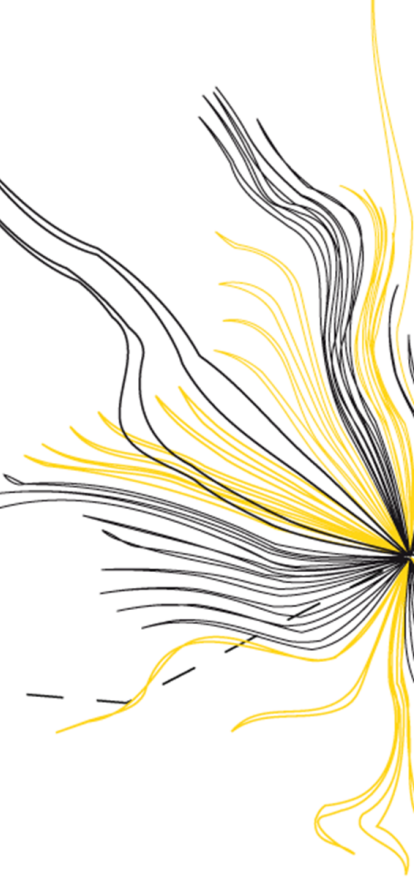


UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering,
Mathematics & Computer Science

Exploring The Relation Between Sense Of Embodiment & Task Performance in Telerobotics

Nils Rublein

MSc. Thesis Interaction Technology
October 2022

Supervisors:
dr. ir. D. Dresscher
dr. ir. G. Engelbienne
Sara Falcone, MSc.

Human Media Interaction Group
Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Summary

Telerobotics is the remote control of a robot and requires high transparency for natural and intuitive interaction. It has been proposed that increased sense of embodiment leads to higher transparency and in turn increases task performance, however this has not been explicitly investigated yet in the domain of telerobotics. This paper aimed to explore the link between the level of SoE and task performance in a user study with a telerobotic setup where in a mixed-subjects design, participants performed a peg & hole task. Two levels of embodiment (supportive & suppressive) have been created by manipulating perceptual cues such as haptic feedback, depth perception and camera perspective. To measure the level of embodiment, surveys, interviews and proprioceptive drift measurements have been administered. Task performance was measured in terms of effectiveness (inserted pegs per trial) and efficiency (time per inserted peg).

Results show that 1), no learning behaviour could be observed on the participant level for either effectiveness nor efficiency, 2) there was no relevant difference in the time needed to insert the pegs between the embodiment groups and 3), the supportive embodiment group was 16 % more likely to insert a peg successfully, and 25 to 30 % more likely to insert all pegs in one trial as compared to the suppressive group, thus being generally more task effective. To analyze the relation between task performance measures and SoE sub-components, a correlation analysis has been performed, in which only self-location correlated significantly with effectiveness.

Based on the results of the present study and other related studies in the field of embodiment and task performance, it has been shown that neither ownership, nor agency, nor self-location strictly modulate task performance. Instead, it is proposed that the contribution of embodiment to task performance is dependent on the task at hand and to what extent egocentric and allocentric perception generate motor commands and sensory predictions in the internal models of the brain that are critical to executing the task.

The present study contributes to the understanding of when and how embodiment can increase task performance in the domain of telerobotics. As such, the results of the present study may guide the design of more effective and efficient telerobotic systems, which may benefit to a plethora of complex and challenging domains, such as tele-surgery or remote rescue & search missions. Additionally, the present study investigated depth perception as a novel perceptual cue to manipulate SoS, which led to significant differences between the embodiment groups in perceived self-location.

Contents

List of Acronyms

1PP	1st Person Perspective
3PP	3rd Person Perspective
BF	Bayes Factor
CLU	Center Lower Upper
CI	Confidence Interval
CMM	Crossmodal Congruency Effect
DoF	Degrees of Freedom
GLM	Generalized Linear Model
HDI	Highest Density Interval
HMD	Head Mounted Display
IPA	Interpretative Phenomenological Analysis
LOO-CV	Leave One Out Cross Validation
MLE	Maximum Likelihood Estimation
OLR	Ordinal Logistic Regression
RHI	Rubber Hand Illusion
SCR	Skin Conductance Response
SD	Standard Difference
SoA	Sense of Agency
SoE	Sense of Embodiment
SoO	Sense of Ownership
SoS	Sense of Self-Location
TFM	Tweak Finder Model
VR	Virtual Reality

1 Introduction

Telerobotics aims at combining the physical strengths of robots and the cognitive abilities of humans over an arbitrary distance to execute tasks in remote environments that are difficult to reach or dangerous for humans to operate in [1]. A wide range of use cases can be found in areas such as rescue & search missions, surgeries, bomb disposals, social interactions or inspection & maintenance [2]. The main goal in telerobotics is to achieve full transparency, i.e. the feeling of directly being present and interacting in the remote environment, as it is expected that increased transparency also leads to an increase of control, task performance and a reduction of cognitive load [3, 4].

Ideal transparency is technically obtained when a perfect match between the position and force signals of the master and slave device exists, or when a perfect match between the environment impedance and the perceived impedance by the operator is created [5]. However, one of the reasons why ideal transparency is not achievable in real systems due to the presence of time delays [5] which have generally two main effects that are highly undesirable: First, time delays can destabilize the system and lead to a severe safety hazard; second, time delays will decrease transparency as any control command as well as any feedback (audio, visual, haptic) is being delayed and out of sync. Overall, time delays degrade the interaction quality and decrease task performance while increasing cognitive load of the operator [6]. The problem of time delays surrounding stability and transparency gave rise to a line of research that is primarily focused on technical solutions such as passivity based control [7, 8] or model mediated control [9]; however, it appears that none of the existing approaches is at a stage where they allow for ideal transparent control in complex, real environments [5]. Toet et al. [6] propose that ideal transparency may also be formulated from a psychological perspective, where the operator has the illusory experience that the robot's body is their own body and hands, making them forget that the interaction is mediated via a robot. This feeling is typically referred to as the Sense of Embodiment (SoE), where the feeling of an external body or part of it is perceived as one's own [10]. SoE may be categorized in three sub-components [10], namely the senses of ownership, agency and self-location. These senses describe the feeling of self-attribution over an external body the feeling of being able to control the external body and the feeling of being located where the external body is located, respectively.

Toet et al. furthermore suggest that the SoE is positively related with dexterous task performance as several neuroscientific studies show that the structures underlying body ownership significantly correlate with those mediating motor control [11–13]. Several studies show that a link exists between SoE and task performance in the domains of prosthetics and Virtual Reality (VR): Prostheses with peripheral neurotactile stimulation providing sensory feedback to elicit somatotopic sensations have been shown to lead to higher prosthesis embodiment [14,15], better prosthesis movement control [16] and better object discrimination and manipulation [14]. Valle et al. [17] showed that an intraneurally controlled bidirectional hand prosthesis induced a higher level of SoE and resulted in higher manual accuracy and richer tactile sensitivity. In addition, the study by Marasco et al. [18] developed a neural-machine interface for amputees to stimulate their kinesthetic sense, which lead to an increased sense of agency and improved movement control. Regarding SoE and task performance in VR, the study by Grechuta et al. [19] showed that induced ownership over a virtual limb directly increased task performance in a sensorimotor task, leading to faster reaction times. In addition, the study by Gorisse et al. [20] showed that first person perspective lead to higher ownership and self-location and more efficient manipulation of objects as compared to a third person perspective. Finally, the paper by Falcone et al. [4] investigated the relative contributions of different perceptual cues on SoE and task performance. In this factorial study, participants had to hit a target in a VR environment,

where the participants' arm and hand motions were mapped to a virtual arm and hand. In this study, each perceptual cue had two levels, supportive and suppressive SoE. Results show a lack of interaction between sensory cues, favoring the predictive encoding theory [21–23], and a significant but weak correlation between the level of SoE and task performance. The authors postulate that the correlation might have been stronger for a more a complex task, that demands a higher workload of the operator. They base this hypothesis on the assumption that a higher SoE would reduce the cognitive workload of the operator, which could in turn increase the learning speed.

Overall, these findings seem to support the hypothesis that SoE influences task performance, however, all of these studies are limited to the domains of prosthetics and VR. The present study therefore investigates the relation between embodiment and dexterous task performance in telerobotics. The goal of this study is therefore to explicitly link the level of embodiment to the level of dexterous task performance in the context of telerobotics. Moreover, based on the proposal by Falcone et al. [4], task performance will be analyzed in terms of learning behaviour. The main research questions is thus:

> To what extent does the level of Sense of Embodiment affect the level of dexterous task performance in telerobotics?

By answering the main research question, this study would contribute to our general understanding of SoE and how it could be used as a framework to design more transparent telerobotic systems, which would bring advances to a plethora of complex and challenging domains. For instance, more efficient and transparent motor control as a consequence of higher agency could lead to safer tele-surgery applications, whereas higher ownership through mulit-sensory integration could lead to higher telepresence and a more personal connection in remote, social interaction.

The remainder of this thesis is structured as follows: First, in chapter 2 the necessary literature to answer the research questions is discussed, which forms the basis for the methodology, which is presented in chapter 3. Then, in chapter 4 the resulting findings of the user study will be presented and will be discussed in chapter 5. In chapter 6 a conclusion will be formulated, summarizing relevant findings and insights, and finally in chapter 7, recommendations will be given on how the current study could be reiterated and directions for future research are proposed.

2 Background

This chapter discusses the necessary background literature to follow the remainder of the report. Specifically, the first section introduces the concept of SoE and methods for measuring and manipulating it. The second section discusses Generalized Linear Models (GMLs), frequentist and Bayesian statistics, modelling of GLMs in `R` as well as how specific data types inform about suitable distributions for analyzing the respective data type. This second section forms then the foundation for the third section, which is about modelling learning curve data in `R`.

2.1 Sense of Embodiment

This section introduces the concept of embodiment and discusses methods for measuring and manipulating it that are relevant in context of the present study. Note that this section is an excerpt of prior work [24] leading up to this study.

SoE may be described as the experience of perceiving an external body (or part thereof) as one's own [10]. Currently there is neither a standardized definition of SoE, nor a standardized framework to asses the level of SoE [2]. Two commonly used definitions for SoE are presented by Kilteni et al. [10] and Gonzalez-Franco & Peck [25, 26]. Kilteni et al. define SoE via three sub-components, namely 1) the Sense of Ownership (SoO), the feeling of self-attribution of an external body or object; 2) the Sense of Agency (SoA), the feeling of being able to control the external object and to interact with the remote environment; and 3) the Sense of Self-Location (SoS), the feeling of being located in a volume of space in the remote environment. In contrast, Gonzalez-Franco & Peck [25] divide SoE in six sub-components, namely 1) ownership, 2) agency and motor control, 3) location of the body, 4) external appearance, 5) tactile sensations and 6) response to external stimuli. Falcone et al. [27] propose to consider external appearance, tactile sensations response to external stimuli as part of SoO. While both definitions of Kilteni et al. and Gonzalez-Franco & Peck consider SoE in the context of VR, the former is more general and can also be used in the context of telerobotics, as proposed by Toet et al. [6]. For this reason, SoE is being considered in the line of Kilteni et al. for the remainder of this paper. Note however that for the present study the definition of SoS that is typically used in RHI studies (perceived location in a volume of space), is extended to also consider spatial presence (or awareness) [28] as done in [29]. This is because spatial presence is critical in telerobotics, as it enables the operator to effectively navigate and manipulate objects in the mediated environment [30]. As such, spatial awareness emphasizes the relocation into another environment and the perceived action possibilities that arise from perceiving spatial dimensions and localizing objects in said environment [29]. It has been shown that spatial presence is influenced by similar cues as SoS ratings [31], which is in line with the theoretical framework by Kilteni et al. [10].

2.1.1 Measuring Sense of Embodiment

There are many different ways to measure the level of SoE, see [2, 6, 32] for a review. In this section, the most relevant explicit and implicit measures in context of the present study are being discussed, see table 2.1 for an overview.

Measure	SoE Component
Surveys*	SoO, SoA, SoS
Interviews*	SoO, SoA, SoS
Proprioceptive drift	SoO, SoA
Physiological measures	SoO, SoA

Table 2.1: SoE measures and corresponding sub-components. Measures marked with a * sign are explicit measures whereas the other measures are implicit.

Explicit Measures

Explicit measurements of SoE include self-reports such as surveys and interviews, which can be customized as desired to measure all three SoE sub-components [27].

Surveys
The probably most influential SoE survey is the one of the original Rubber Hand Illusion (RHI) experiment of Botvinick and Cohen [33], which has been adapted by many other studies such as [34–36]. In the RHI, a rubber hand is placed in a plausible location in front of the participant, whose hand is occluded by a wall. The artificial and real hand are then being stroked synchronously with two identical paint brushes, creating a multi-sensory conflict between visual and proprioceptive information about the real hand's position [37]. The survey of the RHI experiment included questions such as "I felt as if the rubber hand were my hand" and "It seemed as if the touch I was feeling came from somewhere between my own hand and the rubber hand". Later, the RHI has been extended to include a moving hand to study the effect of visuo-proprioceptive stimulation [34] and the survey was extended to explicitly evaluate SoA. Typically, such combined surveys are based on the study by Longo et al. [38], in which a psychometric analysis (i.e., Principal Component Analysis (PCA)) was conducted to systematically investigate the structure of bodily self-consciousness and identify the latent factor structure underlying the RHI experience [39]. In this study, several themes emerged from the PCA, one of them being the theme embodiment in which SoO, SoA and SoS have been grouped together, suggesting that these three components of embodiment, while dissociable in a focused analysis, form a coherent cluster of experience [38]. While most studies focus on SoO and SoA, SoS is commonly less examined on its own [27]. Surveys that specifically address SoS include the survey of David et al. [40] as well as the survey by Debarba et al. [41], where the latter has been adapted by Gorisse et al. [20].

