ABSTRACT

Title of Thesis: Improving Selection of Analogical Inspirations with Chunking and Recombination

Arvind Srinivasan
Masters of Science, 2023

Thesis Directed by: Assistant Professor, Dr. Joel Chan
College of Information Studies

Innovation is vital in various fields, and analogical thinking is a powerful tool for generating creative solutions to complex problems. However, recognizing analogies can be time-consuming, and successful recognition doesn't guarantee their adoption in innovation. In this thesis, A novel computational support system for analogical innovation is proposed that employs the cognitive mechanisms for chunking and recombination as mediums of interaction. Chunking involves identifying and extracting meaningful chunks or segments from a design problem into interactive tiles called *magnets* while recombination involves combining these *magnets* to generate insightful questions that elicit divergent thinking. In this way, the proposed system aims to streamline the process of recognizing and selecting analogical inspirations for innovation while avoiding premature rejection and design fixation. To evaluate the effectiveness of the system, a within-subjects study involving 23 participants was conducted, comparing the proposed interface with a baseline. The study found that using chunking and recombination as interactive mecha-

nisms helped prevent premature rejection of useful analogical leads, resulting in 4 times fewer ignored analogical leads. Participants were also found to make 12 times fewer changes to their decisions, given a minor increment in processing time in the order of 1.5 minutes. Overall, these results suggest that our proposed intervention is an effective tool for facilitating the selection of beneficial analogies, fostering analogical innovation through computational support.

Improving Selection of Analogical Inspirations with Chunking and Recombination

by

Arvind Srinivasan

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Masters in Human Computer Interaction
2023

Advisory Committee:
Professor Joel Chan, Chair
Professor Niklas Elmqvist
Professor Susannah Paletz

Contents

List of Tables

List of Figures

Chapter 1: Introduction

Innovation is crucial for progress and growth in various domains, from technology and science to business and the arts. To stay ahead of the curve and adapt to ever-changing circumstances, individuals and organizations need to generate innovative solutions to complex problems. Analogical thinking is a powerful tool (Holyoak and Thagard, 1997) for developing innovative solutions (Gassmann and Zeschky, 2008), involving drawing connections between seemingly disparate domains to generate new ideas and approaches by identifying their structural similarities (Gentner, 1983; Holyoak and Koh, 1987; Holyoak and Thagard, 1989; Gentner and Markman, 1997).

Far analogies in particular, have been instrumental in some of the most significant scientific breakthroughs throughout history. For example, Kepler's discovery of the laws of planetary motion (Gentner et al., 1997) and universal laws of gravitation and centrifugal force by Isaac Newton was inspired by the analogy between the fall of an apple and the moon's orbit around the earth (Cohen and Stachel, 1979; Herivel, 1960) fundamentally changed the way we perceived the world. Similarly, groundbreaking advancements in technology and science, such as Mestral (1961)'s burr-inspired invention of the "*separable fastening device*" (later popularized as Velcro) and the analogy-inspired design of early computer networks based on the human brain (Winograd and Flores, 2008) were made possible thanks to the discovery of relationships between far analogies that span multiple domains and disciplines.

However, analogical innovation is a time-consuming process. Evidence suggests that memory retrieval is highly sensitive to surface similarity, showing a clear preference for near, within-domain analogies that share object attributes over far, structurally similar analogies that share object relations (Gick and Holyoak, 1980; Hammond et al., 1991; Gentner et al., 1993; Keane, 1997). Furthermore, analogical processing can prove to be a demanding cognitive task, as it can quickly exhaust working memory resources, particularly when multiple relations need to be processed at once (Halford et al., 2005). There have been cases when this cognitive demand took a toll on innovation, leading to sub-optimal outcomes. One such example is the development of the first microwave oven, which was inspired by the magnetron technology in radar. While the microwave oven was a significant innovation, the analogy between radar and cooking did not result in optimal outcomes. It took several years to develop it into a viable product (Gavetti et al., 2005).

Such examples underscore the challenges involved in recognizing and leveraging far-domain analogies for creative problem-solving, highlighting the need to develop strategies for overcoming these obstacles. By considering the limitations of current approaches to analogical processing, it is clear that a new and more effective approach is needed to better support recognition and adoption of far-domain analogies. Cognitive scientists have attempted to address this problem either by offering some form of assistance either by means of instruction (Gick and Holyoak, 1983; Richland and McDonough, 2010; Kokkalis et al., 2013) or by presenting some form of non-overgeneralized (Spiro et al., 1988) abstracted representation of the problem and corresponding analogical examples to highlight their relationships through underlying functional structures (Yu et al., 2014; Kang et al., 2022) to help make beneficial analogical leads more cognitively accessible (Stapel and Winkielman, 1998) and facilitate their "*exaptation*" (Mastrogiorgio and Gilsing,

2016) and transfer across domains. Several computational models and systems exploring various techniques of abstraction and representation were also proposed to facilitate this process (Gentner, 1983; Kang et al., 2022; Linsey et al., 2012; Fu et al., 2013; Chan et al., 2018).

However, successful recognition of beneficial analogical leads alone is not enough to result in their adoption when solving a problem (Hesse and Klecha, 1990). Several competing factors such as expertise (Novick, 1988; Atilola et al., 2016), presence of usable anchors (Brown and Clement, 1989), optimal representations (Gentner, 1983; Duncker, 1945; Glucksberg and Weisberg, 1966; Frank and Ramscar, 2003) and diversity of solutions (Kang et al., 2022) have been found to affect the premature rejection of potential leads ultimately leading to design fixation (Jansson and Smith, 1991; Agogué et al., 2014; Leahy et al., 2020). While several studies have explored the impact this fixation has on the quality of ideas and suggested ways to mitigate the same (Atilola et al., 2016; Leahy et al., 2020; Ward, 1994), little focus is given to their impact on the selection of beneficial analogical leads. In addition, while many studies have identified the cognitive effects of chunking on thinking, learning, memory and comprehension (Knoblich et al., 1999; Petty et al., 2001; Wu et al., 2017; Thalmann et al., 2019), few studies have explored their applicability as a potential visual mechanism (Linsey et al., 2012) to facilitate analogical adoption. Particularly, while there have been studies exploring the use of Generative Pretrained Models (Brown et al., 2020) as a potential tool to enable exploration of the analogical solution space, there is a lack in studies exploring them as a generative mechanism building upon cognitive factors such as recombination to assist in the selection of inspiring analogical leads.

Thus, in this thesis, I propose a system that marries together the chunking as a visual representation with assisted recombination as a generative mechanism. Through this proposed system, I attempt to explore the question: **How might chunking and recombination mechanisms in-**

fluence people's decision-making about analogical leads?

In the following section of this thesis, I will expand on the literature in detail, followed by a comprehensive explanation of the proposed system and elaborate on the methodology used to evaluate the same in the Systems and Methods sections respectively. In the subsequent sections, I will attempt to consolidate answers to the aforementioned questions, hoping to address the current gaps in research, limitations and future work to achieve a comprehensive understanding of and facilitate further development in the design of computational support systems to enable effective adoption of analogies across domains to accelerate innovation.

Chapter 2: Literature Review

2.1 Triaging in the Context of Analogical Innovation

Invented by Dominique Jean Larrey, a military surgeon during the Napoleonic Wars, *Triaging* refers to the process of sorting and prioritizing patients, tasks, or issues based on their level of urgency or severity (Skandalakis et al., 2006). It is commonly used in medical settings to determine the order in which patients receive medical attention based on the severity of their condition. During triage, medical professionals assess the patient's condition and assign them to one of several priority levels. The most urgent cases are treated first, while less urgent cases may have to wait for treatment. This process helped medical professionals make the best use of their resources, time, and expertise, and ensured that patients with the greatest need with the highest chances of positive outcome receive prompt attention.

The term, loosely based on the French verb *trier*, means to separate, *sort*, *shift*, or *select*. Taking this word literally and transferring it to the context of analogical thinking, *Triaging* can be defined as the sorting and prioritization of analogies based on their relevance and potential usefulness in solving a particular problem. Just as medical professionals triage patients based on the severity of their condition, individuals engaging in analogical thinking could triage potential analogies based on their level of relevance and usefulness.

2.2 Factors affecting the Triaging Process

However, similar to medical triaging, there can be cases when the relevance of an analogy is either missed, overlooked or prematurely rejected (Mastrogiorgio and Gilsing, 2016), leading to poor outcomes in finding creative solutions for a given problem. Gick and Holyoak (1980)'s experiments to understand the use of analogy in problem-solving processes, particularly the use of far analogies discovered that the effectiveness of the analogy depended on a number of factors, including the level of similarity between the problems (Chan et al., 2011) presented in the analogy and the problem being solved and the participant's knowledge of their relevance to the problem (Kokkalis et al., 2013) Furthermore, retrieval and usage of analogies from memory was found to be facilitated when the participant had already generated their own solution to the initial problem presented in the analogy. These findings indicate that analogical adoption could be facilitated through strategies such as providing hints to consider the analogy and encouraging participants to generate their own solutions before retrieving and applying the analogy.

Notably, researchers have sought to enhance spontaneous analogical retrieval by promoting a more abstract encoding of the base analogs to render them more accessible during later encounters with analogous situations lacking surface similarities with the base analogs (Forbus et al., 1995; Hummel and Holyoak, 1997). A number of interventions have been successful in achieving this, including presenting the base analog together with its abstract schema (Goldstone and Wilensky, 2008) or a second analogous situation (Catrambone and Holyoak, 1989), asking participants for their comparison, and even discussing the base analog with another (Schwartz, 1995). Transfer advantages have also been obtained by removing irrelevant information in the base analog (Goldstone and Sakamoto, 2003), and by replacing domain-specific terms of the base

situation with domain-general ones.

These interventions have proven to be effective in enhancing problem solving, particularly by overcoming the problem of functional fixedness. This concept was first demonstrated in Duncker's experiments (Duncker, 1945), where participants were presented with a candle, a box of thumbtacks, and a book of matches and were asked to attach the candle to the wall in such a way that it could be lit without dripping wax on the floor. Participants were given the box of thumbtacks, the matches, and the candle and were asked for a way to attach the candle to the wall. Many participants struggled to solve the problem as they were fixed on the traditional use of the thumbtacks. However, when there was a change in the presentation of thumbtacks so as to highlight the optimal relations, those who were asked to solve for the problem were found more likely to find a solution. This study demonstrated that *functional fixedness*, or the inability to see an object's potential uses beyond its conventional functions, can hinder problem solving and some form of abstraction can help.

2.3 Chunking as a Visual Representation

Subsequent studies that built upon Duncker's experiments have further explored this concept by adapting the experiment to accommodate for verbal behavior (Glucksberg, 1962). It was found that presenting a household item in terms of its conventional use, such as "writing" or "building," made participants less likely to come up with novel uses. In contrast, when presented in a more general way, such as "object to hold things together," participants were more likely to generate innovative solutions. Other studies have confirmed that visual representations highlighting key materials can have a direct impact on the number of optimal solutions (Frank and Ram-

scar, 2003). In the context of analogical thinking, (Knoblich et al., 1999) describes a ”chunk” as ”patterns that capture recurring constellations of features or components”. Researchers that study analogies have documented how analogical reasoning (Gentner and Markman, 1997) can be used to abstract these “chunks” from examples to guide and enable effective ideation (Gick and Holyoak, 1983; Ball et al., 2004) by facilitating the restructuring of problem representations (Tang et al., 2015), and how particular strategies of conceptual combination, such as the generation of novel emergent features connecting disparate attributes of the examples can facilitate novel integration of examples into new ideas (Wilkenfeld and Ward, 2001). Approaching from a cognitive perspective, these ”chunks” improved creative problem-solving by reducing cognitive load, enhancing recall, sometimes causing primacy effects in highly motivated individuals (Petty et al., 2001), and improving comprehension when they were clustered either visually (Wu et al., 2017; Luo et al., 2006) or semantically (Richardson et al., 2003). Furthermore, the decomposability of chunks was also found to have a proportional influence on problem-solving ability (Knoblich et al., 1999; Zhang and Soergel, 2020). In a study of human expertise involving expert cognition, Chase and Simon (1973) found that experts were able to think ”bigger”, and have more complex thoughts by compressing complex ideas into memory chunks to overcome the limitations of working memory (Thalmann et al., 2019; Cowan, 2000).

These studies support that visualized representations such as chunking can be utilized to break down complicated problems into simple components with visual representations that helps individuals overcome the barrier of functional fixedness to enable the development of creative solutions.

2.4 Questions as a Recombination Mechanism

Prior studies have shown that, while analogical retrieval is a process that is greatly facilitated by the addition of examples, breaking them down into their corresponding schemas enabled better retrieval of the hidden relational information. Forbus et al.; Hummel and Holyoak reasoned that this process of filtering out irrelevant information helped emphasize the relational information. While individual examples included specific details, schemas excluded anything that was not common to all examples, so minor differences between examples were not found to hinder the finding of good structural matches or the success of identifying relational information (Clement and Gentner, 1991; Gentner and Medina, 1998) in analogies. Furthermore, using generic relational words instead of specific terms was found to aid analogical retrieval, since it resulted in encoding that was less tied to specific details and was more transferable (Clement et al., 1994). In addition, aligning nonidentical relational predicates has been suggested to be facilitated by re-representation (Gentner et al., 1993; Kurtz and Loewenstein, 2007), which can assist in identifying and retrieving relevant analogies in complex systems. Earlier studies conducted in learning environments have explored questions as a possible mechanism to facilitate this process of re-representation and recombination; attempting to facilitate the creation of new linkages between unlinked or previously linked components targeting specific applications (Hargadon and Sutton, 1997). Holmquist (2008) proposed a novel approach to brainstorming called bootlegging, which was found to be suitable for use in multidisciplinary settings. The technique involved generating ideas in two groups - one focusing on users and usage situations, and the other on a specific technology or domain. The ideas were then combined randomly to create unexpected and innovative juxtapositions, which were then used as the basis for quick application brainstorming. Following

this, promising ideas were then further developed to create complete scenarios. The study found that bootlegging effectively stimulated creativity without deviating from the target domain, and could be implemented without a skilled facilitator. Similarly, according to Herring et al. (2009), designers were found to utilize examples better when re-appropriating and recombining components of existing solutions to generate novel ideas. To explore this process of re-appropriation at the invention level, Mastrogiorgio and Gilsing (2016) conducted a study that analyzed a large collection of US patents at the invention level. Through this exploration, the authors discovered that the likelihood of exapting increased if the components, or *"recombinants"* were difficult to recombine. However, since recombination is a creative process, there was a need to be able to rapidly explore connections between many ideas to hasten the process (McCrickard et al., 2013; Jeppesen and Lakhani, 2010). In his review on the taxonomy of questions in experiential learning environments, Tofade et al. (2013) suggested that lower-order, convergent questions, that relied on on factual recall of prior knowledge, did not promote deep thinking. Notably, Open-ended questions (or *exploratory* questions) were found to increase cognitive flexibility and promote engagement (Li and Li, 2009). A recent 2016 study (Siemon et al., 2016) implemented a creativity support system that semi-automatically generated external stimuli in the form of questions that represented the core issues of a task. The experiment the authors conducted showed that participants who were exposed to the semi-automated questions produced better and more versatile ideas than individuals that were not exposed to these questions. The authors utilized Natural Language Processing mechanisms such as tokenization, normalization, and extraction of synonyms from Vector Space Models to automatically generate word stimuli, which were then manually selected by a task generator to form new questions. This was the closest study that implemented automated-question generation as a mechanism for eliciting creative ideation outcomes through

an interface. However, this thesis focuses on utilizing Generative Pretrained Models instead of manual task-generation to generate insightful questions that boost the creative ideation process while seeking to understand their impact on the triaging and selection of analogical leads.

2.5 Computational Systems for Triaging Analogies

There have been previous systems that have allowed innovators to benefit from analogies utilizing mechanisms such as "*chunking*" and "*recombination*" albeit in isolation. For instance, Hope et al. (2017) and Chan et al. (2018) combined crowdsourcing and recurrent neural networks to extract "purpose" and "mechanism" chunks from product descriptions to facilitate analogical retrieval. By conducting ideation experiments over their proposed analogical search engine, they found that analogies retrieved using this method significantly increased people's likelihood of generating novel and creative ideas compared to analogies retrieved by traditional methods. In a similar study conducted with engineering students, a word-tree method of brainstorming was proposed to produce innovative ideas and create analogies (Linsey et al., 2012). This method used key-problem descriptors to abstract the problem into single verbs to aid in formulating analogies and novel ideation. However, both of these studies applied chunking either as a retrieval mechanism or mechanism for brainstorming but not as a cognitive representation. However, there is an evident lack of systems that utilize chunking as a visual representation and recombination as a functional mechanism in their systems.

Large Language Models (LLMs) have been demonstrating immense potential in their ability to generate more complex natural language analogies as evidenced through numerous proof-of-concept demonstrations. For instance, Zhu and Luo (2022) demonstrated that GPT-3's earlier

davinci models could generate analogous design concepts for a given problem by utilizing analogical designs from real-world applications. Furthermore, Webb et al. (2022) replicated the classic "convergence" problem (Gick and Holyoak, 1980) on a more recent version of the model. The researchers demonstrated that when prompted with the analogous story of "*generals seizing a city*," the model was able to converge to a meaningful solution for Duncker's radiation problem (see Duncker, 1945). Additionally, in a study conducted by Bhavya et al. (2022), who examined a more aligned version of these models, it was discovered that language models can generate analogies that are comparable in quality to those created by humans (Ouyang et al., 2022). However, despite the observed potential, there is a lack of studies that are focusing on utlizing these large language models as a tool for developing generative recombinations, which is addressed by this thesis.

In addition, while much of the work on Analogical Transfer made use of an experimental design where people were required to learn about some material before attempting to solve a problem where the learned material could be analogically mapped onto the problem to help solve it, few studies have attempted to address this inherent lack of temporality in these designs either through a targeted design (Tseng et al. (2008)) or employing integrated interfaces with multiple sensors to capture additional context (Hope et al., 2017; Chan et al., 2015; Kim and Horii, 2016). However, in our study, we attempt to operationalize the switching of contexts within interfaces as part of the study measures to gain a deeper understanding of why people make the decisions they do when they interact with the analogies and how our interface aided or hindered them proposing a system that marries together the chunking as a visual representation with assisted recombination as a generative mechanism. A detailed description of our proposed system and methodology will be elaborated in the following sections.

Chapter 3: The Proposed System for Triaging Analogies

3.1 Components of the System

3.1.1 Problem and Analogy Spaces

First, we break down the design problem and analogical leads with solutions into their corresponding functional constraints, inspired by the functional breakdown of research paper abstracts into background, purpose, mechanism and finding proposed by Chan et al. (2018). In this paper, we generalize these functional constraints so that they can be applied across analogies into the following components for an assigned design problem:

Stakeholder This functional constraint highlights the beneficiaries who will be affected by the assigned design problem.

Context This functional constraint highlights the context in which the aforementioned stakeholders face the assigned design problem.

Goal This functional constraint highlights the goal that the stakeholders need to achieve to solve the assigned design problem.

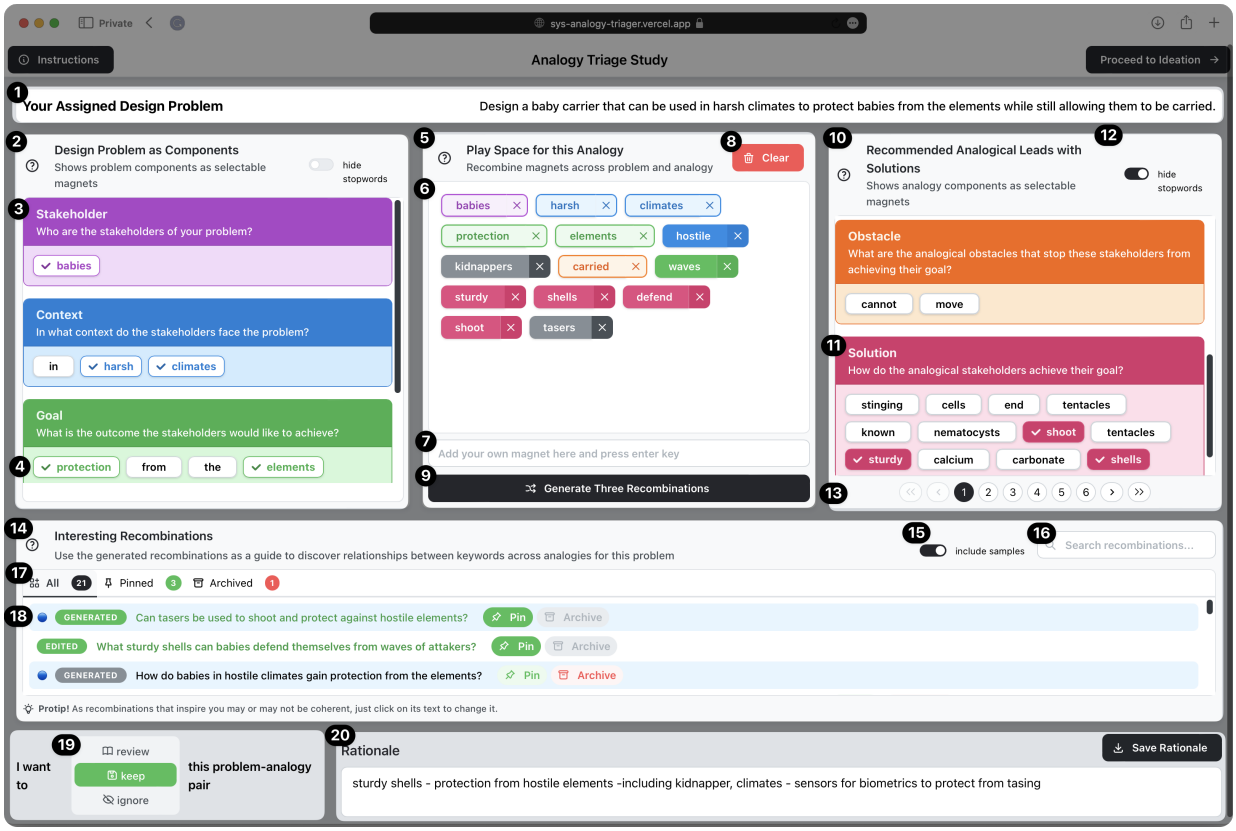


Figure 3.1: ***Components of Analogy Triage Interface***. The figure shows the proposed triage interface designed with the aim to facilitate the processes of recognition and selection of useful analogical leads while providing intuitive interfaces to ease the process of recombination. To address this, the interface contains *four* key components: The PROBLEM SPACE (2) where the assigned design problem (1) is broken down into components representing the functional constraints (3) of *Stakeholder*, ***Context***, ***Goal*** and ***Obstacle*** and the words within each are broken down into selectable words known as "*magnets*" (4); The ANALOGY SPACE (10) where the recommended analogical leads are broken down in a similar manner, with an additional component for the functional constraint of ***Solution*** (11); The PLAY SPACE (5) where the selected magnets (6) from the aforementioned spaces can be *added* (7), *edited*, *deleted*, *cleared* (8) and recombined through spatial maneuvers such as drag-and-drop to mirror the usage of real-world fridge magnets; and the RECOMBINATION SPACE (14) where the magnets are recombined into provocative guiding questions (9) or "*prompts*" through the use of GENERATIVE PRETRAINED MODELS (GPT) with the aim to facilitate divergent thinking while giving users the ability to filter (15, 16, 17) *edit*, *pin* or *archive* (18) interesting recombinations and make decisions to review/keep or ignore (19) based on their exploration so far with a space to state their reasons or rationale (20).

Obstacle This functional constraint highlights an obstacle that hinders the stakeholders from achieving their goal for the assigned design problem.

Alongside these functional constraints, the recommended analogical leads add an additional constraint to highlight the solution:

Solution This functional constraint highlights the solution proposed to solve the goal of the recommended analogical lead for the given design problem.

Below every recommended lead, the user can access the pagination, allowing them to navigate across the different analogical leads recommended for the given design problem.

In addition to these functionalities, based on previous studies on the effects of chunked visual representations (Knoblich et al., 1999) of linguistic features on optimal solution finding (Frank and Ramscar, 2003; Glucksberg, 1962), we propose breaking down the sentences within each of the functional constraints into selectable tiles referred to as *magnets* (as they were designed to be functional analog of real-world fridge magnets). On a related note, Adams et al. (2010) proposed a similar mechanism for collaborative brainstorming but not in the context of chunking as proposed by this thesis.

Furthermore, based on Wu et al. (2017)'s experiments on the role of chunk tightness and familiarity in problem solving that highighted the inverse relationship between chunk tightness and familiarity, we hypothesized that, in such cases, chunk tightness could be reduced by hiding non-functional words. Since participants of this study were randomly assigned different problems, the domain familiarity cannot be predetermined. Hence to account for these familiarity effects, we

provided a feature to hide *stopwords*: words with high frequency automatically omitted during natural language processing-to address them. If the user of this interface feels like getting stuck, the could choose to hide these stopwords enabling them to only focus on key function words.

3.1.2 Playspace and Recombination Space

Prior studies define recombinations as the process of creating new linkages between unlinked or between already linked components (i.e. linking them in new ways), aimed at a specific application (Hargadon and Sutton, 1997). Studying creative design practice, Herring et al. (2009) found that designers used examples by re-appropriating and recombining solution components into novel ideas. Studies exploring the concept of "*exaptation*" or "*the shift in the function of a trait*" (Darwin, 2004) in the context of the evolution of an idea (Mastrogiorgio and Gilsing, 2016) found that the likelihood of "*exapting*" increased if the components/"*recombinants*" of these ideas were difficult to recombine. Based on these insights we hypothesized that, the better the characterization of the components of a complex problem into chunks, the easier it is to be recombined to inspire innovative insights. However, since recombination is a creatively intensive process, to hasten the same, being able to rapidly explore connections between many ideas is an important precondition. This can become hard when the idea units are too large to comprehend quickly (McCrickard et al., 2013; Jeppesen and Lakhani, 2010). In his review on the taxonomy of questions in the context of experiential learning environments, Tofade et al. (2013) highlights that lower-order, convergent questions that rely on students' factual recall of prior knowledge rather than asking higher-order, divergent questions that promote deep thinking, requiring students to analyze and evaluate concepts. Hence, we propose the playspace and recombinations sections

to incorporate a human-in-the-loop type system that utilizes Generative Pretrained Models to quickly construct insightful, divergent questions from combinations of selected ”magnets” to provide guidance towards unique and diverse approaches to the problem. However, given that the generated prompts might not always be relevant or semantically meaningful (Chaturvedi et al.), we also provide the ability that enables the user to edit the generated prompts by building upon the structure in which the insights into these ”forced relationships” (Raudsepp, 1983) hypothesizing that this system will facilitate rapid iteration over existing the solution space while enabling the user to come up with novel, unique, unseen ideas using these generated recombinations as a guide.

Thus, through these features, we hypothesize that the proposed system offers a robust visual interface designed with the objective to facilitate the triaging of analogies and enables its users to quickly identify their relevance in solving for a given problem.

3.2 Implementation Details

3.2.1 Prompt Engineering for Recombinations

Recently, there has been a significant increase in the number of proof-of-concept demonstrations showcasing Language Models (LLMs) ability to generate more complex natural language analogies. For example, Zhu and Luo (2022) demonstrated that GPT-3's earlier *davinci* models could generate analogous design concepts for a given problem when prompted with analogical designs from real-world applications. On a more recent version of the model, Webb et al. (2022) replicated the classic analogical problem solving paradigm proposed by Gick and Holyoak (1980) showing that the model was able to converge to a plausible solution for Duncker's radia-

tion problem (see Duncker, 1945) when prompted with the analogous story of "*generals seizing a city*".

Particularly, in a study on a more aligned version of these models (Ouyang et al., 2022) by Bhavya et al. (2022) found that language models can generate analogies that are comparable in quality to those created by humans if the prompts were engineered properly. The authors found that these models performed better when the prompts were structured as imperative statements that explain the reasoning process instead of using questions as prompts. Reynolds and McDonell (2021) exploring the few-shot paradigm in LLMs found that even the worst performing few-shot prompts such as question-answering tasks used in the original paper (Brown et al., 2020) performed better when applying simple changes to formatting and structuring the in natural language.

For the purposes of this interface at the time of conducting the study, we used the (then) recent LLM from OpenAI: `text-davinci-002`. Based on insights from the aforementioned studies, we decided to align our prompt explorations towards zero-shot learning. When we attempted to include the functional constraints of the problem to contextualize the recombinations under the assumption that the generations would have well-defined relationships, longer prompts tended to confuse the model and the resulting outputs were sub-optimal. However, during this exploration, we found that the adding descriptive words highlighting the kinds of recombinations to prioritize (such as *insightful* and *meaningful*) facilitated the prompt to generate thought-provoking questions. Furthermore, to account for cases where the magnets selected by the users might be from the same component or their own, the prompt structure was simplified thus arriving at the final iteration show below:

Table 3.1: ***Model Parameters for Generating Recombinations***. Contains the model parameters used for generating recombinations using the above prompt.

Prompt Parameters	**Value**
`model`	`text-davinci-002`
`frequencyPenalty`	`0.5`
`presencePenalty`	`0.75`
`temperature`	`0.75`

Final Prompt Template for Generating Recombinations

```
By combining the following list of words together,

generate [n] meaningful questions with insightful

relationships: [word1, word2, ..., wordN]

Output:

1.
```

To further optimize the model we tweaked the parameters (see 3.1) before generating the recombinations for a given set of keywords.

The `temperature` parameter controlled the randomness of the model's responses. A higher temperature value increased the likelihood of the model generating unexpected or creative responses, while a lower temperature value increased the likelihood of the model generating more predictable or safe responses. In this case, experimenting with higher values resulted in questions that were less meaningful. Ultimately, I settled with a temperature value of 0.75 (slightly larger than the default) indicating a moderate level of randomness.

The `frequency_penalty` and `presence_penalty` parameters were used to encourage the model to generate unique and diverse responses by penalizing the model for gener-

ating the same words or phrases multiple times and for not including all the input keywords in its response. For this study, the values used were 0.5 and 0.25, respectively, which means that the penalties are moderate and not too strict, to allow for proper sentence structures. Overall, the model parameters were chosen to balance the need for creativity and diversity in the model's responses with the need for coherence and relevance to the input keywords (magnets) sourced from the playspace.

3.2.2 Technical Implementation

This interface is developed using the React Framework in JavaScript, which is a popular choice for building user interfaces. React provides a set of reusable components that can be used to create interactive web applications. The use of this framework allows for efficient rendering and management of complex user interfaces. React Hooks are used for imperative routing and accessing data from the interface. Hooks are functions that allow developers to use state and other React features without writing a class. This approach simplifies the code and makes it easier to manage and reuse code across components. To store data, the system uses Firestore Backend as a service, which is a cloud-based NoSQL database offered by Google. Firestore provides scalable storage and real-time synchronization for mobile, web, and server applications.

Chapter 4: Methodology

The present study aims to investigate the effectiveness of different interfaces, with or without certain stimuli, in facilitating creative ideation and adoption of multiple analogical leads for a given design problem. Specifically, this study seeks to explore participants' inspirations and choices in finding creative solutions to a design problem based on the recommended analogical leads that they found most helpful. In addition, the study intends to understand participants' preferences for the different interfaces based on their experiences and reasons for those preferences. These information would then be used to answer the key research questions exploring how the intervention leads to changes in user interactions and their triaging decisions.

One particular research question this study seeks to understand is ways in which a participant's decisions about an analogy are changing as they interact with our proposed system–Are people failing to keep a relevant analogical lead only to regret and keep it later? Does the interface have an influence on the quality and the quantity of generated ideas? I hypothesize that, due to increased decomposability due to chunking, complex sources of information such as far-field analogies are more likely to be exapted for ideation (Mastrogiorgio and Gilsing, 2016), possibly by facilitating the identification of the most obscure properties of the assigned design problem (McCaffrey, 2012).

Another research question this study hopes to answer is how participant interactions with

the analogies are changing and how they make sense of these analogies across the triaging and ideation over these analogical leads. I hypothesize that, due to a possible increase in the adoption of far-field analogies, there would either be an increase in the number of ideas or the resulting ideas would be of better quality than the baseline. (Chan et al., 2011; Dahl and Moreau, 2002; Vendetti et al., 2014).

In order to accomplish these objectives and test these hypotheses, the study employs a mixed-methods approach with a within-subjects design. All participants are be assigned to both conditions: an interface with the stimuli (our proposed interface (see chapter 3)) and an interface without the stimuli (a baseline (see section 4.2.1)). Each participant is asked to complete a couple of tasks using both of the assigned interfaces, and their interactions with the system are recorded for later analysis.

To assess the effectiveness of the different interfaces, we examine how participants' decisions about analogical leads change as they interact with the system. To this end, we ask the participants to indicate which analogical leads they found most helpful and why, across the different interfaces. I then analyze these data collected from their interactions with the interfaces, think-alouds and debriefing interviews through quantitative and qualitative analyses to identify patterns or trends that emerge.

Overall, the present study aims to contribute to a better understanding of how better interfaces can be designed to enable triaging and adoption of multiple analogical leads. I believe that this study's findings haveed practical implications for identification, design and development of stimuli that promote creative problem-solving through computational support.

4.1 Study Design

We employ a within-subjects study that is broken down into two phases. This study design was selected with the goal to understand the effects of our interventions at the participant level and facilitate reasoning about how their interactions changed across interventions. The first phase consisted of two 30-minute sessions, during which the participants were asked to interact with both of our proposed interfaces to screen for potential inspirations and come up with novel solutions for the assigned design problems using them (using the Ideation Interface (see 4.2.2)). To account for potential order effects, we randomized the sequence in which the participant interacted with the baseline (see section 4.2.1) and our proposed interface (see chapter 3). We also varied the problems each participant faced in a particular interface to account for learning effects that can be attributed to prior-problem solving. The analogical distance of the recommended analogical leads (near and far analogies) for the corresponding problems were also randomized to control for exposure effects. In the second phase, a short debriefing interview that lasts not more than 10 minutes in conducted to inquire about their experiences with both of the interfaces and understand their reasoning behind it.

4.2 Task Design

To address and evaluate the hypotheses for the key research questions regarding the triaging process using an interface and their benefits during ideation in this study, we break down each interface into two tasks, namely:

The Screening Task During this task, the participants are asked to make decisions on leads that they think would inspire them the most in coming up with creative new solutions for the assigned design problem. To ensure clarity of the task, the participants were explicitly told to select the analogical leads they found "helpful" for "finding a creative new "solution" to the given problem, rather than focusing on their "relevance" to the problem.

The Ideation Task During this task, the participants are asked to come up with solutions that are as creative as possible for a given design problem using the leads they screened in the previous task as a guide. To ensure that the participants were not constrained by the task or the open-ended nature of the selected problems, they were explicitly allowed to add or modify assumptions or aspects of the design problem.

Throughout the duration of the study, heavy emphasis was given to remind the participants to come up with the most creative, novel solutions they can think of for the assigned design problem using the proposed interfaces to perform the aforementioned tasks.

4.2.1 Design of the Baseline Interface

The baseline interface used for this study works very similar to the intervention interface except that it has following significant changes:

(i) The components of the design problem are not further broken down into selectable chunks (magnets).

(ii) The recommended analogical lead is presented, not as chunked components, but as abstracts of text that elaborate on the analogical lead with the proposed solution weaved in the text.

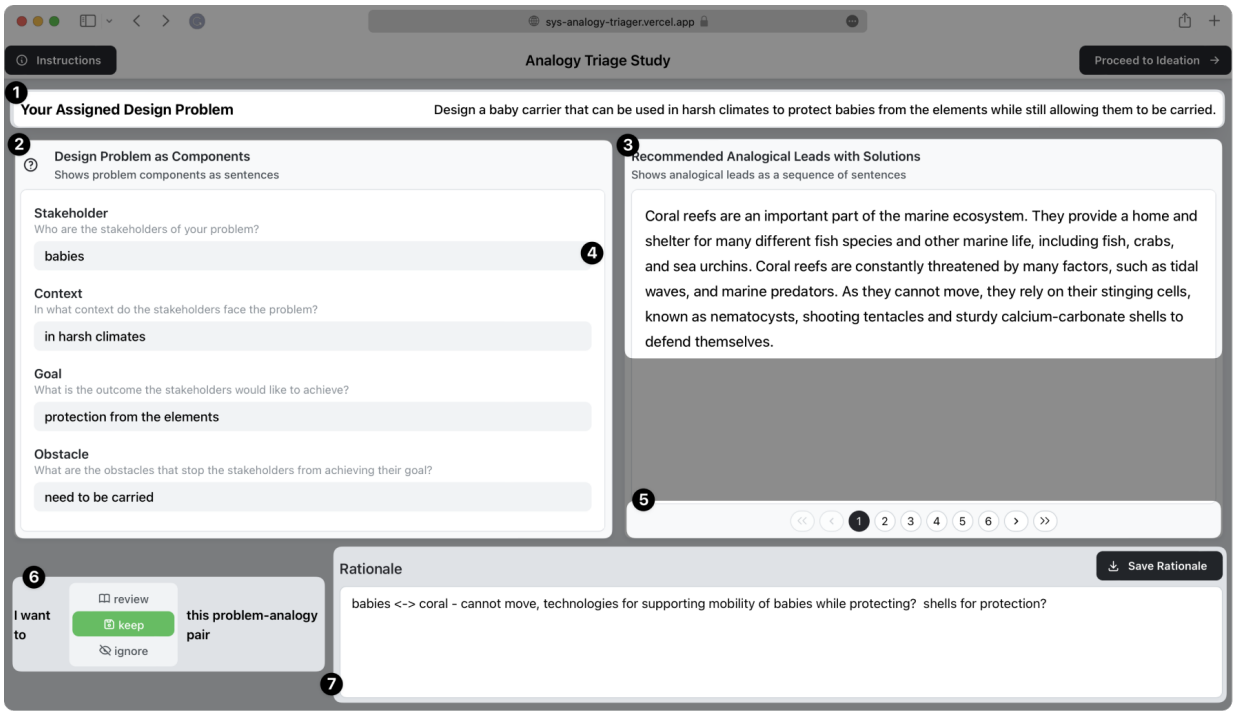


Figure 4.1: ***Design of the Baseline Interface for the Screening Task***. This screenshot highlights the key components in the screening interface used for the baseline condition. The key changes from the intervention intervention discussed in the section 4.2.1 are labelled as (2), (4) and (3)

However, apart from these changes, the pagination across the recommended analogical leads, the controls for decision making and text areas for participants to jot down the rationales function unchanged.

