Measures

Symptom severity was assessed using the washing/contamination subscale of the OCI-R (items 5, 11, and 17). Each item was rated on a 5-point scale from 0 (not at all) to 4 (extremely). Scores range from 0 to 12, with higher scores indicating more severe symptoms. Validated German (Gönner et al., 2007), French (Zermatten et al., 2006), and Italian (Marchetti et al., 2010) versions were used. Individual free-text responses regarding the "most common triggers" were used to identify C-OCS triggers. To prepare these responses for analysis, the first author manually translated the German, French, and Italian entries into English, with the second author validating these translations. The intensity of each trigger was determined from participants' ratings of distress caused by their individual triggers, on a scale from 0 (none) to 10 (maximum).

Creating the C-OCS trigger map

To create a map of trigger categories, we followed a multi-step process. First, we embedded the reported C-OCS triggers using a fine-tuned embedding model (specifically, dwulff/mpnet-cocs on Hugging Face), leveraging open-source platforms that promote accessibility and reproducibility in model sharing (Hussain et al., 2024), to capture their semantic organization. This model was fine-tuned using 20 thousand ratings generated by Llama-3.3-70b-Instruct Grattafiori et al., 2024. Llama was instructed as follows: "We have asked laypeople to name their C-OCD triggers. Your task is to evaluate on a scale from 0 to 100 whether two respondents named exactly the same trigger. Evaluate this in the context of contamination OCD. Briefly reason through your answer. Then return the answer as a number between 0 (fully different) and 100 (fully same). Strictly use the format Answer=[evaluation]."

Based on this model, C-OCS triggers were then grouped into semantically cohesive groups. The average embedding of each group was subsequently projected into a two-dimensional map using the PaCMAP dimensionality reduction algorithm. Each group was labeled using Llama-3.3-70b-Instruct, using the following prompt: "Your task is to provide a short label (1 to 3 words) for the following list of C-OCD trigger responses: List: {trigger_text}. The label should be highly specific, closely capture the elements in the list, and should make sense as a trigger without adding meaning. The best label is often the most frequent element, potentially shortened. Only return the label!"

Finally, this map was partitioned into twelve clusters using hierarchical clustering. The number of clusters was selected manually based on the interpretability of the resulting cluster solution, aiming to identify meaningful trigger categories.

Measuring C-OCS trigger co-occurrence propensity

To identify trigger co-occurrences that exceed chance levels, we employed a simulation-based approach. Specifically, we simulated ten million synthetic respondents. These simulations utilized the inter-participant distribution of response numbers and overall trigger frequencies, under the assumption that these two distributions were independent. Based on this simulated data, we determined the likelihoods of trigger co-occurrences within the responses of the synthetic respondents. This allowed us to establish random expectations (p) for each pair of triggers. We then used these random expectations to calculate a z-scaled co-occurrence score, applying a continuity correction for the binomial distribution. The formula used was $z_{co} = \frac{n-Np}{Np(1-p)}$, where n is the number of observed co-occurrences and N is the total number of possible pairs given the empirical distribution of response numbers.

Data availability

Data and materials are available at https://osf.io/edyg5/.

Code availability

Data and materials are available at https://osf.io/edyg5/.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

DB acknowledges that data collection was carried out within the framework of the Swiss Corona Study, supported by funding from the Transfaculty Research Platform for Molecular and Cognitive Neurosciences, University of Basel and funds from the research fund University Basel granted to DB. The authors thank Nathalie Schicktanz for her contributions to survey implementation and Thomas Schlitt for data management. DUW acknowledges funding from the German Research Foundation (https://gepris.dfg.de/gepris/projekt/546419617?language=en).

Author information

Dorothée Bentz, Department of Biomedicine, Division of Cognitive Neuroscience and Faculty of Psychology, Clinical Psychology, and Translational Psychotherapy Research, University of Basel, Missionsstrasse 62A, 4055 Basel, Switzerland

Dirk U. Wulff, Center for Adaptive Rationality, Max Planck Institute for Human Development, 14195 Berlin, and Cognitive and Decision Sciences, Faculty of Psychology, University of Basel, Missionsstrasse 64A, 4055 Basel, Switzerland.