Interviews
Another commonly used explicit measure for SoE are interviews, which are typically combined with questionnaires in order to allow for further elaboration of participants [27]. Results of the interviews are often simply summarized [42] or incorporated as quotes in order to highlight specific observations [43, 44]. Some papers additionally use forms of emerging coding [45] to facilitate the identification and presentation of themes within the data such as thematic analysis [46] as done in the study by Ziadeh et al. [47] or Interpretative Phenomenological Analysis (IPA) [48] as done in the study by Lewis and Lloyd [49]. The goal of thematic analysis is to minimally organize and describe the data at hand in rich detail [46], which is accomplished via 1) familiarizing yourself with the data; 2) generating initial codes and collating data relevant to each code; 3) searching for themes, i.e. critically collecting and comparing all data relevant to each potential theme; 4) reviewing the identified themes, generating a thematic map of the

analysis; 5) generating clear definitions and names for each theme; and finally 6) producing the final report [46]. Similarly, themes emerge in IPA by analyzing each transcript individually, combining common observations into into larger themes, which illustrate the structure of the experience across the whole sample [49]. Here, the focus lies on the conscious experience as perceived from the participants point of view and how the participant makes sense of that experience.

Implicit Measures

The most common implicit measures include proprioceptive measures, physiological measures and changes in neural activity.

Proprioceptive drift
Proprioceptive drift is the illusory change in perceived location of the own hand towards the artificial hand [37]. Proprioceptive drift has been first induced in the RHI [33] via visuo-tactile stimulation where the real hand and the rubber hand are being stroked synchronously with a set of identical paint brushes. As the strokes are being perceived from a different position from where it is seen, the felt position of the own hand is drifting towards the position of the artificial hand [37]. Proprioceptive drift may be measured via perceptual and motor responses, which are closely related to the concepts of body image and body schema [32]. The body image is defined as a set of perceptions, attitudes, and beliefs towards one's body whereas the body schema is a set of sensory-motor capacities that control movement and posture [50]. In other words, the body image describes how an individual pictures the physical appearance of the their body in their mind, whereas the body schema is the proprioceptive perception of the posture of one's body parts in space, allowing the execution of movements [51].

Various methodologies exist for measuring proprioceptive drift in form of perceptual and motor responses [32]. In perceptual responses, the participant typically indicates the perceived location of the unseen real hand via some measuring devices such as rulers [52] or moving markers [53]. In motor responses, the participant typically executes ballistic pointing with the uncovered hand towards the covered [53, 54] or grasping movements with the covered hand towards a target [55, 56]. Proprioceptive drift induced by visuo-tactile stimulation has been used as implicit measure of ownership [52, 53] but it has also been induced by visuo-tactile stimulation (e.g., voluntary motor control over finger movements of artificial hands), which has been used as a measure for agency next to ownership [34, 57].

Physiological measures
The most commonly used physiological measures include skin conductance [58], body temperature [59] or heart rate [60]. Additionally, pupil dilation has been recently investigated as a novel, implicit measure for embodiment [61]. For higher levels of embodiment, presenting a strong emotional stimulus like a threat to the external body will cause a spike in the respective physiological measure, similar as to threatening the real body of the participant [61]. Examples of such threats include stabbing the rubber hand in the RHI with a needle [62], using a knife to stab an artificial abdomen [63] or splashing water over the artificial hand [61]. While these measures are typically easy to conduct, they are often not reliable as they may be subject to poor signal to noise ratio or may not necessarily be related to a (specific) SoE sub-component but to other external factors [2]. In addition, the devices used to measure these physiological responses may constrain the movements of the operator and can potentially bias the experiment depending on the task that is being used.

2.1.2 Manipulating Sense of Embodiment

Just as there are many ways to measure SoE, there is a plethora of ways to manipulate it, efforts to summarize these ways can be found in [6, 10, 27, 32]. Since most manipulations affect multiple SoE components simultaneously, this section is structured per manipulation rather than

per SoE component. See table 2.2 for an overview of the manipulations and the affected SoE components.

Perceptual Cue	Affected SoE Components
Visuo-tactile stimulation	SoO, SoA, SoS
Visuo-proprioceptive stimulation	SoO, SoA, SoS
Haptic feedback	SoO, SoA
Point of view	SoO, SoS
Depth Perception	SoA, SoS

Table 2.2: Perceptual cues to manipulate SoE and affected sub-components

Visuo-Tactile Stimulation

Visuo-tactile stimulation is typically applied by stroking of the hand and its surrogate (e.g. rubber hand or virtual hand) [6]. Various types of tactile stimuli may be used such as stroking [33], tapping [64] or even painful stimuli like poking with a sharp pin [65]. Factors affecting visual-tactile stimulation include for instance the speed of stroking, the tactile congruency of the stroking tool and the temporal (a)synchrony of the applied tactile stimuli [32]. Lower velocities of stroking are typically rated as more pleasant [66] and induces higher subjective ratings of SoO [67–69] and proprioceptive drift [69] (but see also [68]) as compared to higher velocities. Congruency between the felt touch and the visual features of the stroking tool have been reported to affect ownership [67, 70], for instance, feeling the touch of a toothbrush while seeing the fake hand being touched with a pencil reduces SoO as compared to a congruent condition (however, also see [71]). The effect of asynchronicity has been studied in the domains of the RHI paradigm [11, 33, 38], VR [72] and robotics [73, 74], for all three domains the common observation is that asynchronous stimulation decreases SoO [4]. Here, asynchronicity is typically applied by inserting a delay between the felt and seen touch, which can easily break the embodiment illusion [27]. While various studies have shown that visuo-tactile stimulation can significantly enhance SoO [11, 75, 76], it is not a hard requirement and embodiment can occur even in case of asynchronous visuo-tactile stimulation or in absence of it if other cues are available such as synchronous visuo-proprioceptive cues [35, 74, 77, 78].

Visuo-Proprioceptive Stimulation

Visuo-proprioceptive stimulation is the congruence between efferent (operator's intended motion) and visual afferent (visual perception of actual motion) information, in other words, the match between the commanded movements by the operator and the perceived, resulting movement of the surrogate [79, 80]. Visuo-proprioceptive stimulation has been shown to elicit SoO [79–83], SoA [82, 83] as well as proprioceptive drift [74, 82, 83]. Visuo-proprioceptive stimulation has been shown to be a very effective manipulation of embodiment [60, 84, 85] as the congruency between the operator's intended movement and the visual perception of the actual movement alone is sufficient to create the illusion of ownership [79, 80]. For instance, embodiment may occur in absence of visual-tactile stimuli if sufficient visuo-proprioceptive stimuli are provided [57, 77, 86]. Embodiment that is induced via visuo-proprioceptive stimulation may be affected by human likeness (i.e. realism) of the surrogate as well as the spatial discrepancy between the commanded and resulting position & orientation [6]. For instance, while in [36] an effect of visuo-proprioceptive stimulation with a robotic hand was found for proprioceptive drift but not on SoO; the studies by [82, 83], involving virtual hands with a (presumably) higher degree of realism, found an effect on SoO as well as proprioceptive drift. Next, it has been shown that SoO induced by visuo-proprioceptive stimulation is resistant to smaller errors in position [11, 87, 88] and orientation [89–91], but deteriorates for larger errors [92,]. Asynchronous stimulation generally reduces embodiment as compared to synchronous stim-

ulation, but may increase embodiment as compared to no stimulation at all [74]. Franck et al. [94] showed that temporal mismatches below 150 ms cannot be reliably detected, suggesting that small asynchronies are still perceived as synchronous.

Haptic Feedback

Haptic feedback is typically decomposed into kinesthetic and tactile information where kinesthetic information describes the posture and perception of body parts in space, along with the associated forces; and tactile information characterizes the nature of contact between the skin and a surface, allowing to feel texture, heat, and pressure [95]. This dichotomy allows for different types of stimulation in VR and telerobotics, which are mainly force feedback and vibrotactile feedback [95]. Key factors to consider for haptic feedback include the type of feedback (force vs tactile vs none), time delay and predictability [96].

Several studies have compared the effect of different haptic feedback types on embodiment. For instance, Krogmeier et al. [97] used a vibrotactile vest to simulate collision in VR between the participant's avatar with other virtual agents that walked into them. Results showed a moderate, positive correlation between embodiment and vibrotactile feedback. While Krogmeier et al. only compared no feedback versus vibrotactile feedback, Fröhner et al. [98] compared force feedback, vibrotactile feedback and no haptic feedback in a VR environment and investigated amongst other measures the impact of haptic feedback type on embodiment. Results show superiority of force feedback and vibrotactile feedback over no haptic feedback which was mainly reflected in SoS ratings, however they did not find any significant differences between force feedback and vibrotactile feedback. In the study of Kreimeier et al. [99] participants performed throwing, stacking and object identification tasks in VR, using force feedback, vibrotactile feedback and no haptic feedback. Results show that both force and vibrotactile feedback outperformed no feedback in terms of presence (a similar construct to SoE [29]). Force feedback led to significantly lower execution times for the throwing and the stacking task, whereas for the object identification task vibrotactile feedback increased the detection rates compared to no feedback and force feedback, but also increased the required time of identification. Note that despite hardware issues the haptic devices lead to higher presence scores as compared to no haptic feedback; however differences in task performance between force feedback and vibrotactile feedback may be biased by said hardware issues. Another study that compared the three different haptic feedback types and their impact on embodiment is the study by Richard et al. [95] which compared embodiment ratings over a humanoid avatar during a coloring task. Their results show a significant increase in self-reported embodiment when using force feedback as compared to no feedback and a significant increase in perceived performance when using force feedback as compared to vibrotactile feedback and no feedback. In contrast to the findings of the aforementioned studies [97–99], vibrotactile feedback did not significantly increase embodiment nor ownership. The authors postulate that this effect might be caused by the unpleasantness of the vibrotactile feedback, as the vibrations were perceived as very stressful by some participants. Overall, the general finding regarding the type of feedback seems to be that haptic feedback significantly improves embodiment as compared to no haptic feedback. However, differences between vibrotactile and force feedback seem often to be dependent on the used hardware, which makes it hard to generalize findings.

Another factor to consider is delay and its effect on task performance. Various studies have shown that delay in visual feedback increases the task completion time and error rates [100–103]. In the study of Mackenzie et al. [100] a linear relationship between the delay and task completion time was found. Jay et al. [104] showed that users become aware of latencies at 50 ms, where task performance is affected at around 100-150 ms, which reaches a plateau beyond these values. These observations are supported by the study of by Sengül et al. [105], which investigated the effect of delay on task performance as well as on embodiment in a sim-

ulated surgical-like task. Results show that increased delay led to decreased telepresence and task performance. Specifically, delays between 20 ms and 120 ms significantly impaired the Crossmodal Congruency Effect (CCE), which was used as a proxy to measure embodiment. Moreover, when comparing delayed haptic feedback and delayed visual feedback, no significant difference could be found in the effect of delay on the CCE between the two types of delays, however, task completion time was more affected by visual delays than haptic delays. That said, the force feedback was according to the authors not critical to task, as it could be solved by vision alone, therefore no preemptive conclusions should be drawn here. Overall, these studies highlight the need to consider the effects of delay on task performance when manipulating embodiment through visuo-proprioceptive stimulation in form of delay.

Another key factor to consider is the predictability of the haptic device and its impact on SoA. In [106], Fröhner et al. evaluate the impact of delay as well as predictability of a haptic manipulator on embodiment. Additionally, to assess task performance, they measured jerk, velocity and compensation forces between the participant and the haptic device. Participants performed a reaching task guided by force feedback of the manipulator in a VR environment. Results show good persistence of subjective embodiment under a predictable assistive control while unpredictable assistance deteriorated subjective embodiment. Using the synchronous visual feedback, the predictable control lead to significantly higher scores as compared to the asynchronous visual feedback with unpredictable control. Specifically, SoA was significantly decreased with a delayed visual feedback, which is consistent with the findings of Caspar et al. [107], where agency was significantly decreased for a robotic hand that could be actively controlled but was exposed to delays. Moreover, predictability significantly affected task performance measures jerk and velocity, as higher predictability was reflected in lower jerk and higher velocity, indicating higher performance. This implies that higher controllability due to predictable behavior facilitates the interaction of human and machine, which is supported by the observation of lower jerk levels [96]. Moreover, they report a positive correlation between presence and task performance, which is line with Witmer and Singer [108]. Similarly, Endo et al. [109] found that the assisting haptic feedback force is perceived as collaborative if it matches the intentions of the user, which is reflected in higher SoA ratings and reduced forces demanded from the user [96].