4.2.2 Design of the Ideation Task Interface

After screening for potential inspirations for a given design problem from the presented leads, the participants can proceed to ideation using this interface. The interface for the ideation task follows a simple two-column layout. The left column includes a searchable list of all the recommended analogical leads that were processed by the participant during screening while the right column highlights the assigned design problem with their corresponding components for ease of mapping, while providing a text area for the participants to enter/scrap their ideas as they

come up with them.

The participants can also make changes to their decisions and rationales using this interface. For instance, if a particular reformulation of the design problem explored during ideation renders a particular analogy relevant/irrelevant to the context of solving the problem different from their original intention, they can choose to make the changes as needed without returning to the screening interface. Thus the proposed interface works in conjunction with the screening interface while accommodating for the available stimuli in the baseline (see subsection 4.2.2.2) and intervention (see subsection 4.2.2.1) interfaces with the aim to provide a seamless way for participants to utilize the brainstormed analogies during their adoption and ideation process.

4.2.2.1 Ideation Task Interface for Intervention

To accommodate for the available stimuli in the Intervention during ideation, two additional features were provided. First, the participants could search, filter and unpin their chosen recombinations if found irrelevant or pin archived recombinations if found relevant without transitioning to the screening interface. Second, to facilitate further explorations while tackling situations where the participant is unable to screen through all the analogies within the allocated timeframe, the participants have the option to return to any recommended analogical lead with the click of a button. I believe that these changes would be helpful in identifying and understanding instances of ”regret” from the participant’s interactions with the interface during ideation.

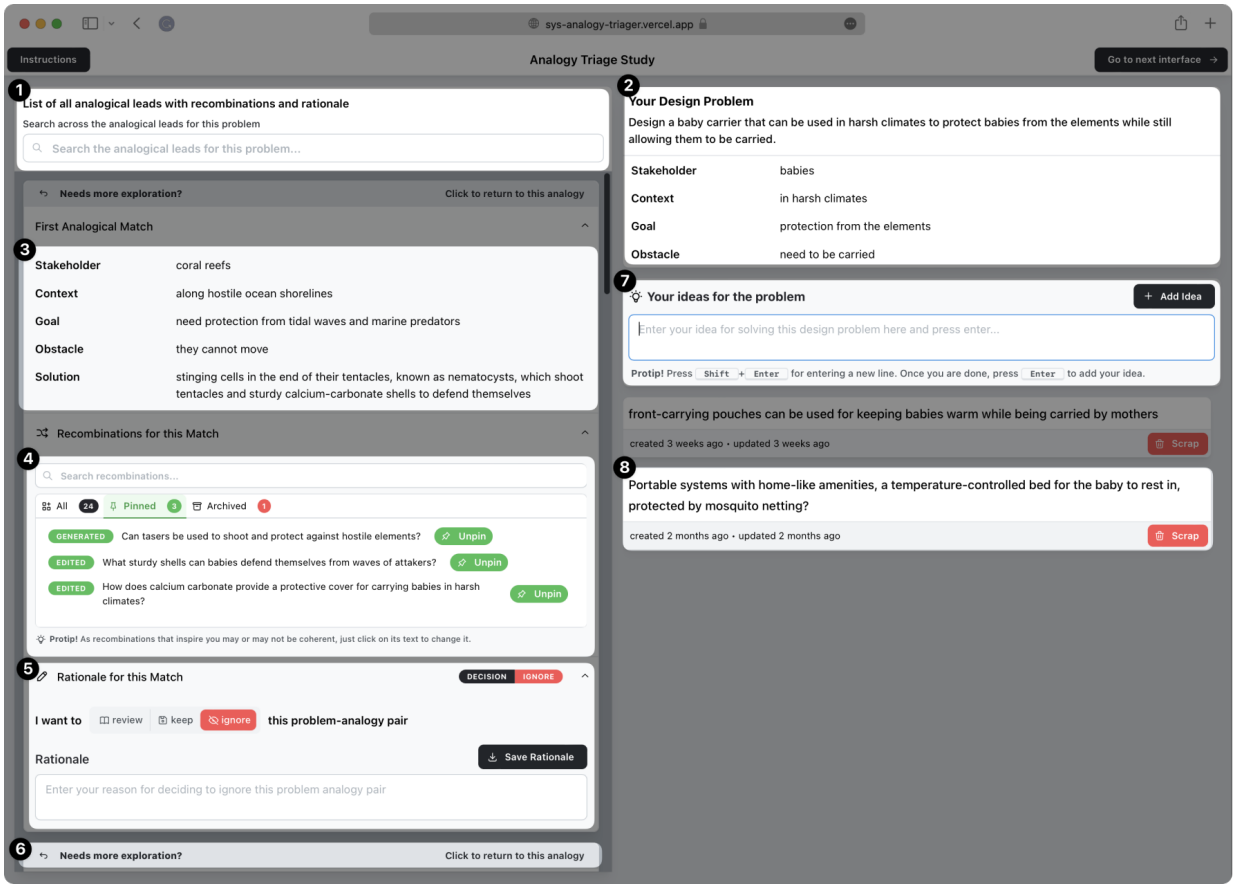


Figure 4.2: ***Design of Ideation Task Interface for Intervention***. This screenshot highlights the key components in the screening interface used for the intervention condition. The key highlights discussed in the section 4.2.2.1 are labelled as (3), (4) and (6)
while (1), (2), (4), (7) and (8) are common for both interfaces (discussed in 4.2.2)

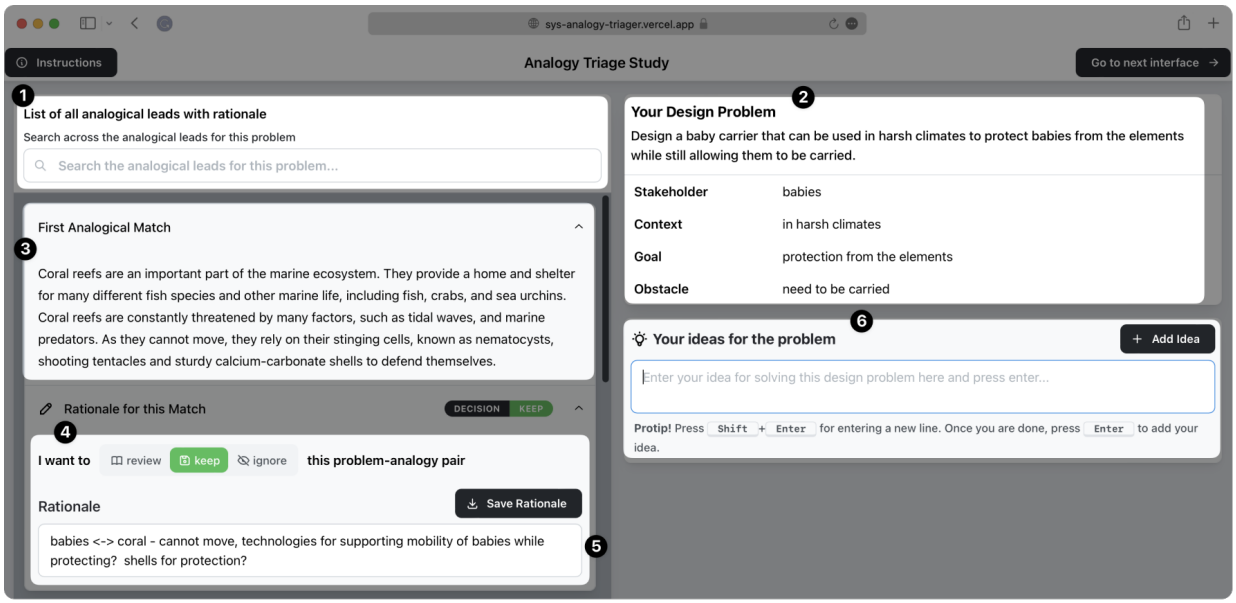


Figure 4.3: ***Design of Ideation Task Interface for Baseline***. This screenshot highlights the key components in the screening interface used for the intervention condition. The key highlight discussed in the section 4.2.2.2 is labelled as (3), while (1), (2), (4), (6) are common for both interfaces (discussed in 4.2.2)

4.2.2.2 Ideation Task Interface for Baseline

To accommodate for the lack of stimuli in the Baseline during ideation, two minor changes are made to the interface. First, unlike the ideation interface for intervention, the analogical leads in the list are presented as abstracts of text. Second, no button is provided to return to the screening task of the baseline interface as there is a lack of need and ”regret” can still be characterized using the existing features without any additions.

4.3 Materials

4.3.1 Source Problems

The design problems were selected so that they are generic enough to be ideated by participants across a diverse range of domains while still requiring creative thinking to come up with innovative solutions. Care was taken to ensure that the shortlisted problems for this study did not require specialized knowledge to develop practical solutions such as a strong expertise in science or technology. Ultimately, we shortlisted three problems shown in Table 4.1. This limitation was to ensure that the participants encountered different problems during the demo and their interactions with the baseline and the interface. The design descriptions of these shortlisted problems were also broken down into its functional constraints (*Stakeholder*, *Context*, *Goal* and *Obstacle*) for usage in our proposed intervention.

4.3.2 Target Analogies

To retrieve the analogies for the shortlisted problems, great care was taken to ensure that the analogies were relevant with varying degrees of structural similarity: there was an equal spread across the number of near and far analogies for each problem. Far analogies were selected if they were from a domain different from the assign problem and were likely to be outside of the participant's prior knowledge. Based on these criteria, we shortlisted six analogical leads for each problem with 3 near and 3 far resulting in a total of 18 analogies as shown in the Table 4.2.

Problem	**Problem Description**	**Problem Components**
Baby Carrier Problem	Design a baby carrier that can be used in harsh climates to protect babies from the elements while still allowing them to be carried.	***Stakeholder***: babies ***Context***: in harsh climates ***Goal***: need protection from the environment ***Obstacle***: need to be carried
Marine Environment Problem	Design a marine environment in which young children and aquatic animals can interact with each other while keeping both safe.	***Stakeholder***: young children and aquatic animals ***Context***: in a marine environment ***Goal***: interact with each other ***Obstacle***: keep both safe
Weight Training Problem	Design a weight training program that can be done while on the go for business people who are busy frequent flyers.	***Stakeholder***: a business person ***Context***: is a busy frequent flyer ***Goal***: do weight training ***Obstacle***: always on the go

Table 4.1: ***Source Problems used for this Study***. This table contains the three source problems that were randomly assigned to the participants of this study across both of the interfaces.

Design Problem	Analogy Distance	Analogy
Baby Carrier Problem	**Near Analogies**	Baby Gates, Mosquito Netting, Kangaroo Pouches
	Far Analogies	Coral Reefs with Stinging Cells, Ground Launched Missiles, Convertible Cars
Marine Environment Problem	**Near Analogies**	Shark Cages, Expert Dolphin Trainers, Deep Sea Submersibles
	Far Analogies	Virtual Reality Psychiatric Treatments, Lane Assist Technologies, Negative Room Pressure
Weight Training Problem	**Near Analogies**	Resistance Bands, Sandbag, Desk Cycle
	Far Analogies	Protein Biosynthesis, Passively Designed Popup Houses, Inflatable Air Mattresses

Table 4.2: ***Target Analogies for the Source Problems used for this Study***. This table contains the 18 target analogies that were randomly assigned to the participants of this study across both of the interfaces based on their assigned design problem

4.4 Participants

Participants were recruited using the Listserv emails and social media platforms like Twitter and Mastodon to reach out to people from outside the university. The only requirement for participation in the study was that the participant had to be 18 years old or over. During recruitment, the participants were asked to answer a questionnaire about their demographics, fields of interest and expertise to get an prior-assessment of their domain knowledge following which, they scheduled a suitable date and time for the study either in-person or online through Zoom. A total of 23 participants from diverse genders (52.2% female, 47.8% male), identities (52.2% as a woman, 43.5% as a man, 4.3% as non-binary), age-groups (65.2% between the ages of 18 and 25, 8.7% between the ages of 30 and 45, 13% between the ages of 46 and 59, 13% over 60), ethnicities (34.8% Caucasian, 39.1% African-American, 17.4% Asian, 4.3% Hispanic, 4.3% Asian American and Caucasian) and education-levels (56.5% with a Bachelors degree, 26.1% with a Masters degree, 13% with a Ph.D. or higher, 4.3% from some high school) took part in the study with one participant opting for an in-person session. There was an equal split between participants within (43.5%) and outside (56.5%) the University of Maryland. Notably, of these outside participants, one had a federal background while another was from the industry.

4.5 Procedure

As outlined in the study design diagram (see Fig. 4.4) and the detailed session-wise breakdown 4.3), Be it a Zoom or a in-person session, at the start of the study, the participant is introduced to the study, what it entails, what we attempt to understand and what is expected of them

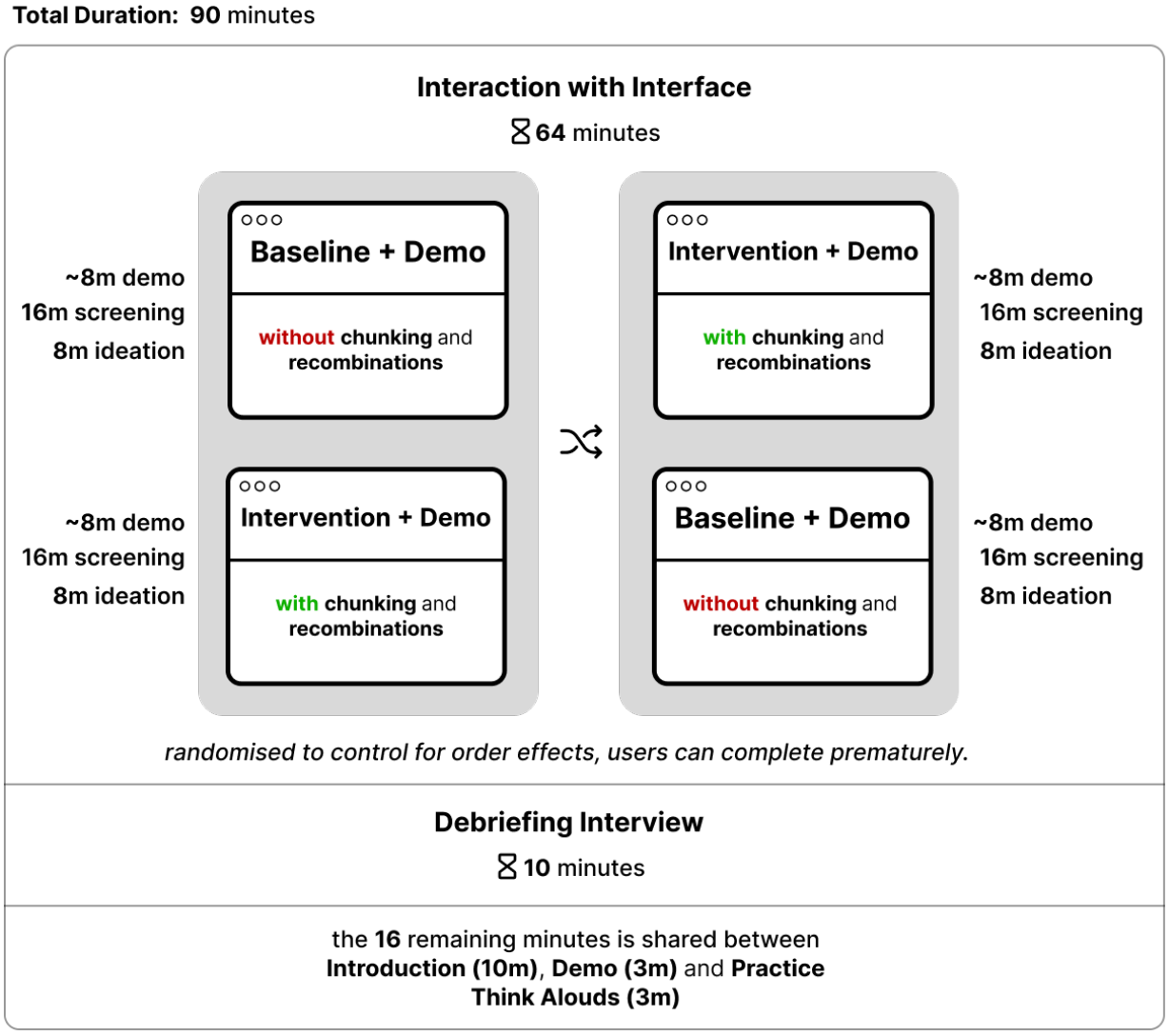


Figure 4.4: This figure shows a visual representation highlighting the key aspects of the study design, the components involved and a broad overview on the duration for each component in the study.

Session Component	Session Purpose	Session Duration
Introduction	Give the participants an overview on the study, its purpose and outline the tasks they was asked to do while offering pointers and clearing high-level questions before the start of the study.	10 minutes
Think Aloud Demo showing participants a YouTube video demo of the think aloud performing a sample task.	To ensure that every participant is familiarized with the concept of think-alouds while clearing possible doubts prior to the study.	3 minutes
Think Aloud Practice asking partipants to visit the Human Computer Interaction Website to perform an practice think-aloud.	To enable the participants to familiarize themselves with the think-aloud before beginning the study.	3 minutes
Demo of the First Interface (can be *Baseline* or *Intervention*)	Run-through on the different features and interactions possible with an example, clearing doubts prior to study.	8 minutes
Interaction with the Screening Task of the First Interface	Evaluate the interactions with the interface for the screening task.	16 minutes
Interaction with the Ideation Task of the First Interface	Evaluate the interactions with the interface for the ideation task.	8 minutes
Demo of the Second Interface (can be *Intervention* or *Baseline*)	Run-through on the different features and interactions possible with an example, clearing doubts prior to study.	8 minutes
Interaction with the Screening Task of the Second Interface	Evaluate the interactions with the interface for the screening task.	16 minutes
Interaction with the Ideation Task of the Second Interface	Evaluate the interactions with the interface for the ideation task.	8 minutes
Debriefing Interview	Understand the interfaces they preferred and reasons for the same.	10 minutes

Table 4.3: ***Session-wise Breakdown of Study Procedure***. This table showcases the session-wise breakdown of the different components in this study and their significance in the context of this study.

and the risks involved highlighting our strong commitment toward ensuring the confidentiality of the recordings. During this explanations, participants were encouraged to interrupt and clarify any doubts they might have. We ensured that the instructions clearly defined what analogical leads were and what their relevance meant in the context of this study. Since we cannot assume every participant to be familiar with think-aloud protocols, a walk-through and practice on the think-aloud process over an unrelated task is administered prior to exposing the participant to the interfaces.

After their consent, a hyperlink to the hosted version of the interface[1] for the study is shared with the participant along with the 6-digit code assigned to them. A seed script is used to randomly generate this code and populate the database with the randomised conditions, problems and analogies that a participant encountered when interacting with the interface.

Before the interaction with every interface, a demo of that interface is performed with an example problem (different from the problems that were assigned for their interactions with the baseline and intervention interfaces) to the different features while showing different ways to use the interface. The participants were encouraged to interrupt and clarify any queries they might have regarding the different available functionalities and its use prior to the study. At the beginning of each task for an interface, the participants were informed about the time allocated for this study. Special emphasis was given on generating "*the most creative, innovative solutions they could think of*" for the design problem. If a participant failed to think-aloud for a consecutive couple of minutes, the participants were explicitly asked to "*speak what is on their mind*".

Once the interaction sessions are completed, a 10-minute debriefing interview is conducted with a participant where they are asked about their experience using the interfaces to select the

[1] https://sys-analogy-triager.vercel.app

analogical leads (to understand their affective responses to each interface) and how useful they were for the goal of selecting inspiring leads that led to the most novel, creative and innovative solutions. In case of vague responses, the participants were asked to verbally rate the interfaces based on their usage and reason about the choice. Once the sessions were complete, the participants were thanked before dispatching their compensation.

4.6 Analysis

This study employs a Mixed Methods approach to investigate the impact of an intervention on participants' decision-making and interaction with analogies. The quantitative data was analyzed using a General Linear Mixed Model (GLMM) to model the effects of multiple variables, such as the type of analogy (near vs. far) and the condition (baseline vs. intervention), on participants' decisions. The transcripts and interaction data was processed to understand the reasons behind the observed trends on the selection of recommended analogical leads during their interaction with the screening phase and instances of regret (changes in decisions) during the ideation phase. Our research question focus on understanding changes in participants' decision-making. We investigate whether there are significant changes in decisions made by participants in the intervention condition compared to the baseline condition, particularly their initial decisions to keep or ignore the recommended analogical leads via the screening interface. We also explore whether there are significant changes in participants' interactions with the analogies (with respect to decisions) in the intervention condition compared to the baseline condition. We use both quantitative and qualitative analyses to arrive at a holistic understanding of the impact of the intervention on decisions.

4.6.1 Effects of Chunking and Recombination on Decision Making over Analogical Leads

To examine the impacts of the proposed intervention on decision making, the **dependent variable** is the initial decisions made by participants in the screening interface while the **independent variables** are the type of *Analogy* (Near vs. Far) and *Condition* (Baseline vs. Intervention). These decisions can take one of the following values, namely: *Keep*, *Ignore* and *Review*. To understand effects of the independent variables on participants' initial decisions to *Keep* or *Ignore*, this categorical dependent variable is converted into a numerical variable where each kind of decision is represented by a binary vector with a length equal to the number of decisions through one-hot encoding. The vector contains zeros in all positions except for the one corresponding to the decision, which is set to one. Similarly, any changes to decisions such as *Ignore*→*Keep/Review*, *Keep*→*Ignore/Review*, and *Review*→*Keep/Ignore*, if found, were coded as *Revisions*.

Performing exploratory analysis of the decision data, when checking for valid decisions across the different participants, we discovered that the decisions of the participant (P313394) were not logged properly. Hence, 24 entries of this participant across conditions and interfaces were ignored to ensure quality. Due to the small sample size, we were unable to test for interaction effects since the models for the computation of interaction effects between analogical distance and condition failed to converge.

The initial information of their decisions during the screening phase were sourced from the interaction logs while their final decisions were taken from the final state of their decisions in the stored data. These decisions were then mapped together with their problem, analogy, interface

and condition prior to processing. The table 4.4 shows a summary of the processed collection of decisions.

Table 4.4: Summary Statistics for Decisions on Analogies

Variable	Valid N	Mean	SD	Min	Max
Condition	263				
... Baseline	131	50%			
... Intervention	132	50%			
interface	263				
... Ideation	0	0%			
... Screening	263	100%			
Analogy Distance	263				
... Far Analogy	132	50%			
... Near Analogy	131	50%			
Initial Ignore Decision	30	0.11	0.32	0	1
Initial Keep Decision	187	0.71	0.45	0	1
Initial Review Decision	46	0.17	0.38	0	1
Final Ignore Decision	28	0.11	0.31	0	1
Final Keep Decision	198	0.75	0.43	0	1
Final Review Decision	37	0.14	0.35	0	1
Ignore→Keep/Review	4	0.015	0.12	0	1
Keep→Ignore/Review	2	0.0076	0.087	0	1

To account for correlation between repeated measures within an individual participant, we use GLMM to model and explore the effects of these independent variables. Furthermore, since the resulting dependent measures is dichotomous, we used a binomial probability distribution to model the decision data with random effects. We used the Full Maximum Likelihood Estimation by the Laplace approximation. The following equations explain the general structure of our models in mixed model notation:

$$\eta_{i(l_v j)} = \gamma_{00} + \sum_{q} \gamma_{q0} X_{qi} + \upsilon_{0j}$$

where,

η_j is the predicted log odds of the i_{th} analogy where a variable $decision$ takes a value v (*keep* and *ignore* for each j_{th} participant.

γ_{00} is the grand mean log odds for all analogies

γ_{q0} is a vector of q predictors ($q = 0$ for our null model)

υ_{0j} is the random effects contribution of variation between participants for mean γ_{00} (i.e., how much a given participant varies from the grand mean)

This allowed us to identify the relationships and help us determine if there are significant changes in participants' decision-making in the intervention condition compared to the baseline condition via the screening interface, while allowing estimation of both within-group and between-group variation. We first build the baseline model to account for variation in l with only the random effects of the participant (υ_{0j}) in the absence of additional predictors to understand the within-participant variation for each decision. We then estimated a model with a fixed effect of **Analogical Distance** (*Near* or *Far*) and another model that added an additional fixed effect of **Experimental Condition** (*Baseline* or *Intervention*) to the model to assess their impact on arriving at each decision (v where l refers to decisions) (*Keep* and *Ignore*). Analysis of Variance (ANOVA) was conducted to evaluate and compare the goodness of fit between each pair of models to select the model with the best fit.

To start with, we examine the impact of analogical distance on decision making. Previous research suggests that relational mapping of analogies are susceptible to surface level similarity (Gick and Holyoak, 1980; Hammond et al., 1991; Gentner et al., 1993; Keane, 1997), and we expect to find a similar relationship in this study. We controlled for other factors such as the order of presenting analogies of varying distance and the order of problems encountered by the

participants across interventions to ensure that any observed effects of analogical distance are not due to other factors.

Later, we explored the impact of our proposed intervention on decision making. Since prior research suggests that optimal mapping and representations (Gentner, 1983; Duncker, 1945; Glucksberg and Weisberg, 1966; Frank and Ramscar, 2003) play an important role in identifying useful analogies/solutions, we expect to find a similar relationship in this study with our intervention that attempts to propose one such representation. We controlled for other factors such as the order of presenting the interface with/without stimuli and the order of problems encountered by the participants across interventions to ensure that any observed effects of our intervention are not due to other factors. Overall, we expect to find that analogy distance and the intervention condition are both important predictors of initial decisions and their changes during final decisions.

Finally, we explored the impact of our proposed intervention on making revisions using GLMM (v where v where l refers to revisions). If our intervention were to help in the selection of beneficial analogical leads, we expect to find far-analogies to be less prematurely rejected with near-analogies to be less prematurely-accepted.

4.6.2 Effects of Chunking and Recombination on Processing time

In addition, we also looked at the time taken to process each analogy of varying distance across the interventions. Based on the differences between available system logs when participants change between analogies, we computed the time taken to process each analogy. Since the logs were specific to the interface and not linked across interfaces, the collected data will include

every log corresponding to a particular interface. If the ending timestamp of a log is missing, the delta calculation was prevented. Only the time taken starting from first interaction until the first click to paginate during screening of an analogy were considered, in the sequence of processing. If a participant returns to an analogy, the time is updated only if it exceeds the current processing time for that analogy. On exploratory analysis of this processed interaction data, we found that the data had 24 NA values, logs that do not contain timestamp information. Apart from these, the logs were not properly logged for 1 participant (P313394). The table 4.5 shows a summary of the collection of processing times.

Table 4.5: Summary Statistics of Processing Time of Analogies

Variable	Valid N	Mean	SD	Min	Max
Condition	220				
... Baseline	110	50%			
... Intervention	110	50%			
Interface	220				
... Screening	220	100%			
Analogy Distance	220				
... Far Analogy	112	51%			
... Near Analogy	108	49%			
Analogy Decision	220				
... Ignore Decision	26	12%			
... Keep Decision	159	72%			
... Review Decision	35	16%			
Actions	197				
... go-to-analogy	187	95%			
... navigate-to-task	2	1%			
... update-no-of-analogies	8	4%			
Seconds per Analogy	36003	184	109	32	528
Return to Analogy	220				
... No	191	87%			
... Yes	29	13%			

4.6.3 Qualitative Analysis to Understand Chunking and Recombination Interactions

Thematic analysis of the transcripts and interviews were also conducted as it helped us understand the reasons behind instances of regret (changes in decisions) during the ideation phase and why participants chose to do so, and understand how the interactive mechanisms of chunking and recombination help.

Chapter 5: Results

5.1 Quantitative Analysis

5.1.1 Decisions on Analogies

For fitting over decisions, we replaced l in our base mixed model equation (see eqn. 4.6.1) with a one-hot encoded categorical variable *decisions* that takes the values v: *keep* or *ignore*, resulting in Null models for Ignore (see table 5.1) and Keep decisions (see table 5.4), Models with Distance as predictor for Ignore (see table 5.2) and Keep decisions (see table 5.5) and finally, Distance with Condition as predictors for Ignore (see table 5.3) and keep decisions (see table 5.6).

All models except the model with both Analogical Distance and Interface Condition as predictors got singular fit, implying that the random effects were either too complex or weren't complex enough that it needed to be addressed by these both the predictors to achieve a non-singular fit. We believe this is due to a smaller sample size and lack of enough data, leading us to over-fit.

As the table 5.3 suggests, **both near analogies and the intervention condition have a significant negative effect on the tendency for participants to ignore analogies**. More specifically, far analogies had 4x higher odds of being ignored compared to near analogies. In addition, participants were 4x less likely to ignore analogies in the intervention interface, supporting our

Table 5.1: Null Model for Ignore Decisions on Analogies

	Dependent variable:
	initial_decision_ignore
Constant	-2.050^{***}
	$(-2.430, -1.670)$
Observations	263
Log Likelihood	-93.349
Akaike Inf. Crit.	190.697
Bayesian Inf. Crit.	197.842
Note:	*p<0.1; **p<0.05; ***p<0.01

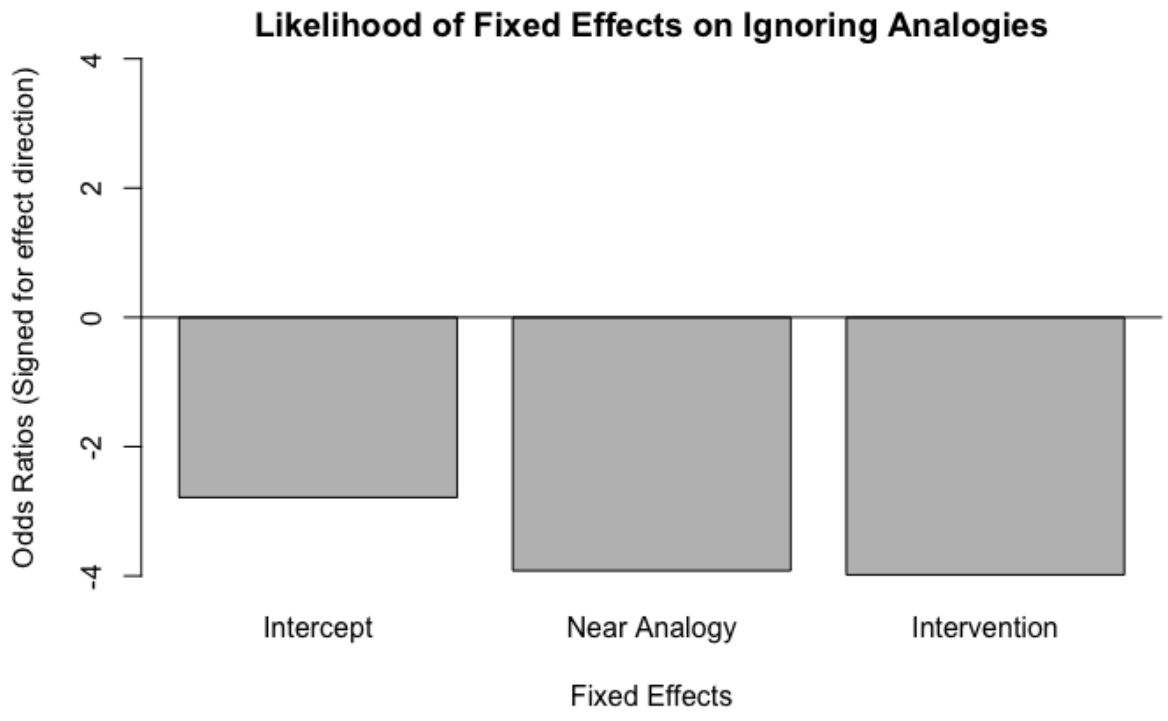


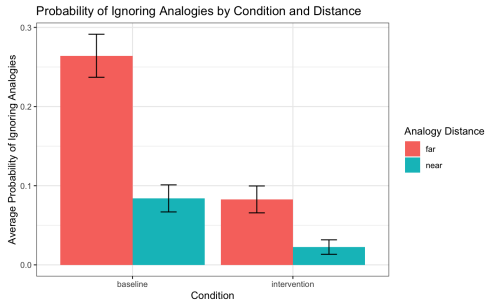
Figure 5.1: Plotting the log odds of fixed effects on ignoring analogies (signed for directionality)

Table 5.2: GLMM Model with Analogical Distance for Ignoring Decisions

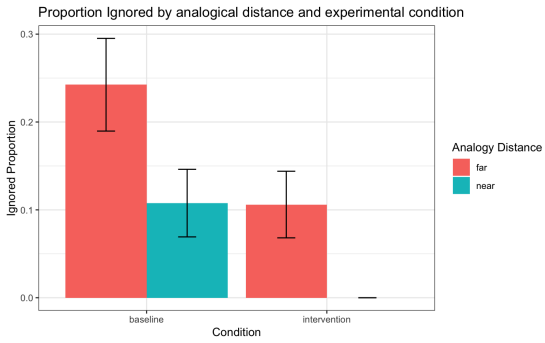
	Dependent variable:
	initial_decision_ignore
Near Analogy	-1.319^{***}
	$(-2.203, -0.434)$
Constant	-1.556^{***}
	$(-2.006, -1.106)$
Observations	263
Log Likelihood	-88.371
Akaike Inf. Crit.	182.742
Bayesian Inf. Crit.	193.459
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.3: GLMM Model with Analogical Distance and Experimental Condition for Ignoring Decisions

	Dependent variable:
	initial_decision_ignore
Near Analogy	-1.366^{***}
	$(-2.271, -0.460)$
Intervention Condition	-1.382^{***}
	$(-2.286, -0.478)$
Constant	-1.025^{***}
	$(-1.578, -0.472)$
Observations	263
Log Likelihood	-83.104
Akaike Inf. Crit.	174.208
Bayesian Inf. Crit.	188.497
Note:	*p<0.1; **p<0.05; ***p<0.01

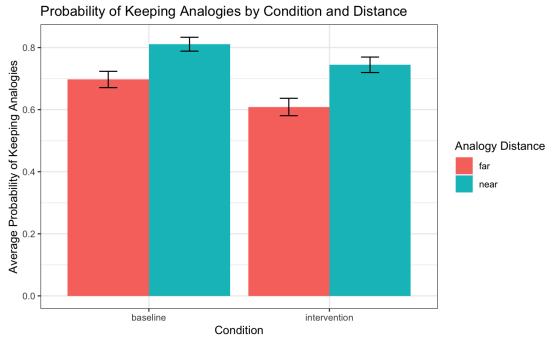


(a) Model-based Plot

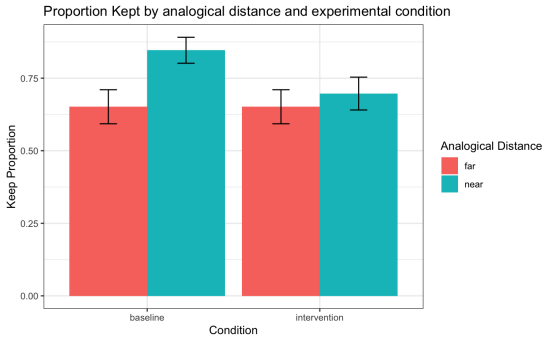


(b) Raw Plot of Proportions

Figure 5.2: Plots describing the Tendency to Ignore Analogies



(a) Model-based Plot



(b) Raw Plot of Proportions

Figure 5.3: Plots describing the Tendency to Keep Analogies

hypothesis (see fig. 5.1).

However, when we tried testing for possible interaction effects, the model could not converge. Looking at the number of ignored analogies, we found that the participants ignored only 11% of the analogies during screening, leading me to believe that this sparsity of data, when processed with other factors defined for the condition due to a lack of variations in the experimental combinations, caused it to misconverge, resulting in large eigen values.