Contributions

DB and DUW designed the study. DB supervised survey implementation and data collection. DUW analyzed and visualized the data. DB and DUW wrote the manuscript.

References

Abramowitz, J., Fabricant, L., Taylor, S., Deacon, B., McKay, D., & Storch, E. (2014). The relevance of analogue studies for understanding obsessions and compulsions. *Clinical Psychology Review*, *34*(3), 206–217. https://doi.org/https://doi.org/10.1016/j.cpr.2014.01.004

Adam, Y., Meinlschmidt, G., Gloster, A., & Lieb, R. (2012). Obsessive-compulsive disorder in the community: 12-month prevalence, comorbidity and impairment. *Social Psychiatry and Psychiatric Epidemiology*, *47*(3), 339–349. https://doi.org/https://doi.org/10.1007/s00127-010-0337-5

Aeschbach, S., Mata, R., & Wulff, D. U. (2025). Mapping mental representations with free associations: A tutorial using the r package associator. *Journal of Cognition*, *8*(1), 3. https://doi.org/https://doi.org/10.5334/joc.407

Andrighetto, G., Szekely, A., Guido, A., Gelfand, M., Abernathy, J., Arikan, G., & . . . Eriksson, K. (2024). Changes in social norms during the early stages of the covid-19 pandemic across 43 countries. *Nature Communications*, *15*, 1436. https://doi.org/https://doi.org/10.1038/s41467-024-44999-5

Angst, J., Gamma, A., Endrass, J., Goodwin, R., Ajdacic, V., Eich, D., & Rössler, W. (2004). Obsessive-compulsive severity spectrum in the community: Prevalence, comorbidity, and course. *European Archives of Psychiatry and Clinical Neuroscience*, *254*(3), 156–164. https://doi.org/https://doi.org/10.1007/s00406-004-0459-4

APA. (2013). *Diagnostic and statistical manual of mental disorders*. American Psychiatric Association.

Binz, M., Alaniz, S., Roskies, A., Aczel, B., Bergstrom, C. T., Allen, C., Schad, D., Wulff, D., West, J. D., Zhang, Q., et al. (2025). How should the advancement of large language models affect the practice of science? *Proceedings of the National Academy of Sciences*, *122*(5), e2401227121. https://doi.org/https://doi.org/10.1073/pnas.2401227121

Bogetto, F., Venturello, S., Albert, U., Maina, G., & Ravizza, L. (1999). Gender-related clinical differences in obsessive-compulsive disorder. *European Psychiatry: The Journal of the Association of European Psychiatrist*, *14*, 434–441. https://doi.org/https://doi.org/10.1016/s0924-9338(99)00224-2

Boyle, S., Roche, B., Dymond, S., & D, H. (2016). Generalisation of fear and avoidance along a semantic continuum. *Cognition and Emotion*, *30*, 340–352. https://doi.org/https://doi.org/10.1080/02699931.2014.1000831

Chandran, D., Robbins, D., Chang, C., Shetty, H., Sanyal, J., & Downs, R., J . . . Hayes. (2019). Use of natural language processing to identify obsessive compulsive symptoms in patients with schizophrenia, schizoaffective disorder or bipolar disorder. *Scientific Reports*, *9*, 14146. https://doi.org/https://doi.org/10.1038/s41598-019-49165-2

Cooper, S., & Dunsmoor, J. (2021). Fear conditioning and extinction in obsessive-compulsive disorder: A systematic review. *Neuroscience Biobehavioral Reviews*, *129*, 75–94. https://doi.org/https://doi.org/10.1016/j.neubiorev.2021.07.026