Point of View

The illusion of ownership can be created both from the point of view of a third person perspective (3PP) [110–112] and a first person perspective (1PP) [60,63,77] but is typically stronger from a 1PP [6,60,63,113] as it is closer to the natural, visual perspective that we humans are used to. Moreover, it has been shown that a 1PP is sufficient in inducing SoE without any further perceptual cues if 1), the avatar has a sufficient level of human likeliness [77,114] and 2), if there is a sufficient match between the position and orientation of the avatar and the human [115]. Furthermore, under these conditions embodiment induced via a 1PP may even persist in case of asynchronous visuo-tactile stimulation [60,77,116], demonstrating the strength of this perceptual cue. In contrast, a 3PP requires additional reinforcement for an embodiment illusion either synchronous visuo-tactile feedback [110,117,118] or spatial overlap between the human and the avatar [88]. Moreover, it has been shown that a 3PP may break the embodiment illusion [60, 63, 119]. In [20] it has been reported that a 1PP and a 3PP create the same level of SoA and SoS, however in [41] it has been shown that a 1PP yields a stronger embodiment in terms of SoS. Moreover, while SoO and SoS are typically highly linked with each other, a 3PP may disassociate these senses [88]. Overall, it appears that a 1PP is a powerful tool to create an embodiment illusion even in absence of other perceptual cues or incongruent visuo-tactile feedback. A 3PP on the other hand creates a weaker embodiment illusion and requires additional perceptual cues to elicit embodiment.

Depth Perception

Depth perception is the ability to identify information in images and correlate it with depth in a scene, using both psychological and physiological cues which include amongst others, monocular and binocular cues [30].

Monocular cues are depth cues that can be estimated with one eye, which include pictorial cues and movement-based cues [30]. Pictorial cues can be estimated from a stationary observer and include occlusion, relative height, relative size, perspective convergence, familiar size, atmospheric perspective, texture gradient, and shadows. Movement-based cues are evoked when we move in the scene and include cues such as motion parallax and deletion & accretion. Motion parallax is an effect where nearby objects seem to move faster than more distant objects, which serves as an important depth cue. Accretion and deletion result from sideways movements of the observer, where some parts of objects become revealed and some occluded. Here, the rate of deletion and accretion also serves as a depth cue.

Binocular cues are depth cues that can be estimated with both eyes, such as stereopsis which is a result from binocular disparity [30]. As our eyes are offset by some distance, they receive a different view of the objects in the scene. While there is a great overlap between these views, the viewpoints are different. This difference is called binocular disparity, which allows us to localize a point in three-dimensional space via triangulation where the resulting feeling of depth is called stereopsis [30].

While depth perception has not been explicitly explored yet in the framework of SoE [120], it is expected that manipulating depth cues influences SoS. For instance, it has been shown that stereoscopic depth cues significantly increase spatial presence [121, 122] while impairing stereopsis leads to reduced spatial presence [123]. Moreover, it is expected that improved depth cues not only increase SoS but in turn also SoA as they provide important sensory information that enables the central nervous system to engage in better predictive motor control [124]. Specifically, binocular depth cues such as stereopsis provide reliable distance cues in the peripersonal space [125, 126], which helps thus to calibrate the arm proprioception, enabling better hand-eye coordination [30]. This is supported by various studies that have shown that decreasing stereopsis leads to reduced performance in grasping tasks [124, 126, 127]. For instance, in [128] it has been shown that monocular viewing disrupted the temporal coordination between the eyes and the hand in a placement task, where the gaze shift was delayed during monocular viewing as compared to binocular viewing. The authors postulate that the shift in gaze behaviour may be caused by the increased uncertainty associated with the performance of the placement task because of increased vergence error during monocular viewing. This hypothesis has been proven in by [124] which showed that stereopsis lead to better predictive control of grasping. In summary, increasing depth perception leads increased spatial awareness, improves calibration of proprioception, which enables better hand-eye coordination and improves predictability and hence agency over the avatar.

2.2 Generalized Linear Models

In this section, we discuss mathematical descriptions of Generalized Linear Models (GLMs) as well as their implementation in `R` syntax, how specific data types inform about suitable distributions for analyzing them and conclude with a discussion on frequentist and Bayesian approaches to statistical analysis. This section serves as foundation for section 2.3, modelling of learning effects.

2.2.1 Basics

GLMs offer great flexibility to model dependencies between predictors and outcome variables for between, within as well mixed (between-within) subjects experiment designs. This section

introduces some basic concepts of GLMs such as intercepts, conditional effects, interaction effects and random effects.

As an example, imagine you are conducting a study in which you want to find out how different keyboard devices (e.g. keyboard, tablet, smart phone) affect the typing speed (words per minute) of users. If we would pool together the data of all users for the dependent variable words per minute `WPM`, we would create a so called grand mean model (or population mean model) which is described by equation 2.1:

$$\begin{aligned} u_i &= \beta_0, \\ \epsilon_i &\sim \text{Gaus}(0, \sigma_\epsilon) \\ y_i &= u_i + \epsilon_i \end{aligned} \tag{2.1}$$

where u_i is the predicted value of a variable for observation i (e.g. `WPM` for participant i) and β_0 is the reference level, the so called intercept which corresponds in this case simply to the population mean. All unexplained effects that induce variation into the data are gathered in error ϵ_i, which are in this case assumed to follow a Gaussian distribution with a mean of 0 and a standard error of σ_ϵ. The final predicted value is described by the u_i and ϵ_i. Note that this way of writing a linear model only works for Gaussian linear models, as only here, the residuals are symmetric and add up to 0 [129]. In section 2.2.3 different error distributions will be introduced to suit data types with different expected distribution shapes.

If we want to add conditional effects for the factor `Device`, we need to add an additional coefficient for each level of this factor. In other words, we need to add an a coefficient for each of the three keyboard devices to equation 2.1 as shown in equation 2.2:

$$u_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}, \tag{2.2}$$

where β_0 is the intercept which serves as reference level (e.g. tablet) and $\beta_1 x_{1i}$ and $\beta_2 x_{2i}$ correspond to the other two keyboard devices. Note how the intercept does not have the x term. This term is a dummy variable that is set for categorical variables to either 1 or 0, which is used to enable or disable certain levels of a factor. For instance, if the tablet is our intercept and we only want to estimate the `WPM` of this device, then effects of the phone and the keyboard would be disabled by setting x_{1i} and x_{2i} to zero. In contrast, if want to estimate the effect of the phone, β_1, on `WPM`, then only x_{2i} is set to zero. Notice how this effect is added to our reference level, the intercept. Based on the sign of β_1, the `WPM` of the phone will be either more or less than the reference level.

We can also add further parameters to this model, for instance, if we suspect that younger users type faster as compared to older users, we could add a second predictor, `Age`, to the model like this:

$$u_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}, \tag{2.3}$$

where $\beta_3 x_i$ corresponds to the parameter for `Age`. Note that 1), the intercept now also includes a reference level for `Age` next to `Device` and 2), x_{3i} is different from the other (categorical) dummy variables as `Age` is a continuous variable that ranges from 0 to the maximum observed value of `Age`. In general, for a number n of coefficients, equation 2.3 can be modelled as

$$u_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_n x_{ni}, \tag{2.4}$$

If we think that older users are more efficient with certain interfaces than younger users, we can add an interaction effect between the variables `Age` and `Interface`, that is, we expect that `Age` does not equally contribute to `WPM` for every `Interface`, but differs in its effect dependent on the level of `Interface`. In other words, we might expect that younger users

perform better with the phone whereas older users might perform better with the keyboard. Such interaction effects are modelled as follows:

$$u_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i}, \tag{2.5}$$

where $\beta_4 x_{4i}$ and $\beta_5 x_{5i}$ represent the added interaction effects between the keyboard devices β_1 and β_2 and `Age`, β_3, respectively. Thus, when adding interaction effects between two factors, we are essentially adding an additional coefficient for each level of the respective factors.

Note that this function only includes fixed effects, which means that this model assumes the same effect of `Interface` and `Age` for every participant. To include individual differences between participants into our model, we can add a random effect as follows:

$$\begin{aligned} u_i &= \beta_0 + \beta_{0j} + \beta_1 x_{1i} + \beta_{1j} x_{1i} + ... + \beta_n x_{ni} + \beta_{nj} x_{ni} \\ \beta_{0j} &\sim \text{Gaus}(0, \sigma_{0j}) \\ &\vdots \\ \beta_{nj} &\sim \text{Gaus}(0, \sigma_{nj}) \\ \epsilon_i &\sim \text{Gaus}(0, \sigma_\epsilon) \end{aligned} \tag{2.6}$$

where every participant j gets their own set of coefficients drawn from a Gaussian distribution with their own standard deviation that is added to the overall predicted value. For readability, u_i may also written as

$$u_i = \beta_0 + \beta_{0j} + x_{1i}(\beta_1 + \beta_{1j}) + ... + x_{ni}(\beta_n + \beta_{nj}) \tag{2.7}$$

Adding a random effect for participants means that participants have their own baseline values, which are spread in the shape of some distribution (e.g. Gaussian) around the intercept. This highlights that random coefficients are additive correction terms to the population-level effect [129]. Moreover, when a model contains both fixed and random effects, the influence of the fixed effects levels, random effects levels and the variance of their distributions are predicted at the same time [129]. This leads to the following benefits: 1), when unequal observations per participant exist, small groups are corrected towards the respective population mean and 2), strong outliers are corrected towards the respective population mean.

2.2.2 Modelling GLMs in R

This section illustrates how the mathematical descriptions introduced in the previous section are written in `R` syntax. Sticking with the keyboard device example, the grand mean model is written as:

```
WPM ~ 1,
```
(2.8)

where the dependent variable `WPM` is on the left side of the ~ symbol and the predictor variable is on the right side. Here, 1 stands for the intercept which in case of the grand mean model simply describes the population mean. Note that it is common to leave out the intercept for non grand mean models, as the intercept is always implied to be present in the model. Next, conditional effects are simply added by including the name of the variable into the equation. Equation 2.3 would be thus written as follows:

```
WPM ~ Interface + Age
```
(2.9)

Next, the interaction effect between `Interface` and `Age` of equation 2.5 is written as:

```
WPM ~ Interface + Age + Interface:Age
```
(2.10)

where `Interface:Age` is the added effect of the interaction. A common shorthand notation for interaction effects is the $\star$ symbol, which summarizes the previous equation to:

$$\texttt{WPM ~ Interface*Age} \tag{2.11}$$

Finally, the random effects of equation 2.6 are written as:

$$\texttt{WPM ~ Interface*Age + (1|Participant)} \tag{2.12}$$

where the term `(1|Participant)` summarises participant specific coefficients.

2.2.3 Data Types and Family Distributions

GLMs utilize families of models, in which every member uses a specific link function to establish linearity and a particular distribution with an appropriate shape and mean-variance relationship [129]. This is especially relevant for design research, as most measures have lower and/or upper boundaries and strictly cannot have linear associations.

Typical data types used in design research include discrete, continuous and ordinal data. To give an example for discrete data, one could consider a classic peg and hole task where a participant needs to insert 10 pegs in a trial, rendering a lower and upper bound of [0,10] for the number of successfully inserted pegs. Similarly for continuous data, one could consider the time needed to insert a peg which has a positive lower bound ($t > 0$), as it is impossible to execute the instantaneously ($t = 0$). Lastly, ordinal data is double bounded by the lowest and highest category. For instance, a seven point Likert scale would have 1 = "strongly disagree" as lowest bound and 7 = "strongly agree" as upper bound.

Additionally one should consider that distributions that have lower and/or upper boundaries are often asymmetric and have mean and variance linked in a particular manner. Most lower-bound distributions (e.g., Poisson, exponential, and Gamma) increase in variance as the mean increases, whereas double-bounded distributions (beta and binomial) generally reach their maximum variance at the location where the distribution is centered and symmetric. This indicates that the researcher's choice of the distribution family determines the shape of randomness as well as the relationship between location and variance [129]. Thus, based on the nature of the measured data, we can select the best fitting family distribution where major criteria are the type (discrete vs continuous) and the bounds of the data. To give an overview, a decision tree of the data types and suitable family distributions is given in figure 2.1:

Continuous Data

Continuous data is numerical data that can take any value, such as height, weight, temperature or time. Typical examples in user research include reaction time, or the total time spend on a task [129]. A variety of upper and lower bounds can apply to continuous data [129]. In case of no bounds, a Gaussian distribution can be used to describe the data. In case of a lower bound of zero, e.g. absolute position difference, a Gamma distribution can be used. When the data has a positive lower bound, exponentially modified Gaussian distribution can be used. Typically. duration based measures such as reaction time and time on task have a positive lower bound as it is impossible to complete a task in zero seconds and thus requires a minimum time to complete the task. When the data has a lower and upper bound, a Beta distribution can be used.