The results of these models exploring the effects of Analogical Distance and Experimental Conditions of keep decisions showed that participants were inclined to keep near analogies irrespective of the interface. Despite a weak association identified (1.5x less likely to keep analogies (see fig. 5.4)), our intervention did not seem to have a significant effect on the participants' deci-

Table 5.4: GLMM Null model for Keep Decisions

	Dependent variable:
	initial_decision_keep
Constant	1.077***
	(0.564, 1.589)
Observations	263
Log Likelihood	−149.790
Akaike Inf. Crit.	303.581
Bayesian Inf. Crit.	310.725
Note:	*p<0.1; **p<0.05; ***p<0.01

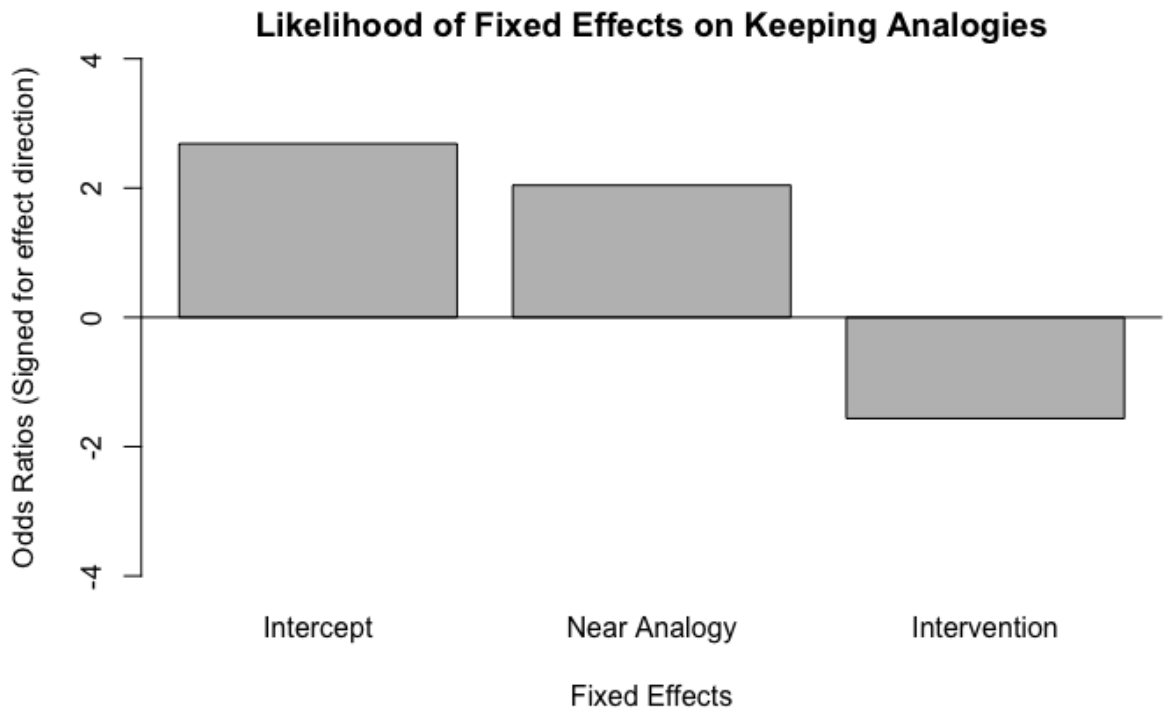


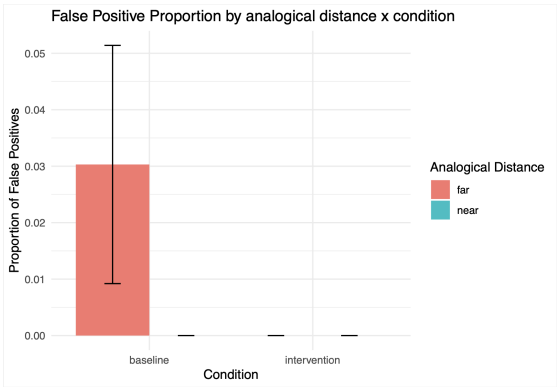
Figure 5.4: Plotting the log odds of fixed effects on keeping analogies (signed for directionality)

Table 5.5: GLMM Analogy Distance model for Keep Decisions

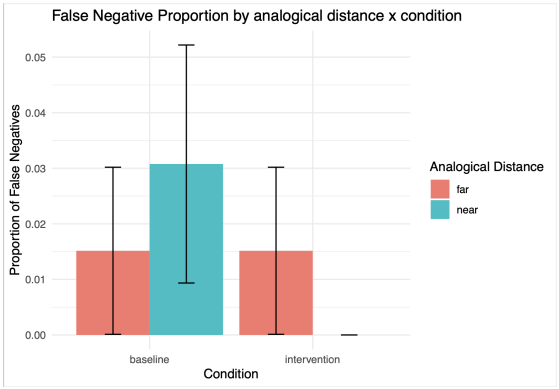
	Dependent variable:
	initial_decision_keep
Near Analogy	0.706**
	(0.117, 1.295)
Constant	0.755**
	(0.174, 1.336)
Observations	263
Log Likelihood	−147.010
Akaike Inf. Crit.	300.020
Bayesian Inf. Crit.	310.736
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.6: GLMM model with Analogy Distance and Experimental Condition for Keep Decisions

	Dependent variable:
	initial_decision_keep
Near Analogy	0.715^{**}
	(0.122, 1.307)
Intervention Condition	-0.448
	$(-1.034, 0.139)$
Constant	0.988^{***}
	(0.322, 1.654)
Observations	263
Log Likelihood	-145.902
Akaike Inf. Crit.	299.804
Bayesian Inf. Crit.	314.093
Note:	*p<0.1; **p<0.05; ***p<0.01



(a) Plot of *Ignore → Keep/Review*



(b) Plot of *Keep → Ignore/Review*

Figure 5.5: Raw plots of *Ignore → Keep/Review* and *Keep → Ignore/Review* across Analogical Distance and Experimental Condition

sions to keep analogies. The plots (see fig. 5.3) showed that participants seemed to have a slightly higher tendency to keep analogies in the baseline interface than in our proposed intervention.

However, upon closer inspection on the data through qualitative analysis of think-alouds, we discovered that this quantitative analysis of kept analogies did not take into consideration the way ”review” decisions were utilized. From the quotes, we identified at-least 6 instances where the participants explicitly stated their plans to process analogies with our screening interface. Of these participants, some shared that they chose this approach to address the constraints on time (P686020, P483347) in processing analogies during the intervention while others felt that they could only screen for relevant analogies when actively thinking about their applicability (P108688, P328365, 948947, P820332). For instance, participant P328365 said, *“I just want to keep it as review because after looking at the other analogies, I might actually make some exposition”*. It is also important to note, if these participants felt sure about the irrelevance of an analogy in solving a problem, they ignored them in the screening interface. We believe that this might explain the differences we found between our quantitative and qualitative observations.

While we attempted to explore the effects of our intervention on *Ignore → Keep/Review*

(when a participant keeps an analogy only to change it back to review or keep) and *Keep* $\rightarrow$ *Ignore/Review* (when a participant ignores an analogy only to change it back to keep or review), we could not perform statistical analysis through GLMM due to lack of observations that could explain for potential experimental variation (2 *Ignore* $\rightarrow$ *Keep/Review* and 4 *Keep* $\rightarrow$ *Ignore/Review* throughout the entire dataset of 263 observed decisions).

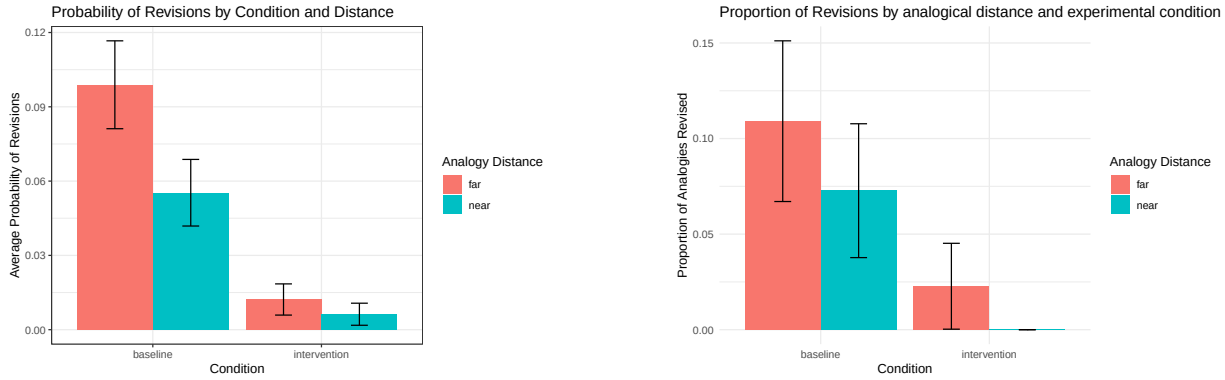
Looking into this qualitatively, we found that a lot of participants who identified the analogical relationships between the assigned design problem and a far analogy, but did not follow-through with their decision during the ideation phase in the baseline interface. Looking for patterns, we found instances where the participants doubted the relevance of far-domain analogies in the baseline interface. For example, when exposed to the lane-assist far analogy for designing a marine environment, participant P598108 initially said, *"I can see some sort of concept of like an outside force supervising physical interactions between creatures...I feel like that would kind of defeat the purpose of having kids interact... So, I ignore?"* but later changed their decision: *"I'm gonna go back... if physically swimming in water for example, physical guard rails or technology could help"*.

Due to the sparsity of variations described in data, a more comprehensive analysis was performed where any changes in decision were coded as *revisions*, including *Ignore* $\rightarrow$ *Keep/Review*, *Keep* $\rightarrow$ *Ignore/Review* and from *Review* $\rightarrow$ *Keep/Ignore*. A GLMM Analysis was performed to fit over the changes in decision making where we replaced l in our mixed model equation (see eqn. 4.6.1) with *revisions* as a the response variable, resulting in models with no predictors (see table 5.7), analogical distance as the predictor (see table 5.8) and analogical distance with condition as predictors (see table 5.9).

Interpreting the goodness of fit between these models, we found that people made revisions

Table 5.7: Null Model for fitting changes in decision (revisionss)

	Dependent variable:
	Revisions
Constant	-3.649^{***}
	(0.788)
Observations	195
Log Likelihood	-39.997
Akaike Inf. Crit.	83.994
Bayesian Inf. Crit.	90.540
Note:	*p<0.1; **p<0.05; ***p<0.01



(a) Model-based Plot

(b) Raw Plot of Proportions

Figure 5.6: Plots describing the Tendency to Revise Analogies

Table 5.8: GLMM Model for fitting changes in decision (revisions) with Analogical Distance as Predictor

	Dependent variable:
	Revisions
Near Analogy	−0.590
	(0.677)
Constant	−3.401***
	(0.824)
Observations	195
Log Likelihood	−39.605
Akaike Inf. Crit.	85.211
Bayesian Inf. Crit.	95.030
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.9: GLMM Model for fitting changes in decision (revisions) with Analogical Distance and Condition as Predictor

	Dependent variable:
	Revisions
Near Analogy	-0.682
	(0.718)
Intervention Condition	-2.525^{**}
	(1.131)
Constant	-3.015^{***}
	(0.941)
Observations	195
Log Likelihood	-35.500
Akaike Inf. Crit.	79.000
Bayesian Inf. Crit.	92.092
Note:	*p<0.1; **p<0.05; ***p<0.01

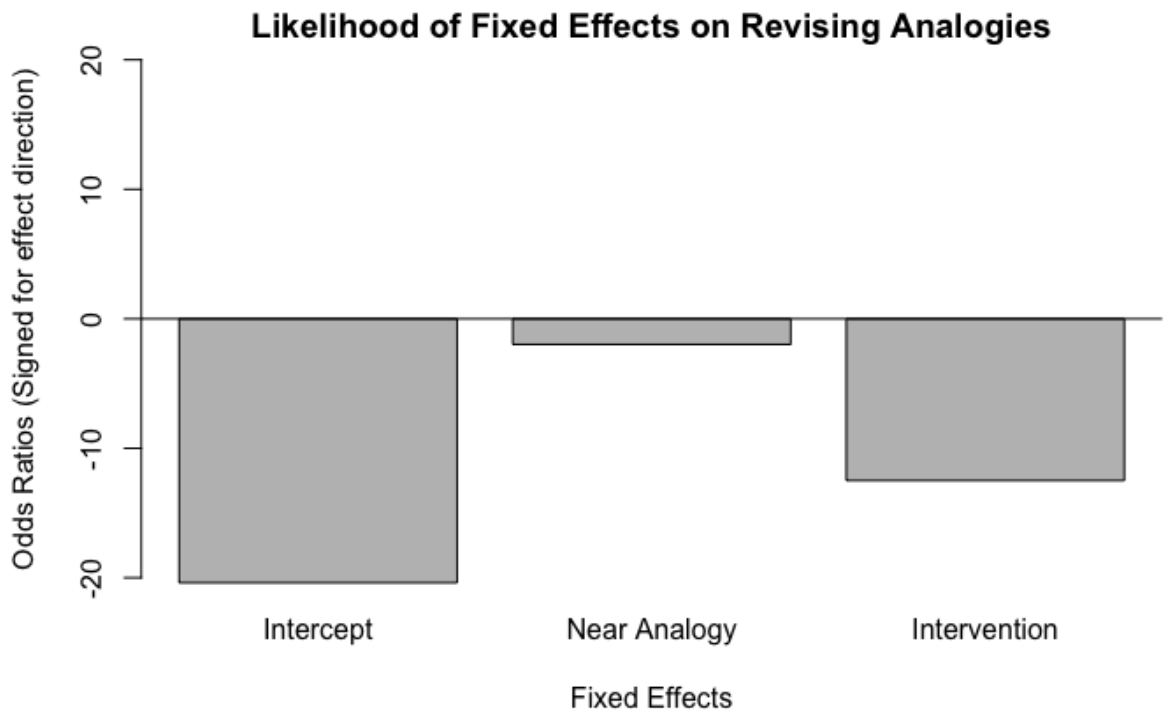
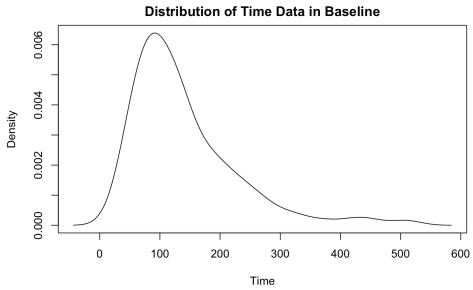
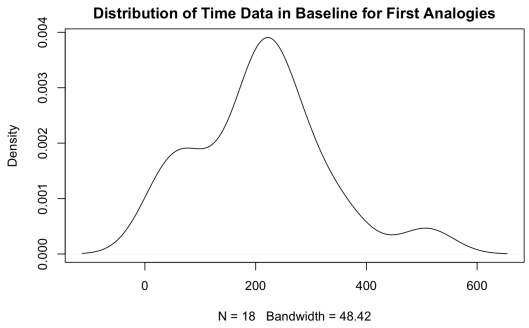


Figure 5.7: Plotting the log odds of fixed effects on mistaking analogies (signed for directionality)

in their initial decision making irrespective of the analogical distance of the recommended leads. We also found that our intervention had a significant effect on the number of revisions committed. By examining the likelihood of the fixed effects (see fig. 5.7) we found that, when interacting with our intervention, the participants were 12x less likely to change their decision about an analogical lead. Looking at the model and plots of proportion (see fig. 5.6), we uncover a pattern where participants were less ambiguous about selecting near analogical leads than farther ones, although unable to substantiate the same in our Quantitative Analysis through the modelling of interaction effects due to insufficient data. Thus, these results show that, while people were generally more inclined to doubt and possibly ignore the relevance of a potential analogical lead, they were less likely to do so when using our intervention.

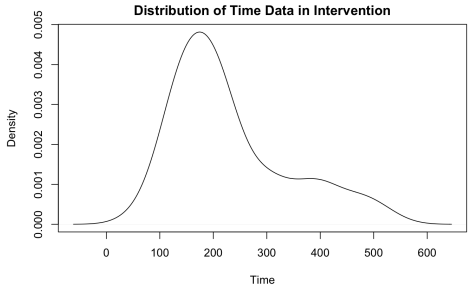


(a) Time Distribution for Baseline

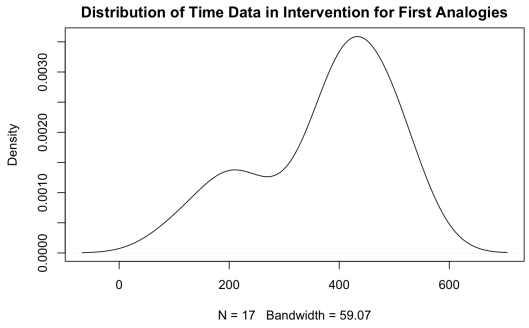


(b) Time Distribution for First Analogies in Baseline

Figure 5.8: Plots describing the time taken to process analogies in Baseline



(a) Time Distribution for Intervention



(b) Time Distribution for First Analogies in Intervention

Figure 5.9: Plots describing the time taken to process analogies in Intervention

5.1.2 Processing Time of Analogies

After accounting for the missing values and data, we found that on average, our participants spent an excess of 1.7 minutes across the interfaces (54 instances in intervention, with 25 in baseline). We also found that this occurred mostly when the participants interacted with the interface for the first time (29 instances found across 22 participants), which we believed to be due to a lack of familiarity. To verify the same, we plotted the time taken to process analogies for the first time for each interface (see time distributions of baseline (fig. 5.8) and intervention (fig. 5.9)) and found that, while participants across both interfaces spent more time during their

first interaction, on average, they took 48.5 seconds longer to process analogies when using our proposed interface for the very first time in comparison with the baseline.

Table 5.10: GLMM of Null Model for effects on processing time of decisions

	Dependent variable:
	Seconds Per Analogy
Constant	184.170^{***}
	(165.093, 203.247)
Observations	196
Log Likelihood	−1,193.095
Akaike Inf. Crit.	2,392.190
Bayesian Inf. Crit.	2,402.024
Note:	*p<0.1; **p<0.05; ***p<0.01

To further assess the relationships between processing time, analogical distance and experimental condition, we replaced l in our base mixed model equation (see equation 4.6.1) with a continuous variable *Seconds Per Analogy* that takes the discrete continuous values of time-differences as its value v, resulting in Null models for Timing (see 5.10), Models with Distance as predictor for Timing (see 5.11) and finally, Distance with Condition as predictors for Timing (see 5.12). The results of these models suggests a significant relationship with the intervention (see 5.10), aligning with the preliminary observations. On average, we found 0.65 times increase in the time spent processing each analogy in intervention in comparison with the baseline (i.e. 1.5 minutes on average). We also found that analogical distance had no association with the time

Table 5.11: GLMM for effects of analogical distance on processing time of decisions

	Dependent variable:
	Seconds Per Analogy
Near Analogy	11.351
	(−18.328, 41.030)
Constant	178.538***
	(154.253, 202.824)
Observations	196
Log Likelihood	−1,189.180
Akaike Inf. Crit.	2,386.359
Bayesian Inf. Crit.	2,399.471
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.12: GLMM for effects of Analogical Distance and Experimental Condition on processing time of decisions

	Dependent variable:
	Seconds Per Analogy
Near Analogy	6.311
	(−20.414, 33.036)
Intervention Condition	89.730***
	(63.035, 116.424)
Constant	138.105***
	(112.014, 164.195)
Observations	196
Log Likelihood	−1,166.130
Akaike Inf. Crit.	2,342.259
Bayesian Inf. Crit.	2,358.650
Note:	*p<0.1; **p<0.05; ***p<0.01

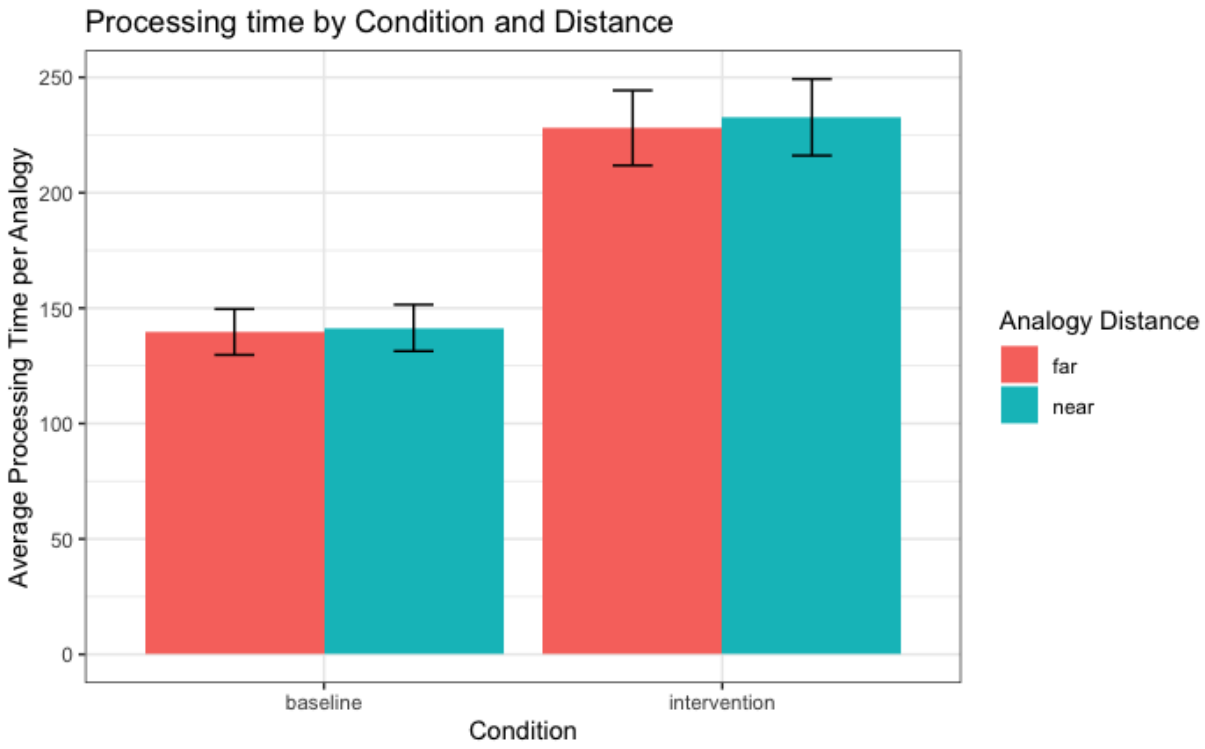


Figure 5.10: Model Plot of time taken to process analogies across Condition and Analogical Distance

taken to process them.

5.2 Qualitative Analysis of Interactions with Analogies

5.2.1 How did people interact with the chunks in our intervention?

Through qualitative analysis of participants' interactions during the screening interface, we found differences in the ways in which participants processed the analogies. In particular, when participants interacted with our intervention, some participants felt that they were more focused on chunks when interacting with our intervention interface. For instance, participant P598108 said *"I, seeing all the words separated out for the second, that, like they were in the second one, made it a little bit hard to for me to visualize like the entire situation, I kind of like the first one a bit better"*. Some participants found that helpful and helped them feel less fixated on the sentence structure. For example, while hiding the *stopwords* in the intervention interface, participant P427161 said '*If I have the stopwords, I feel like it's a full sentence, and it kind of hinders my creativity somehow*'. On the other hand, some participants felt that words prevented them from getting a *"holistic view"* of the analogies as there were a "*lot of things going on*" (P175527) while others attributed it to *"visual clutter"*. For example, P526900 said *I see much more value for magnets... but then the visual clutter on that screen was distracting, but the magnets grabbed the key idea* while participant P948947 said, '*the connections between things... I was a little too focused on the words and less focused on the content, or the contact and content of the ideas from the analogies*'. In general, most participants felt that the intervention was intuitive to a certain degree, and felt that their processing became quicker as they became more familiar with it. A few participants also mentioned that they could intuitively understand the relationship and map connections between different components across the design problem and

recommended analogies thanks to the color-coding of components (P427161, P505404, P598108 to name a few). For instance, participant P948947 said *"It felt like a friendlier interface. But it wasn't immediately intuitive to me... the colors had a definite meaning and that was helpful to me*.

5.2.2 How did people interact with the recombinations in our intervention?

Participants were generally impressed by the quality of recombinations that were generated using large language models. However, when the recombinations were either "*too creative*" or "*repetitive*", participants were less inclined to continue processing. For instance, one participant (P727211) said "*it was helping me to put boundaries around (the problem), but it also caused me to spend some time on things that just were non-starters*". The affective responses of participants we processed ranged from them feeling *humoured* and *inspired* to them feeling *frustrated*. For instance, participant P328365 said, *"(HAHAHA) baby carrying enthusiasts.. I kind of like I didn't think that might be a group... If that is... I want to research that!"* after being inspired by a recombination to add new stakeholders to the design problem. On the other hand, P948947 felt frustrated by poor, repetitive generations and felt less inclined to process an analogical lead further.

Overall, even though they felt that the baseline helped them process analogies quicker, they felt more engaged with interacting with the intervention often mentioning that they would have loved to explore more if not for the constraint of time (P328365, P225318, P259274 to name a few). One participant (P408230) explicitly stated the impact the interfaces had on their ideation process: *"I think, once I found that idea I wanted, I was pretty much stuck on it, so (baseline) did*

help me think more to devote that idea. I wouldn't say it was easier for me, but it helped me think more deeply. In (intervention), I was exposed to more ideas". Another participant (P598108) highlighted the usefulness of our intervention '*..when I was moving the magnets around... I did like that! The questions appeared because they did make me to think about things a little bit differently'*.

Chapter 6: Discussion

The discussion section provides an overview of the study conducted to explore how interfaces can support analogical innovation by incorporating chunking and recombination. The study included 23 participants who were given design problems to solve with the help of recommended analogies. The analysis focused on how our interface impacted participants' selection of relevant analogical leads to come up with creative solutions. The findings suggest that the proposed interface positively impacted the users' interactions with the system and their triaging decisions, resulting in ideas that integrated solutions across different analogies. Participants found the proposed interface intuitive and engaging, and felt that it facilitated the generation of ideas. However, prior knowledge constrained how the participants benefited from the analogies.

Participants' comfort and experience with visual representation seemed to have an impact on their perception of our intervention's usefulness. For instance, participant P175527 expressed discomfort with the magnets interface, preferring a more familiar block of text ('*It felt like with the magnets there's like too much going on at once for me. I'm more familiar with just blocks of text at this point*'). In contrast, participant P344424, experienced in software engineering, found the pace of the brainstorming with our intervention to be fast and enjoyable (*"If it was at the pace that we were going at today, I like the kind of fast brainstorming (with intervention). And I work at a software company with most convoluted interfaces. So I wouldn't take it as*

representative, but I thought it was fine for me"). Participant P328365, an older participant with prior brainstorming experience, had difficulty adjusting to the magnets interface but recognized its underlying inspiration ('*You know, maybe, if it been physical, like the magnets would maybe be a little more, you know, like the fridge poetry stuff definitely, that sparks a lot of creativity for myself. But I wasn't really used to it on screen*').

Prior knowledge, including analogies from previous studies, personal experience, and professional expertise, also affected participants' interaction with our intervention. Participants often ignored analogies that they found repetitive or aversive. For example, when designing a safe marine environment for children and marine animals, participants P505404 and P408230 ignored the shark cage analogy due to its association with the controversy at the Mexican border.

However, our intervention allowed for the identification of more abstract relationships during screening, leading to interesting ideas during ideation. Some participants who had previously ignored repetitive relationships felt more inclined to integrate them when exposed to prior analogies for a different problem in the intervention. For instance, participant P177527 integrated prior analogies about submersible devices and dolphin trainers, recognizing the usefulness of vehicles for facilitating safe interactions in a marine environment. Similarly, an expert from the industry who had rejected far analogies due to their "draconian assumptions" integrated the air mattress and sand analogies when designing for travel, recognizing the resources that are available everywhere.

In conclusion, our study investigated how interfaces can support analogical innovation by incorporating chunking and recombination. The results showed that our proposed interface positively impacted users' interactions with the system, leading to ideas that integrated solutions across different analogies. However, prior knowledge constrained how participants benefited

from the analogies, although there were instances where some participants broke through this mindset when interacting with our proposed interface with the aid of recombinations generated by large language models. Our proposed interface was also perceived as intuitive and engaging, facilitating an exploratory mindset and the generation of diverse ideas that integrated different aspects of the screened analogies.

It is important to note that our study had many limitations, and further research is needed to explore the full potential of designing interfaces that support triaging for analogical innovation. Nonetheless, our study contributes to the broader research area of human-computer interaction and opens up possibilities for designing interfaces that better support analogical innovation in design.

6.1 Limitations and Future Work

As our study had a relatively diverse population pool of participants across a range of ages, genders, cultures, ethnicities, and education levels, including people from academia and the industry, we believe that our findings on their interactions with our interface can be generalized to a broader population. However, the present study had many limitations.

Firstly, due to the small set of problems, our study could not say anything about the changes in interaction across different kinds of problems. During our debriefing interviews with a couple of participants, we got a hint of a possible relationship. For instance, while interacting with our intervention, participant 175527 said, "*...I felt like there was a logical end for myself versus like, the children interacting with the aquatic animals like, there's one million things that that I could go with*".

In addition, having a limited number of participants meant higher difficulty for GLMM models to converge, particularly when looking for significant interaction effects. Future work can explore on how different design problems with multiple number of analogies (with varying degrees of diversity) can help understand the effectiveness of our proposed interface in selecting analogical leads, if possible, with a larger population pool.

Furthermore, several participants felt that the current study which constrained them to integrate and generate recombinations for only one analogy at a time was too limiting. For example, participant 328365 said, *"If it was the ability to generate more unique words when I was playing in that one instance versus just moving on doing, moving on, I think it would have been. It would have led me to different places."*. In addition, from a design perspective, although inviting, our intervention did not take into consideration the accessibility and usability for color-blind participants, which can also be a possible avenue for later evaluation.

Secondly, the present study design gave great importance to the screening interface while the ideation interface only served a means to record changes in decision. Since the ideation in the present study design could integrate multiple analogies, quantity alone could not describe the impacts of the screening process on ideation. Furthermore, due to the within-subject nature of this study, the impact of ideation might not be isolated to a particular interface. Future studies can explore this by adding a way for the participants to self-rate their ideas and using a between-subjects design might give better insights on the same. Qualitatively, in the scope of this thesis, only a thematic analysis was performed on the the think-alouds and interviews of participants. A more rigorous qualitative analysis on the collected data with inter-rater reliability testing can help in improving this analysis even further, currently outside the scope of this thesis

Additionally, participants frequently complained about the occasional nonsensical or repet-

itive generations. This happened because this study used `text-davinci-002` with lower temperature to ensure some degree of control, enabling the participants to have control over the generations of the model. However, lower temperature led to deterministic but less creative generations. When we tested our exact prompt in later versions of GPT (GPT-3 and ChatGPT), we found that we were able to generate guiding questions that were more informative, intuitive, and proactive in nature. Furthermore, given the dialogue-like nature of ChatGPT, we could simplify the interface facilitating iterative recombination (i.e. using prior recombinations as starting points to generate new ones). It would be interesting to see how such a system would offer a better way to triage analogies more quickly and effectively.

6.2 Implications

While the aforementioned insights from the participants pave the way for further improving this interface, they also point towards a possible future, where innovations can be accelerated through effective and efficient analogical reasoning through computational support. Instead of having problems baked-in such as the ones in the current study, a proposed system could source analogical problems relevant to a source problem provided by a participant through analogical search mechanisms (similar to those proposed by Chan et al. (2018) and Kang et al. (2022)) while integrating large language models to create conversational systems of computational support that foster analogical innovation. If taken in a sociotechnical context, interesting generated recombinations could be used as conversation starters when trying to tie in the ideas across people and disciplines. If implemented, such a system could cause a paradigm shift in the ways in which current scientific progress is achieved. Later studies could explore how participants of such as

system could interface within a larger sociotechnical structure. If we could achieve significant insights and breakthroughs in understanding this, Information and Knowledge silos could become a thing of the past, ultimately breaking down boundaries and accelerating innovation.

Appendix A: R Code for Quantitative Analysis

A.1 R Code for Qualitative Analysis of Decisions

Quantitative Analysis of Decisions

Arvind Srinivasan

2023-03-12

Qualitative Analysis

Preprocessing for Analysis

First, load the collected dataset of decisions (`decisions.csv`) to begin processing.

```
library(readr)
setwd("~/Research/exp-analogy-triage/analysis/quantitative")
decisions <- read_csv("decisions.csv")
```

```
library(tidyr)
decisions$condition = as.factor(decisions$condition)
decisions$interface = as.factor(decisions$interface)
decisions$analogy_distance = as.factor(decisions$analogy_distance)
decisions$first_decision_logged = as.factor(decisions$first_decision_logged)
decisions$final_decision = as.factor(decisions$final_decision)
summary(decisions)
```

```
##  participantID      problemID    analogy_order   analogy_shortname
##  Min.   :108688   Min.   :1.00   Min.   :1.000   Length:550
##  1st Qu.:328365   1st Qu.:1.00   1st Qu.:2.000   Class :character
##  Median :483347   Median :2.00   Median :4.000   Mode  :character
##  Mean   :481322   Mean   :2.04   Mean   :3.509
##  3rd Qu.:649102   3rd Qu.:3.00   3rd Qu.:5.000
##  Max.   :948547   Max.   :3.00   Max.   :6.000
##        condition       interface       key            analogy_distance
##  baseline    :274   ideation :275   Length:550         far :276
##  intervention:276   screening:275   Class :character   near:274
##                                     Mode  :character
##
##
##
##  first_decision_logged final_decision     review           keep
##  ignore: 32            ignore: 56     Min.   :0.0000   Min.   :0.0000
##  keep  :200            keep  :394     1st Qu.:0.0000   1st Qu.:0.0000
##  review:318            review:100     Median :1.0000   Median :0.0000
##                                       Mean   :0.5782   Mean   :0.3636
##                                       3rd Qu.:1.0000   3rd Qu.:1.0000
##                                       Max.   :1.0000   Max.   :1.0000
##      ignore        decision_changed    rationale
##  Min.   :0.00000   Length:550         Length:550
##  1st Qu.:0.00000   Class :character   Class :character
##  Median :0.00000   Mode  :character   Mode  :character
##  Mean   :0.05818
```

```
##  3rd Qu.:0.00000
##  Max.   :1.00000
```

```
sum(is.na(decisions))
```

```
## [1] 158
```

Before performing the qualitative analysis, we clean up the columns.

Looking at the data, we find that no decisions are logged for `313394` across all analogies in both interfaces. Hence, this data does nothing but adds non-useful data that might affect the observed effects if taken into consideration.

```
# 12 analogies in baseline and 12 analogies in intervention results in a total of 24 analogies.
# We now check if the decisions across these interfaces don't vary.
nrow(subset(decisions, participantID == 313394 & first_decision_logged == final_decision)) == 24
```

```
## [1] TRUE
```

```
# if TRUE, then this data does not help in processing and needs to be removed.
# if FALSE, then this data can be kept.
```

```
df <- data.frame(decisions)
df <- subset(decisions, participantID != 313394 )
# remove unnecessary columns
df <- within(df, rm(keep, review, ignore, decision_changed, rationale, key))
# rename the decision columns to faciliate one-hot encoding
colnames(df)[8] <- "initial_decision_"
colnames(df)[9] <- "final_decision_"
```

Now we perform **One Hot Encoding** of the **initial** and **final** decisions.

```
# one-hot encode the "initial_decision" column
encoded_df <- model.matrix(~ initial_decision_ - 1, data = df)
df <- cbind(df, encoded_df)

# one-hot encode the "final_decision" column
encoded_df <- model.matrix(~ final_decision_ - 1, data = df)
df <- cbind(df, encoded_df)

# remove the decision columns
df <- within(df, rm(initial_decision_, final_decision_))
```

Code for the change in decisions by comparing the `initial_decision` and `final_decision` columns. If the `initial_decision` is `ignore`, code it as a `ignore_to_keep.review`. If the `initial_decision` is `keep`, code it as a `keep_to_ignore.review`.

```
# create a new column for false negative results
df$decision_ignore_to_keep.review <- ifelse(
  df$initial_decision_ignore == 1 & df$final_decision_ignore != 1,
  1, 0)

# create a new column for false positive results
df$decision_keep_to_ignore.review <- ifelse(
  df$initial_decision_keep == 1 & df$final_decision_keep != 1,
  1, 0)
```

This will be later used for analyzing the changes to decisions across the different interfaces. For the convenience of analysis, split the interface by their categories into their own data frames.

```r
# split the data frame into based on interface
fit.screening <- subset(df, interface == "screening")
fit.ideation <- subset(df, interface == "ideation")
```

```r
library(vtable)
```

```
## Loading required package: kableExtra
```

```r
sample_df = as.data.frame(fit.screening)
sample_df = within(sample_df, rm(participantID, problemID, analogy_order))
st(sample_df, out="latex", summ = list(
     c('sum(x, na.rm=TRUE)','mean(x)','sd(x)','min(x)','max(x)')
   ),
   summ.names = list(
     c('Valid N','Mean','SD','Min','Max')
   ))
```

Table 1: Summary Statistics

Variable	Valid N	Mean	SD	Min	Max
condition	263				
... baseline	131	50%			
... intervention	132	50%			
interface	263				
... ideation	0	0%			
... screening	263	100%			
analogy_distance	263				
... far	132	50%			
... near	131	50%			
initial_decision_ignore	30	0.11	0.32	0	1
initial_decision_keep	187	0.71	0.45	0	1
initial_decision_review	46	0.17	0.38	0	1
final_decision_ignore	28	0.11	0.31	0	1
final_decision_keep	198	0.75	0.43	0	1
final_decision_review	37	0.14	0.35	0	1
decision_ignore_to_keep.review	4	0.015	0.12	0	1
decision_keep_to_ignore.review	2	0.0076	0.087	0	1

Now, these data frames are ready for analysis.