Cullen, A., Dowling, N., Segrave, R., Carter, A., & Yücel, M. (2021). Exposure therapy in a virtual environment: Validation in obsessive compulsive disorder. *Journal of Anxiety Disorders*, *80*. https://doi.org/https://doi.org/10.1016/j.janxdis.2021.102404

de Quervain, D., Aerni, A., Amini, E., Bentz, D., Coynel, D., Gerhards, C., Fehlmann, B., Freytag, V., Papassotiropoulos, A., Schicktanz, N., Schlitt, T., Zimmer, A., & Zuber, P. (2020a). The swiss corona stress study. *OSF Preprints*. https://doi.org/https://doi.org/10.31219/osf.io/jqw6a

de Quervain, D., Aerni, A., Amini, E., Bentz, D., Coynel, D., Gerhards, C., Fehlmann, B., Freytag, V., Papassotiropoulos, A., Schicktanz, N., Schlitt, T., Zimmer, A., & Zuber, P. (2020b). The swiss corona stress study: Second pandemic wave. *OSF Preprints*. https://doi.org/https://doi.org/10.31219/osf.io/6cseh

Dunsmoor, J., White, A., & LaBar, K. (2011). Conceptual similarity promotes generalization of higher order fear learning. *Learning Memory*, *18*, 156–160. https://doi.org/10.1101/lm.2016411

Feuerriegel, S., Maarouf, A., Bär, D., Geissler, D., Schweisthal, J., Pröllochs, N., & . . . Van Bavel, J. (2025). Using natural language processing to analyse text data in behavioural science. *Nature Reviews Psychology*, *4*, 96–111. https://doi.org/https://doi.org/10.1038/s44159-024-00392-z

for Mental Health (UK). Leicester (UK): British Psychological Society (UK), N. C. C. (2006). Nice clinical guidelines, no. 31: Obsessive-compulsive disorder: Core interventions in the treatment of obsessive-compulsive disorder and body dysmorphic disorder. *Obsessive-Compulsive Disorder: Core Interventions in the Treatment of Obsessive-Compulsive Disorder and Body Dysmorphic Disorder*.

Freytag, V., Schicktanz, N., Coynel, D., Schlitt, T., Amini, E., Papassotiropoulos, A., & de Quervain, D. (2022). The swiss corona stress study: Long covid symptoms in relation to stress and depressive symptoms. *OSF Preprints*. https://doi.org/https://doi.org/10.31219/osf.io/2x4pg

García-Soriano, G., Belloch, A., Morillo, C., & Clark, D. (2011). Symptom dimensions in obsessive–compulsive disorder: From normal cognitive intrusions to clinical obsessions. *Journal of Anxiety Disorders*, *25*, 474–482. https://doi.org/https://doi.org/10.1016/j.janxdis.2010.11.012

Goñi, J., Arrondo, G., Sepulcre, J., Martincorena, I., Vélez de Mendizábal, N., Corominas-Murtra, B., Bejarano, B., Ardanza-Trevijano, S., Peraita, H., Wall, D. P., et al. (2011). The semantic organization of the animal category: Evidence from semantic verbal fluency and network theory. *Cognitive processing*, *12*, 183–196. https://doi.org/https://doi.org/10.1007/s10339-010-0372-x

Gönner, S., Leonhart, R., & Ecker, W. (2007). The german version of the obsessive-compulsive inventory-revised: A brief self-report measure for the multidimensional assessment of obsessive-compulsive symptoms. *Psychotherapie, Psychosomatik, medizinische Psychologie*, *57*(9–10), 395–404. https://doi.org/https://doi.org/10.1055/s-2007-970894

Grattafiori, A., Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Vaughan, A., et al. (2024). The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*. https://doi.org/https://doi.org/10.48550/arXiv.2407.21783

Hall, G. (1996). Learning about associatively activated stimulus representations: Implications for acquired equivalence and perceptual learning. *Animal Learning Behavior*, *24*, 233–255. https://doi.org/https://doi.org/10.3758/BF03198973