Discrete Data

Discrete data is countable data that cannot be broken into smaller parts, such as number of errors or number of success in a trial. Discrete has an inherent lower bound which can be zero or higher. When there is no well defined upper bound, or it is very large, (e.g. number of visitors

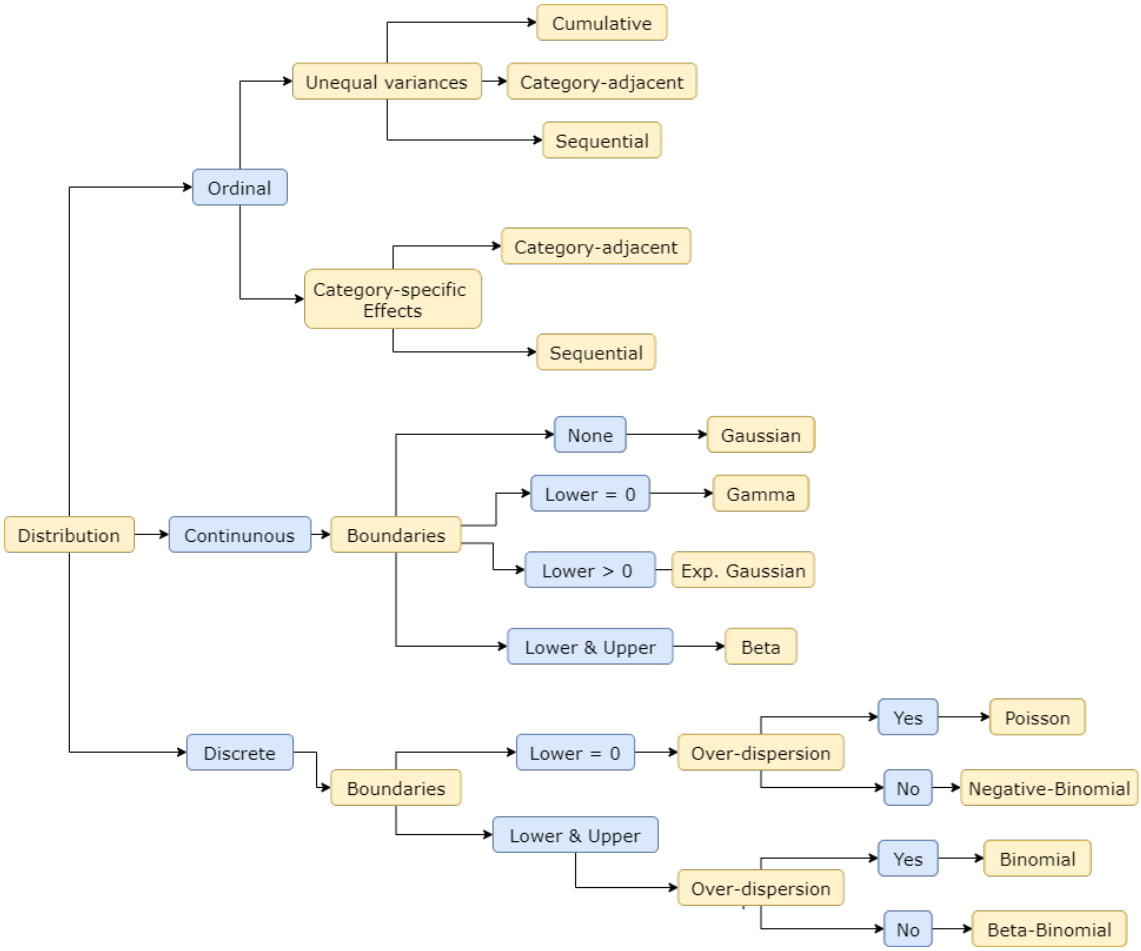


Figure 2.1: Decision tree for GLM family distributions for continuous, discrete and ordinal data. Adapted from [130] and [129].

on a website), a Poisson distribution can be used to describe the shape of the data [129]. Note that the Poisson distribution expects the data to have a mean equal to its variance. However, often in user research the variance is greater than the mean, which leads to an over-dispersion of the model (the variance is greater than expected). In such cases, a negative Binomial distribution can be used. When there is a well defined upper bound (e.g. determined by the research design), then a Binomial distribution can be used. In case of over-dispersion, a Beta-Binomial distribution can be used.

Ordinal Data

Ordinal data is typically described as data whose categories exhibit a natural order [130, 131]. For instance, Likert scales are considered ordinal as the categories of the Likert items have a clear ranking. Despite the fact that ordinal data is generally not considered metric, it is still often analyzed via approaches that expect metric data. This common practice however may lead to severe misinterpretation of the data, as statistical methods that assume metric responses (e.g. t-tests and analysis of variance), may give rise to a decrease in correct detection rates, erroneous effect-size estimates, increased type I error rates, and even inversions of differences between groups [130, 132]. According to [132], this is caused by the following three reasons: 1), Metric responses need to be equidistant in statistical models, however this assumption may not be met for the response categories of ordinal variables, due to the subjective interpretation of the respondent. For instance, the difference between “strongly disagree” and “moderately disagree” may be much smaller for someone than the difference between “moderately disagree”

and “strongly agree”; 2), ordinal responses may not follow a Gaussian distribution, especially if values at the extremes of the spectrum are frequently chosen and 3), unobserved variables that are rooted in the observed ordinal variable may differ in variance between groups, condition, etc.. When treated as metric variables, these unequal variances may not be detected in the first place.

Fortunately, recent advances in statistical methods and computation have given rise to appropriate models for ordinal variables, such as Ordinal Logistic Regression (OLR) [130]. In `BRMS`, OLR for Likert scales may be implemented via the cumulative or category-adjacent models, see [130] for details. Both models assume the predictor to have the same effect on all response categories, which may not be true in all cases. In case of category-specific effects, certain predictors may influence responses that are higher on the rating scale differently than responses that are lower on the rating scale. For instance, the predictor "embodiment condition" may have little influence over the responses "strongly disagree" over "moderately disagree" but may strongly predict whether participants choose “moderately agree” over “strongly agree” for a specific Likert item. Instead of modeling just one coefficient for the predictor, K coefficients, one for each category, can be estimated to model category-specific effects. While this can be done easily with category-adjacent models, it may lead to modelling problems for cumulative models, caused by negative probabilities [130].

Next to category specific effects, unequal variances in the latent variables should also be accounted for, as these can lead to problems such as inflated error rates and distorted effect sizes [132]. For example, between different embodiment groups, there might be unequal variances in answering the question "The hand felt like my own hand". Unequal variances can be modelled by both cumulative and category specific models, by varying the variance as a function of a desired predictor, provided that the baseline variance is fixed to some value.

Note that entire Likert scales (multiple Likert items) can be estimated simultaneously, by adding a random factor for each Likert item to the model, as shown in equation 3.4:

```
Data ~ Condition*Trial + (1|Participant) + (1|item)
```
(2.13)

This way, we can control via `(1|Participant)` dependencies between ratings from the same person while at the same time controlling for dependencies between the ratings of the same item via `(1|Item)`, by varying the thresholds by the same amount for each person and item respectively. Additionally, if one would want the categories to vary differently across items, the term `(1|item)` can be changed to `(cs(1)|item)`. This way each item receives its own set of thresholds, accounting for differences between Likert items. Overall, including Likert in such a way avoids the typical averaging or summing of Likert items, which treats them as metric variables, and inflates the number of parameters in ordinal logistic regression. For instance, instead of using one seven-binned Likert item, summing up three of such items would add up to 21 bins. Using multi-level modelling by including Likert items as random factor circumvents this problem in elegant way [129].

Finally, one may ask the question how do we interpret the results of a OLR model? Briefly, to capture the ordinal nature of the data, OLR models assume that there are K intercepts, which are called thresholds. These thresholds partition the latent (not observable) continuous variable $\tilde{Y}$ into $K+1$ ordered categories. For instance, a 7 point Likert item would have $K+1=7$ categories and $K=6$ thresholds. We then assume that that $\tilde{Y}$ follows some distribution (e.g. Gaussian) with cumulative distribution function F, applying the probit link function to each threshold then obtains the cumulative probability of responses below that threshold if all predictor variables were zero, see [130] for more details. While it is important to understand the meaning of these thresholds, they are rarely the central focus for the actual analysis, instead, the regression coefficients for the respective of the levels of the predictor that we wish to estimate are of relevance. Differences between these regression coefficients may be interpreted

analogous to Cohen's d [133], indicating to what extent participants of one group are giving higher or lower scores, as compared to the reference group.

2.2.4 Frequentist vs Bayesian Statistics

GLMs are available in both the frequentist and Bayesian statistical framework, however they differ in the way they interpret probability as well as the answers they provide to hypothesis testing and parameter estimation. This section first describes both frameworks in terms of their concepts of probability, hypothesis testing and parameter estimation; and argues then why the Bayesian framework has been used in the present study instead of the frequentist framework.

The Frequentist View on Probability

In frequentist statistics, probability is viewed as relative frequency of events occurring in a long running series of experiments, such that in the long term the range of the true mean of a statistic may be inferred [134]. In frequentist null hypothesis significance testing (NHST), a p value is calculated where the null hypothesis is typically rejected if $p < 0.05$. Here, the p value is obtained by repeatedly sampling descriptive test summaries (e.g. the t statistic) of simulated data and comparing it to the relative frequency of the actual data, the likelihood. Note that the only information provided in the frequentist hypothesis test is the p value, which allows the researcher to decide whether or not the null hypothesis should be rejected [135]. Notice that from the p value itself it does not follow that the null hypothesis can be accepted (e.g. when $p > 0.05$) or that the alternative hypothesis is true (e.g. when $p < 0.05$). Thus, the frequentist hypothesis test cannot favor a hypothesis, it can only reject or fail to reject a hypothesis. Furthermore, the p value alone does not indicate how large the estimated magnitude of the parameters or the certainty thereof is. The frequentist confidence interval does merely describe the range of parameters which would not be rejected by $p < 0.05$. Note that this confidence interval does not carry any distributional information, that is, it does not answer how probable parameters are, in for instance the middle of the confidence interval as compared to the boundaries of the confidence interval [135].

The Bayesian View on Probability

In contrast, Bayesian statistics views probability as re-allocation of certainty across possibilities [135], i.e. the parameter values of the intercept and the regression coefficients of a GLM. Rather than simply relying on the likelihood of the data as in frequentist statistics, prior knowledge is used to update the beliefs about the data [136]. Specifically, the so called prior distribution, which describes one's beliefs over the distribution of the data before observing it, is used to express the (un)certainty about the regression parameters. This prior distribution is then multiplied with the likelihood of the observed data to generate the posterior distribution, which describes the probability of parameter values, given the observed data and the prior. This is illustrated visually in figure 2.2 and mathematically described by Bayes theorem in equation 2.14:

$$P(\theta|Y) = \frac{P(Y|\theta) \cdot P(\theta)}{P(Y)} \tag{2.14}$$

Here, the (un)certainty about the estimates of the GLM parameters is manifested in the spread of the posterior distribution, where a wider spread would indicate higher uncertainty and a smaller spread would indicate higher certainty [135]. To summarize this (un)certainty, the 95 % Highest Density Interval (HDI) can be used, which consists of the parameter with the highest probability (i.e., density), and spreads over the 95 % most probable values. Thus, whereas the frequentist confidence interval 1), describes only the range of parameters that would not be rejected by $p < 0.05$ and 2), does not allow inference about the probability for any of those

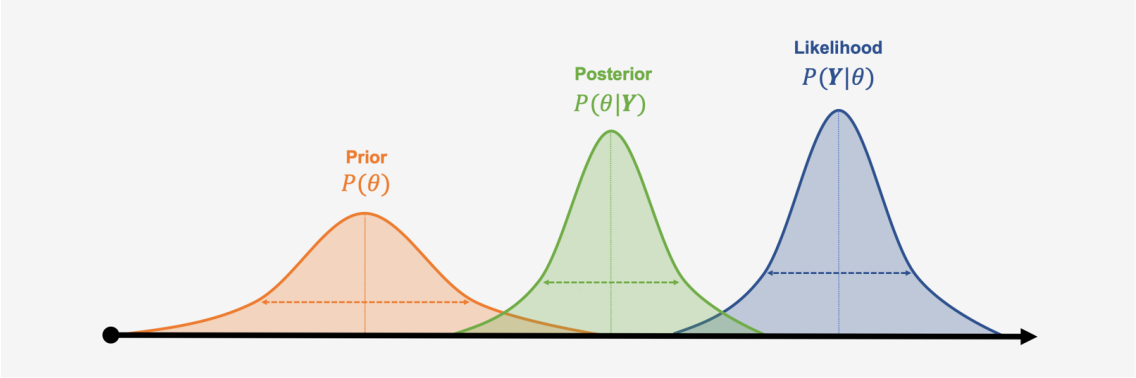


Figure 2.2: Prior, posterior and likelihood distribution [137].

parameters, the Bayesian HDI does allow to calculate the full distribution of parameters and their probabilities.

There are mainly three kinds of prior distributions, namely 1) non-informative priors, 2) weakly-informative priors and 3) informative priors [136]. Non-informative priors (e.g. flat prior distributions) reflect a great amount of uncertainty of the probability of the estimated parameter values. Weakly-informative priors reflect some information about the nature of the data under consideration and express more certainty than compared to non-informative priors, however they typically have no large impact on the final posterior distribution. Finally, the informative prior incorporates the most amount of certainty into the model and has also the most influence on the posterior distribution. Informative prior knowledge may exist in the form of data of earlier experiments or for (supposedly) known models of some phenomena.