Looking at Initial Ignore Decisions in Screening

First, we define a **null model with no predictors** to analyze how participants as a group **initially ignored analogies** during screening. This will help us see the improvement in fit as we add predictors to identify their effects.

```r
# load the package
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```r
# define the null model
fit.screening.ignore.null <- glmer(
                      as.factor(initial_decision_ignore)
                      ~ (1| participantID),
                    data = fit.screening,
                    family = binomial())
```

```
## boundary (singular) fit: see help('isSingular')
```

```r
# show summary of model
summary(fit.screening.ignore.null)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: as.factor(initial_decision_ignore) ~ (1 | participantID)
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##    190.7    197.8    -93.3    186.7      261
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.3588 -0.3588 -0.3588 -0.3588  2.7869
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 0        0
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.050      0.194  -10.57   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

Adding analogical_distance as a predictor

```r
# load the package
library(lme4)

# define the model adding analogical distance as a predictor
```

Table 2:

	Dependent variable:
	as.factor(initial_decision_ignore)
Constant	-2.050^{***}
	$(-2.430, -1.670)$
Observations	263
Log Likelihood	-93.349
Akaike Inf. Crit.	190.697
Bayesian Inf. Crit.	197.842
Note:	*p<0.1; **p<0.05; ***p<0.01

```
fit.screening.ignore.nearfar <- glmer(
                      as.factor(initial_decision_ignore)
                      ~ (1| participantID)
                      + as.factor(analogy_distance),
                    data = fit.screening,
                    family = binomial())
```

```
## boundary (singular) fit: see help('isSingular')
```

```
# show summary of model
summary(fit.screening.ignore.nearfar)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## as.factor(initial_decision_ignore) ~ (1 | participantID) + as.factor(analogy_distance)
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##    182.7    193.5    -88.4    176.7      260
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.4594 -0.4594 -0.2376 -0.2376  4.2088
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 0        0
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      -1.5559     0.2295  -6.780  1.2e-11 ***
## as.factor(analogy_distance)near  -1.3185     0.4512  -2.922  0.00347 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
```

```
##             (Intr)
## as.fctr(n_) -0.509
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

Table 3:

	Dependent variable:
	as.factor(initial_decision_ignore)
as.factor(analogy_distance)near	-1.319^{***}
	$(-2.203, -0.434)$
Constant	-1.556^{***}
	$(-2.006, -1.106)$
Observations	263
Log Likelihood	−88.371
Akaike Inf. Crit.	182.742
Bayesian Inf. Crit.	193.459
Note:	*p<0.1; **p<0.05; ***p<0.01

From the above data, we can see that there is a **near analogies** have a **significant inverse relationship** on a **participant's tendency to ignore.** In other words, participants in this data have a tendency to ignore far analogies.

Use **ANOVA** to check for improvements in fit over the null model.

```
eval.nearfar.null.fit = anova(
    fit.screening.ignore.nearfar,
    fit.screening.ignore.null,
    test='LRT')

eval.nearfar.null.fit
```

```
## Data: fit.screening
## Models:
## fit.screening.ignore.null: as.factor(initial_decision_ignore) ~ (1 | participantID)
## fit.screening.ignore.nearfar: as.factor(initial_decision_ignore) ~ (1 | participantID) + as.factor(a
##                              npar    AIC    BIC  logLik deviance Chisq Df
## fit.screening.ignore.null       2 190.70 197.84 -93.349   186.70
## fit.screening.ignore.nearfar    3 182.74 193.46 -88.371   176.74 9.955  1
##                              Pr(>Chisq)
## fit.screening.ignore.null
## fit.screening.ignore.nearfar   0.001604 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This shows that, the adding the `analogical_distance` as predictor improves the fitness of data.

Exploring the effects of `condition` with `analogical_distance`

Hypothesizing that this effect of `analogical_distance` differs between the `baseline` and our `intervention` interface, we add `condition` as an additional predictor.

```r
# load the package
library(lme4)

# define the model adding analogical distance as a predictor
fit.screening.ignore.nearfar.x.condition <- glmer(
                      initial_decision_ignore
                      ~ (1| participantID)
                      + analogy_distance
                      + condition,
                    data = fit.screening,
                    family = binomial())

# show summary of model
summary(fit.screening.ignore.nearfar.x.condition)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: initial_decision_ignore ~ (1 | participantID) + analogy_distance +
##     condition
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##    174.2    188.5    -83.1    166.2      259
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.6096 -0.3054 -0.2991 -0.1519  3.3437
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 0.02157  0.1469
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -1.0245     0.2822  -3.631 0.000282 ***
## analogy_distancenear    -1.3658     0.4619  -2.957 0.003107 **
## conditionintervention   -1.3821     0.4612  -2.997 0.002727 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) anlgy_
## anlgy_dstnc -0.426
## cndtnntrvnt -0.430  0.067
```

```r
sign(fixef(fit.screening.ignore.nearfar.x.condition)) * exp(abs(fixef(fit.screening.ignore.nearfar.x.co
```

```
##           (Intercept)  analogy_distancenear conditionintervention
##             -2.785834             -3.918746             -3.983077
```

```r
barplot(sign(fixef(fit.screening.ignore.nearfar.x.condition)) * exp(abs(fixef(fit.screening.ignore.near
abline(a=0,b=0)
```

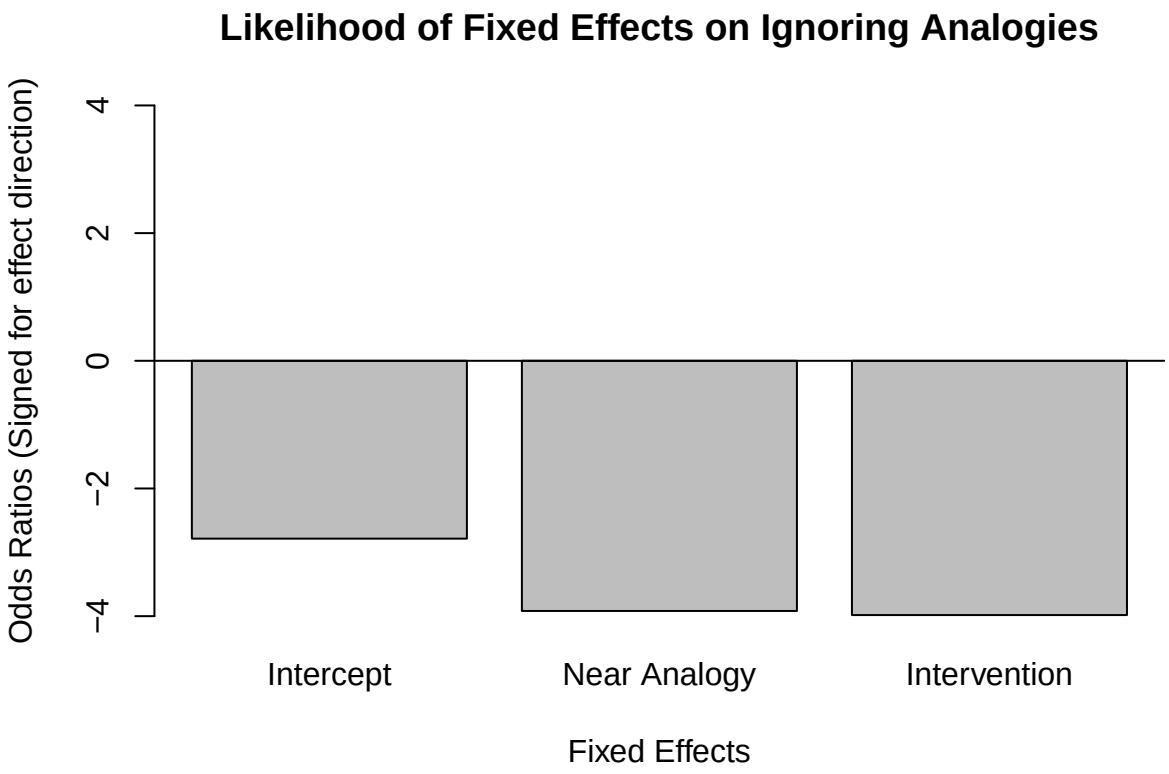


Table 4:

	Dependent variable:
	initial_decision_ignore
analogy_distancenear	-1.366^{***}
	$(-2.271, -0.460)$
conditionintervention	-1.382^{***}
	$(-2.286, -0.478)$
Constant	-1.025^{***}
	$(-1.578, -0.472)$
Observations	263
Log Likelihood	−83.104
Akaike Inf. Crit.	174.208
Bayesian Inf. Crit.	188.497
Note:	*p<0.1; **p<0.05; ***p<0.01

```
library(ggplot2)
library(dplyr)
library(tidyr)

fit.screening.ignore.nearfar.condition.plot <- select(fit.screening, analogy_distance, condition)
```

```r
pred_prob <- predict(fit.screening.ignore.nearfar.x.condition, type = "response")
fit.screening.ignore.nearfar.condition.plot$probs <- pred_prob
fit.screening.ignore.nearfar.condition.plot$se_prop <- sqrt(pred_prob*(1-pred_prob)/nrow(fit.screening)

summarize_data <- function(data) {
  summary_data <- data %>%
    group_by(analogy_distance, condition) %>%
    summarise(avg_probs = mean(probs), avg_se_prop = mean(se_prop))
  return(summary_data)
}

plot_grouped_barchart <- function(data) {
  plot <- ggplot(data, aes(x = condition, y = avg_probs, fill = analogy_distance)) +
    geom_bar(stat = "identity", position = "dodge") +
    geom_errorbar(aes(ymin = avg_probs - avg_se_prop, ymax = avg_probs + avg_se_prop), width = 0.2, posi
    labs(title = "Probability of Ignoring Analogies by Condition and Distance",
         x = "Condition",
         y = "Average Probability of Ignoring Analogies",
         fill = "Analogy Distance") +
    theme_bw()
  return(plot)
}

data_summary <- summarize_data(fit.screening.ignore.nearfar.condition.plot)
plot_grouped_barchart(data_summary)
```

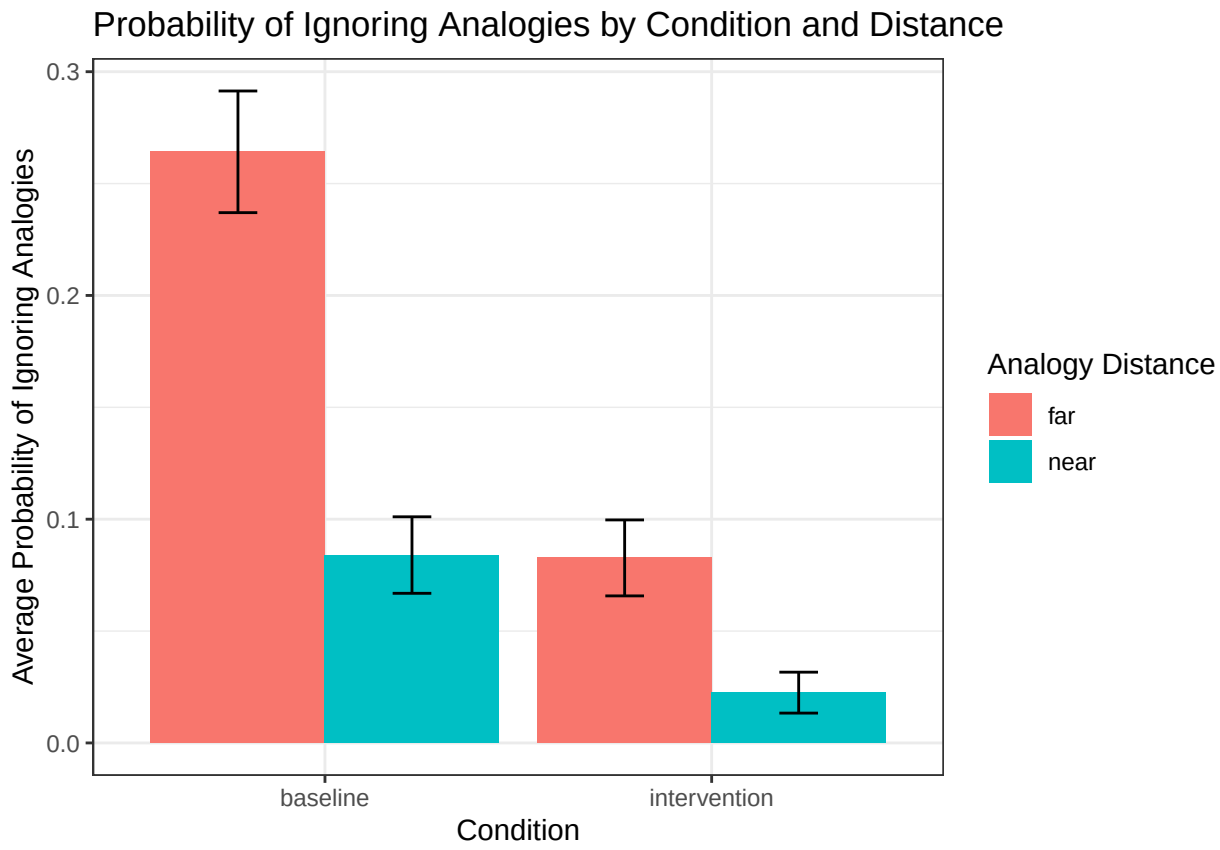


Table 5:

	Dependent variable:
	initial_decision_ignore
analogy_distancenear	-1.366^{***}
	$(-2.271, -0.460)$
conditionintervention	-1.382^{***}
	$(-2.286, -0.478)$
Constant	-1.025^{***}
	$(-1.578, -0.472)$
Observations	263
Log Likelihood	−83.104
Akaike Inf. Crit.	174.208
Bayesian Inf. Crit.	188.497
Note:	*p<0.1; **p<0.05; ***p<0.01

This shows that the participants had a decreased tendency to ignore analogies in the **intervention.**

Using **ANOVA** to check for improvements in fit over the model with only **analogical distance** as the predictor, we find that the adding condition as an additional predictor improves the goodness of fit of the model.

```
eval.nearfar.condition.fit = anova(
    fit.screening.ignore.nearfar.x.condition,
    fit.screening.ignore.nearfar,
    test='LRT')

eval.nearfar.condition.fit
```

```
## Data: fit.screening
## Models:
## fit.screening.ignore.nearfar: as.factor(initial_decision_ignore) ~ (1 | participantID) + as.factor(ar
## fit.screening.ignore.nearfar.x.condition: initial_decision_ignore ~ (1 | participantID) + analogy_dis
##                                          npar    AIC    BIC  logLik deviance
## fit.screening.ignore.nearfar                3 182.74 193.46 -88.371   176.74
## fit.screening.ignore.nearfar.x.condition    4 174.21 188.50 -83.104   166.21
##                                           Chisq Df Pr(>Chisq)
## fit.screening.ignore.nearfar
## fit.screening.ignore.nearfar.x.condition 10.534  1   0.001172 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing for Interaction Effects

```
library(lme4)

# define the model adding analogical distance as a predictor
fit.screening.ignore.nearfar.condition.interaction <- glmer(
                      initial_decision_ignore
```

```
                      ~ (1| participantID)
                      + analogy_distance
                      + condition
                      + analogy_distance:condition,
                    data = fit.screening,
                      family = binomial())
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Hessian is numerically singular: parameters are not uniquely determined
```

```
summary(fit.screening.ignore.nearfar.condition.interaction)
```

```
## Warning in vcov.merMod(object, use.hessian = use.hessian): variance-covariance matrix computed from
## not positive definite or contains NA values: falling back to var-cov estimated from RX
```

```
## Warning in vcov.merMod(object, correlation = correlation, sigm = sig): variance-covariance matrix co
## not positive definite or contains NA values: falling back to var-cov estimated from RX
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: initial_decision_ignore ~ (1 | participantID) + analogy_distance +
##     condition + analogy_distance:condition
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##    172.2    190.0    -81.1    162.2      258
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.5689 -0.3477 -0.3448  0.0000  2.9123
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 0.008319 0.09121
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##                                             Estimate Std. Error z value
## (Intercept)                                -1.142e+00  2.880e-01  -3.963
## analogy_distancenear                       -9.764e-01  4.930e-01  -1.980
## conditionintervention                      -9.933e-01  4.928e-01  -2.016
## analogy_distancenear:conditionintervention -2.866e+01  9.769e+05   0.000
##                                            Pr(>|z|)
## (Intercept)                                7.39e-05 ***
## analogy_distancenear                         0.0477 *
## conditionintervention                        0.0438 *
## analogy_distancenear:conditionintervention   1.0000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) anlgy_ cndtnn
```

```
## anlgy_dstnc -0.582
## cndtnntrvnt -0.582  0.340
## anlgy_dstn:  0.000  0.000  0.000
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## unable to evaluate scaled gradient
##  Hessian is numerically singular: parameters are not uniquely determined
```

Scaling of the values did not help (since the dependent variable was already one-hot encoded).

However if we look at the number of ignored analogies, we can see that the participants ignored only 30 analogies of 263 -leading me to believe that this sparsity of data, when processed with other factors defined for the condition due to a lack of variations in the experimental combinations, it could not converge or caused it to misconverge, resulting in large eigen values.

Visual Analysis

Given that there seems to be an interaction, we attempt to plot this relationship graphically.

```
library(pivottabler)

# get a pivot table of counts
pt.ignore.count <- PivotTable$new()
pt.ignore.count$addData(fit.screening)
pt.ignore.count$addColumnDataGroups("condition")
pt.ignore.count$addRowDataGroups("analogy_distance")
pt.ignore.count$defineCalculation(
  calculationName="TotalIgnored",
  summariseExpression="n()")
cat(pt.ignore.count$getLatex())
```

	baseline	intervention	Total
far	66	66	132
near	65	66	131
Total	131	132	263

```
# get a pivot table of means
pt.ignore.mean <- PivotTable$new()
pt.ignore.mean$addData(fit.screening)
pt.ignore.mean$addColumnDataGroups("condition")
pt.ignore.mean$addRowDataGroups("analogy_distance")
pt.ignore.mean$defineCalculation(
  calculationName="MeanIgnored",
  summariseExpression="mean(initial_decision_ignore, na.rm=TRUE)")
cat(pt.ignore.mean$getLatex())
```

	baseline	intervention	Total
far	0.242424242424242	0.106060606060606	0.174242424242424
near	0.107692307692308	0	0.0534351145038168
Total	0.175572519083969	0.053030303030303	0.114068441064639

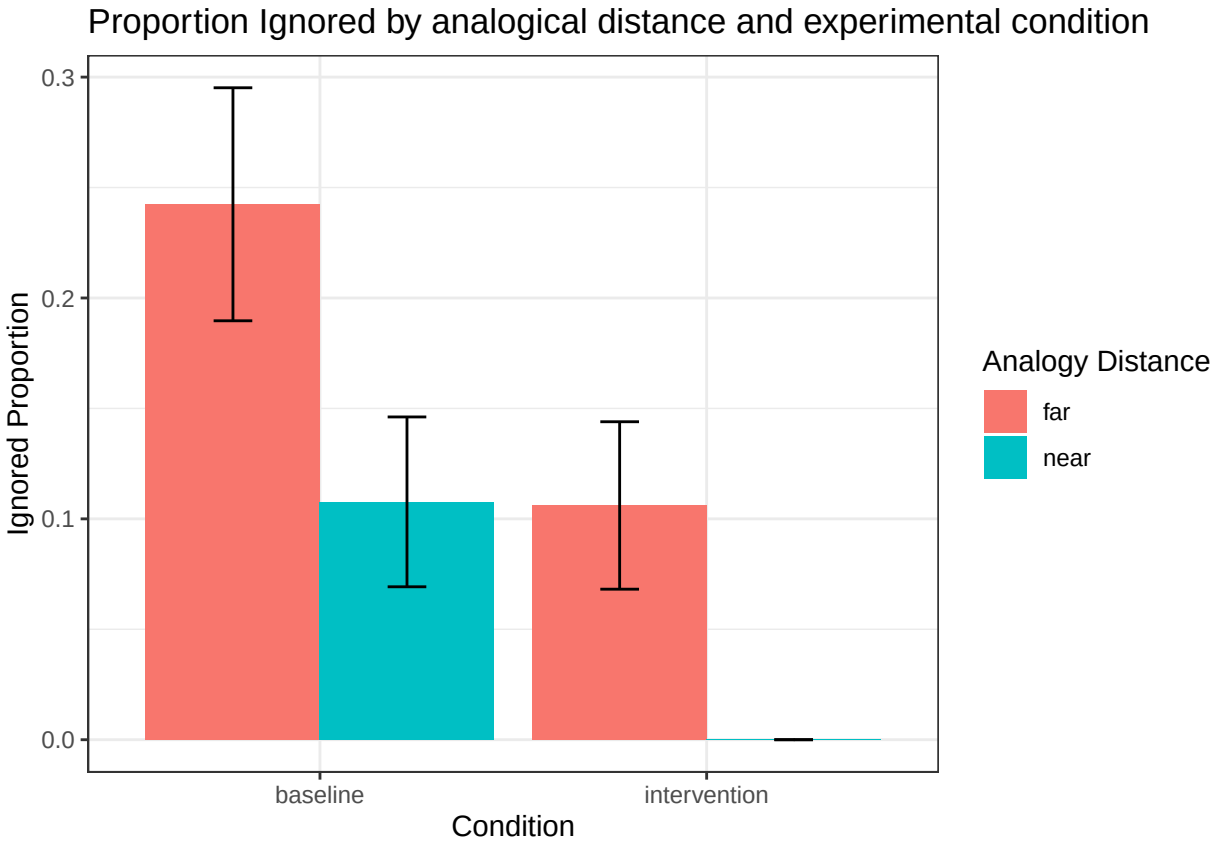
```
# function to compute standard error of proportion
se_prop <- function(p, n) {
  sqrt(p*(1-p)/n)
}
```

```r
# apply function to every cell
ignore.condition.error.pivot <- mapply(
    se_prop,
    ignore.condition.mean.pivot,
    ignore.condition.count.pivot)

# convert to dataframe
ignore.condition.error.pivot <- as.data.frame(ignore.condition.error.pivot)
knitr::kable(ignore.condition.error.pivot, "latex")
```

baseline	intervention	Total
0.0527508	0.0379017	0.0330154
0.0384497	0.0000000	0.0196496
0.0332406	0.0195049	0.0196022

```r
library(ggplot2)

# Create sample data frames
df_prob <- data.frame(
  row = c("far", "near", "Total"),
  baseline = ignore.condition.mean.pivot$baseline,
  intervention = ignore.condition.mean.pivot$intervention
)
# Remove the Total column from the probabilities
df_prob <- df_prob[df_prob$row != "Total", ]

# contains calculated error terms
df_error <- data.frame(
  row = c("far", "near", "Total"),
  baseline = ignore.condition.error.pivot$baseline,
  intervention = ignore.condition.error.pivot$intervention
)
# Remove the Total column from the error terms
df_error <- df_error[df_prob$row != "Total", ]

# Reshape data frames to long format
df_prob_long <- tidyr::pivot_longer(
  df_prob,
  cols = c("baseline", "intervention"),
  names_to = "column",
  values_to = "value")
df_error_long <- tidyr::pivot_longer(
  df_error,
  cols = c("baseline", "intervention"),
  names_to = "column",
  values_to = "value")

# Merge data frames
df_merged <- merge(df_prob_long,
                   df_error_long,
                   by = c("row", "column"))

# Plot the data using ggplot2
ggplot(df_merged, aes(x = column,
```

```
                    y = value.x,
                    fill = row)) +
geom_bar(position = "dodge",
         stat = "identity") +
labs(x = "Condition",
     y = "Ignored Proportion",
     fill = "Analogy Distance") +
ggtitle("Proportion Ignored by analogical distance and experimental condition") +
theme_bw() +
geom_errorbar(aes(
  ymin = value.x - value.y,
  ymax = value.x + value.y),
  width = 0.2,
  position = position_dodge(0.9))
```



Looking at Keep Decisions in Screening

First, we define a **null model with no predictors** to analyze how participants as a group **initially keep analogies** during screening. This will help us see the improvement in fit as we add predictors to identify their effects.

```
# load the package
library(lme4)

# define the null model
fit.screening.keep.null <- glmer(
```

```
                      as.factor(initial_decision_keep)
                      ~ (1| participantID),
                    data = fit.screening,
                    family = binomial())

# show summary of model
summary(fit.screening.keep.null)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: as.factor(initial_decision_keep) ~ (1 | participantID)
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##    303.6    310.7   -149.8    299.6      261
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.4266 -0.7983  0.4121  0.5794  1.4361
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 0.9384   0.9687
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.0766     0.2616   4.116 3.86e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 6:

	Dependent variable:
	as.factor(initial_decision_keep)
Constant	1.077^{***}
	(0.564, 1.589)
Observations	263
Log Likelihood	−149.790
Akaike Inf. Crit.	303.581
Bayesian Inf. Crit.	310.725
Note:	*p<0.1; **p<0.05; ***p<0.01

Adding `analogical_distance` as a predictor

```
# load the package
library(lme4)

# define the model adding analogical distance as a predictor
```

```
fit.screening.keep.nearfar <- glmer(
                      as.factor(initial_decision_keep)
                      ~ (1| participantID)
                      + as.factor(analogy_distance),
                    data = fit.screening,
                    family = binomial())

# show summary of model
summary(fit.screening.keep.nearfar)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## as.factor(initial_decision_keep) ~ (1 | participantID) + as.factor(analogy_distance)
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##    300.0    310.7   -147.0    294.0      260
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0957 -0.8127  0.4772  0.5550  1.7514
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 1.008    1.004
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       0.7553     0.2965   2.547   0.0109 *
## as.factor(analogy_distance)near   0.7060     0.3004   2.350   0.0188 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## as.fctr(n_) -0.423
```

From the above data, we can see that **near analogies** also has a **significant effect** on a **participant's tendency to keep.**

```
eval.nearfar.null.fit = anova(
    fit.screening.keep.nearfar,
    fit.screening.keep.null,
    test='LRT')

eval.nearfar.null.fit
```

```
## Data: fit.screening
## Models:
## fit.screening.keep.null: as.factor(initial_decision_keep) ~ (1 | participantID)
## fit.screening.keep.nearfar: as.factor(initial_decision_keep) ~ (1 | participantID) + as.factor(analog
##                            npar    AIC    BIC  logLik deviance  Chisq Df
```

Table 7:

	Dependent variable:
	as.factor(initial_decision_keep)
as.factor(analogy_distance)near	0.706**
	(0.117, 1.295)
Constant	0.755**
	(0.174, 1.336)
Observations	263
Log Likelihood	−147.010
Akaike Inf. Crit.	300.020
Bayesian Inf. Crit.	310.736
Note:	*p<0.1; **p<0.05; ***p<0.01

```
## fit.screening.keep.null       2 303.58 310.73 -149.79   299.58
## fit.screening.keep.nearfar    3 300.02 310.74 -147.01   294.02 5.5611  1
##                             Pr(>Chisq)
## fit.screening.keep.null
## fit.screening.keep.nearfar    0.01836 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(
    fit.screening.keep.nearfar,
    fit.screening.ignore.nearfar,
    test='LRT')
```

```
## Data: fit.screening
## Models:
## fit.screening.keep.nearfar: as.factor(initial_decision_keep) ~ (1 | participantID) + as.factor(analog
## fit.screening.ignore.nearfar: as.factor(initial_decision_ignore) ~ (1 | participantID) + as.factor(an
##                               npar    AIC    BIC   logLik deviance  Chisq Df
## fit.screening.keep.nearfar       3 300.02 310.74 -147.010   294.02
## fit.screening.ignore.nearfar     3 182.74 193.46  -88.371   176.74 117.28  0
##                               Pr(>Chisq)
## fit.screening.keep.nearfar
## fit.screening.ignore.nearfar
```

As the `fit.screening.ignore.nearfar` model has a lower AIC, BIC, log-likelihood, and deviance compared to the `fit.screening.keep.nearfar` model, the better fit. The `fit.screening.ignore.nearfar` model also has a significantly higher chi-square value and lower p-value, indicating a significant improvement in model fit compared to the `fit.screening.keep.nearfar` model. This shows that, there is a **stronger relationship between** the **tendency to ignore** and **analogical distance** than their **tendency to keep.**

Exploring the effects of `condition` with `analogical_distance`

Hypothesizing that this effect of `analogical_distance` differs between the `baseline` and our `intervention` interface, we add `condition` as an additional predictor.

```
# load the package
library(lme4)
```

```r
# define the model adding analogical distance as a predictor
fit.screening.keep.nearfar.x.condition <- glmer(
                      as.factor(initial_decision_keep)
                      ~ (1| participantID)
                      + as.factor(analogy_distance)
                      + as.factor(condition),
                    data = fit.screening,
                    family = binomial())

# show summary of model
summary(fit.screening.keep.nearfar.x.condition)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## as.factor(initial_decision_keep) ~ (1 | participantID) + as.factor(analogy_distance) +
##     as.factor(condition)
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##    299.8    314.1   -145.9    291.8      259
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3681 -0.7387  0.4223  0.6058  1.9810
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 1.038    1.019
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                        0.9879     0.3399   2.907  0.00365 **
## as.factor(analogy_distance)near    0.7146     0.3022   2.365  0.01804 *
## as.factor(condition)intervention  -0.4476     0.2993  -1.495  0.13482
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) as.(_)
## as.fctr(n_) -0.354
## as.fctr(cn) -0.472 -0.036
```

```r
barplot(sign(fixef(fit.screening.keep.nearfar.x.condition)) * exp(abs(fixef(fit.screening.keep.nearfar.
abline(a=0,b=0)
```

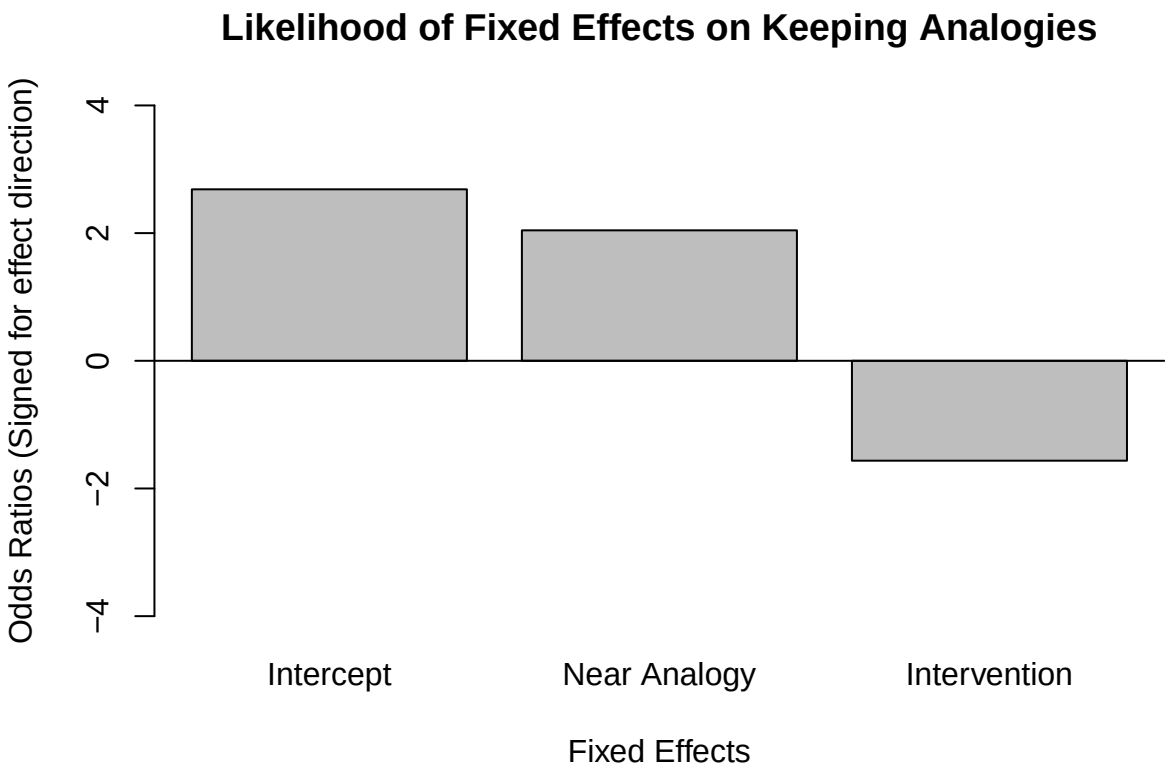


Table 8:

	Dependent variable:
	as.factor(initial_decision_keep)
as.factor(analogy_distance)near	0.715**
	(0.122, 1.307)
as.factor(condition)intervention	−0.448
	(−1.034, 0.139)
Constant	0.988***
	(0.322, 1.654)
Observations	263
Log Likelihood	−145.902
Akaike Inf. Crit.	299.804
Bayesian Inf. Crit.	314.093
Note:	*p<0.1; **p<0.05; ***p<0.01

This shows that the participants had a tendency to **keep near analogies** in general although much in the **intervention condition**.

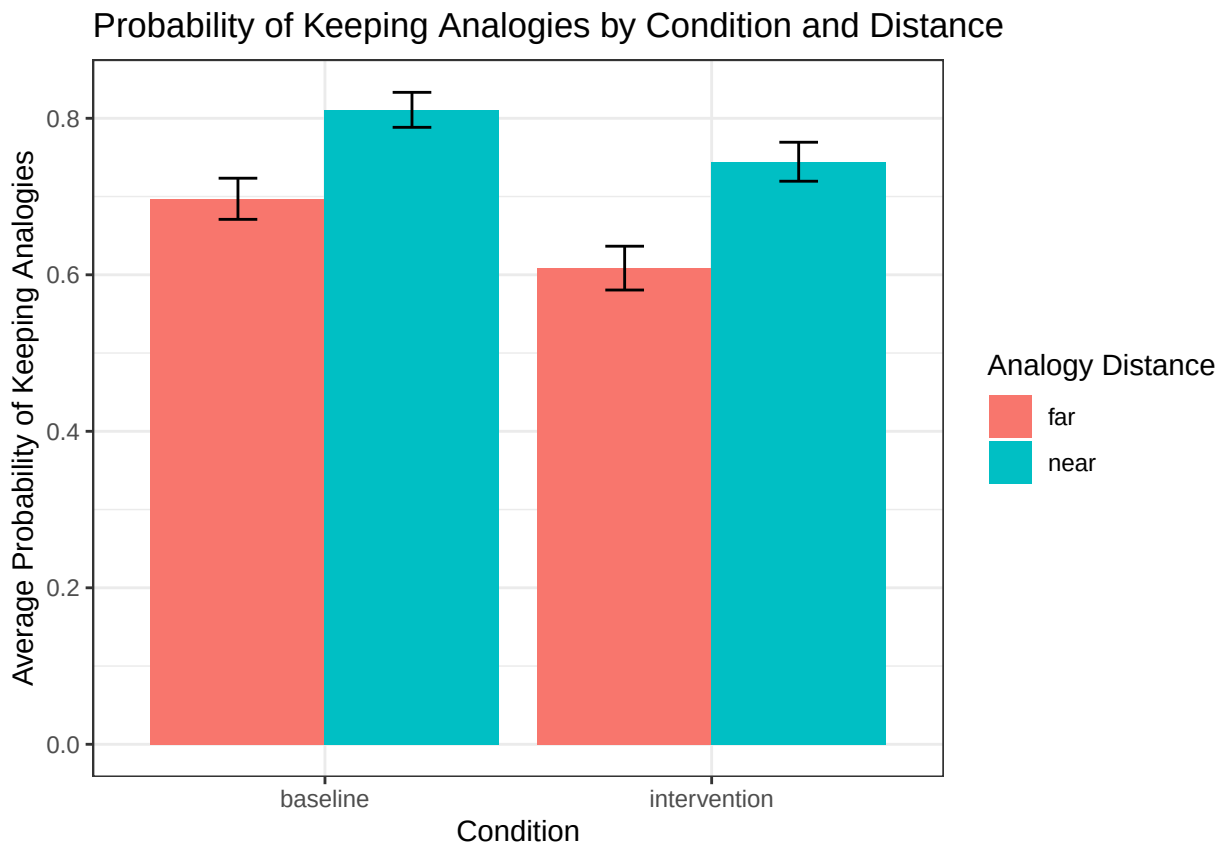
Using **ANOVA** to check for improvements in fit over the model with only **analogical distance** as the predictor, we find that the adding condition as an additional predictor improves the goodness of fit of the

model.