Haslam, N., Williams, B., Kyrios, M., McKay, D., & Taylor, S. (2005). Subtyping obsessive-compulsive disorder: A taxometric analysis. *Behavior Therapy*, *36*, 381–391. https://doi.org/https://doi.org/10.1016/S0005-7894(05)80120-0

Hussain, Z., Binz, M., Mata, R., & Wulff, D. U. (2024). A tutorial on open-source large language models for behavioral science. *Behavior Research Methods*, *56*(8), 8214–8237. https://doi.org/https://doi.org/10.3758/s13428-024-02455-8

Jacoby, R., & Abramowitz, J. (2016). Inhibitory learning approaches to exposure therapy: A critical review and translation to obsessive-compulsive disorder. *Clinical Psychology Review*, *49*, 28–40. https://doi.org/https://doi.org/10.1016/j.cpr.2016.07.001

Karno, M., Golding, J., Sorenson, S., & Burnam, M. (1988). The epidemiology of obsessive-compulsive disorder in five us communities. *Archives of General Psychiatry*, *45*, 1094–1099. https://doi.org/10.1001/archpsyc.1988.01800360042006

Kodzaga, I., Heistermann, J, & Zlomuzica, A. (2025). Generalization of exposure therapy: Systematic review and recommendations for future research. *Behaviour Research and Therapy*, *190*, 104751. https://doi.org/https://doi.org/10.1016/j.brat.2025.104751

Labad, J., Menchon, J., Alonso, P., Segalas, C., Jimenez, S., Jaurrieta, N., & … Vallejo, J. (2008). Gender differences in obsessive–compulsive symptom dimensions. *Depression and Anxiety*, *25*, 832–838. https://doi.org/https://doi.org/10.1002/da.20332

Leiner, D. J. (2019). *Sosci survey, version version 3.1.06* (tech. rep.). https://www.soscisurvey.de.

Marchetti, I., Chiri, L., Ghisi, M., & Sica, C. (2010). Obsessive-compulsive inventory-revised (oci-r): Presentazione e indicazioni di utilizzo nel contesto italiano. *Psicoterapia Cognitiva Comportamentale*, *16*(1), 69–84. https://doi.org/http://hdl.handle.net/1854/LU-1148230

Mataix-Cols, D., Cullen, S., Lange, K., Zelaya, F., Andrew, C., Amaro, E., Brammer, M., Williams, S., Speckens, A., & ML, P. (2003). Neural correlates of anxiety associated with obsessive-compulsive symptom dimensions in normal volunteers. *Biological Psychiatry*, *53*, 482–93. https://doi.org/https://doi.org/10.1016/S0006-3223(02)01504-4

Mataix-Cols, D., Lawrence, N., Wooderson, S., Speckens, A., & Phillips, M. (2009). The maudsley obsessive-compulsive stimuli set: Validation of a standardized paradigm for symptom-specific provocation in obsessive-compulsive disorder. *Psychiatry Research*, *168*, 238–241. https://doi.org/https://doi.org/10.1016/j.psychres.2008.05.007

Mataix-Cols, D., Rosario-Campos, M., & Leckman, J. (2005). A multidimensional model of obsessive-compulsive disorder. *Am J Psychiatry*, *162*, 228–38. https://doi.org/https://doi.org/10.1176/appi.ajp.162.2.228

Ong, C., Clyde, J., Bluett, E., Levin, M., & Twohig, M. (2016). Dropout rates in exposure with response prevention for obsessive-compulsive disorder: What do the data really say? *Journal of Anxiety Disorders*, *40*, 8–17. https://doi.org/https://doi.org/10.1016/j.janxdis.2016.03.006

Öst, L.-G., Havnen, A., Hansen, B., & G, K. (2015). Cognitive behavioral treatments of obsessive–compulsive disorder. a systematic review and meta-analysis of studies published 1993–2014. *Clinical Psychology Review*, *40*, 156–169. https://doi.org/https://doi.org/10.1016/j.cpr.2015.06.003