The Bayesian counterpart to the frequentist NHST may be formulated either in terms of parameter estimation or model comparison. In the former approach, a supposedly informed prior distribution is taken, and the posterior distribution given the prior and the likelihood of the data is estimated. Then, for a one-sided hypothesis in form of $a > b$, one can sample from the posterior distribution and simply calculate the probability of this hypothesis. Additionally, one can calculate the evidence ratio for this hypothesis and its alternative, which would be the ratio of the posterior probabilities $a > b$ and $a < b$, where values greater than one would indicate that the evidence is in favor for the hypothesis in the numerator and vice versa for the hypothesis in the denominator. However, one may also approach Bayesian hypothesis testing in terms of model comparison. Here, the focus is not on estimating the magnitude of the parameter but on comparing two prior distributions and estimating which one is the most credible. In this case, one prior is formulated such that the parameter value is the null value whereas the alternative prior is such formulated that the parameter could be any value (depending on the chosen distribution). Using these two priors and the likelihood of the observed data, the ratio of the two resulting posterior distributions is calculated, which is called the Bayes Factor (BF) [138]. This ratio quantifies the relative evidence for either hypothesis by indicating how much more likely the observed data are under the respective rival model [139]. Thus, whereas the parameter estimation approach considers the posterior probability of a parameter given a chosen prior and the data, the model comparison estimates the relative probability for priors describing the null hypothesis and the alternative hypothesis, given the data. Note that neither of these approaches is uniquely 'correct', as they test the null value in different ways [140].

Justification for the Bayesian framework in the present study

The Bayesian framework has been chosen in the present study for the following reasons:

First, it provides overall a richer amount of information as compared to the frequentist framework. For estimation, the frequentist framework only provides an interval of parameters that

would not be rejected whereas the Bayesian framework provides via the posterior probability distributions for all parameters within the GLM, from which differences, ratios, effect sizes and predictions can be directly computed [135, 141]. In addition, for hypothesis testing the frequentist framework can only reject or fail to reject a hypothesis, but not prove the truthfulness of a hypothesis. In contrast, the Bayesian framework allows to compare the hypotheses relatively to each other (for parameter estimation as well as model comparison), thus allowing it to quantify support in favour of the null hypothesis and not just against it [135, 141].

Second, frequentist regression engines for complex models of realistic data are often limited by hill climbing algorithms that may fail to converge, as well as large-N approximations to sampling distributions that result in extremely optimistic p values and confidence intervals [135]. Specifically for multi-level models (models incorporating fixed and random effects), conventional statistical methods often fail as numerical integration to compute the maximum likelihood estimation (MLE) is intractable due to the high dimensionality [136]. This leads to an overall limited amount of model types and data distributions for frequentist analysis [135]. In contrast, Bayesian techniques are generally more robust, do not rely on large-N approximations [141] and a wider range of data distributions is available, especially for multi-level models [129]. Moreover, frequentist GLMs allow to only model linear predictors, but not non-linear models such as learning curve data. Bayesian GLMs on the other hand do allow to model such non-linear model by either 1), fully formulating a non-linear predictor term with the relevant parameters, each of which can be estimated via GLMs or 2), predicting the form of the non-linear relationship using splines or Gaussian processes [142].

Third, the posterior includes distributions and credibility interval for all parameters within the model, allowing researchers to make inferences about individual parameters, such as investigating differences in variance between groups [129]. This is interesting specifically for telerobotics, as it is a domain that uses typically very expensive equipment to carry out sensitive and/or dangerous tasks; thus the least amount of variance in task performance is desired.

2.3 Modelling Learning Effects

In this section we will first briefly discuss how learning behaviour is typically being modelled. Then, as learning curve data is typically too complex for traditional statistical approaches as it is both longitudinal and non-linear at the same time, we will discuss how GLMs within the Bayesian framework of statistical thinking can be used to elegantly estimate and predict such complex data.

2.3.1 Modelling Learning Behaviour

General skill acquisition may be described in three stages, namely the cognitive, associative, and autonomous stage [143]. The cognitive stage is also known as the initial stage of learning. Here, the learner is trying to find out "what to do", by cognitively understanding the requirements and parameters of a specific movement. This phase is characterized by slow and error prone performance, while being high in demand of mental effort. In the associative stage, the learner shifts the focus from "what to do" to how "how to do", that is, they are trying to adapt and improve the newly learned skill. In this phase, the number of errors is reduced and patterns of performance start to materialize. However, for tasks requiring high mental load, performance can still significantly decrease. Finally, in the autonomous stage, the learner does not require any attention for executing the task and could execute other tasks simultaneously. The performance in this learning phase is stable and efficient and characterized by low mental load. Via a set of internal cognitive, perceptual, motor, and perceptual-motor processes that are associated to practice, a learner gradually progresses through the three stages and obtains skillful performance [144]. Anderson [145] separates the cognitive stage in three further segments, namely the declarative, knowledge compilation and procedural stage. In the declarative stage,

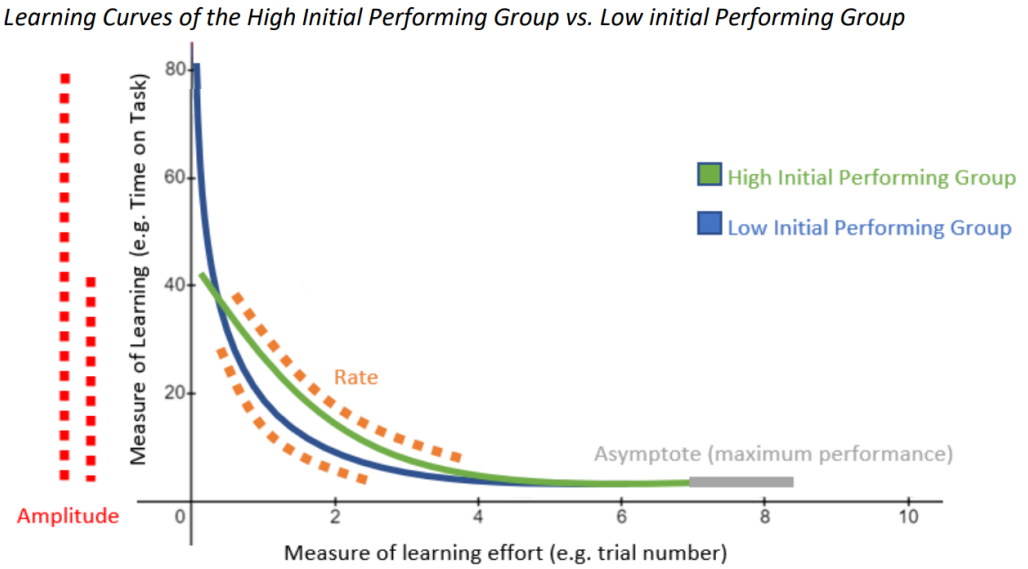


Figure 2.3: Exponential learning curve, where the asymptote describes the maximum performance, the rate reflects the learning speed and the amplitude reflects the potential for improvement [147].

chunks of information are separately being decoded whereas in the knowledge compilation stage, these chunks are combined together, leading to the recognition of similar patterns of information. Lastly, in the procedural stage, the recognition of these patterns is integrated in the procedures required for performing the skill under consideration.

Learning in the models of Pots & Fitts [143] and Anderson [145] are an amorphous mass, that is they have no clearly defined shape of form. To visualize and predict the skill acquisition over time of an individual, exponential learning curves [146] can be used as depicted in figure 2.3.

These curves are parameterized into three variables: the amplitude, describing the learning potential; the learning rate, describing the speed of learning; and the asymptote, reflecting the maximum possible performance. Mathematically, this exponential learning curve is described by the function:

$$P(t) = Asymp + Ampl * e^{-rate*t} \tag{2.15}$$

where P is the performance in trial t. In the beginning, there is a great amount of potential for improvement, where the amplitude is typically lower for learners that already have previous experience as compared to learners with less previous experience. With every trial the amplitude decreases until it amplitude becomes zero and the asymptote, the maximum performance, has been reached.

Based on the exponential learning curve model, Schmettow [129] has developed a new model, the Tweak Finder Model (TFM). This model defines learning as the process of exploring a pool of undiscovered tweaks and is based on the following assumptions: 1), the pool of possible tweaks is finite. 2), finding a tweak is irreversible; and 3), there is a fixed probability of finding a tweak. As a consequence, there are many tweaks to be found in the beginning of the learning process, where you will find less tweaks with every trial until the pull of possible tweaks is exhausted and you have reached your maximum performance. For a beginner, the pool of tweaks is full and there are many tweaks to be found whereas for an expert most tweaks are already found. Converting these assumptions into a mathematical model is done as follows:

$$P(t) = Asymp + Ampl * c^{t} \tag{2.16}$$

where the amplitude equals βN_0, a rate β times the number of all possible tweaks N_0. The term c describes the catch rate, the constant probability of discovering new tweaks during a trial. This model has been used by several students at the University of Twente to compare and predict learning under various conditions in the surgical domain [147–151], but also to describe effectiveness of driving simulators [152, 153].

Learning curve data is generally complex as it is longitudinal and non-linear at the same time, i.e. data is measured at different points in time and participants do typically not improve linearly with every trial [147]. Moreover, the data analysis increases in complexity when the goal is to estimate differences on the participant level, rather than on the group level where individual parameters, such as amplitude or catch rate, may be lost during averaging over the respective groups [129, 154]. To adequately capture such complex data, traditional statistical methods are most likely going to struggle [142] – enter Generalized Linear Models (GLMs) in conjunction with the Bayesian framework of statistics.

2.3.2 Modelling Learning Curves in R

To model non-linear learning curve data, the Tweak Finding Model as described in [129] can be used. Here, the dependent variable, for instance Time on Task `ToT`, can be estimated over various trials, by incorporating the variables `amplitude`, `asymptote` and `Chrt` (catch rate):

```
ToT ~ Asymp + Ampl*Chrt^Trial,                    (2.17)
```

Note that for such non-linear models, we are required to specify priors for each of the predictors (`amplitude`, `asymptote`, `Chrt`), separately. Using expected maximum and minimum bounds, we can create can weakly informative priors, see [129] for a tutorial. Once such priors have been defined, linearization can be integrated into the model formula, to make the model fitting itself easier. This can be done by specifying the priors on the desired scale of the link function, using the respective inverse functions in the model formula itself, and using the identity as link function for the family distribution. To fit the model above, one could specify the priors on the log scale, use the Gamma distribution as family distribution and convert equation 2.17 to equation C.1:

```
ToT ~ exp(Asymp) + exp(Ampl)*inv_logit^-1(Chrt^Trial),    (2.18)
```

To distinguish between effects of different factors, one can add suitable predictors to each of the learning model predictors, which are then inserted as a list into the non-linear model of equation C.1. For instance, the following would model 1) differences between different groups of a condition (e.g. high vs low SoE) by adding the term `SoE` to `Chrt`, 2), fatigue by adding the term `Trial` to the `asymptote` and 3) individual differences to all three variables, `amplitude`, `asymptote`, and `Chrt` by adding the random effect `(1|Participant)`:

```
   Ampl ~1 + (1|Participant),
  Chrt ~SoE + (1|Participant),                    (2.19)
Asymp ~Trial + (1|Participant)
```

Once the model is fit, we obtain posterior distributions for each of the parameters `amplitude`, `asymptote`, and `Chrt` and their conditional effects. We can then make inferences about differences between groups in learning speed, maximum possible performance. We can also use this model to predict performance beyond the observed data. Overall, employing the TFM in the `BRMS` engine presents a promising approach to estimate non-linear multi-level models of

learning curve data, which would not be possible with conventional, frequentist approaches. However, note that the observed data however do needs to follow the learning curve on the participant level in order for this model to be representative [129]. In case of flat or wobbly lines of the observed data, the amplitude is either zero or the catch rate is so small, that in practice no learning is observed, thus a basic GLM with linear predictors, including the predictor `Trial` as predictor for independent trials, can be used. This then allows to compare absolute performance per trial, but does not allow to make any inferences within the framework of the TFM.

3 Methodology

The goal of this study was to investigate the relation between Sense of Embodiment (SoE) and task performance in the context of telerobotics. This section describes the methodology of the present study, that is, how the research question and hypotheses gave rise to the overall experiment design, such as measures, user tasks and the experiment procedure.

3.1 Research Question & Hypotheses

The goal of this study was to explicitly link the level of SoE to the level of dexterous task performance in the context of telerobotics. The main research question as formulated in the introduction is

> To what extent does the level of Sense of Embodiment affect the level of dexterous task performance in telerobotics?

To answer this question, two levels of embodiment, suppressive and supportive, have been created, similar as in [4]. Based on these levels of embodiment, two hypotheses have been formulated, namely:

- **H1:** A supportive embodiment leads to better dexterous task performance when compared to a suppressive embodiment.
- **H2:** A supportive embodiment leads to an increased learning rate for dexterous tasks when compared to a suppressive embodiment.

3.1.1 H1, Embodiment correlates with dexterous task performance

The first hypothesis is taken from Toet et al. [6], in which the authors cite various studies showing that SoE significantly correlates with various motor regions in the brain [11–13]. Based on these studies, Toet et al. postulate that SoE correlates with task performance which has been proven to some extent in the domains of prosthetics [14, 15, 17, 18] and VR [4, 19, 20] and is therefore expected to also transfer to the domain of telerobotics.

The free energy principle & the predictive coding framework

The hypothesis by Toet et al. may also be supported by the *free energy principle* [155] which forms the basis for the *predictive coding theory* [21–23]. The *free energy principle* has been proposed as a principle to unify various theories that describe the inner workings of the brain. This principle states that the brain tries to minimize surprise, that is, the difference between a brain's current and predicted states, to maintain a systemic homeostasis in case of destabilizing factors in an environment [155, 156].