```
eval.nearfar.condition.fit = anova(
    fit.screening.keep.nearfar.x.condition,
    fit.screening.keep.nearfar,
    test='LRT')

eval.nearfar.condition.fit
```

```
## Data: fit.screening
## Models:
## fit.screening.keep.nearfar: as.factor(initial_decision_keep) ~ (1 | participantID) + as.factor(analog
## fit.screening.keep.nearfar.x.condition: as.factor(initial_decision_keep) ~ (1 | participantID) + as.
##                                        npar    AIC    BIC  logLik deviance
## fit.screening.keep.nearfar                3 300.02 310.74 -147.01   294.02
## fit.screening.keep.nearfar.x.condition    4 299.80 314.09 -145.90   291.80
##                                         Chisq Df Pr(>Chisq)
## fit.screening.keep.nearfar
## fit.screening.keep.nearfar.x.condition 2.2153  1     0.1366
```

```
library(ggplot2)
library(dplyr)
library(tidyr)

fit.screening.keep.nearfar.condition.plot <- select(fit.screening, analogy_distance, condition)
pred_prob <- predict(fit.screening.keep.nearfar.x.condition, type = "response")
fit.screening.keep.nearfar.condition.plot$probs <- pred_prob
fit.screening.keep.nearfar.condition.plot$se_prop <- sqrt(pred_prob*(1-pred_prob)/nrow(fit.screening))
summarize_data <- function(data) {
  summary_data <- data %>%
    group_by(analogy_distance, condition) %>%
    summarise(avg_probs = mean(probs), avg_se_prop = mean(se_prop))
  return(summary_data)
}

plot_grouped_barchart <- function(data) {
  plot <- ggplot(data, aes(x = condition, y = avg_probs, fill = analogy_distance)) +
    geom_bar(stat = "identity", position = "dodge") +
    geom_errorbar(aes(ymin = avg_probs - avg_se_prop, ymax = avg_probs + avg_se_prop), width = 0.2, posi
    labs(title = "Probability of Keeping Analogies by Condition and Distance",
         x = "Condition",
         y = "Average Probability of Keeping Analogies",
         fill = "Analogy Distance") +
    theme_bw()
  return(plot)
}

data_summary <- summarize_data(fit.screening.keep.nearfar.condition.plot)
plot_grouped_barchart(data_summary)
```



Testing for Interaction Effects

```
library(lme4)

# define the model adding analogical distance as a predictor
fit.screening.keep.nearfar.condition.interaction <- glmer(
                       initial_decision_keep
                       ~ (1| participantID)
                       + analogy_distance
                       + condition
                       + analogy_distance:condition,
                     data = fit.screening,
                       family = binomial())
summary(fit.screening.keep.nearfar.condition.interaction)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: initial_decision_keep ~ (1 | participantID) + analogy_distance +
##     condition + analogy_distance:condition
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##    298.9    316.8   -144.5    288.9      258
##
## Scaled residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -2.8137 -0.7313  0.3722  0.6270  1.8127
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 1.073    1.036
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##                                             Estimate Std. Error z value
## (Intercept)                                 7.620e-01  3.621e-01   2.104
## analogy_distancenear                        1.291e+00  4.655e-01   2.773
## conditionintervention                      -1.441e-05  3.983e-01   0.000
## analogy_distancenear:conditionintervention -1.040e+00  6.152e-01  -1.691
##                                            Pr(>|z|)
## (Intercept)                                 0.03534 *
## analogy_distancenear                        0.00555 **
## conditionintervention                       0.99997
## analogy_distancenear:conditionintervention  0.09078 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) anlgy_ cndtnn
## anlgy_dstnc -0.454
## cndtnntrvnt -0.550  0.428
## anlgy_dstn:  0.346 -0.753 -0.647
```

In this case, although the model converges, there does not seem to be a significant interaction effect.

Visual Analysis

However, lets try to plot the raw data to see if we can see any trends.

```
library(pivottabler)

# get a pivot table of counts
pt.keep.count <- PivotTable$new()
pt.keep.count$addData(fit.screening)
pt.keep.count$addColumnDataGroups("condition")
pt.keep.count$addRowDataGroups("analogy_distance")
pt.keep.count$defineCalculation(
  calculationName="TotalKeep",
  summariseExpression="n()")
cat(pt.keep.count$getLatex())
```

	baseline	intervention	Total
far	66	66	132
near	65	66	131
Total	131	132	263

```
# get a pivot table of means
pt.keep.mean <- PivotTable$new()
pt.keep.mean$addData(fit.screening)
pt.keep.mean$addColumnDataGroups("condition")
```

```
pt.keep.mean$addRowDataGroups("analogy_distance")
pt.keep.mean$defineCalculation(
  calculationName="MeanKeep",
  summariseExpression="mean(initial_decision_keep, na.rm=TRUE)")
cat(pt.keep.mean$getLatex())
```

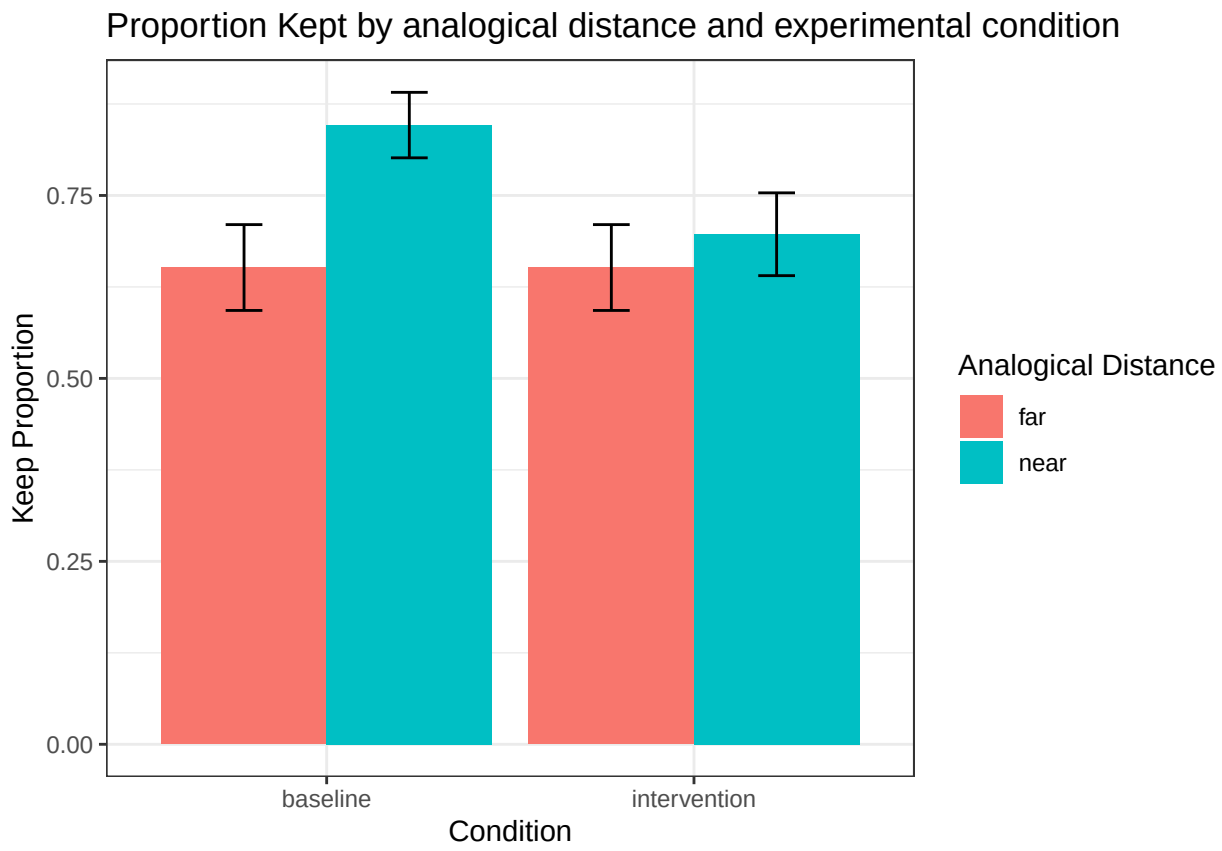
	baseline	intervention	Total
far	0.651515151515151	0.651515151515151	0.651515151515151
near	0.846153846153846	0.696969696969697	0.770992366412214
Total	0.748091603053435	0.674242424242424	0.711026615969582

```
# function to compute standard error of proportion
se_prop <- function(p, n) {
  sqrt(p*(1-p)/n)
}

# apply function to every cell
keep.condition.error.pivot <- mapply(
    se_prop,
    keep.condition.mean.pivot,
    keep.condition.count.pivot)

# convert to dataframe
keep.condition.error.pivot <- as.data.frame(keep.condition.error.pivot)
knitr::kable(keep.condition.error.pivot, "latex")
```

baseline	intervention	Total
0.0586519	0.0586519	0.0414732
0.0447519	0.0565689	0.0367125
0.0379283	0.0407914	0.0279508

```
library(ggplot2)

# Create sample data frames
df_prob <- data.frame(
  row = c("far", "near", "Total"),
  baseline = keep.condition.mean.pivot$baseline,
  intervention = keep.condition.mean.pivot$intervention
)
# Remove the Total column from the probabilities
df_prob <- df_prob[df_prob$row != "Total", ]

# contains calculated error terms
df_error <- data.frame(
  row = c("far", "near", "Total"),
  baseline = keep.condition.error.pivot$baseline,
  intervention = keep.condition.error.pivot$intervention
)
# Remove the Total column from the error terms
df_error <- df_error[df_prob$row != "Total", ]

# Reshape data frames to long format
df_prob_long <- tidyr::pivot_longer(
  df_prob,
```

```
  cols = c("baseline", "intervention"),
  names_to = "column",
  values_to = "value")
df_error_long <- tidyr::pivot_longer(
  df_error,
  cols = c("baseline", "intervention"),
  names_to = "column",
  values_to = "value")

# Merge data frames
df_merged <- merge(df_prob_long,
                   df_error_long,
                   by = c("row", "column"))

# Plot the data using ggplot2
ggplot(df_merged, aes(x = column,
                      y = value.x,
                      fill = row)) +
  geom_bar(position = "dodge",
           stat = "identity") +
  labs(x = "Condition",
       y = "Keep Proportion",
       fill = "Analogical Distance") +
  ggtitle("Proportion Kept by analogical distance and experimental condition") +
  theme_bw() +
  geom_errorbar(aes(
    ymin = value.x - value.y,
    ymax = value.x + value.y),
    width = 0.2,
    position = position_dodge(0.9))
```



Looking at Changes in Decisions from Ignore to Keep/Review in Decision Making

First, we define a **null model with no predictors** to analyze how participants as a group changed their decisions from **ignore** to **review/keep** (I guess I didn't think of this earlier! My bad! It's relevant!). This will help us see the improvement in fit as we add predictors to identify their effects.

Note that, from the 263 observations, participants faced a false negative 4 times - leading me to believe that this sparsity of data, when processed with other factors such as analogical distance and condition, due to a lack of variations in the experimental combinations, the models will not converge or cause it to misconverge, resulting in large eigen values.

Visual Analysis

Lets attempt to plot the raw data to see if there are any trends.

```
library(pivottabler)

# get a pivot table of counts
pt.ignore_to_keep.review.count <- PivotTable$new()
pt.ignore_to_keep.review.count$addData(fit.screening)
pt.ignore_to_keep.review.count$addColumnDataGroups("condition")
pt.ignore_to_keep.review.count$addRowDataGroups("analogy_distance")
pt.ignore_to_keep.review.count$defineCalculation(
  calculationName="Total Decisions changed from Ignored to Keep/Review",
  summariseExpression="n()")
cat(pt.ignore_to_keep.review.count$getLatex())
```

	baseline	intervention	Total
far	66	66	132
near	65	66	131
Total	131	132	263

```
# get a pivot table of means
pt.ignore_to_keep.review.mean <- PivotTable$new()
pt.ignore_to_keep.review.mean$addData(fit.screening)
pt.ignore_to_keep.review.mean$addColumnDataGroups("condition")
pt.ignore_to_keep.review.mean$addRowDataGroups("analogy_distance")
pt.ignore_to_keep.review.mean$defineCalculation(
  calculationName="Mean Decisions changed from Ignored to Keep/Review",
  summariseExpression="mean(decision_ignore_to_keep.review, na.rm=TRUE)")
cat(pt.ignore_to_keep.review.mean$getLatex())
```

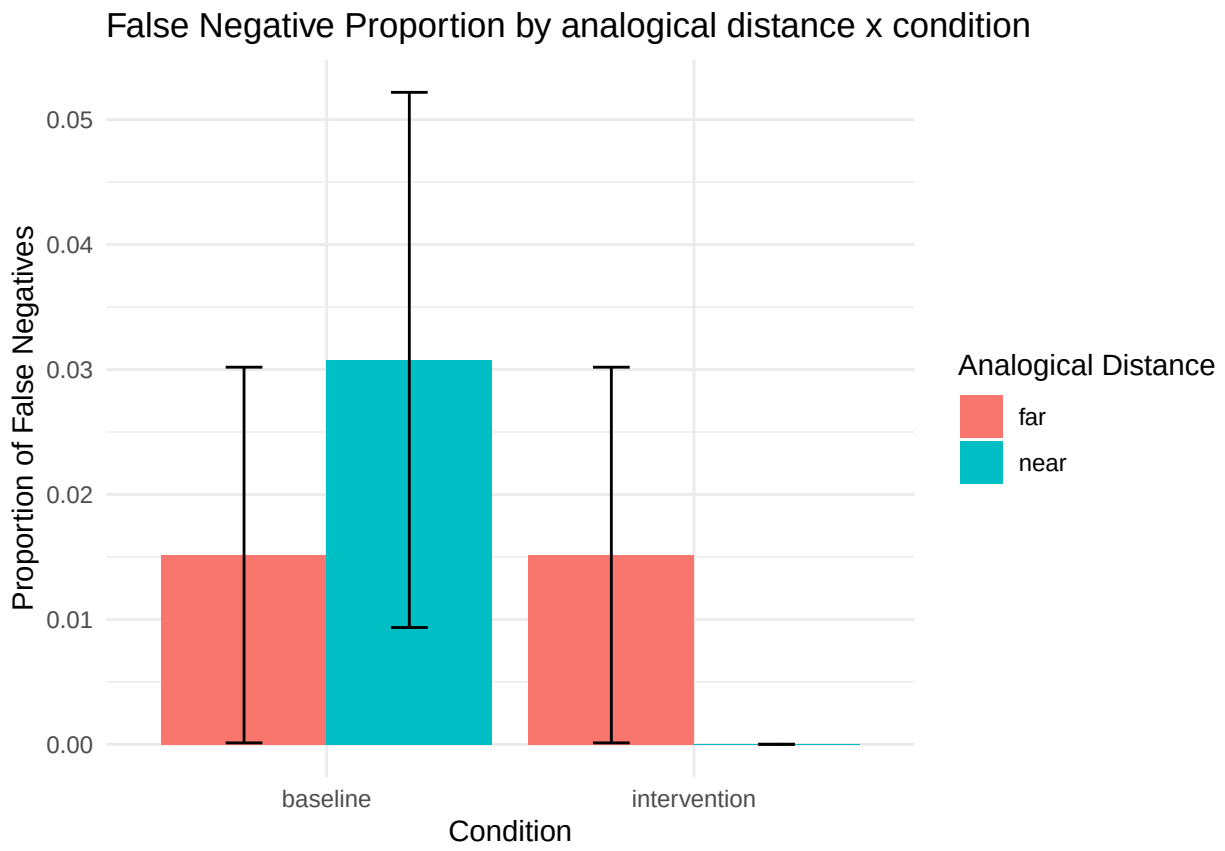
	baseline	intervention	Total
far	0.0151515151515152	0.0151515151515152	0.0151515151515152
near	0.0307692307692308	0	0.0152671755725191
Total	0.0229007633587786	0.00757575757575758	0.0152091254752852

```
# function to compute standard error of proportion
se_prop <- function(p, n) {
  sqrt(p*(1-p)/n)
}

# apply function to every cell
ignore_to_keep.review.condition.error.pivot <- mapply(
    se_prop,
    ignore_to_keep.review.condition.mean.pivot,
    ignore_to_keep.review.condition.count.pivot)

# convert to dataframe
ignore_to_keep.review.condition.error.pivot <- as.data.frame(ignore_to_keep.review.condition.error.pivo
knitr::kable(ignore_to_keep.review.condition.error.pivot, "latex")
```

baseline	intervention	Total
0.0150363	0.0150363	0.0106323
0.0214198	0.0000000	0.0107128
0.0130695	0.0075470	0.0075465

```
library(ggplot2)

# Create sample data frames
df_prob <- data.frame(
  row = c("far", "near", "Total"),
  baseline = ignore_to_keep.review.condition.mean.pivot$baseline,
  intervention = ignore_to_keep.review.condition.mean.pivot$intervention
)
# Remove the Total column from the probabilities
df_prob <- df_prob[df_prob$row != "Total", ]

# contains calculated error terms
df_error <- data.frame(
```

```
  row = c("far", "near", "Total"),
  baseline = ignore_to_keep.review.condition.error.pivot$baseline,
  intervention = ignore_to_keep.review.condition.error.pivot$intervention
)
# Remove the Total column from the error terms
df_error <- df_error[df_prob$row != "Total", ]

# Reshape data frames to long format
df_prob_long <- tidyr::pivot_longer(
  df_prob,
  cols = c("baseline", "intervention"),
  names_to = "column",
  values_to = "value")
df_error_long <- tidyr::pivot_longer(
  df_error,
  cols = c("baseline", "intervention"),
  names_to = "column",
  values_to = "value")

# Merge data frames
df_merged <- merge(df_prob_long,
                   df_error_long,
                   by = c("row", "column"))

# Plot the data using ggplot2
ggplot(df_merged, aes(x = column,
                      y = value.x,
                      fill = row)) +
  geom_bar(position = "dodge",
           stat = "identity") +
  labs(x = "Condition",
       y = "Proportion of False Negatives",
       fill = "Analogical Distance") +
  ggtitle("False Negative Proportion by analogical distance x condition") +
  theme_minimal() +
  geom_errorbar(aes(
    ymin = value.x - value.y,
    ymax = value.x + value.y),
    width = 0.2,
    position = position_dodge(0.9))
```



Looking at Changes in Decisions from Keep to Ignore/Review in Decision Making

First, we define a **null model with no predictors** to analyze how participants as a group changed their decisions from **keep** to **ignore/review** (I guess what I selected was irrelevant!). This will help us see the improvement in fit as we add predictors to identify their effects.

Note that, from the 263 observations, participants faced a false positive 2 times - leading me to believe that this sparsity of data, when processed with other factors such as analogical distance and condition, due to a lack of variations in the experimental combinations, the models will not converge or cause it to misconverge, resulting in large eigen values.

Visual Analysis

Lets attempt to plot the raw data to see if we can find any trends.

```
library(pivottabler)

# get a pivot table of counts
pt.keep_to_ignore.review.count <- PivotTable$new()
pt.keep_to_ignore.review.count$addData(fit.screening)
pt.keep_to_ignore.review.count$addColumnDataGroups("condition")
pt.keep_to_ignore.review.count$addRowDataGroups("analogy_distance")
pt.keep_to_ignore.review.count$defineCalculation(
  calculationName="Total decisions changed from Keep to Ignore/Review",
  summariseExpression="n()")
cat(pt.keep_to_ignore.review.count$getLatex())
```

	baseline	intervention	Total
far	66	66	132
near	65	66	131
Total	131	132	263

```
# get a pivot table of means
pt.keep_to_ignore.review.mean <- PivotTable$new()
pt.keep_to_ignore.review.mean$addData(fit.screening)
pt.keep_to_ignore.review.mean$addColumnDataGroups("condition")
pt.keep_to_ignore.review.mean$addRowDataGroups("analogy_distance")
pt.keep_to_ignore.review.mean$defineCalculation(
  calculationName="Mean decisions changed from Keep to Ignore/Review",
  summariseExpression="mean(decision_keep_to_ignore.review, na.rm=TRUE)")
cat(pt.keep_to_ignore.review.mean$getLatex())
```

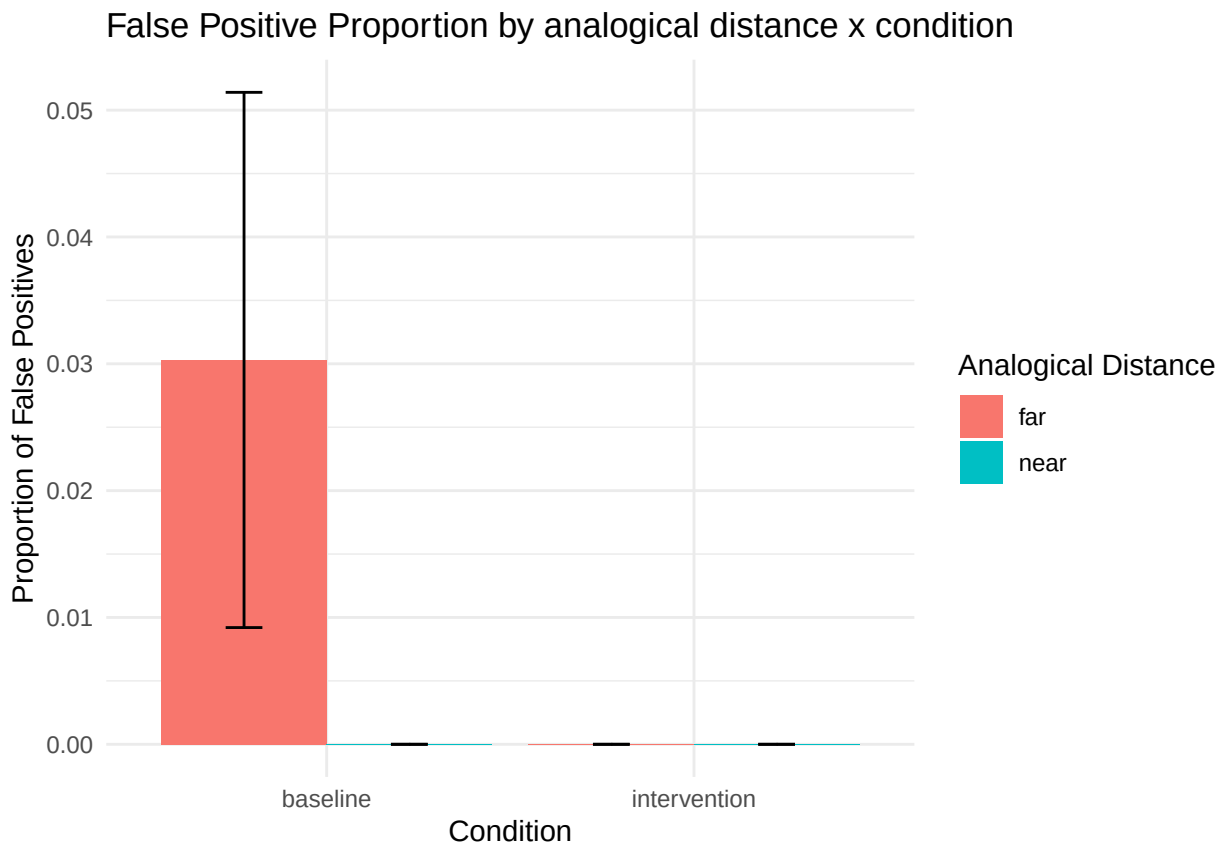
	baseline	intervention	Total
far	0.0303030303030303	0	0.0151515151515152
near	0	0	0
Total	0.0152671755725191	0	0.00760456273764259

```
# function to compute standard error of proportion
se_prop <- function(p, n) {
  sqrt(p*(1-p)/n)
}

# apply function to every cell
keep_to_ignore.review.condition.error.pivot <- mapply(
    se_prop,
    keep_to_ignore.review.condition.mean.pivot,
    keep_to_ignore.review.condition.count.pivot)

# convert to dataframe
keep_to_ignore.review.condition.error.pivot <- as.data.frame(keep_to_ignore.review.condition.error.pivo
knitr::kable(keep_to_ignore.review.condition.error.pivot, "latex")
```

baseline	intervention	Total
0.0211003	0	0.0106323
0.0000000	0	0.0000000
0.0107128	0	0.0053568

```
library(ggplot2)

# Create sample data frames
df_prob <- data.frame(
  row = c("far", "near", "Total"),
  baseline = keep_to_ignore.review.condition.mean.pivot$baseline,
  intervention = keep_to_ignore.review.condition.mean.pivot$intervention
)
# Remove the Total column from the probabilities
df_prob <- df_prob[df_prob$row != "Total", ]

# contains calculated error terms
df_error <- data.frame(
```

```
  row = c("far", "near", "Total"),
  baseline = keep_to_ignore.review.condition.error.pivot$baseline,
  intervention = keep_to_ignore.review.condition.error.pivot$intervention
)
# Remove the Total column from the error terms
df_error <- df_error[df_prob$row != "Total", ]

# Reshape data frames to long format
df_prob_long <- tidyr::pivot_longer(
  df_prob,
  cols = c("baseline", "intervention"),
  names_to = "column",
  values_to = "value")
df_error_long <- tidyr::pivot_longer(
  df_error,
  cols = c("baseline", "intervention"),
  names_to = "column",
  values_to = "value")

# Merge data frames
df_merged <- merge(df_prob_long,
                   df_error_long,
                   by = c("row", "column"))

# Plot the data using ggplot2
ggplot(df_merged, aes(x = column,
                      y = value.x,
                      fill = row)) +
  geom_bar(position = "dodge",
           stat = "identity") +
  labs(x = "Condition",
       y = "Proportion of False Positives",
       fill = "Analogical Distance") +
  ggtitle("False Positive Proportion by analogical distance x condition") +
  theme_minimal() +
  geom_errorbar(aes(
    ymin = value.x - value.y,
    ymax = value.x + value.y),
    width = 0.2,
    position = position_dodge(0.9))
```



Models that failed to converge for Keep->Ignore/Review and Ignore->Review/Keep

Looking at when the decisions were initilly kept but changed

```
# load the package
library(lme4)

# define the null model
fit.false.positive.null <- glmer(
                      decision_keep_to_ignore.review
                      ~ (1| participantID),
                    data = fit.screening,
                    family = binomial())
```

```
## boundary (singular) fit: see help('isSingular')
```

```
# show summary of model
summary(fit.false.positive.null)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: decision_keep_to_ignore.review ~ (1 | participantID)
##    Data: fit.screening
##
```

```
##      AIC      BIC   logLik deviance df.resid
##     27.5     34.6    -11.8     23.5      261
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.0875 -0.0875 -0.0875 -0.0875 11.4237
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 0        0
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.8714     0.7098  -6.863 6.75e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

Table 9:

	Dependent variable:
	decision_keep_to_ignore.review
Constant	-4.871^{***}
	$(-6.263, -3.480)$
Observations	263
Log Likelihood	−11.750
Akaike Inf. Crit.	27.501
Bayesian Inf. Crit.	34.645
Note:	*p<0.1; **p<0.05; ***p<0.01

Adding `analogical_distance` as a predictor

```
# load the package
library(lme4)

# define the model adding analogical distance as a predictor
fit.false.positive.nearfar <- glmer(
                      decision_keep_to_ignore.review
                      ~ (1| participantID)
                      + analogy_distance,
                    data = fit.screening,
                    family = binomial())
```

```
## boundary (singular) fit: see help('isSingular')
```

```
# show summary of model
summary(fit.false.positive.nearfar)
```

```
## Warning in vcov.merMod(object, use.hessian = use.hessian): variance-covariance matrix computed from
## not positive definite or contains NA values: falling back to var-cov estimated from RX
```

```
## Warning in vcov.merMod(object, correlation = correlation, sigm = sig): variance-covariance matrix co
## not positive definite or contains NA values: falling back to var-cov estimated from RX
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## decision_keep_to_ignore.review ~ (1 | participantID) + analogy_distance
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##     26.7     37.4    -10.4     20.7      260
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -0.124 -0.124  0.000  0.000  8.062
##
## Random effects:
##  Groups        Name        Variance  Std.Dev.
##  participantID (Intercept) 3.807e-17 6.17e-09
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -4.174e+00  7.125e-01  -5.859 4.67e-09 ***
## analogy_distancenear -3.175e+01  5.531e+06   0.000        1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## anlgy_dstnc 0.000
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

Adding analogical distance as a predictor fails to converge due to sparse data (only 2 instances).

Exploring the effects of `condition` with `analogical_distance`

```
# load the package
library(lme4)

# define the model adding analogical distance as a predictor
fit.false.positive.nearfar.x.condition <- glmer(
                      decision_keep_to_ignore.review
                      ~ (1| participantID)
                      + analogy_distance
                      + condition,
                    data = fit.screening,
                    family = binomial())
```

```
## boundary (singular) fit: see help('isSingular')
```

```
# show summary of model
summary(fit.false.positive.nearfar.x.condition)
```

```
## Warning in vcov.merMod(object, use.hessian = use.hessian): variance-covariance matrix computed from
## not positive definite or contains NA values: falling back to var-cov estimated from RX

## Warning in vcov.merMod(object, correlation = correlation, sigm = sig): variance-covariance matrix co
## not positive definite or contains NA values: falling back to var-cov estimated from RX

## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## decision_keep_to_ignore.review ~ (1 | participantID) + analogy_distance +
##     condition
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##     25.9     40.2     -9.0     17.9      259
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.1768  0.0000  0.0000  0.0000  5.6569
##
## Random effects:
##  Groups        Name        Variance  Std.Dev.
##  participantID (Intercept) 8.596e-14 2.932e-07
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -3.466e+00  7.181e-01  -4.826 1.39e-06 ***
## analogy_distancenear  -3.370e+01  6.779e+06   0.000        1
## conditionintervention -3.813e+01  6.753e+06   0.000        1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) anlgy_
## anlgy_dstnc  0.000
## cndtnntrvnt  0.000 -0.502
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

Also fails to converge due to lack of observations.

Testing for Interaction Effects

```
library(lme4)

# define the model adding analogical distance as a predictor
fit.false.positive.nearfar.condition.interaction <- glmer(
                      decision_keep_to_ignore.review
                      ~ (1| participantID)
                      + analogy_distance
                      + condition
                      + analogy_distance:condition,
                    data = fit.screening,
```

Table 10:

	Dependent variable:
	decision_keep_to_ignore.review
analogy_distancenear	−33.701
	(−13,286,667.000, 13,286,599.000)
conditionintervention	−38.133
	(−13,236,247.000, 13,236,171.000)
Constant	−3.466***
	(−4.873, −2.058)
Observations	263
Log Likelihood	−8.962
Akaike Inf. Crit.	25.925
Bayesian Inf. Crit.	40.213
Note:	*p<0.1; **p<0.05; ***p<0.01

```
                      family = binomial())
```

```
## boundary (singular) fit: see help('isSingular')
```

```
summary(fit.false.positive.nearfar.condition.interaction)
```

```
## Warning in vcov.merMod(object, use.hessian = use.hessian): variance-covariance matrix computed from
## not positive definite or contains NA values: falling back to var-cov estimated from RX
```

```
## Warning in vcov.merMod(object, correlation = correlation, sigm = sig): variance-covariance matrix co
## not positive definite or contains NA values: falling back to var-cov estimated from RX
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## decision_keep_to_ignore.review ~ (1 | participantID) + analogy_distance +
##     condition + analogy_distance:condition
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##     27.9     45.8     -9.0     17.9      258
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.1768  0.0000  0.0000  0.0000  5.6569
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 0        0
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##                                          Estimate Std. Error z value
```

```
## (Intercept)                                -3.466e+00  7.181e-01  -4.826
## analogy_distancenear                       -3.462e+01  8.324e+06   0.000
## conditionintervention                      -4.581e+01  8.261e+06   0.000
## analogy_distancenear:conditionintervention  8.841e+00  1.434e+07   0.000
##                                           Pr(>|z|)
## (Intercept)                               1.39e-06 ***
## analogy_distancenear                             1
## conditionintervention                            1
## analogy_distancenear:conditionintervention       1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) anlgy_ cndtnn
## anlgy_dstnc  0.000
## cndtnntrvnt  0.000  0.000
## anlgy_dstn:  0.000 -0.580 -0.576
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

The results from the model show that, there are no interaction effects, which is reasonable given the lack of data and a large sample size.

Looking at when Decisions were Initially Ignored but changed

```
# load the package
library(lme4)

# define the null model
fit.false.negative.null <- glmer(
                     decision_ignore_to_keep.review
                     ~ (1| participantID),
                   data = fit.screening,
                   family = binomial())

# show summary of model
summary(fit.false.negative.null)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: decision_ignore_to_keep.review ~ (1 | participantID)
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##     37.7     44.8    -16.8     33.7      261
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -0.42319 -0.00814 -0.00814 -0.00814  2.36298
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 44.36    6.661
```

```
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -9.586      3.140  -3.053  0.00226 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 11:

	Dependent variable:
	decision_ignore_to_keep.review
Constant	−9.586***
	(−15.740, −3.432)
Observations	263
Log Likelihood	−16.833
Akaike Inf. Crit.	37.666
Bayesian Inf. Crit.	44.810
Note:	*p<0.1; **p<0.05; ***p<0.01

Adding `analogical_distance` as a predictor

Did not result in a significant effect better than the null.

```
# load the package
library(lme4)

# define the model adding analogical distance as a predictor
fit.false.negative.nearfar <- glmer(
                      decision_ignore_to_keep.review
                      ~ (1| participantID)
                      + analogy_distance,
                    data = fit.screening,
                    family = binomial())

# show summary of model
summary(fit.false.negative.nearfar)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## decision_ignore_to_keep.review ~ (1 | participantID) + analogy_distance
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##     39.7     50.4    -16.8     33.7      260
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -0.42329 -0.00814 -0.00814 -0.00814  2.36352
##
```

```
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 44.36    6.661
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -9.5864231  3.1893125  -3.006  0.00265 **
## analogy_distancenear  0.0009122  1.1196762   0.001  0.99935
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## anlgy_dstnc -0.174
```

Table 12:

	Dependent variable:
	decision_ignore_to_keep.review
analogy_distancenear	0.001
	(−2.194, 2.195)
Constant	−9.586***
	(−15.837, −3.335)
Observations	263
Log Likelihood	−16.833
Akaike Inf. Crit.	39.666
Bayesian Inf. Crit.	50.382
Note:	*p<0.1; **p<0.05; ***p<0.01

```
eval.nearfar.null.fit = anova(
    fit.false.negative.nearfar,
    fit.false.negative.null,
    test='LRT')

eval.nearfar.null.fit
```

```
## Data: fit.screening
## Models:
## fit.false.negative.null: decision_ignore_to_keep.review ~ (1 | participantID)
## fit.false.negative.nearfar: decision_ignore_to_keep.review ~ (1 | participantID) + analogy_distance
##                            npar    AIC    BIC  logLik deviance Chisq Df
## fit.false.negative.null       2 37.666 44.810 -16.833   33.666
## fit.false.negative.nearfar    3 39.666 50.382 -16.833   33.666     0  1
##                            Pr(>Chisq)
## fit.false.negative.null
## fit.false.negative.nearfar     0.9994
```

Exploring the effects of `condition` with `analogical_distance`

Did not result in a significant effect better than the null.

```r
# load the package
library(lme4)

# define the model adding analogical distance as a predictor
fit.false.negative.nearfar.x.condition <- glmer(
                      decision_ignore_to_keep.review
                      ~ (1| participantID)
                      + analogy_distance
                      + condition,
                    data = fit.screening,
                    family = binomial())
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0194339 (tol = 0.002, component 1)
```

```r
# show summary of model
summary(fit.false.negative.nearfar.x.condition)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## decision_ignore_to_keep.review ~ (1 | participantID) + analogy_distance +
##     condition
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##     40.4     54.6    -16.2     32.4      259
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.5511 -0.0098 -0.0098 -0.0050  3.5853
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 47.37    6.882
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##                          Estimate Std. Error   z value Pr(>|z|)
## (Intercept)           -9.221478   0.003639 -2534.358   <2e-16 ***
## analogy_distancenear   0.001507   0.003638     0.414    0.679
## conditionintervention -1.360648   0.003638  -373.972   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) anlgy_
## anlgy_dstnc 0.000
## cndtnntrvnt 0.000  0.000
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.0194339 (tol = 0.002, component 1)
```

```r
eval.nearfar.condition.fit = anova(
    fit.false.negative.nearfar.x.condition,
```

Table 13:

	Dependent variable:
	decision_ignore_to_keep.review
analogy_distancenear	0.002
	(−0.006, 0.009)
conditionintervention	−1.361***
	(−1.368, −1.354)
Constant	−9.221***
	(−9.229, −9.214)
Observations	263
Log Likelihood	−16.179
Akaike Inf. Crit.	40.359
Bayesian Inf. Crit.	54.647
Note:	*p<0.1; **p<0.05; ***p<0.01

```
    fit.false.negative.nearfar,
    test='LRT')

eval.nearfar.condition.fit
```

```
## Data: fit.screening
## Models:
## fit.false.negative.nearfar: decision_ignore_to_keep.review ~ (1 | participantID) + analogy_distance
## fit.false.negative.nearfar.x.condition: decision_ignore_to_keep.review ~ (1 | participantID) + analog
##                                        npar    AIC    BIC  logLik deviance
## fit.false.negative.nearfar                3 39.666 50.382 -16.833   33.666
## fit.false.negative.nearfar.x.condition    4 40.359 54.647 -16.179   32.359
##                                         Chisq Df Pr(>Chisq)
## fit.false.negative.nearfar
## fit.false.negative.nearfar.x.condition 1.3068  1      0.253
```

Testing for Interaction Effects

```
library(lme4)

# define the model adding analogical distance as a predictor
fit.false.negative.nearfar.condition.interaction <- glmer(
                      decision_ignore_to_keep.review
                      ~ (1| participantID)
                      + analogy_distance
                      + condition
                      + analogy_distance:condition,
                    data = fit.screening,
                      family = binomial())
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Hessian is numerically singular: parameters are not uniquely determined
```

```
summary(fit.false.negative.nearfar.condition.interaction)
```

```
## Warning in vcov.merMod(object, use.hessian = use.hessian): variance-covariance matrix computed from
## not positive definite or contains NA values: falling back to var-cov estimated from RX
```

```
## Warning in vcov.merMod(object, correlation = correlation, sigm = sig): variance-covariance matrix co
## not positive definite or contains NA values: falling back to var-cov estimated from RX
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## decision_ignore_to_keep.review ~ (1 | participantID) + analogy_distance +
##     condition + analogy_distance:condition
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##     40.4     58.2    -15.2     30.4      258
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -0.67877 -0.01157 -0.00709  0.00000  2.40313
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 48.22    6.944
## Number of obs: 263, groups:  participantID, 22
##
## Fixed effects:
##                                              Estimate Std. Error z value
## (Intercept)                                -9.863e+00  4.404e+00  -2.239
## analogy_distancenear                        9.786e-01  1.445e+00   0.677
## conditionintervention                       1.002e-06  1.624e+00   0.000
## analogy_distancenear:conditionintervention -7.715e+02  8.261e+06   0.000
##                                            Pr(>|z|)
## (Intercept)                                  0.0251 *
## analogy_distancenear                         0.4981
## conditionintervention                        1.0000
## analogy_distancenear:conditionintervention   0.9999
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) anlgy_ cndtnn
## anlgy_dstnc -0.213
## cndtnntrvnt -0.184  0.562
## anlgy_dstn:  0.000  0.000  0.000
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## unable to evaluate scaled gradient
##  Hessian is numerically singular: parameters are not uniquely determined
```