Otte, J., Schicktanz, N., & Bentz, D. (2025). Impact of the covid-19 pandemic on obsessive-compulsive symptoms in the swiss general population. *Frontiers in Psychology*, *14*, 1071205. https://doi.org/https://doi.org/10.3389/fpsyg.2023.1071205

Pittig, A., Van Den Berg, L., & Vervliet, B. (2016). The key role of extinction learning in anxiety disorders: Behavioral strategies to enhance exposure-based treatments. *Current Opinion in Psychiatry*, *29*, 39–47. https://doi.org/http://journals.lww.com/00001504-201601000-00008

Rachman, S. (2004). Fear of contamination. *Behaviour Research and Therapy*, *42*, 1227–1255. https://doi.org/https://doi.org/10.1016/j.brat.2003.10.009

Rasmussen, S., & Eisen, J. (1992). The epidemiology and clinical features of obsessive compulsive disorder. *Psychiatric Clinics of North America*, *15*, 743–758. https://doi.org/https://doi.org/10.1016/S0193-953X(18)30205-3

Rosa-Alcázar, A., Sánchez-Meca, J., & Gómez-Conesa, F., A ADN Marín-Martínez. (2008). Psychological treatment of obsessive–compulsive disorder: A meta-analysis. *Clinical Psychology Review*, *28*, 1310–1325. https://doi.org/https://doi.org/10.1016/j.cpr.2008.07.001

Shepard, R. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323. https://doi.org/10.1126/science.3629243

Shepard, R. (2004). How a cognitive psychologist came to seek universal laws. *Psychonomic Bulletin Review*, *11*, 1–23. https://doi.org/https://doi.org/10.3758/BF03206455

Simon, D., Kischkel, E., Spielberg, R., & Kathmann, N. (2012). A pilot study on the validity of using pictures and videos for individualized symptom provocation in obsessive–compulsive disorder. *Psychiatry Research*, *198*, 81–88. https://doi.org/https://doi.org/10.1016/j.psychres.2011.12.022

Sousa, M., Costa, A., Almeida, A., Soriano-Mas, C., Silva Moreira, P., & Morgado, P. (2024). Symptom provocation in obsessive-compulsive disorder: Validation of the braga obsessive compulsive image set (bocis). *Journal of Psychiatric Research*, *175*, 144–152. https://doi.org/10.1016/j.jpsychires.2024.04.046

Vervliet, B., Kindt, M., Vansteenwegen, D., & Hermans, D. (2010). Fear generalization in humans: Impact of verbal instructions. *Behaviour Research and Therapy*, *48*, 38–43. https://doi.org/https://doi.org/10.1016/j.brat.2009.09.005

Vervliet, B., Vansteenwegen, D., Baeyens, F., Hermans, D., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a stimulus change after extinction. *Behaviour Research and Therapy*, *43*, 357–371. https://doi.org/https://doi.org/10.1016/j.brat.2004.02.005

Vervoort, E., Vervliet, B., Bennet, M., & F, B. (2014). Generalization of human fear acquisition and extinction within a novel arbitrary stimulus category. *PLoS ONE*, *9*, e96569. https://doi.org/https://doi.org/10.1371/journal.pone.0096569

Wulff, D. U., Hills, T. T., & Mata, R. (2022). Structural differences in the semantic networks of younger and older adults. *Scientific Reports*, *12*(1), 21459. https://doi.org/https://doi.org/10.1038/s41598-022-11698-4

Wulff, D. U., Hussain, Z., & Mata, R. (2024). The behavioral and social sciences need open llms. https://doi.org/https://doi.org/10.31219/osf.io/ybvzs

Zermatten, A., Van der Linden, M., Jermann, F., & Ceschi, G. (2006). Validation of a french version of the obsessive–compulsive inventory-revised in a non-clinical sample. *European Review of Applied Psychology*, *56*(3), 151–5. https://doi.org/https://doi.org/10.1016/j.erap.2005.07.003