Naturally, the question arises how can an agent minimize surprise as it cannot know whether its sensations are surprising and could not avoid them even if it did know. Free energy tries to answer this question by serving as a proxy for surprise: as an agent reduces its free energy, it also reduces its surprise. According to Friston, free energy can be estimated by an agent because it has access to its sensory states and a recognition density that is encoded by its internal states [155]. The latter is a probabilistic function of what caused a particular sensation. Thus, to minimize surprise, an agent can change its sensory input by acting on the world or it can change its recognition density by changing its internal states. The former method describes selectively sampling the sensory inputs the agent expects, which is also known as active inference [157]. Here, Friston [155] gives the example of feeling our way in darkness where we anticipate what we might touch next and then try to confirm those expectations. The latter

method describes the process of updating the internal model of the world, which is expressed in terms of the probability of a sensation and its causes occurring together. These two processes are used to update posterior beliefs in the Bayesian brain hypothesis [158] in which the brain represents sensory information probabilistically. Here, the brain has a model of the world that generates explanations and predictions about its sensations, against which sensory samples are tested to update beliefs about their causes [155]. This model consists of a likelihood (the probability of sensory data, given their causes) and a prior (the a priori probability of those causes). The Bayesian brain hypothesis describes perception then as the inversion of the likelihood to access the posterior probability of the causes, given sensory data.

The *predictive coding theory* is an extension of the free energy principle, which postulates that the brain formulates models at each stage of perceptual and cognitive processing to decide what information it expects from the level below (top-down) [6]. The actual bottom-up sensory information is then compared to the predicted information, where prediction errors are transmitted to higher levels of processing, updating the model of the world, which corresponds to the body image and schema in the context of SoE. In [156], the authors investigated the relationship between mental effort, information-resource processing costs, and free energy. They demonstrated that free energy differences negatively correlates with the increased allocation of information-processing resources as well as with subjective workload ratings, providing the first empirical evidence of a relationship between mental effort, brain free energy, and neurocognitive information-processing. In other words, they showed that less prediction errors lead to less mental workload in the framework of the free energy principle. In context of the predictive coding theory, a higher level of SoE would be associated with less prediction errors over body image and schema as sensory states match prior beliefs about their causes. Based on [156] it is expected that higher levels of SoE lead to lower levels in mental workload which in turn is expected to lead to an increase in task performance as more processing power can be allocated to the task at hand.

Studies on embodiment & task performance

While there have not been any studies conducted explicitly on the relation between embodiment, task performance and cognitive workload, there is a variety of studies demonstrating higher task performance for increased levels of embodiment. The following sub-sections discuss studies that investigated the contribution of specific embodiment sub-components on task performance.

Ownership
Various studies have shown that SoO correlates with motor regions in the brain [11, 13, 159], specifically the insula, the temporo-parietal junction, and premotor cortices. In [19, 160, 161] it has been shown that ownership directly mediates motor performance. Kilteni et al. [160] explain this relationship with the body schema, where the "central multisensory representation of the owned limb in space provides critical input information about the body state to the forward model that generates the predictions about the sensory consequences of the limb's voluntary action". Therefore, perceptual ambiguity (e.g. through delay or reduced depth perception) would create a distorted multisensory representation of the body, leading to less accurate predictions of how the limb would respond to commanded movements.

Agency
In both of the studies Laha et al. [162] and Egeberg et al. [163], the sense of ownership and agency coincided with greater performance, however both studies were lacking correlation analysis to study the relation between the phenomena [164]. In Rosa et al. [164], the connectedness to a virtual avatar was altered to create different levels of SoO SoA. They observed a direct correlation between SoA and task performance. The study by Falcone et al. [4], investigated the effect of various perceptual cues on embodiment as well as on task performance.

The results show a significant (but weak) correlation between SoE (i.e., SoO and SoA) and task performance, where visuo-proprioceptive synchronicity affected SoO and SoA the most.

Self-Localization
In Gorisse et al. [112], the effect of different view points (1PP vs 3PP) on embodiment and task performance was compared for a navigation task in VR, where 1) participants had to deflect projectiles flying from a distance towards their avatar and 2) navigate through three sections without deviating from the given path, where access to each section is activated by manipulation of a terminal. The authors found an increased spatial awareness for the 3PP, which the authors explain by the wider field of view created by the 3PP, enabling participants to react faster to objects coming from the periphery of the avatar. However, although the 3PP enabled an increased perception (i.e. faster detection) of the projectiles, the actual task performance remained similar for both view points. Additionally, a majority of the participants reported that the deflection was easier for the 1PP condition, as it allowed for better depth perception and as the avatar did not occlude the approaching projectiles. Similar findings were reported for the second task. Here, objective and subjective results show that participants with the 1PP obtained higher effectiveness for the manipulation task as well as lower navigation times.

Medeiros et al. [165] investigated the effect of 1PP and 3PP with various degrees of realism (abstract, mesh-based, and a real-time point-cloud avatar) on SoE and task performance in a navigation task in a VR environment. Their results show that a 3PP creates a similar level of embodiment and spatial awareness when a realistic representation of the avatar is used. In all other cases they found that the a 1PP, regardless of the realism level of the avatar, induced higher SoE (i.e., SoO, SoA and SoS) and lead to significantly higher task performance (total task time, number of collisions, collision time), showing that a 1PP enabled participants to execute the task faster and with more precision as compared to a 3PP. These findings differ from the studies by Gorisse et al. [112] in which a 3PP was found to improve spatial awareness. Medeiros et al. argue that this may be only true when the 3PP allows for a wider field of view and when the task can benefit from it (e.g. when objects are not always visible). In the study by Medeiros et al. the objects were always visible for the participant, to account for the bias of a wider field of view.

3.1.2 H2, Embodiment correlates with learning speed

The second hypothesis is also based on the free energy principle and the predictive coding framework, specifically in relation to the concept of reinforcement learning and the exploitation-exploration (bias-variance) trade-off. Based on optimal control theory [166], the brain tries to optimizes value, which is expected reward (or its complement, expected cost), which corresponds to reinforcement learning in behavioural psychology [155]. Given a particular policy that labels states as rewarding or costly, a set of actions is prescribed to the agent. According to the principle of optimality [166], cost is the rate of change of value which depends on changes in sensory states suggesting that optimal policies are formed via prior expectations about the motion of sensory states [155]. In the predictive coding framework, free energy is minimized, which effectively optimizes empirical priors (i.e., the probability of causes at one level, given those in the level above). Due to the hierarchical link between empirical priors, they are updated by sensory states, enabling the brain to optimize its prior expectations online [155].

Note that learning always faces an exploitation-exploration trade-off: the stronger the prior beliefs are, the less sampling of states is needed (i.e., less exploration) for learning of the given task [167], and the more exploitation is done. In contrast, an agent with weak prior beliefs will explore a wide range of states and thus be less sample efficient. This has been proven for the predictive coding paradigm in [168], where two scenarios have been compared in which amongst others, two active inference agents (one with and one without prior beliefs) have been compared. Here an agent with no prior beliefs was significantly less effective in completing

the task as compared to an agent with prior beliefs, as it was constantly exploring the state space whereas the agent with prior beliefs reached the exploitation phase after a few training episodes, which constitutes faster learning.

In this context, good embodiment corresponds to strong prior beliefs or certainty about the body image and body schema the surrogate which therefore requires less exploration to learn how to use the surrogate. For instance, higher predictability visuo-proprioceptive feedback would increase the certainty of how the surrogate responds to commanded motions, thus decreasing exploration and exploitation, i.e. effective use of the surrogate to execute a task. Put simply, the system feels more intuitive because it resembles what we are used to from every day, how our human body works. It is therefore expected that a higher level of embodiment will result in faster learning of the system as compared to a lower level of embodiment.

3.2 Experiment Setup & Design

This section describes the experiment setup as well as the experiment design. These sections have been merged together as the used setup contributed to a large part to the choices made in the manipulations of the experiment design while specific parts of the setup have been added based on the experiment design.

3.2.1 Experiment Setup

The telerobotic setup used for this experiment consists out of a Virtuose 6D [169] and HGlove [170] on the master side and a Franka Emika Panda robotic arm [171] and QB Softhand [172] on the slave side, as depicted in figures 3.1 and 3.2.

Figure 3.1: Telerobotic setup on the master side, including a Virtuose 6D and HGlove.

Haptic feedback has been implemented for the HGlove and QB Softhand in previous work by Nadgere [173] based on the papers of Ajoudani et al. [174] and Brygo et al. [175]. The haptic feedback is based on the concept of synergies [176], simplifying the grasping procedure to one Degree of Freedom (DoF). In [173], the existing methods have been extended with a pose tracking controller that guides the finger tips of the operator along the first synergy to reduce the pose mismatch between the remote and local hand. To account for fluctuating haptic feedback of the HGlove for low (close to zero) gripping velocities, the controller of the Nadgere has been extended with a dead band controller, as described in [24].

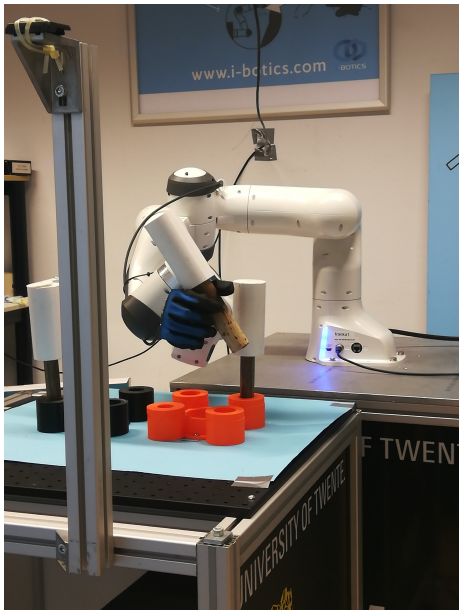


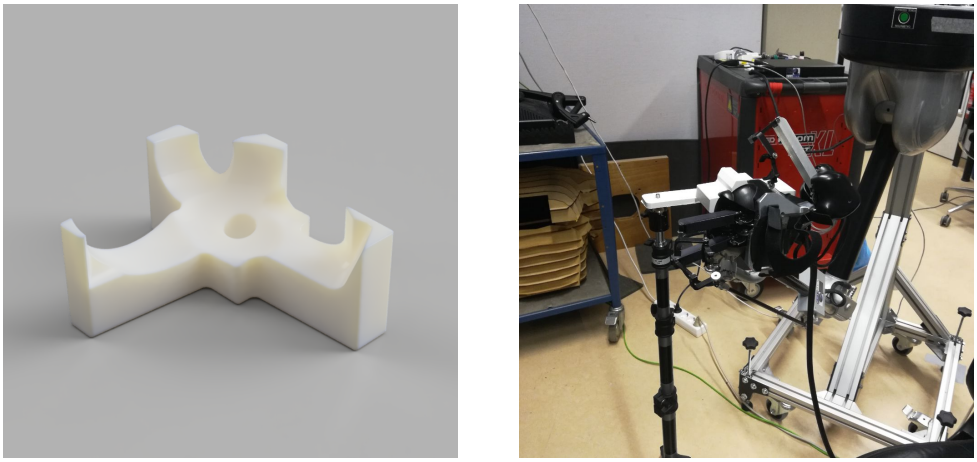
Figure 3.2: Telerobotic setup on the slave side, including a Franka Emika Panda and QB Softhand.

The Virtuose 6D provides haptic feedback by reflecting the estimated impedance of the remote environment via a passive bi-directional impedance controller that has been implemented by Lieftink [177]. The Virtuose 6D has additionally gravity compensation implemented in order to let its end-effector feel weightless and thus reduce fatigue caused by its weight. Based on the configuration of the Franka arm for a given task, the robot arm may easily move into its joint limits, causing the arm to enter an error mode that locks all of its joints, interrupting the experience. The controller of the Virtuose and Franka arm has therefore been adopted to model rotational springs into each joint, pushing the arm out of the joint limits based on the distance to the respective joint, as described in [24]. Note that control and haptic feedback of the QB hand & HGlove is independent from the Virtuose & Franka arm and vice versa, as they are running on different PC's. This means that any delays of these systems and the associated stability issues are independent from each other. To "clutch" the participants reliably in at the same position, a negative 3D print of the HGlove has been mounted onto a tripod, as depicted in figure 3.3. The term "clutching" refers to the process of matching the world frames of the master and slave device [178]. It is important to clutch in the participant at the right position as it otherwise may limit the available workspace in the remote environment. For example, when the participant's arm is too high when clutching in, there will be less workspace available in the remote scene as the participant will reach the limit of the local workspace to soon.

Finally, to stream the remote environment to the operator, a HTC VIVE Pro Eye [179] and a Zed Mini stereo camera [180] have been used. Using the Zed SKD, the camera images of the Zed Mini have been projected onto two planes in Unity, which have then been rendered onto lenses of the HMD.

3.2.2 Experiment Design

The experiment design is a between-within subjects design with two factors, namely `SoE` and `trial`, where the former constitutes the between-subjects part and the latter the within-subjects part of the study design. The `SoE` factor has two levels, namely 1) supportive, enhancing the level of SoE and 2) suppressive, diminishing the level of SoE. The sensory cues to manipulate the SoE have been chosen mainly based on the available setup for the study and include 1) point of view, 2) depth cues, and 3) haptic feedback. See table 3.1 for an overview, including the affected SoE sub-components.