A.2 R Code for Qualitative Analysis of Revisions

Analysis of Revisions

Arvind Srinivasan

2023-03-15

Preprocessing for Analysis

First, load the collected dataset of decisions (`decision-with-review-and-mistake.csv`) to begin processing.

```
library(readr)
setwd("~/Research/exp-analogy-triage/analysis/quantitative")
decisions <- read_csv("decisions-with-review-and-mistake.csv")
```

```
library(tidyr)
decisions$condition = as.factor(decisions$condition)
decisions$interface = as.factor(decisions$interface)
decisions$analogy_distance = as.factor(decisions$analogy_distance)
decisions$first_decision_logged = as.factor(decisions$first_decision_logged)
decisions$final_decision = as.factor(decisions$final_decision)
decisions <- drop_na(decisions)
```

```
library(vtable)
```

```
## Loading required package: kableExtra
```

```
sample_df = as.data.frame(decisions)
sample_df = within(sample_df, rm(participantID, problemID, analogy_order))
st(sample_df, out="latex", summ = list(
     c('sum(x, na.rm=TRUE)','mean(x)','sd(x)','min(x)','max(x)')
   ),
   summ.names = list(
     c('Valid N','Mean','SD','Min','Max')
   ))
```

Before performing the qualitative analysis, we clean up the columns.

```
library(tidyr)
decisions = drop_na(decisions)
summary(decisions)
```

```
##  participantID      problemID     analogy_order   analogy_shortname
##  Min.   :108688   Min.   :1.000   Min.   :1.000   Length:390
##  1st Qu.:344424   1st Qu.:1.000   1st Qu.:2.000   Class :character
##  Median :483347   Median :2.000   Median :3.000   Mode  :character
##  Mean   :474726   Mean   :2.103   Mean   :3.436
##  3rd Qu.:598108   3rd Qu.:3.000   3rd Qu.:5.000
##  Max.   :948547   Max.   :3.000   Max.   :6.000
##         condition       interface       key            analogy_distance
##  baseline    :220   ideation :195   Length:390         far :198
##  intervention:170   screening:195   Class :character   near:192
##                                     Mode  :character
```

Table 1: Summary Statistics

Variable	Valid N	Mean	SD	Min	Max
condition	390				
... baseline	220	56%			
... intervention	170	44%			
interface	390				
... ideation	195	50%			
... screening	195	50%			
analogy_distance	390				
... far	198	51%			
... near	192	49%			
first_decision_logged	390				
... ignore	28	7%			
... keep	169	43%			
... review	193	49%			
final_decision	390				
... ignore	44	11%			
... keep	305	78%			
... review	41	11%			
review	193	0.49	0.5	0	1
keep	169	0.43	0.5	0	1
ignore	28	0.072	0.26	0	1
mistake	160	0.41	0.49	0	1
decision_changed	390				
... no	160	41%			
... yes	230	59%			

```
##
##
##
##  first_decision_logged final_decision     review           keep
##  ignore: 28            ignore: 44     Min.   :0.0000   Min.   :0.0000
##  keep  :169            keep  :305     1st Qu.:0.0000   1st Qu.:0.0000
##  review:193            review: 41     Median :0.0000   Median :0.0000
##                                       Mean   :0.4949   Mean   :0.4333
##                                       3rd Qu.:1.0000   3rd Qu.:1.0000
##                                       Max.   :1.0000   Max.   :1.0000
##      ignore           mistake       decision_changed    rationale
##  Min.   :0.00000   Min.   :0.0000   Length:390         Length:390
##  1st Qu.:0.00000   1st Qu.:0.0000   Class :character   Class :character
##  Median :0.00000   Median :0.0000   Mode  :character   Mode  :character
##  Mean   :0.07179   Mean   :0.4103
##  3rd Qu.:0.00000   3rd Qu.:1.0000
##  Max.   :1.00000   Max.   :1.0000
```

```
df <- data.frame(decisions)
# remove unnecessary columns
df <- within(df, rm(keep, review, ignore, decision_changed, rationale, key))
# rename the decision columns to faciliate one-hot encoding
colnames(df)[8] <- "initial_decision_"
colnames(df)[9] <- "final_decision_"
```

Now we perform **One Hot Encoding** of the **initial** and **final** decisions.

```
# one-hot encode the "initial_decision" column
encoded_df <- model.matrix(~ initial_decision_ - 1, data = df)
df <- cbind(df, encoded_df)

# one-hot encode the "final_decision" column
encoded_df <- model.matrix(~ final_decision_ - 1, data = df)
df <- cbind(df, encoded_df)

# remove the decision columns
df <- within(df, rm(initial_decision_, final_decision_))
```

Code for the change in decisions by comparing the `initial_decision` and `final_decision` columns. If the `initial_decision` is `ignore`, code it as a `false_negative`. If the `initial_decision` is `keep`, code it as a `false_positive`.

```
# create a new column for false negative results
df$decision_false_negative <- ifelse(
  df$initial_decision_ignore == 1 & df$final_decision_ignore != 1,
  1, 0)

# create a new column for false positive results
df$decision_false_positive <- ifelse(
  df$initial_decision_keep == 1 & df$final_decision_keep != 1,
  1, 0)
```

This could be later used for analyzing the changes to decisions across the different interfaces. For the convenience of analysis, split the interface by their categories into their own data frames.

```
# split the data frame into based on interface
fit.screening <- subset(df, interface == "screening")
```

```
df.ideation <- subset(df, interface == "ideation")
```

Now, these data frames are ready for analysis.

Looking at "Revisions" in Screening

A point to note is that, since we have addressed the difference in usecase of review in intervention, the data is ready for processing mistakes (any changes in decisions).

First, we define a **null model with no predictors** to analyze how participants as a group **initially mistaked analogies** during screening. This will help us see the improvement in fit as we add predictors to identify their effects.

```
# load the package
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
# define the null model
fit.screening.mistake.null <- glmer(
                      mistake
                      ~ (1| participantID),
                    data = fit.screening,
                    family = binomial())

# show summary of model
summary(fit.screening.mistake.null)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: mistake ~ (1 | participantID)
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##     84.0     90.5    -40.0     80.0      193
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.5473 -0.2400 -0.1351 -0.1317  4.1675
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 1.98     1.407
## Number of obs: 195, groups:  participantID, 21
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.6488     0.7883  -4.629 3.68e-06 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 2:

	Dependent variable:
	mistake
Constant	−3.649***
	(0.788)
Observations	195
Log Likelihood	−39.997
Akaike Inf. Crit.	83.994
Bayesian Inf. Crit.	90.540
Note:	*p<0.1; **p<0.05; ***p<0.01

Adding `analogical_distance` as a predictor

```
# load the package
library(lme4)

# define the model adding analogical distance as a predictor
fit.screening.mistake.nearfar <- glmer(
                      mistake
                      ~ (1| participantID)
                      + analogy_distance,
                    data = fit.screening,
                    family = binomial())

# show summary of model
summary(fit.screening.mistake.nearfar)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: mistake ~ (1 | participantID) + analogy_distance
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##     85.2     95.0    -39.6     79.2      192
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.6267 -0.2016 -0.1490 -0.1128  3.6929
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 1.996    1.413
## Number of obs: 195, groups:  participantID, 21
##
## Fixed effects:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -3.4009     0.8245  -4.125 3.71e-05 ***
```

```
## analogy_distancenear  -0.5897     0.6773  -0.871    0.384
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## anlgy_dstnc -0.278
```

Table 3:

	Dependent variable:
	mistake
analogy_distancenear	−0.590
	(0.677)
Constant	−3.401***
	(0.824)
Observations	195
Log Likelihood	−39.605
Akaike Inf. Crit.	85.211
Bayesian Inf. Crit.	95.030
Note:	*p<0.1; **p<0.05; ***p<0.01

Use **ANOVA** to check for improvements in fit over the null model.

```
eval.nearfar.null.fit = anova(
    fit.screening.mistake.nearfar,
    fit.screening.mistake.null,
    test='LRT')

eval.nearfar.null.fit
```

```
## Data: fit.screening
## Models:
## fit.screening.mistake.null: mistake ~ (1 | participantID)
## fit.screening.mistake.nearfar: mistake ~ (1 | participantID) + analogy_distance
##                               npar    AIC   BIC  logLik deviance  Chisq Df
## fit.screening.mistake.null       2 83.994 90.54 -39.997   79.994
## fit.screening.mistake.nearfar    3 85.211 95.03 -39.605   79.211 0.7836  1
##                               Pr(>Chisq)
## fit.screening.mistake.null
## fit.screening.mistake.nearfar      0.376
```

This shows that, the adding the `analogical_distance` as predictor does not improve the fitness of data over null.

Exploring the effects of `condition` with `analogical_distance`

Hypothesizing that this effect of `analogical_distance` differs between the `baseline` and our `intervention` interface, we add `condition` as an additional predictor.

```
# load the package
library(lme4)
```

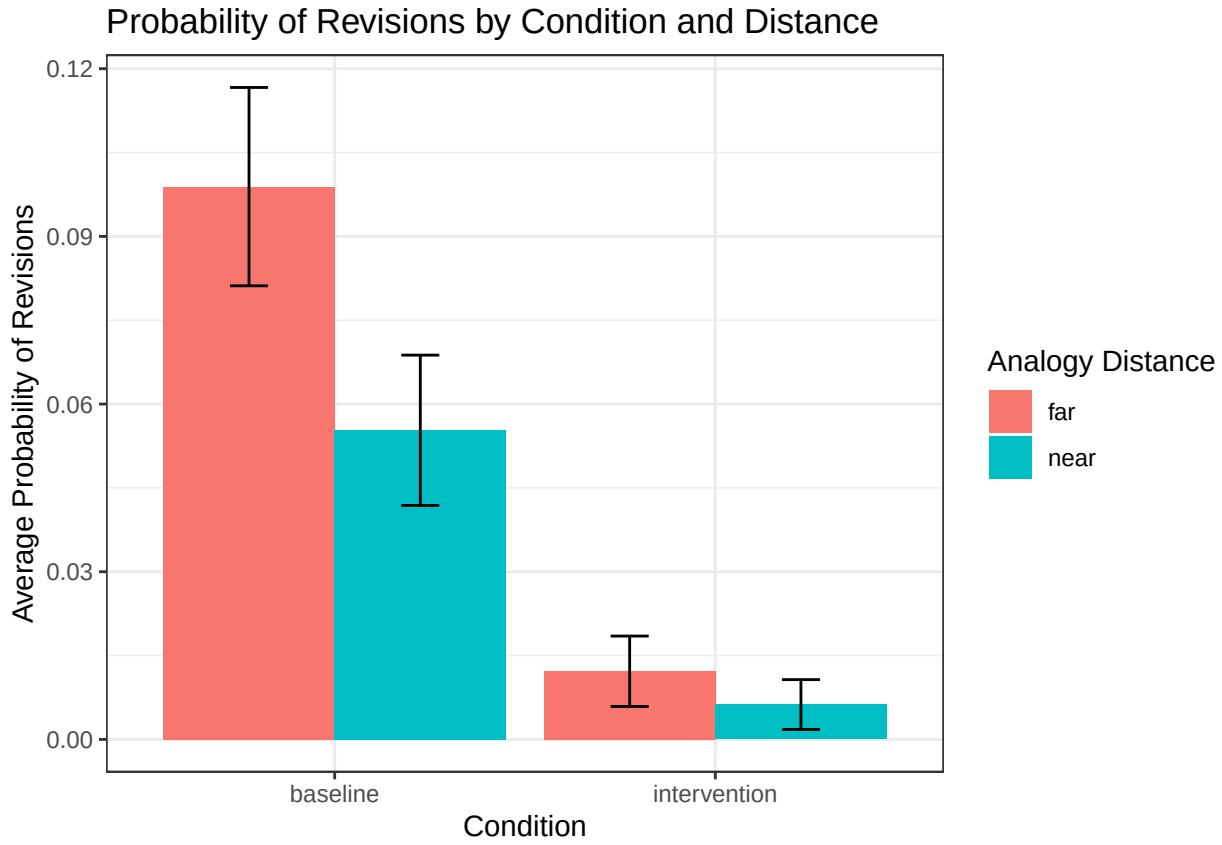
```r
# define the model adding analogical distance as a predictor
fit.screening.mistake.nearfar.x.condition <- glmer(
                      mistake
                      ~ (1| participantID)
                      + analogy_distance
                      + condition,
                    data = fit.screening,
                    family = binomial())

# show summary of model
summary(fit.screening.mistake.nearfar.x.condition)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: mistake ~ (1 | participantID) + analogy_distance + condition
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##     79.0     92.1    -35.5     71.0      191
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.0213 -0.1791 -0.1270 -0.0505  5.7457
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 2.832    1.683
## Number of obs: 195, groups:  participantID, 21
##
## Fixed effects:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -3.0150     0.9413  -3.203  0.00136 **
## analogy_distancenear    -0.6818     0.7182  -0.949  0.34246
## conditionintervention   -2.5253     1.1312  -2.232  0.02559 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) anlgy_
## anlgy_dstnc -0.239
## cndtnntrvnt  0.000  0.063
```

Interestingly, the intervention condition seems to have a significant effect on the mistakes.

```r
library(ggplot2)
library(dplyr)
library(tidyr)

fit.screening.mistake.nearfar.condition.plot <- select(fit.screening, analogy_distance, condition)
pred_prob <- predict(fit.screening.mistake.nearfar.x.condition, type = "response")
fit.screening.mistake.nearfar.condition.plot$probs <- pred_prob
fit.screening.mistake.nearfar.condition.plot$se_prop <- sqrt(pred_prob*(1-pred_prob)/nrow(fit.screening
```

```r
summarize_data <- function(data) {
  summary_data <- data %>%
    group_by(analogy_distance, condition) %>%
    summarise(avg_probs = mean(probs), avg_se_prop = mean(se_prop))
  return(summary_data)
}

plot_grouped_barchart <- function(data) {
  plot <- ggplot(data, aes(x = condition, y = avg_probs, fill = analogy_distance)) +
    geom_bar(stat = "identity", position = "dodge") +
    geom_errorbar(aes(ymin = avg_probs - avg_se_prop, ymax = avg_probs + avg_se_prop), width = 0.2, pos
    labs(title = "Probability of Revisions by Condition and Distance",
         x = "Condition",
         y = "Average Probability of Revisions",
         fill = "Analogy Distance") +
    theme_bw()
  return(plot)
}

data_summary <- summarize_data(fit.screening.mistake.nearfar.condition.plot)
plot_grouped_barchart(data_summary)
```



Using **ANOVA** to check for improvements in fit over the model with only **analogical distance** as the predictor, we find that the adding condition as an additional predictor improves the goodness of fit of the model.

Table 4:

	Dependent variable:
	mistake
analogy_distancenear	−0.682
	(0.718)
conditionintervention	−2.525**
	(1.131)
Constant	−3.015***
	(0.941)
Observations	195
Log Likelihood	−35.500
Akaike Inf. Crit.	79.000
Bayesian Inf. Crit.	92.092
Note:	*p<0.1; **p<0.05; ***p<0.01

```
eval.nearfar.condition.fit = anova(
    fit.screening.mistake.nearfar.x.condition,
    fit.screening.mistake.nearfar,
    test='LRT')

eval.nearfar.condition.fit
```

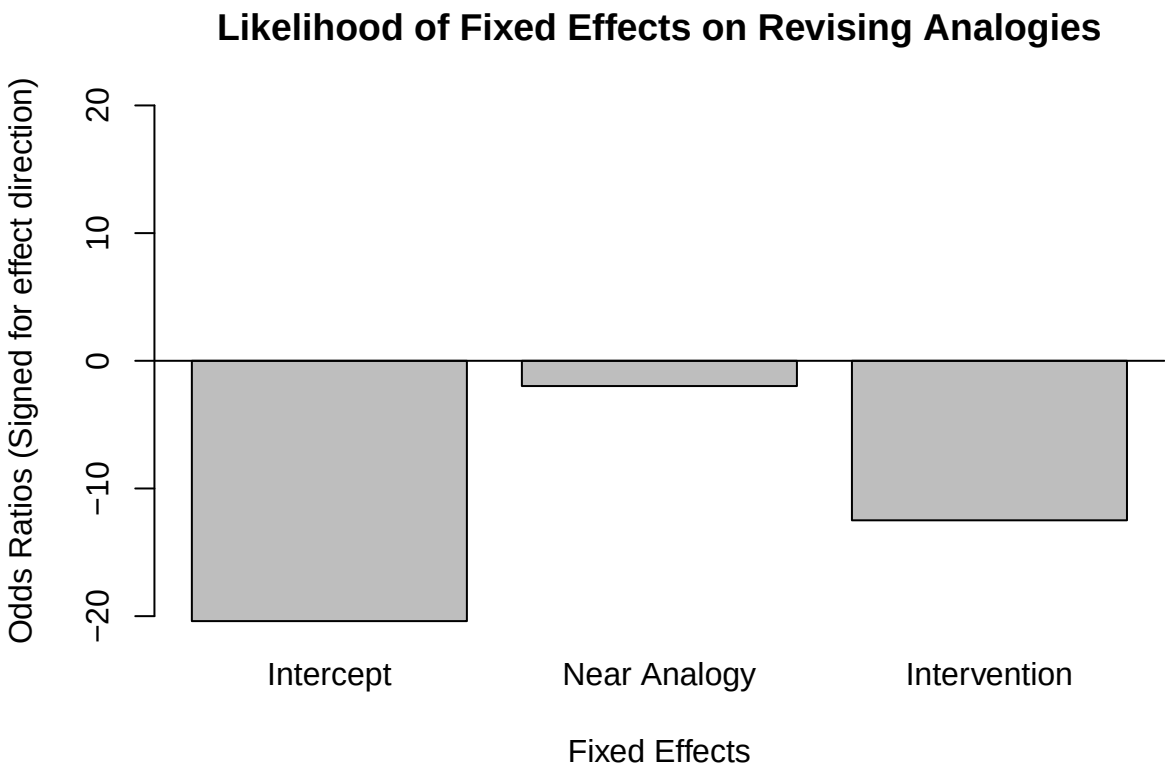
```
## Data: fit.screening
## Models:
## fit.screening.mistake.nearfar: mistake ~ (1 | participantID) + analogy_distance
## fit.screening.mistake.nearfar.x.condition: mistake ~ (1 | participantID) + analogy_distance + conditi
##                                           npar    AIC    BIC  logLik deviance
## fit.screening.mistake.nearfar                3 85.211 95.030 -39.605   79.211
## fit.screening.mistake.nearfar.x.condition    4 79.000 92.092 -35.500   71.000
##                                            Chisq Df Pr(>Chisq)
## fit.screening.mistake.nearfar
## fit.screening.mistake.nearfar.x.condition 8.2109  1   0.004164 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This shows that the `nearfar.x.condition` model is significant!

```
sign(fixef(fit.screening.mistake.nearfar.x.condition)) * exp(abs(fixef(fit.screening.mistake.nearfar.x.
```

```
##           (Intercept)  analogy_distancenear conditionintervention
##            -20.388888             -1.977395            -12.494033
```

```
barplot(sign(fixef(fit.screening.mistake.nearfar.x.condition)) * exp(abs(fixef(fit.screening.mistake.nea
abline(a=0,b=0)
```



Testing for Interaction Effects

```r
library(lme4)

# define the model adding analogical distance as a predictor
fit.screening.mistake.nearfar.condition.interaction <- glmer(
                      mistake
                      ~ (1| participantID)
                      + analogy_distance
                      + condition
                      + analogy_distance:condition,
                    data = fit.screening,
                      family = binomial())
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Hessian is numerically singular: parameters are not uniquely determined
```

```
summary(fit.screening.mistake.nearfar.condition.interaction)
```

```
## Warning in vcov.merMod(object, use.hessian = use.hessian): variance-covariance matrix computed from
## not positive definite or contains NA values: falling back to var-cov estimated from RX
```

```
## Warning in vcov.merMod(object, correlation = correlation, sigm = sig): variance-covariance matrix co
```

```
## not positive definite or contains NA values: falling back to var-cov estimated from RX
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: mistake ~ (1 | participantID) + analogy_distance + condition +
##     analogy_distance:condition
##    Data: fit.screening
##
##      AIC      BIC   logLik deviance df.resid
##     80.1     96.5    -35.1     70.1      190
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.9814 -0.1733 -0.1345  0.0000  4.6269
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  participantID (Intercept) 2.816    1.678
## Number of obs: 195, groups:  participantID, 21
##
## Fixed effects:
##                                             Estimate Std. Error z value
## (Intercept)                                -3.086e+00  6.830e-01  -4.518
## analogy_distancenear                       -4.982e-01  7.941e-01  -0.627
## conditionintervention                      -2.026e+00  1.276e+00  -1.588
## analogy_distancenear:conditionintervention -3.375e+01  1.048e+07   0.000
##                                            Pr(>|z|)
## (Intercept)                                6.24e-06 ***
## analogy_distancenear                          0.530
## conditionintervention                         0.112
## analogy_distancenear:conditionintervention    1.000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) anlgy_ cndtnn
## anlgy_dstnc -0.468
## cndtnntrvnt -0.268  0.261
## anlgy_dstn:  0.000  0.000  0.000
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## unable to evaluate scaled gradient
##  Hessian is numerically singular: parameters are not uniquely determined
```

Scaling of the values did not help (since the dependent variable was already one-hot encoded).

However if we look at the number of mistaked analogies, we can see that the participants mistaked only 11 analogies of 195 -leading me to believe that this sparsity of data, when processed with other factors defined for the condition due to a lack of variations in the experimental combinations, it could not converge or caused it to misconverge, resulting in large eigen values.

Visual Analysis

Given that there seems to be an interaction, we attempt to plot this relationship graphically.

```
library(pivottabler)

# get a pivot table of counts
pt.mistake.count <- PivotTable$new()
pt.mistake.count$addData(fit.screening)
pt.mistake.count$addColumnDataGroups("condition")
pt.mistake.count$addRowDataGroups("analogy_distance")
pt.mistake.count$defineCalculation(
  calculationName="TotalMistakes",
  summariseExpression="n()")
cat(pt.mistake.count$getLatex())
```

	baseline	intervention	Total
far	55	44	99
near	55	41	96
Total	110	85	195

```
# get a pivot table of means
pt.mistake.mean <- PivotTable$new()
pt.mistake.mean$addData(fit.screening)
pt.mistake.mean$addColumnDataGroups("condition")
pt.mistake.mean$addRowDataGroups("analogy_distance")
pt.mistake.mean$defineCalculation(
  calculationName="MeanMistakes",
  summariseExpression="mean(mistake, na.rm=TRUE)")
cat(pt.mistake.mean$getLatex())
```

	baseline	intervention	Total
far	0.109090909090909	0.0227272727272727	0.0707070707070707
near	0.0727272727272727	0	0.0416666666666667
Total	0.0909090909090909	0.0117647058823529	0.0564102564102564

```
mistake.condition.count.pivot <- pt.mistake.count$asDataFrame()
mistake.condition.count.pivot
```

```
##       baseline intervention Total
## far         55           44    99
## near        55           41    96
## Total      110           85   195
```

```
mistake.condition.mean.pivot <- pt.mistake.mean$asDataFrame()
mistake.condition.mean.pivot
```

```
##         baseline intervention      Total
## far   0.10909091   0.02272727 0.07070707
## near  0.07272727   0.00000000 0.04166667
## Total 0.09090909   0.01176471 0.05641026
```

```
# function to compute standard error of proportion
se_prop <- function(p, n) {
  sqrt(p*(1-p)/n)
}

# apply function to every cell
mistake.condition.error.pivot <- mapply(
```

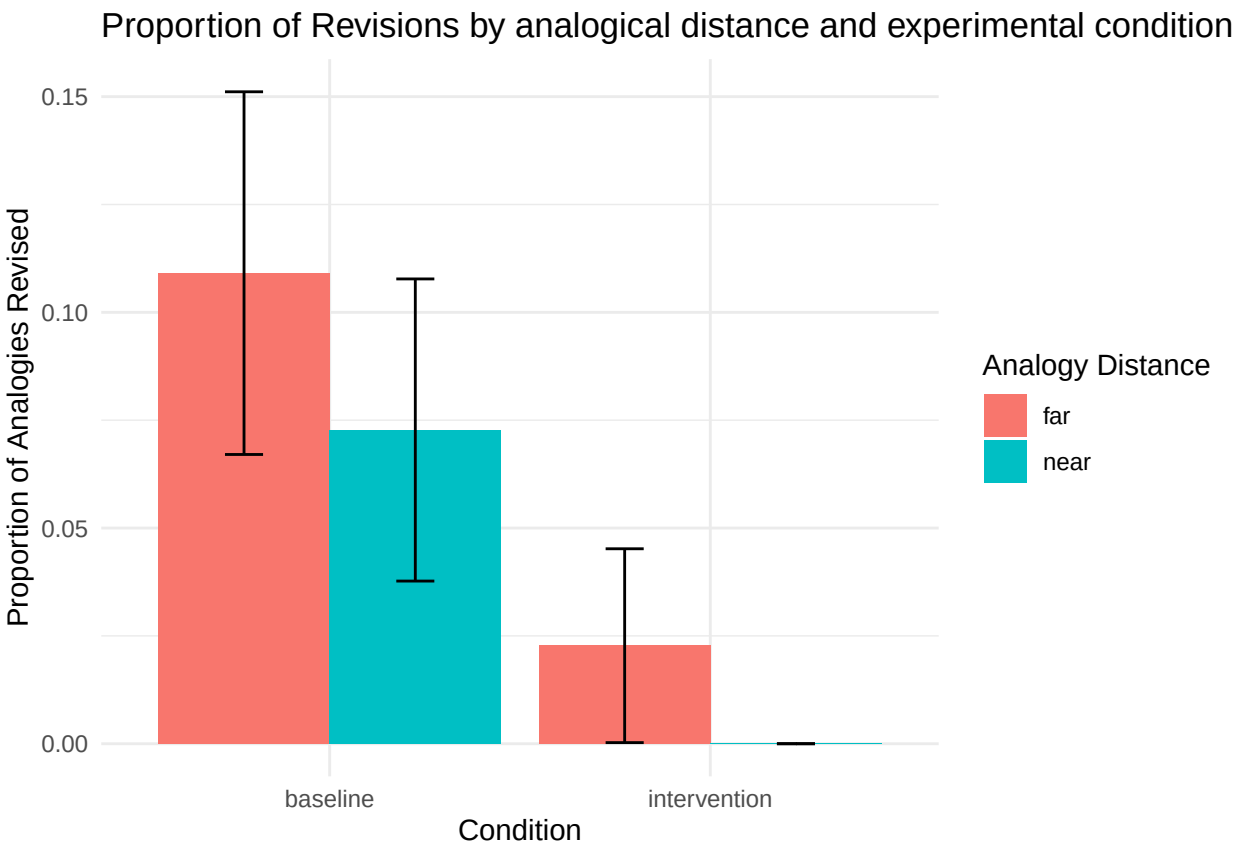
```
    se_prop,
    mistake.condition.mean.pivot,
    mistake.condition.count.pivot)

# convert to dataframe
mistake.condition.error.pivot <- as.data.frame(mistake.condition.error.pivot)
mistake.condition.error.pivot
```

```
baseline intervention      Total
```

1 0.04203680 0.02246752 0.02576263 2 0.03501636 0.00000000 0.02039469 3 0.02741012 0.01169530 0.01652165

```
library(ggplot2)

# Create sample data frames
df_prob <- data.frame(
  row = c("far", "near", "Total"),
  baseline = mistake.condition.mean.pivot$baseline,
  intervention = mistake.condition.mean.pivot$intervention
)
# Remove the Total column from the probabilities
df_prob <- df_prob[df_prob$row != "Total", ]

# contains calculated error terms
df_error <- data.frame(
  row = c("far", "near", "Total"),
  baseline = mistake.condition.error.pivot$baseline,
  intervention = mistake.condition.error.pivot$intervention
)
# Remove the Total column from the error terms
df_error <- df_error[df_error$row != "Total", ]

# Reshape data frames to long format
df_prob_long <- tidyr::pivot_longer(
  df_prob,
  cols = c("baseline", "intervention"),
  names_to = "column",
  values_to = "value")
df_error_long <- tidyr::pivot_longer(
  df_error,
  cols = c("baseline", "intervention"),
  names_to = "column",
  values_to = "value")

# Merge data frames
df_merged <- merge(df_prob_long,
                   df_error_long,
                   by = c("row", "column"))

# Plot the data using ggplot2
ggplot(df_merged, aes(x = column,
                      y = value.x,
                      fill = row)) +
  geom_bar(position = "dodge",
           stat = "identity") +
```

```
labs(x = "Condition",
     y = "Proportion of Analogies Revised",
     fill = "Analogy Distance") +
ggtitle("Proportion of Revisions by analogical distance and experimental condition") +
theme_minimal() +
geom_errorbar(aes(
  ymin = value.x - value.y,
  ymax = value.x + value.y),
  width = 0.2,
  position = position_dodge(0.9))
```



A.3 R Code for Qualitative Analysis of Processing Time

Analysis of Processing Time

Arvind Srinivasan

2023-03-13

Preprocessing

First, load the collected dataset of decisions (`timestamps.csv`) to begin processing.

```
library(readr)
setwd("~/Research/exp-analogy-triage/analysis/quantitative")
timestamps <- read_csv("timestamps.csv")
timestamps
```

```
## # A tibble: 220 x 13
##       pid condition inter~1 problem key   analo~2 analo~3 action curre~4 next_~5
##     <dbl> <chr>     <chr>     <dbl> <chr> <chr>   <chr>   <chr>    <dbl>   <dbl>
##  1 108688 baseline  screen~       3 1086~ near    keep    go-to~       1       2
##  2 108688 baseline  screen~       3 1086~ far     keep    go-to~       2       3
##  3 108688 baseline  screen~       3 1086~ near    ignore  go-to~       3       4
##  4 108688 baseline  screen~       3 1086~ far     review  go-to~       4       5
##  5 108688 baseline  screen~       3 1086~ near    keep    go-to~       5       6
##  6 108688 interven~ screen~       2 1086~ far     ignore  go-to~       1       2
##  7 108688 interven~ screen~       2 1086~ near    keep    go-to~       2       3
##  8 108688 interven~ screen~       2 1086~ near    keep    go-to~       3       4
##  9 108688 interven~ screen~       2 1086~ far     keep    navig~       4       5
## 10 108688 interven~ screen~       2 1086~ far     ignore  go-to~       5       6
## # ... with 210 more rows, 3 more variables: timestamp <dbl>, delta_s <dbl>,
## #   return_to_analogy <lgl>, and abbreviated variable names 1: interface,
## #   2: analogy_distance, 3: analogy_decision, 4: current_analogy,
## #   5: next_analogy
```

```
library(vtable)
```

```
## Loading required package: kableExtra
```

```
library(tidyr)
sample_df = as.data.frame(timestamps)
sample_df <- drop_na(sample_df)
sample_df = within(sample_df, rm(pid, problem, key, current_analogy, next_analogy, timestamp))
st(sample_df, out="latex", summ = list(
     c('sum(x, na.rm=TRUE)','mean(x)','sd(x)','min(x)','max(x)')
   ),
   summ.names = list(
     c('Valid N','Mean','SD','Min','Max')
   ))
```

We can see that the data has 24 NA values - logs that do not contain timestamp information. There is also logs missing for 1 participant (`313394`).

Table 1: Summary Statistics

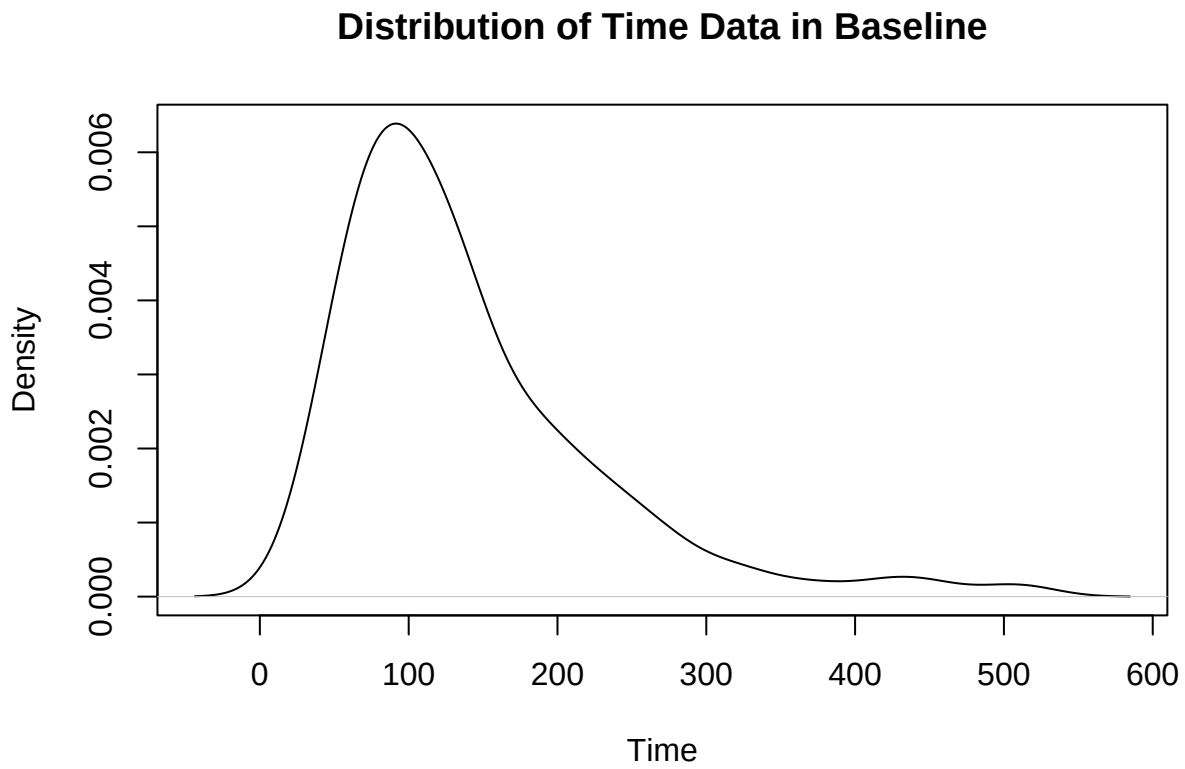
Variable	Valid N	Mean	SD	Min	Max
condition	196				
... baseline	102	52%			
... intervention	94	48%			
interface	196				
... screening	196	100%			
analogy_distance	196				
... far	98	50%			
... near	98	50%			
analogy_decision	196				
... ignore	23	12%			
... keep	147	75%			
... review	26	13%			
action	196				
... go-to-analogy	187	95%			
... navigate-to-task	2	1%			
... update-no-of-analogies	7	4%			
delta_s	36003	184	109	32	528
return_to_analogy	196				
... No	170	87%			
... Yes	26	13%			

```
library(tidyr)
df <- as.data.frame(timestamps)
df <- drop_na(df)
df$delta_s
```

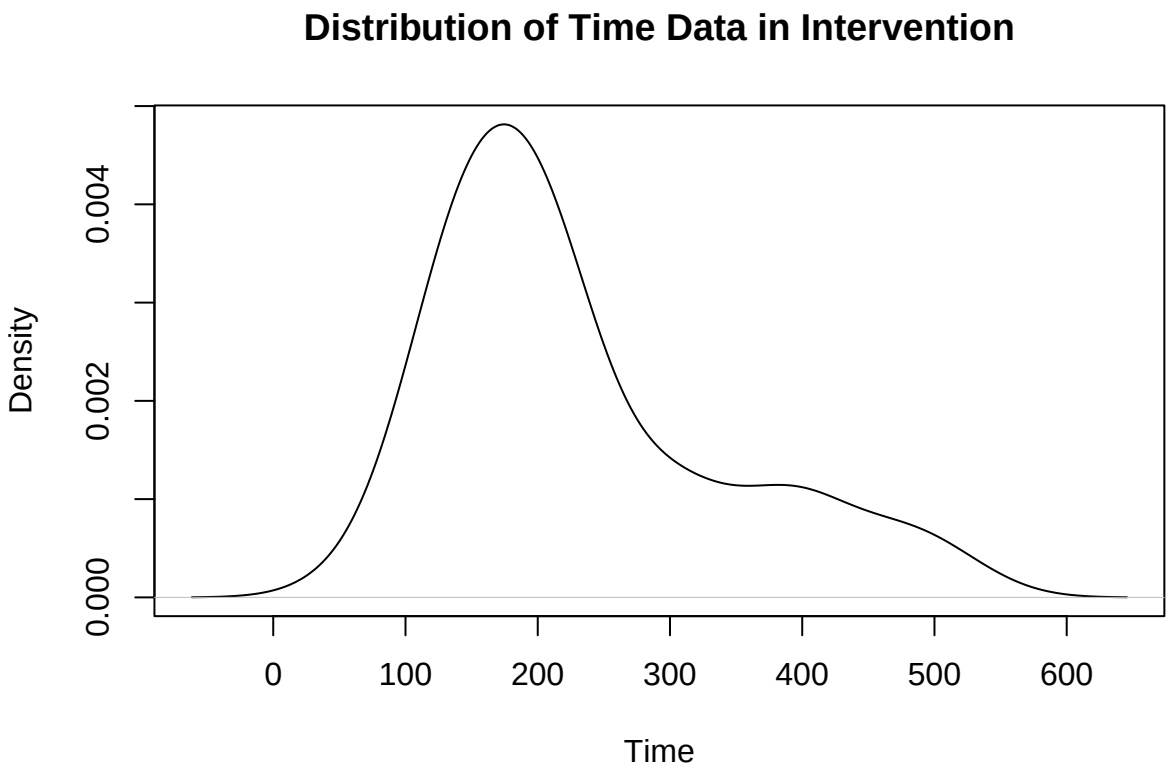
```
##   [1]  34 123  57  62  97 451 231 132 133 319 191 246 182 154 120 215  72  78
##  [19]  92 136 275 271 175 138 158 389 351 288 182 208 321 321 213 131 163 410
##  [37] 177 190 221 217 214  92  96  48  89 397 180 147 125 137 371 227 449  99
##  [55]  81 130 203 180 160 109 528 134 402 231  96  89 102  68 376  56 278 190
##  [73] 184 195  79  52 162 378 175 151 116 138  88  86  64  52 103 494 150 101
##  [91] 197 205 239 142  93  57  44 159 136 238 141 227 181 509 198  84 104 147
## [109] 484 135 222 195 110 170 124 136 503 316 138 203 164 405 315 301 307 294
## [127] 369 419 138 197 262 122  89 133 124  78 164 146 143 146 110 246 153 198
## [145] 284 184 103  80 124  65 245  50 125 109 418 232 214 204 184 190 175 141
## [163] 117  74 182 140 107  79  56  75  84  65 459 229 176 205 235 101  72  50
## [181] 136 304 474 192 127 136  32 115 102  57  49 202 239 241 137 191
```

21 participants spent an excess of 1.710672 minutes on average across the interfaces (54 instances in intervention, with 25 in baseline). We also found that this occured mostly when the participants interacted with the interface for the first time (29 instances found across 22 participants).

```
# Create a density plot of the time data
plot(density(subset(df, condition == "baseline")$delta_s), main = "Distribution of Time Data in Baselin
```

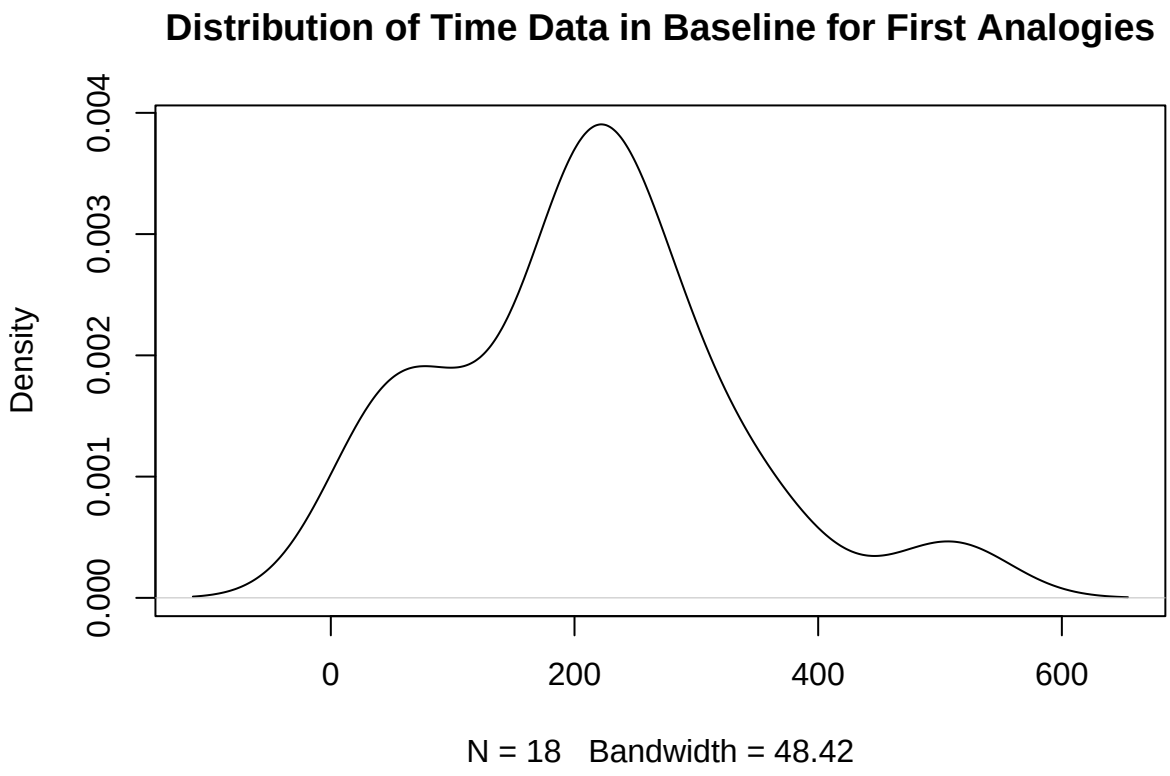



```
# Create a density plot of the time data
plot(density(subset(df, condition == "intervention")$delta_s), main = "Distribution of Time Data in Int
```

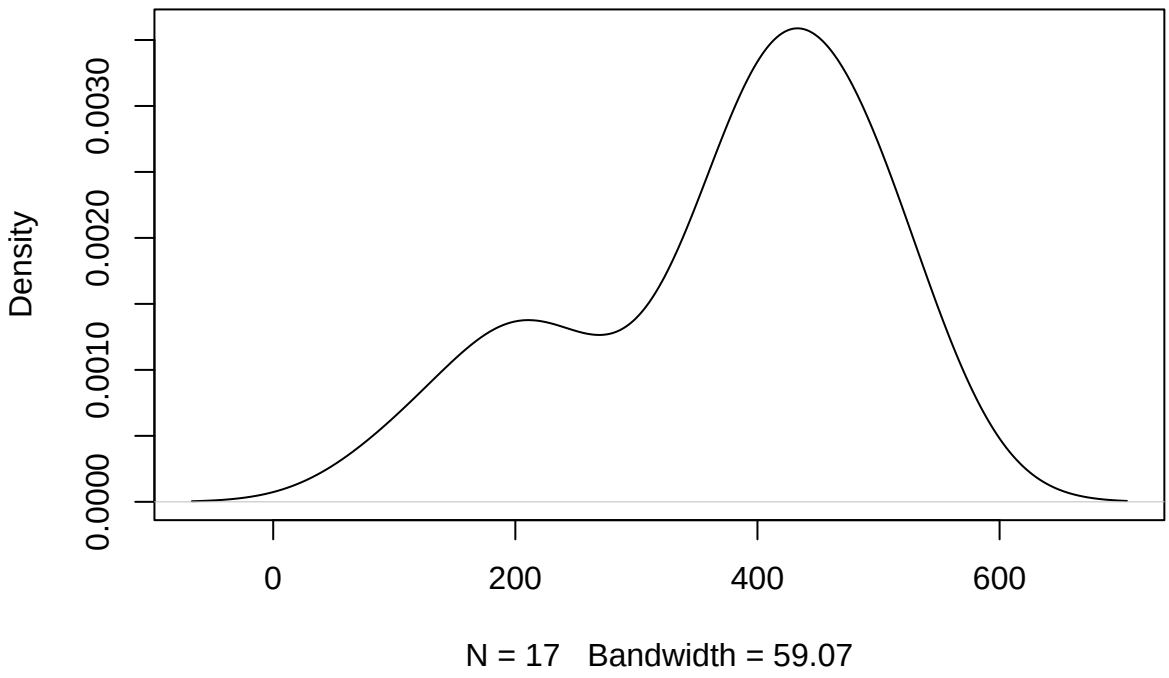


Now let's check for possible familiarity effects on time, particularly first analogies.

```
plot(density(subset(df, condition=="baseline" & current_analogy == 1 & next_analogy == 2)$delta_s),  ma
```




```
plot(density(subset(df, condition=="intervention" & current_analogy == 1 & next_analogy == 2)$delta_s),
```



As we can see, there is a familiarity effect when dealing with first analogies in general, irrespective of the interface, with more time taken with the intervention interface.

```
baseline.first <- density(subset(df, condition=="baseline" & current_analogy == 1 & next_analogy == 2)$
intervention.first <- density(subset(df, condition=="intervention" & current_analogy == 1 & next_analogy
# average time taken to process first analogies in baseline
mean(baseline.first$x)
```

```
## [1] 270.5
```

```
# average time taken to process first analogies in intervention
mean(intervention.first$x)
```

```
## [1] 319
```

```
# difference in processing time for first analogies in average - intervention took 48 seconds more when
mean(intervention.first$x) - mean(baseline.first$x)
```

```
## [1] 48.5
```

Hence to accomodate for this effect, we remove all instances of data where the participant processed the first analogy across the interfaces.

```
sum(is.na(df))
```

```
## [1] 0
```

Now, that the data is addressed for their discrepancies, we can perform GLMM analysis on the time taken to process analogies across condition and analogical_distance.

Constructing a Null Model to check Timing of Decisions

First, we define a **null model with no predictors** to analyze how long participants as a group took to process analogies and ignore them. This will help us see the improvement in fit as we add predictors to identify their effects. We use `gaussian()` as the family since we are dealing with continuous variables. The `delta_s` column is normalized while the `analogy_decision` is one-hot encoded.

```
# load the package
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
sdf <- df
#sdf$pid <- as.factor(df$pid)
#sdf$delta_s <- scale(sdf$delta_s) scaling before lmer does not make sense
# define the null model -> glmer with gaussian() is lmer()
fit.timing.null <- lmer(
                      delta_s ~ (1| pid),
                    data = sdf)

# show summary of model
summary(fit.timing.null, tTable=TRUE)
```

```
## Warning in summary.merMod(fit.timing.null, tTable = TRUE): additional arguments
## ignored
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: delta_s ~ (1 | pid)
##    Data: sdf
##
## REML criterion at convergence: 2386.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.3418 -0.7097 -0.2268  0.3773  3.1800
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  pid      (Intercept)   805.9   28.39
##  Residual             11146.4  105.58
## Number of obs: 196, groups:  pid, 22
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  184.170      9.733   18.92
```

Adding `analogical_distance` as a predictor

```
# load the package
library(lme4)
```

Table 2:

	Dependent variable:
	delta_s
Constant	184.170***
	(165.093, 203.247)
Observations	196
Log Likelihood	−1,193.095
Akaike Inf. Crit.	2,392.190
Bayesian Inf. Crit.	2,402.024
Note:	*p<0.1; **p<0.05; ***p<0.01

```
sdf <- df
#sdf$delta_s = scale(sdf$delta_s) scaling does not make sense

# define the null model -> glmer with gaussian() is lmer()
fit.timing.nearfar <- lmer(
                      delta_s ~ (1|pid) + analogy_distance,
                    data = sdf)

# show summary of model
summary(fit.timing.nearfar, tTable=TRUE)
```

```
## Warning in summary.merMod(fit.timing.nearfar, tTable = TRUE): additional
## arguments ignored
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: delta_s ~ (1 | pid) + analogy_distance
##    Data: sdf
##
## REML criterion at convergence: 2378.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.3934 -0.6610 -0.2110  0.3692  3.2308
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  pid      (Intercept)   846.6   29.1
##  Residual             11148.7  105.6
## Number of obs: 196, groups:  pid, 22
##
## Fixed effects:
##                      Estimate Std. Error t value
## (Intercept)            178.54      12.39   14.41
## analogy_distancenear    11.35      15.14    0.75
##
## Correlation of Fixed Effects:
##             (Intr)
## anlgy_dstnc -0.609
```

Table 3:

	Dependent variable:
	delta_s
analogy_distancenear	11.351
	(−18.328, 41.030)
Constant	178.538***
	(154.253, 202.824)
Observations	196
Log Likelihood	−1,189.180
Akaike Inf. Crit.	2,386.359
Bayesian Inf. Crit.	2,399.471
Note:	*p<0.1; **p<0.05; ***p<0.01

This result suggests that there is a statistically insignificant possibility that near analogies have some sort of relationship with the time taken. However nothing can be said without exploring their decisions and the conditions.

```
eval.nearfar.null.fit = anova(
    fit.timing.nearfar,
    fit.timing.null,
    test='LRT')
```

```
## refitting model(s) with ML (instead of REML)
```

```
eval.nearfar.null.fit
```

```
## Data: sdf
## Models:
## fit.timing.null: delta_s ~ (1 | pid)
## fit.timing.nearfar: delta_s ~ (1 | pid) + analogy_distance
##                     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## fit.timing.null        3 2398.6 2408.4 -1196.3   2392.6
## fit.timing.nearfar     4 2400.0 2413.1 -1196.0   2392.0 0.5416  1     0.4618
```

Checking the anova shows that there is little to no difference between the null model and analogical distance.

Checking `analogical_decision` as a predictor

```
# load the package
library(lme4)

sdf <- df
# sdf$delta_s = scale(sdf$delta_s) scaling does not make sense

# define the null model -> glmer with gaussian() is lmer()
fit.timing.decision <- lmer(
                      delta_s ~ (1| pid) + analogy_decision,
                    data = sdf)

# show summary of model
```

```
summary(fit.timing.decision, tTable=TRUE)
```

```
## Warning in summary.merMod(fit.timing.decision, tTable = TRUE): additional
## arguments ignored
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: delta_s ~ (1 | pid) + analogy_decision
##    Data: sdf
##
## REML criterion at convergence: 2367.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4012 -0.6564 -0.2650  0.3814  3.1495
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  pid      (Intercept)   748.7   27.36
##  Residual             11166.1  105.67
## Number of obs: 196, groups:  pid, 22
##
## Fixed effects:
##                         Estimate Std. Error t value
## (Intercept)               152.77      23.13   6.604
## analogy_decisionkeep       34.80      24.07   1.446
## analogy_decisionreview     39.55      31.00   1.275
##
## Correlation of Fixed Effects:
##             (Intr) anlgy_dcsnk
## anlgy_dcsnk -0.900
## anlgy_dcsnr -0.701  0.669
```

```
eval.decision.null.fit = anova(
    fit.timing.decision,
    fit.timing.null,
    test='LRT')
```

```
## refitting model(s) with ML (instead of REML)
```

```
eval.decision.null.fit
```

```
## Data: sdf
## Models:
## fit.timing.null: delta_s ~ (1 | pid)
## fit.timing.decision: delta_s ~ (1 | pid) + analogy_decision
##                      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## fit.timing.null         3 2398.6 2408.4 -1196.3   2392.6
## fit.timing.decision     5 2400.3 2416.7 -1195.1   2390.3 2.2998  2     0.3167
```

There seems to be a significant correlation between the time taken to process analogies and their decision to review a particular analogy later. In particular, the participants seemed to take more time when deciding to review than keep. The improvement in fit over the null model shows that adding `analogical_decision` as predictor better describes the data. However, we still do not know the effects of `condition`.

```
eval.decision.nearfar.fit = anova(
    fit.timing.decision,
    fit.timing.nearfar,
```

```
    test='LRT')
```

```
## refitting model(s) with ML (instead of REML)
```

```
eval.decision.nearfar.fit
```

```
## Data: sdf
## Models:
## fit.timing.nearfar: delta_s ~ (1 | pid) + analogy_distance
## fit.timing.decision: delta_s ~ (1 | pid) + analogy_decision
##                       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## fit.timing.nearfar       4 2400.0 2413.1 -1196.0   2392.0
## fit.timing.decision      5 2400.3 2416.7 -1195.1   2390.3 1.7582  1     0.1849
```

Comparing between `analogical_distance` and `analogical_decision` over fit, we find that `analogical_decision` as a predictor performs significantly better in describing the data.

Adding `condition` as an additional predictor

```
# load the package
library(lme4)

sdf <- df
#sdf$delta_s = scale(sdf$delta_s)

# define the null model -> glmer with gaussian() is lmer()
fit.timing.nearfar.x.condition <- glmer(
                      delta_s ~ (1| pid) + analogy_distance + condition,
                    data = sdf)
```

```
## Warning in glmer(delta_s ~ (1 | pid) + analogy_distance + condition, data =
## sdf): calling glmer() with family=gaussian (identity link) as a shortcut to
## lmer() is deprecated; please call lmer() directly
```

```
# show summary of model
summary(fit.timing.nearfar.x.condition)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: delta_s ~ (1 | pid) + analogy_distance + condition
##    Data: sdf
##
## REML criterion at convergence: 2332.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7802 -0.6329 -0.2878  0.3155  3.6082
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  pid      (Intercept) 1028     32.07
##  Residual             8984     94.78
## Number of obs: 196, groups:  pid, 22
##
## Fixed effects:
##                       Estimate Std. Error t value
## (Intercept)            138.105     13.312  10.375
```

```
## analogy_distancenear     6.311     13.636   0.463
## conditionintervention   89.730     13.620   6.588
##
## Correlation of Fixed Effects:
##             (Intr) anlgy_
## anlgy_dstnc -0.478
## cndtnntrvnt -0.460 -0.064
```

Table 4:

	Dependent variable:
	delta_s
analogy_distancenear	6.311
	(−20.414, 33.036)
conditionintervention	89.730***
	(63.035, 116.424)
Constant	138.105***
	(112.014, 164.195)
Observations	196
Log Likelihood	−1,166.130
Akaike Inf. Crit.	2,342.259
Bayesian Inf. Crit.	2,358.650
Note:	*p<0.1; **p<0.05; ***p<0.01

```
library(ggplot2)
library(dplyr)
library(tidyr)

fit.timing.nearfar.x.condition.plot <- select(sdf, analogy_distance, condition)
fit.timing.nearfar.x.condition.plot
```

```
##     analogy_distance    condition
## 1               near     baseline
## 2                far     baseline
## 3               near     baseline
## 4                far     baseline
## 5               near     baseline
## 6                far intervention
## 7               near intervention
## 8               near intervention
## 9                far intervention
## 10               far intervention
## 11               far intervention
## 12               far intervention
## 13              near intervention
## 14              near intervention
## 15              near intervention
## 16              near     baseline
## 17               far     baseline
```

```
## 18              near     baseline
## 19              near     baseline
## 20               far     baseline
## 21              near     baseline
## 22              near     baseline
## 23               far     baseline
## 24               far     baseline
## 25               far     baseline
## 26               far intervention
## 27              near intervention
## 28               far intervention
## 29              near intervention
## 30              near intervention
## 31              near     baseline
## 32               far     baseline
## 33              near     baseline
## 34               far     baseline
## 35               far     baseline
## 36               far intervention
## 37               far intervention
## 38              near intervention
## 39              near intervention
## 40              near intervention
## 41               far     baseline
## 42               far     baseline
## 43               far     baseline
## 44              near     baseline
## 45              near     baseline
## 46              near intervention
## 47              near intervention
## 48               far intervention
## 49               far intervention
## 50              near intervention
## 51               far intervention
## 52               far intervention
## 53              near     baseline
## 54              near     baseline
## 55               far     baseline
## 56              near     baseline
## 57               far intervention
## 58               far intervention
## 59              near intervention
## 60              near intervention
## 61               far intervention
## 62              near intervention
## 63              near intervention
## 64              near     baseline
## 65               far     baseline
## 66               far     baseline
## 67               far     baseline
## 68              near     baseline
## 69              near intervention
## 70               far intervention
## 71              near intervention
```

```
## 72              far intervention
## 73             near     baseline
## 74              far     baseline
## 75              far     baseline
## 76              far     baseline
## 77             near     baseline
## 78              far intervention
## 79             near intervention
## 80             near intervention
## 81             near intervention
## 82              far intervention
## 83              far     baseline
## 84              far     baseline
## 85             near     baseline
## 86              far     baseline
## 87             near     baseline
## 88             near intervention
## 89              far intervention
## 90              far intervention
## 91             near intervention
## 92             near intervention
## 93             near     baseline
## 94              far     baseline
## 95             near     baseline
## 96              far     baseline
## 97              far     baseline
## 98              far     baseline
## 99             near     baseline
## 100             far     baseline
## 101            near     baseline
## 102            near intervention
## 103             far intervention
## 104             far     baseline
## 105             far     baseline
## 106             far     baseline
## 107            near     baseline
## 108            near     baseline
## 109            near intervention
## 110             far intervention
## 111            near intervention
## 112            near     baseline
## 113             far     baseline
## 114             far     baseline
## 115            near     baseline
## 116             far     baseline
## 117            near intervention
## 118             far intervention
## 119            near intervention
## 120             far intervention
## 121             far intervention
## 122             far intervention
## 123            near intervention
## 124             far intervention
## 125            near intervention
```

```
## 126             far intervention
## 127            near     baseline
## 128             far     baseline
## 129            near     baseline
## 130             far     baseline
## 131             far     baseline
## 132            near     baseline
## 133             far     baseline
## 134             far     baseline
## 135            near     baseline
## 136            near     baseline
## 137            near intervention
## 138             far intervention
## 139             far intervention
## 140            near intervention
## 141             far intervention
## 142            near intervention
## 143             far intervention
## 144            near intervention
## 145            near     baseline
## 146             far     baseline
## 147             far     baseline
## 148             far     baseline
## 149            near     baseline
## 150            near     baseline
## 151             far     baseline
## 152             far     baseline
## 153            near     baseline
## 154             far     baseline
## 155             far intervention
## 156            near intervention
## 157            near intervention
## 158            near intervention
## 159             far intervention
## 160            near     baseline
## 161            near     baseline
## 162            near     baseline
## 163             far     baseline
## 164             far     baseline
## 165            near intervention
## 166             far intervention
## 167            near intervention
## 168            near intervention
## 169            near     baseline
## 170             far     baseline
## 171             far     baseline
## 172            near     baseline
## 173            near intervention
## 174            near intervention
## 175             far intervention
## 176             far intervention
## 177             far     baseline
## 178            near     baseline
## 179            near     baseline
```

```
## 180              far     baseline
## 181              far     baseline
## 182              far intervention
## 183             near intervention
## 184              far intervention
## 185             near intervention
## 186             near intervention
## 187              far     baseline
## 188             near     baseline
## 189             near     baseline
## 190              far     baseline
## 191             near     baseline
## 192              far intervention
## 193              far intervention
## 194             near intervention
## 195             near intervention
## 196              far intervention
```

```
pred_prob <- predict(fit.timing.nearfar.x.condition, data=fit.timing.nearfar.x.condition.plot, type = "
```

```
## Warning in predict.merMod(fit.timing.nearfar.x.condition, data =
## fit.timing.nearfar.x.condition.plot, : unused arguments ignored
```

```
fit.timing.nearfar.x.condition.plot$probs <- pred_prob
fit.timing.nearfar.x.condition.plot$se_prop <- sqrt(abs(pred_prob)*abs(1-pred_prob)/nrow(df))
fit.timing.nearfar.x.condition.plot
```

```
##     analogy_distance    condition    probs   se_prop
## 1               near     baseline 132.5531  9.432299
## 2                far     baseline 126.2419  8.981493
## 3               near     baseline 132.5531  9.432299
## 4                far     baseline 126.2419  8.981493
## 5               near     baseline 132.5531  9.432299
## 6                far intervention 215.9716 15.390789
## 7               near intervention 222.2829 15.841592
## 8               near intervention 222.2829 15.841592
## 9                far intervention 215.9716 15.390789
## 10               far intervention 215.9716 15.390789
## 11               far intervention 207.4683 14.783407
## 12               far intervention 207.4683 14.783407
## 13              near intervention 213.7795 15.234211
## 14              near intervention 213.7795 15.234211
## 15              near intervention 213.7795 15.234211
## 16              near     baseline 124.0498  8.824915
## 17               far     baseline 117.7386  8.374108
## 18              near     baseline 124.0498  8.824915
## 19              near     baseline 124.0498  8.824915
## 20               far     baseline 117.7386  8.374108
## 21              near     baseline 175.0400 12.467095
## 22              near     baseline 175.0400 12.467095
## 23               far     baseline 168.7288 12.016290
## 24               far     baseline 168.7288 12.016290
## 25               far     baseline 168.7288 12.016290
## 26               far intervention 258.4585 18.425575
## 27              near intervention 264.7698 18.876378
```

```
## 28              far intervention 258.4585 18.425575
## 29             near intervention 264.7698 18.876378
## 30             near intervention 264.7698 18.876378
## 31             near     baseline 171.2504 12.196404
## 32              far     baseline 164.9391 11.745599
## 33             near     baseline 171.2504 12.196404
## 34              far     baseline 164.9391 11.745599
## 35              far     baseline 164.9391 11.745599
## 36              far intervention 254.6689 18.154884
## 37              far intervention 254.6689 18.154884
## 38             near intervention 260.9801 18.605687
## 39             near intervention 260.9801 18.605687
## 40             near intervention 260.9801 18.605687
## 41              far     baseline 120.1571  8.546860
## 42              far     baseline 120.1571  8.546860
## 43              far     baseline 120.1571  8.546860
## 44             near     baseline 126.4683  8.997667
## 45             near     baseline 126.4683  8.997667
## 46             near intervention 216.1980 15.406962
## 47             near intervention 216.1980 15.406962
## 48              far intervention 209.8868 14.956158
## 49              far intervention 209.8868 14.956158
## 50             near intervention 216.1980 15.406962
## 51              far intervention 241.0928 17.185164
## 52              far intervention 241.0928 17.185164
## 53             near     baseline 139.3779  9.919785
## 54             near     baseline 139.3779  9.919785
## 55              far     baseline 133.0667  9.468979
## 56             near     baseline 139.3779  9.919785
## 57              far intervention 222.7964 15.878272
## 58              far intervention 222.7964 15.878272
## 59             near intervention 229.1076 16.329076
## 60             near intervention 229.1076 16.329076
## 61              far intervention 242.8165 17.308287
## 62             near intervention 249.1278 17.759090
## 63             near intervention 249.1278 17.759090
## 64             near     baseline 159.3980 11.349804
## 65              far     baseline 153.0868 10.898999
## 66              far     baseline 153.0868 10.898999
## 67              far     baseline 153.0868 10.898999
## 68             near     baseline 159.3980 11.349804
## 69             near intervention 231.0385 16.466997
## 70              far intervention 224.7273 16.016193
## 71             near intervention 231.0385 16.466997
## 72              far intervention 224.7273 16.016193
## 73             near     baseline 141.3088 10.057706
## 74              far     baseline 134.9975  9.606901
## 75              far     baseline 134.9975  9.606901
## 76              far     baseline 134.9975  9.606901
## 77             near     baseline 141.3088 10.057706
## 78              far intervention 200.5995 14.292775
## 79             near intervention 206.9107 14.743579
## 80             near intervention 206.9107 14.743579
## 81             near intervention 206.9107 14.743579
```

```
## 82              far intervention 200.5995 14.292775
## 83              far     baseline 110.8698  7.883473
## 84              far     baseline 110.8698  7.883473
## 85             near     baseline 117.1810  8.334280
## 86              far     baseline 110.8698  7.883473
## 87             near     baseline 117.1810  8.334280
## 88             near intervention 226.7130 16.158034
## 89              far intervention 220.4018 15.707230
## 90              far intervention 220.4018 15.707230
## 91             near intervention 226.7130 16.158034
## 92             near intervention 226.7130 16.158034
## 93             near     baseline 136.9833  9.748742
## 94              far     baseline 130.6721  9.297936
## 95             near     baseline 136.9833  9.748742
## 96              far     baseline 130.6721  9.297936
## 97              far     baseline 130.6721  9.297936
## 98              far     baseline 141.8358 10.095352
## 99             near     baseline 148.1471 10.546158
## 100             far     baseline 141.8358 10.095352
## 101            near     baseline 148.1471 10.546158
## 102            near intervention 237.8768 16.955446
## 103             far intervention 231.5655 16.504643
## 104             far     baseline 167.0096 11.893487
## 105             far     baseline 167.0096 11.893487
## 106             far     baseline 167.0096 11.893487
## 107            near     baseline 173.3208 12.344292
## 108            near     baseline 173.3208 12.344292
## 109            near intervention 263.0505 18.753575
## 110             far intervention 256.7393 18.302772
## 111            near intervention 263.0505 18.753575
## 112            near     baseline 155.3077 11.057636
## 113             far     baseline 148.9965 10.606831
## 114             far     baseline 148.9965 10.606831
## 115            near     baseline 155.3077 11.057636
## 116             far     baseline 148.9965 10.606831
## 117            near intervention 245.0374 17.466923
## 118             far intervention 238.7262 17.016120
## 119            near intervention 245.0374 17.466923
## 120             far intervention 238.7262 17.016120
## 121             far intervention 238.7262 17.016120
## 122             far intervention 289.3262 20.630411
## 123            near intervention 295.6374 21.081214
## 124             far intervention 289.3262 20.630411
## 125            near intervention 295.6374 21.081214
## 126             far intervention 289.3262 20.630411
## 127            near     baseline 205.9077 14.671935
## 128             far     baseline 199.5965 14.221131
## 129            near     baseline 205.9077 14.671935
## 130             far     baseline 199.5965 14.221131
## 131             far     baseline 199.5965 14.221131
## 132            near     baseline 116.8765  8.312529
## 133             far     baseline 110.5652  7.861722
## 134             far     baseline 110.5652  7.861722
## 135            near     baseline 116.8765  8.312529
```

```
## 136            near     baseline 116.8765  8.312529
## 137            near intervention 206.6062 14.721828
## 138             far intervention 200.2950 14.271024
## 139             far intervention 200.2950 14.271024
## 140            near intervention 206.6062 14.721828
## 141             far intervention 219.6526 15.653717
## 142            near intervention 225.9638 16.104520
## 143             far intervention 219.6526 15.653717
## 144            near intervention 225.9638 16.104520
## 145            near     baseline 136.2341  9.695228
## 146             far     baseline 129.9229  9.244422
## 147             far     baseline 129.9229  9.244422
## 148             far     baseline 129.9229  9.244422
## 149            near     baseline 136.2341  9.695228
## 150            near     baseline 143.6019 10.221499
## 151             far     baseline 137.2906  9.770694
## 152             far     baseline 137.2906  9.770694
## 153            near     baseline 143.6019 10.221499
## 154             far     baseline 137.2906  9.770694
## 155             far intervention 227.0204 16.179986
## 156            near intervention 233.3316 16.630789
## 157            near intervention 233.3316 16.630789
## 158            near intervention 233.3316 16.630789
## 159             far intervention 227.0204 16.179986
## 160            near     baseline 119.9037  8.528763
## 161            near     baseline 119.9037  8.528763
## 162            near     baseline 119.9037  8.528763
## 163             far     baseline 113.5925  8.077956
## 164             far     baseline 113.5925  8.077956
## 165            near intervention 209.6334 14.938061
## 166             far intervention 203.3222 14.487257
## 167            near intervention 209.6334 14.938061
## 168            near intervention 209.6334 14.938061
## 169            near     baseline 136.0502  9.682094
## 170             far     baseline 129.7390  9.231289
## 171             far     baseline 129.7390  9.231289
## 172            near     baseline 136.0502  9.682094
## 173            near intervention 225.7800 16.091386
## 174            near intervention 225.7800 16.091386
## 175             far intervention 219.4687 15.640583
## 176             far intervention 219.4687 15.640583
## 177             far     baseline 136.2765  9.698255
## 178            near     baseline 142.5877 10.149061
## 179            near     baseline 142.5877 10.149061
## 180             far     baseline 136.2765  9.698255
## 181             far     baseline 136.2765  9.698255
## 182             far intervention 226.0062 16.107547
## 183            near intervention 232.3175 16.558351
## 184             far intervention 226.0062 16.107547
## 185            near intervention 232.3175 16.558351
## 186            near intervention 232.3175 16.558351
## 187             far     baseline 111.6170  7.936849
## 188            near     baseline 117.9282  8.387656
## 189            near     baseline 117.9282  8.387656
```

```
## 190              far     baseline 111.6170  7.936849
## 191             near     baseline 117.9282  8.387656
## 192              far intervention 201.3467 14.346151
## 193              far intervention 201.3467 14.346151
## 194             near intervention 207.6580 14.796954
## 195             near intervention 207.6580 14.796954
## 196              far intervention 201.3467 14.346151
```

```
summarize_data <- function(data) {
  summary_data <- data %>%
    group_by(analogy_distance, condition) %>%
    summarise(avg_probs = mean(probs), avg_se_prop = mean(se_prop))
  return(summary_data)
}

plot_grouped_barchart <- function(data) {
  plot <- ggplot(data, aes(x = condition, y = avg_probs, fill = analogy_distance)) +
    geom_bar(stat = "identity", position = "dodge") +
    geom_errorbar(aes(ymin = avg_probs - avg_se_prop, ymax = avg_probs + avg_se_prop), width = 0.2, posi
    labs(title = "Processing time by Condition and Distance",
         x = "Condition",
         y = "Average Processing Time per Analogy",
         fill = "Analogy Distance") +
    theme_bw()
  return(plot)
}

data_summary <- summarize_data(fit.timing.nearfar.x.condition.plot)
data_summary
```

```
## # A tibble: 4 x 4
## # Groups:   analogy_distance [2]
##   analogy_distance condition    avg_probs avg_se_prop
##   <chr>            <chr>            <dbl>       <dbl>
## 1 far              baseline          140.        9.94
## 2 far              intervention      228.       16.3
## 3 near             baseline          141.       10.1
## 4 near             intervention      233.       16.6
```

```
plot_grouped_barchart(data_summary)
```

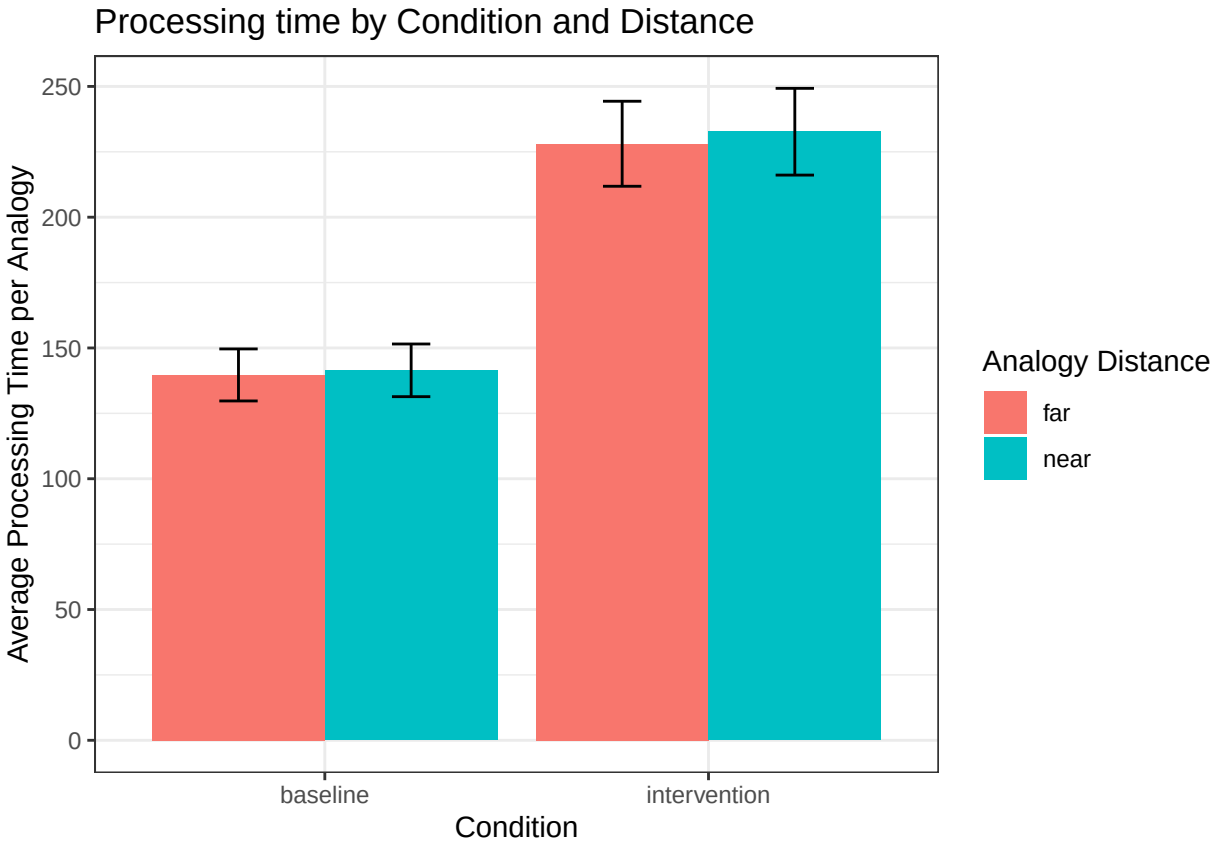


Table 5:

	Dependent variable:
	delta_s
analogy_distancenear	6.311
	(13.636)
conditionintervention	89.730***
	(13.620)
Constant	138.105***
	(13.312)
Observations	196
Log Likelihood	−1,166.130
Akaike Inf. Crit.	2,342.259
Bayesian Inf. Crit.	2,358.650
Note:	*p<0.1; **p<0.05; ***p<0.01