(a) **(b)**

Figure 3.3: Negative 3D model of the HGlove (a) and resulting print holding the HGlove (b) that has been used for clutching in participants always at the same position.

Manipulation	SoE Components	Suppressive SoE	Supportive SoE
Point of view	SoO, SoS	3PP	1PP
Haptic Feedback	SoO, SoA	No	Yes
Depth Cues	SoA, SoS	Monocular	Binocular

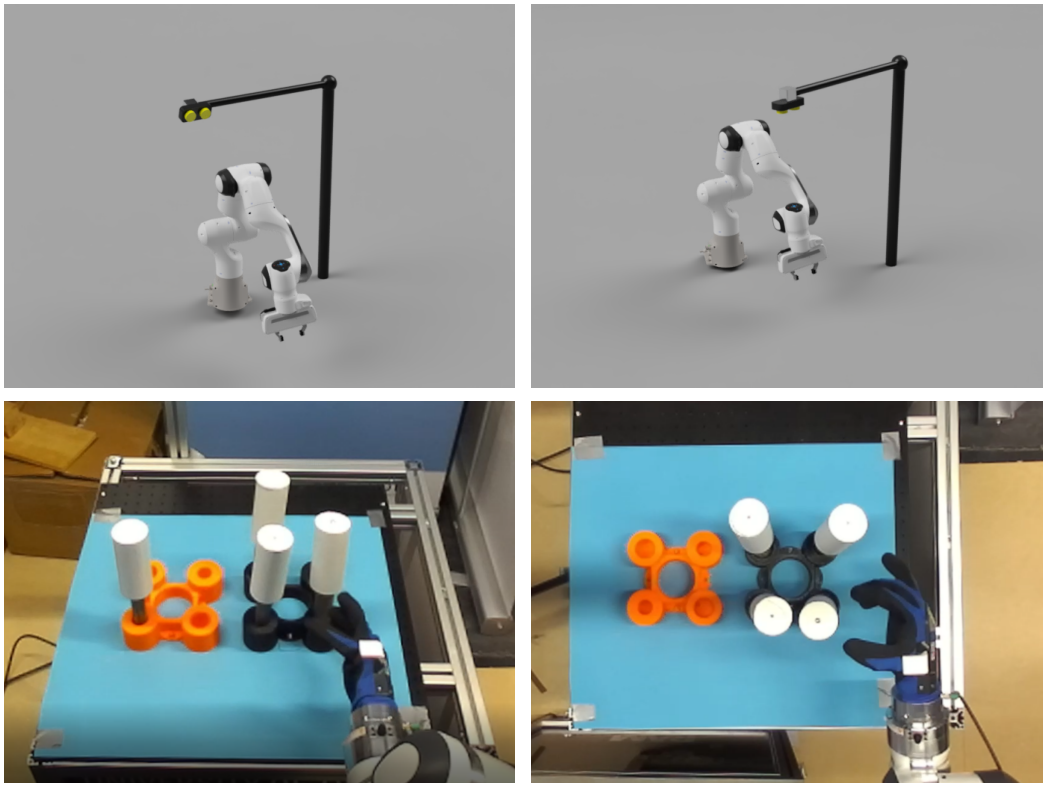
Table 3.1: Chosen perceptual cues to manipulate SoE.

Point of View

As discussed in section 2.1.2, the point of view is a powerful tool to manipulate SoE, as a 1PP leads to significantly stronger levels of SoO [6, 60, 63, 113] and SoS [41]. Moreover, embodiment can be elicited via a 1PP without any further perceptual cues [77, 114] and persist even under asynchronous stimuli [60, 77, 116], whereas a 3PP requires additional reinforcement in form of either synchronous visuo-tactile feedback [110, 117, 118] or spatial overlap between the human and the avatar [88]. Thus, for the present study it is expected that a 1PP enhances the levels of perceived ownership and self-location whereas a 3PP is expected to diminish them. To create a 1PP and 3PP, two different camera perspective have been chosen as depicted in figure 3.4. Here, the top row illustrates the position of the camera in relation to the robot arm whereas the bottom row shows the resulting HMD scenes.

Haptic Feedback

As discussed in section 2.1.2, congruent haptic feedback can significantly improve embodiment levels as compared to no feedback due to multisensory integration of visual and motor information [95]. In the present study, participants of the supportive embodiment group experienced force feedback via the HGlove whereas participants of the suppressive embodiment group did not experience any force feedback. Note that the force feedback rendered via the HGlove consists out of force feedback that 1), provides force feedback based on interaction forces estimated on the QB Softhand, thus reflecting forces experienced from grasping objects in the remote scene and 2), provides pose tracking feedback, i.e. force feedback guiding the finger tips of the user along the first synergy, and thus along the available grasping trajectory of the QB Softhand. Note that no user experience studies have yet been conducted on the added effect of the pose tracking controller by Nadgere [173]. It is expected that the force feedback provided to the supportive embodiment group elicits higher SoO and SoA levels as compared to the suppresssive embodiment group: First, based on the RHI paradigm [33, 83] and previ-



(a) 1PP **(b)** 3PP

Figure 3.4: Camera positions used to create a 1PP (a) and 3PP (b).

ous studies investigating the effect of force feedback on ownership [95, 181], it is expected that the multisensory integration of displaying synchronous visual and force feedback elicits higher subjective SoO ratings as compared to experiencing no force feedback. Second, it is expected that the pose tracking feedback enhances the predictability of the movements of the QB Softhand and therefore increase agency for grasping movements.

Note that both groups experienced the haptic feedback provided by the Virtuose 6D as to preserve the experimental setup, i.e. prevent participants from colliding the robot arm with objects present in the remote environment.

Depth Perception

Depth perception has been used as a novel manipulation to affect SoS as well as SoA. As discussed in section 2.1.2, increased levels of depth perception have been shown to improve spatial awareness in the remote environment [121–123]. Additionally, binocular cues such as stereopsis have been shown to increase depth perception significantly in the peripersonal space [125], enhancing task performance for tasks like grasping or reaching [126]. Binocular cues therefore calibrate proprioception of the arm, enabling better hand-eye coordination which ultimately leads to higher agency for arm movements [124]. In the present study, depth perception has been manipulated by providing binocular (and monocular) depth cues to the supportive embodiment group while providing only monocular depth cues to the suppressive embodiment group. Specifically, stereopsis has been used as a binocular cue by rendering each of the stereo camera images onto a plane in Unity which is then streamed to the lenses of the HMD [31]. Similarly, to decrease the depth perception of the suppressive embodiment group, only one of the camera images has been used and streamed to both lenses of the HMD. Therefore, the suppressive group had to rely on monocular cues only, such as using shadows or occlusion as depth references and experienced reduced depth perception in the peripersonal

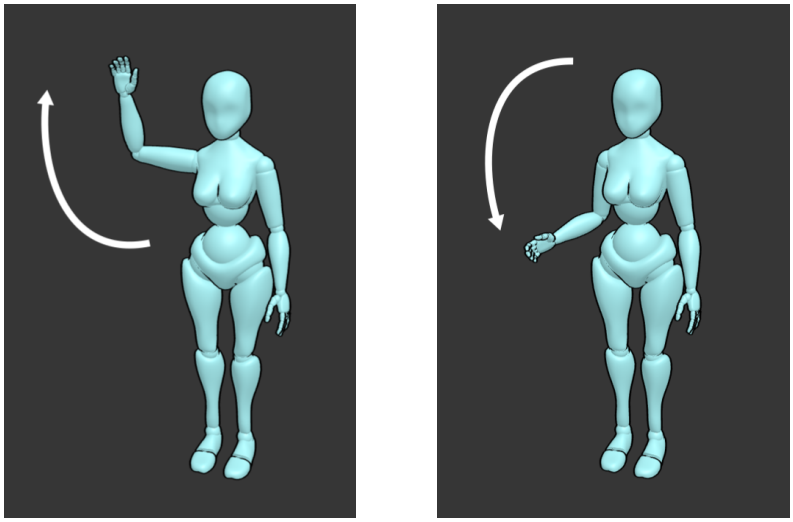
space. Combined with the 3PP, it is expected that it will be especially difficult to predict movements in the Z plane for the suppressive group, whereas the supportive group is expected to undergo a visual experience that is relatively close to the natural, human vision.

3.3 User Tasks

This section describes the user tasks that have been employed in the current study, namely a 1) position Judgement, 2) exploration, and 3) peg & hole task.

3.3.1 Position Judgement

The first task, position judgment, is a motor response task [32] to measure proprioceptive drift and is based on the study of Lenggenhager et al. [110]. In this task, the participant's arm has been displaced by an experimenter to a random position, from which the participant was then asked to move their arm back to their initial location as illustrated in figure 3.5. The final position of the Virtuose end-effector was recorded once the participant verbally expressed that they found their initial location. Based on previous studies [33, 52, 182], this task has been repeated five times to measure proprioceptive drift. This task has been conducted in two trials: the first trial serves as a baseline and has been performed at the very beginning of the experiment before any visuo-proprioceptive stimulation has taken place to induce SoE. The second trial has been conducted after the exploration task and the peg & hole tasks have been completed.



(a) **(b)**

Figure 3.5: Position judgement task. First, the hand of the participant is displaced by the experimented to a random location (a). Then, the participant has to find their initial location back (b).

During this task, the participant was wearing the HMD and was seeing the remote robot hand moving, and thus not their real hand. As the participants hand was moved, the remote hand was moved as well.

3.3.2 Exploration

The second task was an exploration task that was meant to get the participants accustomed to the controls of the tele-operation setup and induce SoE via visuo-proprioceptive stimulation [73]. In this task, the participants have gone through a guided exploration phase in which they had to touch the corners of the mounting plate in a random sequence which has been verbally expressed via the experimenter. The participants were guided by means of the sequence to avoid any sudden movements and to preserve the experimental setup. Additionally, this task was intended to prime the participant and induce a SoE over the robot arm via visuo-

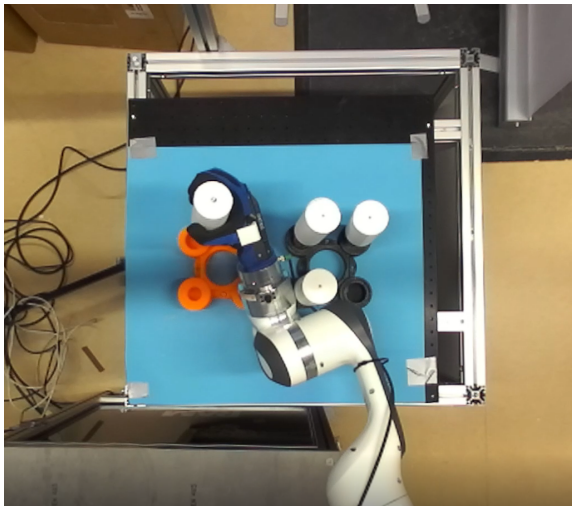


Figure 3.6: Choosing the order of inserting the pegs in important. Here, the Franka arm being blocked by the lower left peg.

proprioceptive stimulation [73], which has been shown to be comparable to visuo-tactile synchronization for generating SoE for robot arms. The latter method has not been used due to practical reasons, as the finger mounts of exoskeleton (HGlove) worn by the participant exert a constant pressure on finger tips, which would create a visuo-tactile conflict with the remote hand seen in the HMD.

3.3.3 Peg & Hole

The main task of the experiment was a peg & hole task [183]. In this task, participants have been asked to move pegs from mount A to mount B by using the telerobotic setup, as depicted in figure 3.7. When a participant dropped a peg, it was out of the "game" and has been removed from the remote environment via the experimenter. Additionally, once a peg has been inserted into a hole, it could not be removed. Note that the size of the holes and peg shafts was varied such that the bigger peg did not fit into the smaller hole. This has been done in order to increase the complexity of the task and make learning effects more visible as participants need to first recognize that are different sizes and then choose their strategy in moving the pegs in the right order. For instance, not starting with the lower left peg, will block the arm movement to the right as shown in figure 3.6.

This task has been chosen because it challenges the operator to precisely manipulate objects using the robotic surrogate and tests the operator's spatial perception of remote environment. Moreover, it allowed to easily quantify task performance, as effectiveness can be measured as successfully moved pegs from mount A to mount B whereas efficiency may be measured in terms of the total completion time per trial or time per successfully moved peg. Additionally, this task has been chosen based on the capabilities of the present system, in particular the haptic feedback of the QB softhand and the HGlove, which is dependent on the estimated external interaction forces when grasping an object, thus if a grasp does not produce sufficient external force towards the remote hand, the operator will feel a limited amount of haptic feedback via the HGlove. In other words, if an object that is being grasped is too small, or does not fit the shape of the grasping trajectory of the QB hand, the operator will feel an insufficient amount of haptic feedback. Appropriately sized pegs that fit the QB soft hand have been 3D printed in order to yield suitable haptic feedback as depicted in figure 3.7.