After adding `condition` as an additional predictor, we find that, in general, people tended to take more time when interacting with the intervention interface, irrespective of analogical distance.

```
# load the package
library(lme4)
```

```
# define the null model -> glmer with gaussian() is lmer()
fit.timing.nearfar.x.condition <- glmer(
                      delta_s ~ (1| pid) + analogy_distance + condition + analogy_distance:condition,
                    data = df)
```

```
## Warning in glmer(delta_s ~ (1 | pid) + analogy_distance + condition +
## analogy_distance:condition, : calling glmer() with family=gaussian (identity
## link) as a shortcut to lmer() is deprecated; please call lmer() directly
```

```
# show summary of model
summary(fit.timing.nearfar.x.condition, tTable=TRUE)
```

```
## Warning in summary.merMod(fit.timing.nearfar.x.condition, tTable = TRUE):
## additional arguments ignored
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## delta_s ~ (1 | pid) + analogy_distance + condition + analogy_distance:condition
##    Data: df
##
## REML criterion at convergence: 2323.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8176 -0.6158 -0.2584  0.2932  3.6305
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  pid      (Intercept) 1035     32.17
##  Residual             9017     94.96
## Number of obs: 196, groups:  pid, 22
##
## Fixed effects:
##                                            Estimate Std. Error t value
## (Intercept)                                  134.98      14.74   9.160
## analogy_distancenear                          12.87      18.94   0.680
## conditionintervention                         96.63      19.42   4.976
## analogy_distancenear:conditionintervention   -13.69      27.40  -0.500
##
## Correlation of Fixed Effects:
##             (Intr) anlgy_ cndtnn
## anlgy_dstnc -0.606
## cndtnntrvnt -0.594  0.460
## anlgy_dstn:  0.425 -0.693 -0.712
```

Testing for interaction effects with `analogy_decision`

```
# load the package
library(lme4)

# define the null model -> glmer with gaussian() is lmer()
fit.timing.decision.x.condition.interaction <- glmer(
                      delta_s ~ (1| pid) + analogy_decision + condition + analogy_decision:condition,
                    data = df)
```

```
## Warning in glmer(delta_s ~ (1 | pid) + analogy_decision + condition +
## analogy_decision:condition, : calling glmer() with family=gaussian (identity
## link) as a shortcut to lmer() is deprecated; please call lmer() directly
```

```
# show summary of model
summary(fit.timing.decision.x.condition.interaction, tTable=TRUE)
```

```
## Warning in summary.merMod(fit.timing.decision.x.condition.interaction, tTable =
## TRUE): additional arguments ignored
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## delta_s ~ (1 | pid) + analogy_decision + condition + analogy_decision:condition
##    Data: df
##
## REML criterion at convergence: 2300.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8594 -0.6259 -0.2569  0.2472  3.5307
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  pid      (Intercept) 1010     31.78
##  Residual             8973     94.73
## Number of obs: 196, groups:  pid, 22
##
## Fixed effects:
##                                             Estimate Std. Error t value
## (Intercept)                                   118.95      23.73   5.013
## analogy_decisionkeep                           27.53      25.33   1.087
## analogy_decisionreview                         20.18      44.67   0.452
## conditionintervention                         168.78      49.32   3.422
## analogy_decisionkeep:conditionintervention    -82.49      51.90  -1.589
## analogy_decisionreview:conditionintervention  -96.77      67.50  -1.434
##
## Correlation of Fixed Effects:
##              (Intr) anlgy_dcsnk anlgy_dcsnr cndtnn anlgy_dcsnk:
## anlgy_dcsnk  -0.864
## anlgy_dcsnr  -0.490  0.449
## cndtnntrvnt  -0.442  0.416       0.247
## anlgy_dcsnk:  0.423 -0.490      -0.226      -0.952
## anlgy_dcsnr:  0.323 -0.295      -0.678      -0.743  0.695
```

Checking for the goodness of fit across interactions:

```
eval.decision.nearfar.interaction.fit = anova(
    fit.timing.nearfar.x.condition,
    fit.timing.decision.x.condition.interaction,
    test='LRT')
```

```
## refitting model(s) with ML (instead of REML)
```

```
eval.decision.nearfar.interaction.fit
```

```
## Data: df
## Models:
```

```
## fit.timing.nearfar.x.condition: delta_s ~ (1 | pid) + analogy_distance + condition + analogy_distanc
## fit.timing.decision.x.condition.interaction: delta_s ~ (1 | pid) + analogy_decision + condition + ana
##                                              npar    AIC    BIC  logLik deviance
## fit.timing.nearfar.x.condition                  6 2364.5 2384.1 -1176.2   2352.5
## fit.timing.decision.x.condition.interaction     8 2365.2 2391.5 -1174.6   2349.2
##                                              Chisq Df Pr(>Chisq)
## fit.timing.nearfar.x.condition
## fit.timing.decision.x.condition.interaction 3.242  2     0.1977
```

```
library(lme4)

# define the null model -> glmer with gaussian() is lmer()
fit.timing.distance.x.decision.x.condition.interaction <- glmer(
                      delta_s ~ (1| pid) + analogy_decision + condition + analogy_distance:condition + a
                    data = df)
```

```
## Warning in glmer(delta_s ~ (1 | pid) + analogy_decision + condition +
## analogy_distance:condition + : calling glmer() with family=gaussian (identity
## link) as a shortcut to lmer() is deprecated; please call lmer() directly
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

```
# show summary of model
summary(fit.timing.distance.x.decision.x.condition.interaction, tTable=TRUE)
```

```
## Warning in
## summary.merMod(fit.timing.distance.x.decision.x.condition.interaction, :
## additional arguments ignored
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## delta_s ~ (1 | pid) + analogy_decision + condition + analogy_distance:condition +
##     analogy_distance:analogy_decision + analogy_decision:condition +
##     analogy_distance:analogy_decision:condition
##    Data: df
##
## REML criterion at convergence: 2250.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8793 -0.6071 -0.2497  0.2637  3.5570
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  pid      (Intercept) 1219     34.92
##  Residual             8916     94.43
## Number of obs: 196, groups:  pid, 22
##
## Fixed effects:
##                                                      Estimate
## (Intercept)                                           122.131
## analogy_decisionkeep                                   17.582
## analogy_decisionreview                                  9.175
## conditionintervention                                 166.436
## conditionbaseline:analogy_distancenear                 -8.171
## conditionintervention:analogy_distancenear           -105.764
## analogy_decisionkeep:analogy_distancenear              20.355
```

```
## analogy_decisionreview:analogy_distancenear                        35.303
## analogy_decisionkeep:conditionintervention                        -88.157
## analogy_decisionreview:conditionintervention                      -42.255
## analogy_decisionkeep:conditionintervention:analogy_distancenear  112.871
##                                                                  Std. Error
## (Intercept)                                                         27.825
## analogy_decisionkeep                                                31.325
## analogy_decisionreview                                              52.975
## conditionintervention                                               51.513
## conditionbaseline:analogy_distancenear                              51.124
## conditionintervention:analogy_distancenear                         105.595
## analogy_decisionkeep:analogy_distancenear                           55.773
## analogy_decisionreview:analogy_distancenear                         95.454
## analogy_decisionkeep:conditionintervention                          56.629
## analogy_decisionreview:conditionintervention                        79.479
## analogy_decisionkeep:conditionintervention:analogy_distancenear     98.144
##                                                                  t value
## (Intercept)                                                        4.389
## analogy_decisionkeep                                               0.561
## analogy_decisionreview                                             0.173
## conditionintervention                                              3.231
## conditionbaseline:analogy_distancenear                            -0.160
## conditionintervention:analogy_distancenear                        -1.002
## analogy_decisionkeep:analogy_distancenear                          0.365
## analogy_decisionreview:analogy_distancenear                        0.370
## analogy_decisionkeep:conditionintervention                        -1.557
## analogy_decisionreview:conditionintervention                      -0.532
## analogy_decisionkeep:conditionintervention:analogy_distancenear    1.150
##
## Correlation of Fixed Effects:
##               (Intr) anlgy_dcsnk anlgy_dcsnr cndtnn cndtnb:_ cndtnn:_
## anlgy_dcsnk   -0.828
## anlgy_dcsnr   -0.507  0.442
## cndtnntrvnt   -0.507  0.451       0.296
## cndtnbsln:_   -0.512  0.455       0.307       0.289
## cndtnntrv:_   -0.246  0.217       0.483       0.155  0.480
## anlgy_dcsnk:_  0.473 -0.565      -0.283      -0.264 -0.921   -0.439
## anlgy_dcsnr:_  0.274 -0.239      -0.534      -0.169 -0.543   -0.899
## anlgy_dcsnk:   0.466 -0.558      -0.264      -0.912 -0.265   -0.142
## anlgy_dcsnr:   0.340 -0.298      -0.687      -0.674 -0.207   -0.493
## anlgy_dc::_   -0.006  0.090      -0.355      -0.014  0.010   -0.825
##               anlgy_dcsnk:_ anlgy_dcsnr:_ anlgy_dcsnk: anlgy_dcsnr:
## anlgy_dcsnk
## anlgy_dcsnr
## cndtnntrvnt
## cndtnbsln:_
## cndtnntrv:_
## anlgy_dcsnk:_
## anlgy_dcsnr:_  0.496
## anlgy_dcsnk:   0.322         0.154
## anlgy_dcsnr:   0.190         0.366         0.608
## anlgy_dc::_   -0.100         0.683        -0.083        0.418
## fit warnings:
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

The fixed effects coefficients indicate that there is a **significant main effect of condition** on `delta_s`, with participants in the intervention condition taking longer to process the analogies than participants in the baseline condition (Estimate = 144.35 seconds (~2 minutes more), t-value = 2.570 (statistically significant)).

Visual Analysis

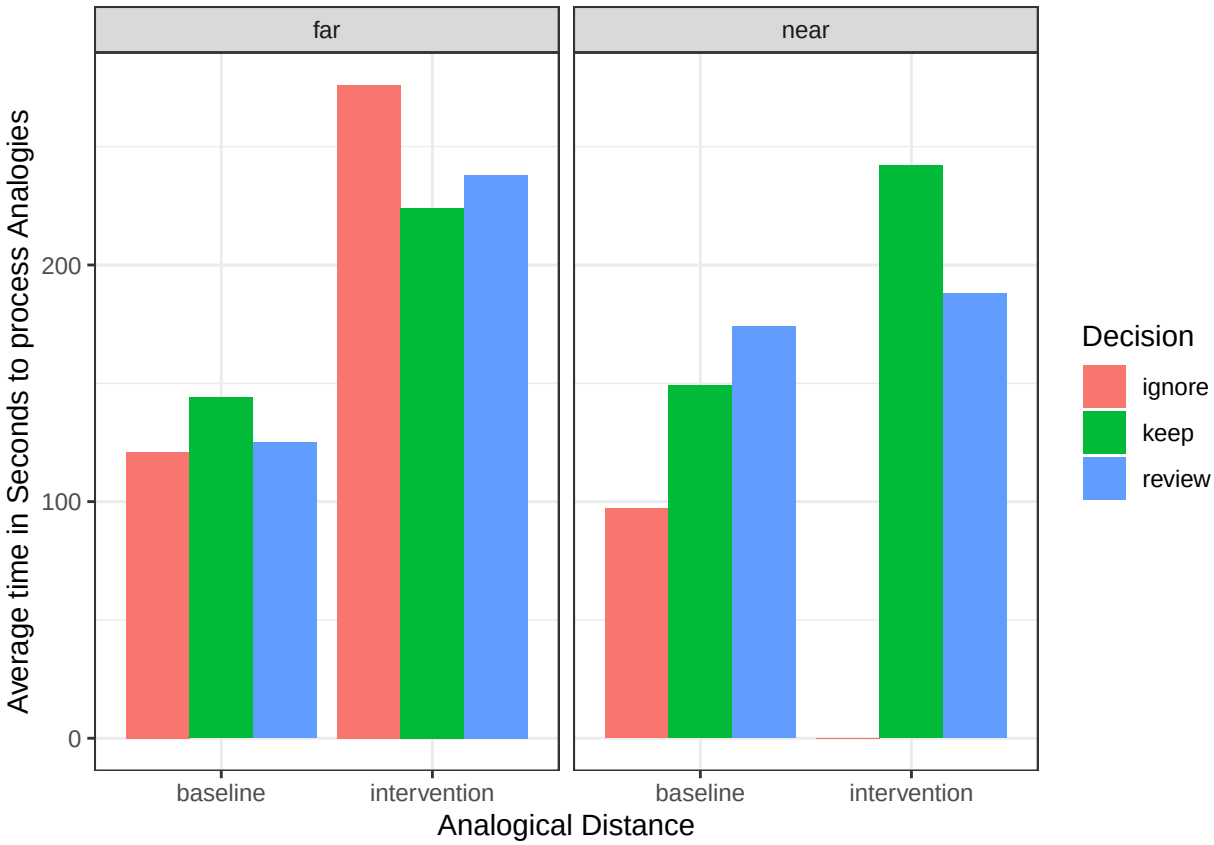
Lets take a look at the raw plots:

```
library(pivottabler)

# get a pivot table of counts
pt.timing <- PivotTable$new()
pt.timing$addData(df)
pt.timing$addColumnDataGroups("condition")
pt.timing$addRowDataGroups("analogy_distance")
pt.timing$addRowDataGroups("analogy_decision")
pt.timing$defineCalculation(
  calculationName="MeanSeconds",
  summariseExpression="round(mean(delta_s, na.rm=TRUE))")
pt.timing$evaluatePivot()
out.timing <- pt.timing$asDataFrame()
out.timing <- out.timing[-c(4,8,9),c("baseline", "intervention")]
out.timing
```

```
        baseline intervention
```

far ignore 121 276 far keep 144 224 far review 125 238 near ignore 97 NA near keep 149 242 near review 174 188

```
library(ggplot2)

# create sample data
dat <- c(out.timing$baseline, out.timing$intervention)
dat[is.na(dat)] <- 0
data <- data.frame(
  analogy_type = factor(rep(c("baseline", "intervention"), each = 6)),
  group = factor(rep(c("far", "near"), each = 3, times = 2)),
  decision = factor(rep(c("ignore", "keep", "review"), times = 4)),
  time = dat
)

# plot
ggplot(data, aes(x = analogy_type, y = time, fill = decision)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_grid(. ~ group) +
  labs(x = "Analogical Distance", y = "Average time in Seconds to process Analogies", fill = "Decision"
  theme_bw()
```


far
near
Average time in Seconds to process Analogies
200
100
0
baseline
intervention
baseline
intervention
Analogical Distance
Decision
ignore
keep
review

Bibliography

Keith J. Holyoak and Paul Thagard. The analogical mind. *American Psychologist*, 52(1):35–44, Jan 1997. ISSN 1935-990X, 0003-066X. doi: 10.1037/0003-066X.52.1.35.

Oliver Gassmann and Marco Zeschky. Opening up the solution space: The role of analogical thinking for breakthrough product innovation. *Creativity and Innovation Management*, 17(2): 97–106, Jun 2008. ISSN 0963-1690, 1467-8691. doi: 10.1111/j.1467-8691.2008.00475.x.

Dedre Gentner. Structure-mapping: A theoretical framework for analogy*. *Cognitive Science*, 7 (2):155–170, Apr 1983. ISSN 03640213. doi: 10.1207/s15516709cog0702_3. URL http://doi.wiley.com/10.1207/s15516709cog0702_3.

Keith J. Holyoak and Kyunghee Koh. Surface and structural similarity in analogical transfer. *Memory & Cognition*, 15(4):332–340, Jul 1987. ISSN 0090-502X, 1532-5946. doi: 10.3758/BF03197035. URL http://link.springer.com/10.3758/BF03197035.

Keith J. Holyoak and Paul Thagard. Analogical mapping by constraint satisfaction. *Cognitive Science*, 13(3):295–355, Jul 1989. ISSN 03640213. doi: 10.1207/s15516709cog1303_1. URL http://doi.wiley.com/10.1207/s15516709cog1303_1.

Dedre Gentner and Arthur B. Markman. Structure mapping in analogy and similarity. *American Psychologist*, 52(1):45–56, Jan 1997. ISSN 1935-990X, 0003-066X. doi: 10.1037/0003-066X.

52.1.45. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.52.1.45.

D. Gentner, S. Brem, R. W. Ferguson, P. Wolff, A. B. Markman, and K. D. Forbus. *Analogy and Creativity in the Works of Johannes Kepler*, page 403–459. American Psychological Association, Washington D.C., 1997. Citation Key: gentnerAnalogyCreativityWorks1997.

Robert S. Cohen and John J. Stachel. Newton and the Law of Gravitation [1965d]. In Robert S. Cohen, Marx W. Wartofsky, Robert S. Cohen, and John J. Stachel, editors, *Selected Papers of Léon Rosenfeld*, volume 21, pages 58–87. Springer Netherlands, Dordrecht, 1979. ISBN 978-90-277-0652-2 978-94-009-9349-5. doi: 10.1007/978-94-009-9349-5_9.

J. W. Herivel. Newton's discovery of the law of centrifugal force. *Isis*, 51(4):546–553, Dec 1960. ISSN 0021-1753, 1545-6994. doi: 10.1086/349412. URL https://www.journals.uchicago.edu/doi/10.1086/349412.

George Mestral. Separable fastening device, Nov 1961. URL https://patents.google.com/patent/US3009235A/en.

Terry Winograd and Fernando Flores. *Understanding computers and cognition: a new foundation for design*. Addison-Wesley, Boston, 24th printing edition, 2008. ISBN 978-0-201-11297-9.

Mary L. Gick and Keith J. Holyoak. Analogical problem solving. *Cognitive Psychology*, 12(3): 306–355, 1980. doi: 10.1016/0010-0285(80)90013-4. 03379.

Kristian J. Hammond, Colleen M. Seifert, and Kenneth C. Gray. Functionality in analogical transfer: A hard match is good to find. *The Journal of the Learning Sciences*, 1(2):111–152, 1991. ISSN 1050-8406. URL https://www.jstor.org/stable/1466674.

D. Gentner, M.J. Rattermann, and K.D. Forbus. The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, 25(4):524–575, Oct 1993. ISSN 00100285. doi: 10.1006/cogp.1993.1013.

Mark T. Keane. What makes an analogy difficult? the effects of order and causal structure on analogical mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(4):946–967, 1997. ISSN 1939-1285, 0278-7393. doi: 10.1037/0278-7393.23.4.946.

Graeme S. Halford, Rosemary Baker, Julie E. McCredden, and John D. Bain. How many variables can humans process? *Psychological Science*, 16(1):70–76, Jan 2005. ISSN 0956-7976, 1467-9280. doi: 10.1111/j.0956-7976.2005.00782.x.

Giovanni Gavetti, Daniel A. Levinthal, and Jan W. Rivkin. Strategy making in novel and complex worlds: the power of analogy. *Strategic Management Journal*, 26(8):691–712, Aug 2005. ISSN 0143-2095, 1097-0266. doi: 10.1002/smj.475.

Mary L. Gick and Keith J. Holyoak. Schema induction and analogical transfer. *Cognitive Psychology*, 15(1):1–38, Jan 1983. ISSN 00100285. doi: 10.1016/0010-0285(83)90002-6.

Lindsey E. Richland and Ian M. McDonough. Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology*, 35(1):28–43, Jan 2010. ISSN 0361476X. doi: 10.1016/j.cedpsych.2009.09.001.

Nicolas Kokkalis, Thomas Köhn, Johannes Huebner, Moontae Lee, Florian Schulze, and Scott R. Klemmer. Taskgenies: Automatically providing action plans helps people complete tasks. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(5):27, 2013. doi: 10.1145/2513560. 00000 Citation Key: kokkalisTaskgeniesAutomaticallyProviding2013.

Rand J. Spiro, Richard L. Coulson, P. J. Feltovich, and Daniel K. Anderson. *Cognitive flexibility theory advanced knowledge acquisition in ill-structured domains /.* Number 441 in Reading Research and Education Center Report. Champaign, IL, 1988. 01887 Citation Key: spiroCognitiveFlexibilityTheory1988.

Lixiu Yu, Aniket Kittur, and Robert E. Kraut. Searching for analogical ideas with crowds. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, page 1225–1234, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2557378. URL http://doi.acm.org/10.1145/2556288.2557378. Citation Key: yuSearchingAnalogicalIdeas2014.

Hyeonsu B. Kang, Xin Qian, Tom Hope, Dafna Shahaf, Joel Chan, and Aniket Kittur. Augmenting scientific creativity with an analogical search engine. *ACM Transactions on Computer-Human Interaction*, Mar 2022. ISSN 1073-0516. doi: 10.1145/3530013. URL https://doi.org/10.1145/3530013. Just Accepted Citation Key: kangAugmentingScientificCreativity2022.

Diederik A. Stapel and Piotr Winkielman. Assimilation and contrast as a function of context-target similarity, distinctness, and dimensional relevance. *Personality and Social Psychology Bulletin*, 24(6):634–646, Jun 1998. ISSN 0146-1672, 1552-7433. doi: 10.1177/0146167298246007.

Mariano Mastrogiorgio and Victor Gilsing. Innovation through exaptation and its determinants: The role of technological complexity, analogy making & patent scope. *Research Policy*, 45

(7):1419–1435, Sep 2016. ISSN 0048-7333. doi: 10.1016/j.respol.2016.04.003. Citation Key: mastrogiorgioInnovationExaptationIts2016.

J. S. Linsey, A. B. Markman, and K. L. Wood. Design by analogy: A study of the wordtree method for problem re-representation. *Journal of Mechanical Design*, 134(4):041009, 2012. ISSN 10500472. doi: 10.1115/1.4006145. Citation Key: linseyDesignAnalogyStudy2012.

Katherine Fu, Joel Chan, Jonathan Cagan, Kenneth Kotovsky, Christian Schunn, and Kristin Wood. The meaning of "near" and "far": The impact of structuring design databases and the effect of distance of analogy on design output. *Journal of Mechanical Design*, 135(2):021007, Feb 2013. ISSN 1050-0472, 1528-9001. doi: 10.1115/1.4023158.

Joel Chan, Joseph Chee Chang, Tom Hope, Dafna Shahaf, and Aniket Kittur. Solvent: A mixed initiative system for finding analogies between research papers. 2018. doi: 10/ggv7np.

F.W. Hesse and D. Klecha. Use of analogies in problem solving. *Computers in Human Behavior*, 6(1):115–129, Jan 1990. ISSN 07475632. doi: 10.1016/0747-5632(90)90034-E.

Laura R. Novick. Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3):510–520, 1988. ISSN 1939-1285, 0278-7393. doi: 10.1037/0278-7393.14.3.510.

Olufunmilola Atilola, Megan Tomko, and Julie S. Linsey. The effects of representation on idea generation and design fixation: A study comparing sketches and function trees. *Design Studies*, 42:110–136, Jan 2016. ISSN 0142694X. doi: 10.1016/j.destud.2015.10.005.

David E. Brown and John Clement. Overcoming misconceptions via analogical reasoning: ab-

stract transfer versus explanatory model construction. *Instructional Science*, 18(4):237–261, Dec 1989. ISSN 0020-4277, 1573-1952. doi: 10.1007/BF00118013.

Karl Duncker. On problem-solving. *Psychological Monographs*, 58(5):i–113, 1945. ISSN 0096-9753. doi: 10.1037/h0093599. Citation Key: dunckerProblemsolving1945.

Sam Glucksberg and Robert W. Weisberg. Verbal behavior and problem solving: Some effects of labeling in a functional fixedness problem. *Journal of Experimental Psychology*, 71(5): 659–664, 1966. ISSN 0022-1015. doi: 10.1037/h0023118. Citation Key: glucksbergVerbalBehaviorProblem1966.

Michael C Frank and Michael Ramscar. How do presentation and context influence representation for functional fixedness tasks? In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2003. Citation Key: frankHowPresentationContext2003.

David G. Jansson and Steven M. Smith. Design fixation. *Design Studies*, 12(1):3–11, 1991. doi: 10.1016/0142-694X(91)90003-F. Citation Key: janssonDesignFixation1991.

Marine Agogué, Akin Kazakçi, Armand Hatchuel, Pascal Le Masson, Benoit Weil, Nicolas Poirel, and Mathieu Cassotti. The impact of type of examples on originality: Explaining fixation and stimulation effects. *The Journal of Creative Behavior*, 48(1):1–12, 2014. ISSN 2162-6057. doi: 10.1002/jocb.37.

Keelin Leahy, Shanna R. Daly, Seda McKilligan, and Colleen M. Seifert. Design fixation from initial examples: Provided versus self-generated ideas. *Journal of Mechanical Design*, 142 (101402), Apr 2020. ISSN 1050-0472. doi: 10.1115/1.4046446. URL `https://doi.org/10.1115/1.4046446`. 00000 Citation Key: leahyDesignFixationInitial2020.

T. B. Ward. Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology*, 27(1):1–40, 1994. ISSN 00100285. doi: 10.1006/cogp.1994.1010. 00000 Citation Key: wardStructuredImaginationRole1994.

G. Knoblich, S. Ohlsson, H. Haider, and D. Rhenius. Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(6):1534–1555, 1999. doi: 10.1037/0278-7393.25.6.1534. 00691 Citation Key: knoblichConstraintRelaxationChunk1999.

Richard E. Petty, Zakary L. Tormala, Chris Hawkins, and Duane T. Wegener. Motivation to think and order effects in persuasion: The moderating role of chunking. *Personality and Social Psychology Bulletin*, 27(3):332–344, Mar 2001. ISSN 0146-1672. doi: 10.1177/0146167201273007. Citation Key: pettyMotivationThinkOrder2001.

Xiaofei Wu, Mei He, Yinglu Zhou, Jing Xiao, and Jing Luo. Decomposing a Chunk into Its Elements and Reorganizing Them As a New Chunk: The Two Different Sub-processes Underlying Insightful Chunk Decomposition. *Frontiers in Psychology*, 8:2001, November 2017. ISSN 1664-1078. doi: 10.3389/fpsyg.2017.02001.

Mirko Thalmann, Alessandra S. Souza, and Klaus Oberauer. How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(1): 37–55, Jan 2019. ISSN 1939-1285, 0278-7393. doi: 10.1037/xlm0000578.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh,

Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. *arXiv:2005.14165 [cs]*, Jun 2020. URL http://arxiv.org/abs/2005.14165. 00030 arXiv: 2005.14165 Citation Key: brownLanguageModelsAre2020.

Panagiotis N. Skandalakis, Panagiotis Lainas, Odyseas Zoras, John E. Skandalakis, and Petros Mirilas. "to afford the wounded speedy assistance": Dominique jean larrey and napoleon. *World Journal of Surgery*, 30(8):1392–1399, Aug 2006. ISSN 0364-2313, 1432-2323. doi: 10.1007/s00268-005-0436-8.

Joel Chan, Katherine Fu, Christian. D. Schunn, Jonathan Cagan, Kristin L. Wood, and Kenneth Kotovsky. On the benefits and pitfalls of analogies for innovative design: Ideation performance based on analogical distance, commonness, and modality of examples. *Journal of Mechanical Design*, 133:081004, 2011. doi: 10.1115/1.4004396. 00304 Citation Key: chanBenefitsPitfallsAnalogies2011.

Kenneth D. Forbus, Dedre Gentner, and Keith Law. MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, 19(2):141–205, April 1995. doi: 10.1207/s15516709cog1902_1. URL https://doi.org/10.1207/s15516709cog1902_1.

John E. Hummel and Keith J. Holyoak. Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, 104(3):427–466, July 1997. doi: 10.1037/0033-295x.104.3.427. URL https://doi.org/10.1037/0033-295x.104.3.427.

Robert L. Goldstone and Uri Wilensky. Promoting transfer by grounding complex systems

principles. *Journal of the Learning Sciences*, 17(4):465–516, October 2008. doi: 10.1080/10508400802394898. URL https://doi.org/10.1080/10508400802394898.

Richard Catrambone and Keith J. Holyoak. Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15 (6):1147–1156, Nov 1989. ISSN 1939-1285, 0278-7393. doi: 10.1037/0278-7393.15.6.1147.

Daniel L. Schwartz. The emergence of abstract representations in dyad problem solving. *Journal of the Learning Sciences*, 4(3):321–354, July 1995. doi: 10.1207/s15327809jls0403_3. URL https://doi.org/10.1207/s15327809jls0403_3.

Robert L Goldstone and Yasuaki Sakamoto. The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, 46(4):414–466, June 2003. doi: 10.1016/s0010-0285(02)00519-4. URL https://doi.org/10.1016/s0010-0285(02)00519-4.

Sam Glucksberg. The influence of strength of drive on functional fixedness and perceptual recognition. *Journal of Experimental Psychology*, 63(1):36–41, Jan 1962. ISSN 0022-1015. doi: 10.1037/h0044683.

L. J. Ball, T. C. Ormerod, and N. J. Morley. Spontaneous analogising in engineering design: a comparative analysis of experts and novices. *Design Studies*, 25(5):495–508, 2004. ISSN 0142694X. doi: 10.1016/j.destud.2004.05.004. Citation Key: ballSpontaneousAnalogisingEngineering2004.

Xiaochen Tang, Jiaoyan Pang, Qi-Yang Nie, Markus Conci, Junlong Luo, and Jing Luo. Probing the cognitive mechanism of mental representational change during chunk decomposition: a

parametric fmri study. *Cerebral Cortex*, 26(7):2991–2999, 2015. doi: 10.1093/cercor/bhv113. Citation Key: tangProbingCognitiveMechanism2015.

M. J. Wilkenfeld and T. B. Ward. Similarity and emergence in conceptual combination. *Journal of Memory and Language*, 45(1):21–38, 2001. ISSN 0749596X. doi: 10.1006/jmla.2000.2772. Citation Key: wilkenfeldSimilarityEmergenceConceptual2001.

Jing Luo, Kazuhisa Niki, and Guenther Knoblich. Perceptual contributions to problem solving: Chunk decomposition of chinese characters. *Brain research bulletin*, 70(4–6):430–443, 2006. doi: 10.1016/j.brainresbull.2006.07.005. Citation Key: luoPerceptualContributionsProblem2006.

D. C. Richardson, M. J. Spivey, L. W. Barsalou, and K. McRae. Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, 27:767–780, 2003. doi: 10.1207/s15516709cog2705_4. Citation Key: richardsonSpatialRepresentationsActivated2003.

Pengyi Zhang and Dagobert Soergel. Cognitive mechanisms in sensemaking: A qualitative user study. *Journal of the Association for Information Science and Technology*, 71(2):158–171, 2020. ISSN 2330-1643. doi: 10.1002/asi.24221. Citation Key: zhangCognitiveMechanismsSensemaking2020.

William G. Chase and Herbert A. Simon. *The mind's eye in chess*, page 215–281. New York, NY, 1973. 02662.

Nelson Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24:87–185, 2000. doi: 10.1017/S0140525X01003922. Citation Key: cowanMagicalNumberShortterm2000.

Catherine A. Clement and Dedre Gentner. Systematicity as a selection constraint in analogical mapping. *Cognitive Science*, 15(1):89–132, January 1991. doi: 10.1207/s15516709cog1501_3. URL https://doi.org/10.1207/s15516709cog1501_3.

Dedre Gentner and José Medina. Similarity and the development of rules. *Cognition*, 65(2-3): 263–297, January 1998. doi: 10.1016/s0010-0277(98)00002-x. URL https://doi.org/10.1016/s0010-0277(98)00002-x.

C.A. Clement, R. Mawby, and D.E. Giles. The effects of manifest relational similarity on analog retrieval. *Journal of Memory and Language*, 33(3):396–420, June 1994. doi: 10.1006/jmla.1994.1019. URL https://doi.org/10.1006/jmla.1994.1019.

K. J. Kurtz and J. Loewenstein. Converging on a new role for analogy in problem solving and retrieval: When two problems are better than one. *Memory & Cognition*, 35(2):334–341, 2007. doi: 10.3758/BF03193454. Citation Key: kurtzConvergingNewRole2007.

Andrew Hargadon and Robert I. Sutton. Technology brokering and innovation in a product development firm. *Administrative Science Quarterly*, 42(4):716, Dec 1997. ISSN 00018392. doi: 10.2307/2393655.

Lars Erik Holmquist. Bootlegging: Multidisciplinary brainstorming with cut-ups. In *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008*, PDC '08, page 158–161, USA, 2008. Indiana University. ISBN 9780981856100.

Scarlett R. Herring, Chia-Chen Chang, Jesse Krantzler, and Brian P. Bailey. Getting inspired!: understanding how and why examples are used in creative design practice. In *Proceedings of the 27th international conference on Human factors in computing systems*, page 87–96.

ACM, 2009. doi: 10.1145/1518701.1518717. Citation Key: herringGettingInspiredUnderstanding2009.

D. Scott McCrickard, Shahtab Wahid, Stacy M. Branham, and Steve Harrison. *Achieving Both Creativity and Rationale: Reuse in Design with Images and Claims*, page 105–119. Human–Computer Interaction Series. Springer London, 2013. ISBN 978-1-4471-4110-5. doi: 10.1007/978-1-4471-4111-2_6. URL http://link.springer.com/chapter/10.1007/978-1-4471-4111-2_6. Citation Key: mccrickardAchievingBothCreativity2013.

Lars Bo Jeppesen and Karim R. Lakhani. Marginality and problem-solving effectiveness in broadcast search. *Organization science*, 21(5):1016–1033, 2010. doi: 10.1287/orsc.1090.0491. 01163 Citation Key: jeppesenMarginalityProblemsolvingEffectiveness2010.

Toyin Tofade, Jamie Elsner, and Stuart T. Haines. Best practice strategies for effective use of questions as a teaching tool. *American Journal of Pharmaceutical Education*, 77(7):155, Sep 2013. ISSN 0002-9459, 1553-6467. doi: 10.5688/ajpe777155.

Yuwen Li and Dongmei Li. Open-ended questions and creativity education in mathematics. 2009.

Dominik Siemon, Taras Rarog, and Susanne Robra-Bissantz. Semi-automated questions as a cognitive stimulus in idea generation. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*. IEEE, January 2016. doi: 10.1109/hicss.2016.39. URL https://doi.org/10.1109/hicss.2016.39.

Tom Hope, Joel Chan, Aniket Kittur, and Dafna Shahaf. Accelerating innovation through analogy mining. 2017. doi: 10.48550/ARXIV.1706.05585. URL https://arxiv.org/abs/1706.05585.

Q. Zhu and J. Luo. Generative pre-trained transformer for design concept generation: An exploration. *Proceedings of the Design Society*, 2:1825–1834, May 2022. ISSN 2732-527X. doi: 10.1017/pds.2022.185. Citation Key: zhuGenerativePreTrainedTransformer2022.

Taylor Webb, Keith J. Holyoak, and Hongjing Lu. Emergent analogical reasoning in large language models. (arXiv:2212.09196), Dec 2022. URL http://arxiv.org/abs/2212.09196. arXiv:2212.09196 [cs] Citation Key: webbEmergentAnalogicalReasoning2022.

Bhavya Bhavya, Jinjun Xiong, and Chengxiang Zhai. Analogy generation by prompting large language models: A case study of instructgpt. (arXiv:2210.04186), Oct 2022. URL http://arxiv.org/abs/2210.04186. arXiv:2210.04186 [cs] Citation Key: bhavyaAnalogyGenerationPrompting2022.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. (arXiv:2203.02155), Mar 2022. URL http://arxiv.org/abs/2203.02155. arXiv:2203.02155 [cs] Citation Key: ouyangTrainingLanguageModels2022.

Ian Tseng, Jarrod Moss, Jonathan Cagan, and Kenneth Kotovsky. The role of timing and analogical similarity in the stimulation of idea generation in design. *Design Studies*, 29(3):203–221, 2008. ISSN 0142694X. 00234 Citation Key: tsengRoleTimingAnalogical2008.

Joel Chan, Steven P. Dow, and Christian D. Schunn. Do the best design ideas (really) come from

conceptually distant sources of inspiration? *Design Studies*, 36:31–58, 2015. doi: 10.1016/j.destud.2014.08.001. 00105 Citation Key: chanBestDesignIdeas2015.

Eunyoung Kim and Hideyuki Horii. Analogical thinking for generation of innovative ideas: An exploratory study of influential factors. *Interdisciplinary Journal of Information, Knowledge, and Management*, 11:201–214, Jul 2016. doi: 10.28945/3539. Citation Key: kimAnalogicalThinkingGeneration2016.

David J. Adams, Siobhan Hugh-Jones, and Ed Sutherland. Raising awareness of individual creative potential in bioscientists using a web-site based approach. *Bioscience Education*, 15(1): 1–7, Jun 2010. ISSN 1479-7860. doi: 10.3108/beej.15.5.

Charles Darwin. *On the origin of species, 1859*. Routledge, 2004.

Akshay Chaturvedi, Swarnadeep Bhar, Soumadeep Saha, U. Garain, and Nicholas M. Asher. Analyzing semantic faithfulness of language models via input intervention on conversational question answering.

Eugene Raudsepp. Stimulating creative thinking. *IEEE Transactions on Professional Communication*, PC-26(4):209–211, 1983. ISSN 0361-1434. doi: 10.1109/TPC.1983.6448180.

Laria Reynolds and Kyle McDonell. Prompt programming for large language models: Beyond the few-shot paradigm. *arXiv:2102.07350 [cs]*, Feb 2021. URL `http://arxiv.org/abs/2102.07350`. arXiv: 2102.07350 Citation Key: reynoldsPromptProgrammingLarge2021.

Tony McCaffrey. Innovation relies on the obscure: A key to overcoming the classic problem

of functional fixedness. *Psychological Science*, 23(3):215–218, Mar 2012. ISSN 0956-7976, 1467-9280. doi: 10.1177/0956797611429580.

Darren W. Dahl and Page Moreau. The influence and value of analogical thinking during new product ideation. *Journal of Marketing Research*, 39(1):47–60, 2002. doi: 10.1509/jmkr.39.1.47.18930. 00714 Citation Key: dahlInfluenceValueAnalogical2002.

Michael S. Vendetti, Aaron Wu, and Keith J. Holyoak. Far-Out Thinking: Generating Solutions to Distant Analogies Promotes Relational Thinking. *Psychological Science*, 25(4):928–933, April 2014. ISSN 0956-7976, 1467-9280. doi: 10.1177/0956797613518079.