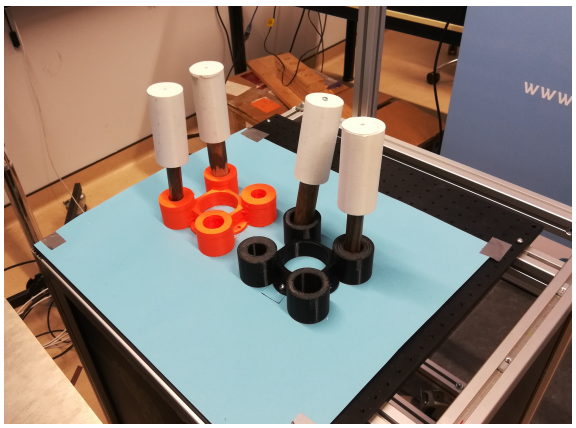


Figure 3.7: Mounts of the peg & hole task.

3.4 Measures

To answer the research question and the corresponding hypotheses, relevant constructs have been operationalized into measurable variables, which are in this case SoE, task performance and learning effects. All variables of interest and their respective measures can be found below in table 3.2:

Variable	Measures
SoE	Surveys, interviews, proprioceptive drift
Task performance	Time, number inserted pegs
Learning	Amplitude, asymptote, catch rate

Table 3.2: Variables and measures.

3.4.1 Sense of Embodiment

A prerequisite of the present study was that the level of SoE has been reliably manipulated. Therefore, a manipulation check has been conducted, that is, the effectiveness of the manipulation has been evaluated by assessing all three sub-components via explicit and implicit measures. The explicit measures consisted out of self reports, namely surveys and interviews whereas the implicit measure was a proprioceptive drift measurement.

Survey

The survey used for this study was adopted from [35, 36], with the questions being on a seven point Likert scale (1 = strongly disagree, 7 = strongly agree). The survey included two questions per SoE component, where one of the questions was a control question. The complete survey can be found in appendix B.1.2.

Interview

The interview has been semi-structured and followed a five phase structure to guide the interview in a seamless manner [45]. Next to questions about general user experience and perceived task performance, it included similar questions about SoE levels as asked in the survey, but due to the semi-structured format of the interview, the participant has been allowed to elaborate on their perceived SoE level. All interview questions, separated into the five phases, can be found in the appendix B.2. The survey has been analyzed via emerging coding, similar to the IPA evaluation as done by Lewis and Lloyd [49]. The findings of the interviews have been used to guide the discussion regarding the results from the SoE questionnaire and the measured task performance.

Proprioceptive Drift

In addition, to the above described explicit measures, proprioceptive drift [33, 52] has been used as an implicit measure for SoE to support the explicit findings from interviews and surveys, as these might be affected by subjective bias of the participants [27]. For a high SoE level, it was expected that the new location of the local hand is now closer, in other words drifting, towards the remote hand [52, 182]. In the present study, proprioceptive drift has been measured as the absolute difference in position between the initial and end position of the position judgement task. The respective positions have been measured by recording the position of the end effector of the Virtuose 6D. As proprioceptive drift varies based on the perceived visual distanced towards the hands [54, 184] and two different levels for camera perspective and depth perception have been used in the present study between the embodiment groups, proprioceptive drift has been measured along the X, Y and Z axis separately, as not to bias the measurement toward a specific axis.

3.4.2 Task Performance

Task performance has been measured in terms of effectiveness and efficiency within the peg & hole task via the number of successfully inserted pegs per trial and the time needed to insert a peg successfully. Learning effects have been measured within the TFM framework as discussed in section 2.3. That is, learning has been assessed in terms of maximum performance (asymptote), learning potential (amplitude) and learning speed (catch rate).

3.5 Participants

31 participants have been recruited for the experiment (21 male, 10 female), ranging from 18-24 years (41.9 %) to 25-34 years (58.1 %). Participation requirements included:

- Being right handed
- No upper body injuries
- Normal to normative vision
- Naive to the purposes of the experiment
- No prior experience with telerobotic setups or SoE experiments

Participants needed to be right handed for two reasons, 1) for results to be comparable and 2), the available Virtuose 6D and HGlove configuration was made for right handed people. Furthermore, all participants needed to have normal to normative vision to reduce variation due to impaired vision. Participants that have injuries/issues with their upper body would have been excluded as these 1) may distract them during the experiment, making it more difficult to create and maintain the illusion of embodiment and 2), they would have a disadvantage in terms of task performance compared to the other participants. Furthermore, participants should have had no prior experience with telerobotics and SoE experiments and should be generally naive to the purposes of the experiment. Participants have been recruited via convenience sampling and have been compensated for their time with vouchers worth 10 euros.

3.6 Procedure

At the start of the session, the participants have been welcomed and the general procedure of the experiment has been explained to them, afterwards they have been asked to sign a consent form (see appendix A). Next, participants have filled out a survey that includes 1) a section on

demographics, 2) questions about gaming & VR experience, and 3) previous experience with tele-robotics [1].

Next, the HGlove has been calibrated by creating a synergy database for the right hand of the participant. To calibrate the system, the participant had to first open and close their hand to its maximum open/closed position and then slowly close their hand from a fully opened pose. Additionally, the participant had to hold a ball in their right hand in order to guide their grasping movement along the 1st synergy.

After the participant's hand has been calibrated, the Virtuose 6D and the Franka arm has been clutched in. Only when the particpant was clutched in, the Franka arm would follow the movements of the Virtuose 6D and only then haptic feedback could be generated via the Virtuose 6D. Once the calibration procedure has been completed and the participant was clutched in, the actual experiments began, that is, the participant was asked to perform the proprioceptive drift measurement task, the exploration task and the peg & hole task as discussed in section 3.3. The latter task has been repeated for four trials where after each trial, the participant had to verbally answer the SoE survey in situ as done in [73, 185, 186] in order to mitigate recall bias and to avoid recalibrating the Hglove after each trial, as the latter could have broken the SoE illusion. At the end of peg & hole task, the interview as described in section 3.4 was conducted.

Finally, at the very end of experiment, the motivation behind the study was explained to the participants and they have been thanked for their participation and given the opportunity to ask questions and give suggestions for further improvement. The total duration of each experiment lasted approximately 1.5 to 2 hours.

3.7 Statistical Analysis

To analyze the SoE questionnaire and the task performance measures, various GLMs have been fitted using the R package `BRMS` [142, 154], which is a Bayesian regression engine that is build upon Stan [187]. This section introduces the models that have been fitted and describes the general workflow that has been used to analyze the fitted models. All data types, their respective properties and suggested GLM candidate distributions based on figure 2.1 are summarized in table 3.3:

Measure	**Properties**	**Candidate distribution**
Survey Response	Ordinal, double bounded [1,7]	Cumulative, category-adjacent
Proprioceptive drift	Continuous, lower bound at 0	Gamma
Number of pegs	Discrete, double bounded [0,4]	Binomial, beta-Binomial
Time per peg	Continuous, positive lower bound	Exponential Gaussian

Table 3.3: Properties of measures and candidate GLM family distributions.

3.7.1 Models

This section first describes the general model structure used and then discusses the specific model and family distributions used for the proprioceptive drift measures, the SoE survey responses and the task performance measures.

For all of the data types under consideration we wish to estimate the difference between the embodiment groups in the respective measurement over a set of trials. For instance, for the proprioceptive drift we want to compare the baseline measurement before the peg & hole experiment with the proprioceptive drift measurement after the experiment. The minimal needed model to measure these differences over a set of trials includes the fixed effects for `Condition` and `Trial` as well as the conditional effects thereof. In R modeling syntax this is

[1] Note that 2) and 3) have not been further used in the present study due to time limitations.

described by the function

```
Data ~ Condition + Trial + Condition:Trial,          (3.1)
```

where the dependent variable is on the left hand side of the ~ sign, and the predictor variables on the left hand sign. Here, `Data` represents any desired dependent variable, e.g. proprioceptive drift. On the right hand side of the function, the predictors are separated by the + sign, where `Condition` and `Trial` are the fixed effects and `Condition:Trial` is conditional effect. This formula can also be summarized as

```
Data ~ Condition * Trial.                            (3.2)
```

In addition, random effects for the participants are added to represent the variation between individuals by adding the term `(1|Participant)` as follows:

```
Data ~ Condition * Trial + (1|Participant).          (3.3)
```

Proprioceptive Drift

Since proprioceptive drift is a continuous variable, possible family distributions include the Gaussian, Gamma and exponentially modified Gaussian distribution, as discussed in section 2.2.3. Since the absolute value of the position difference between the initial and end position of the position judgment task has been taken and the data thus a lower bound at zero has, it is expected that either the Gamma or the exponentially modified Gaussian distribution achieve the best model fit.

Survey Responses

As the survey responses classify as ordinal data, they should be estimated with family distributions that are suitable for ordinal logistic regression as discussed in section 2.2.3. For each set of survey responses (SoO, SoA, SoS), the following set of models has been fitted, where each model has a different objective:

Model	**Objective**
Cumulative	Base model
Category-Adjacent	Category specific effects
Unequal var. between conditions	Unequal variance between SoE groups
Unequal var. between items	Unequal variance between items

Table 3.4: Different OLR models and their objectives.

The first model (cumulative) is the most simple model and serves as base model to compare the other models to. The category-adjacent model is meant to account for category specific effects, whereas the last two models are meant to account for unequal variances between embodiment groups and items, respectively.

Since there are two Likert items per SoE sub-component, this results in six different data sets, for each of which these four models would have needed to be fit. However, as discussed in section 2.2.3, multiple Likert items of a Likert scale can be estimated simultaneously by including a random factor for the Likert items. Thus, to minimize the amount of models needed to be fit, both Likert items per SoE have been estimated simultaneously. Additionally, the internal consistency of the respective Likert items has been checked by calculating Cronbach's α. If the Likert items were consistent, that is $\alpha > 0.7$ [188, 189], then the model structure of equation 3.3 has extended by adding a random factor for the Likert item as follows:

```
Data ~ Condition * Trial + (1|Participant) + (1|item)     (3.4)
```

If the Likert items would have not been consistent, separate models would have been fit for them.

Task Performance

No learning effects have been found on the participant level for either of the task performance measure (see section 4.2), and thus using the non-linear TFM was not applicable[2]. Instead, absolute task performance per trial has been analyzed with the same general model structure of equation 3.3. For the count data, family distributions included the Poisson, Binomial and Beta-Binomial distribution and for the time data, family distributions included the Gaussian, Gamma and exponentially modified distribution.

3.7.2 Workflow

The general workflow used for the Bayesian analysis is compromised of an initial data exploration, followed by the comparison of various models of which one is chosen for the final inference, as depicted in figure 3.8:

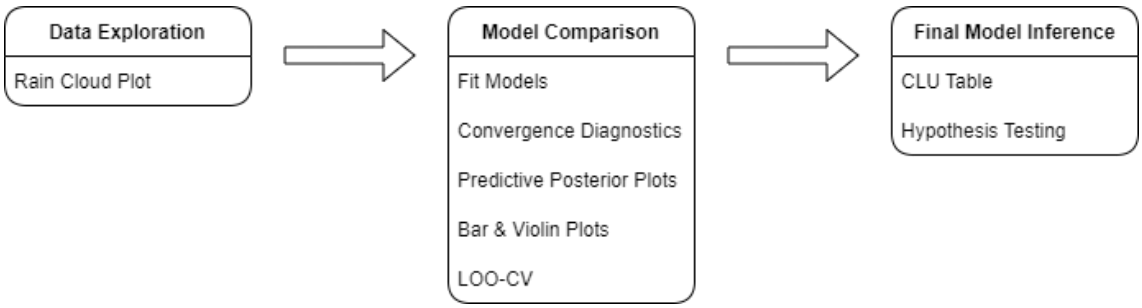


Figure 3.8: General workflow for Bayesian Analysis.

Data Exploration

For every data type, an exploratory data analysis has been conducted in the form of rain cloud plots [190], which is a combination of violin plots, box plots and scatter plots. This summarizes the distribution, summary statistics and the individual data points respectively in one single plot. The rain cloud plot allowed thus to get a detailed initial overview of the data, while also serving reference when evaluating the GLMs.

Model Comparison

Various models with different family distributions for the data type at hand have been fit, as described in section 3.7.1. Every model has been fit with four Markov chain Monte Carlo (MCMC) chains and a minimum of 2000 iterations (1000 warm-up, 1000 sampling). For all models (besides the learning curve models), the default priors generated by `BRMS` have been used, which are half student-t priors with 3 degrees of freedom, where the mean μ is the median of the dependent variable and scale σ is the median absolute deviation of the dependent variable if $\sigma > 2.5$, otherwise it is set to 2.5. Such priors are weakly-informative and were meant to have as little influence on the estimation of the posterior distribution as possible, while improving convergence and sampling efficiency [154]. Additionally, default priors for family specific distributions have been estimated as well such as the `shape` parameter for the Gamma distribution [191]. These additional default priors are also weakly-informative, having only minimal in-

[2]For the interested reader, a tutorial on how to create the learning curve models and priors for both task performance measures has been included in the appendix, see section C.7

fluence on the estimated posterior distribution. In the present study, these weakly-informative default priors have been used for the following reasons: First, there is no existing data from previous experiments that could be used for this specific study to build strong priors. Second, prior specification is a challenging process in itself, where inadequate specification may result in prior distributions that are informative in undesired ways, having a large impact on the posterior and thus distorting the results of the analysis [192]. Third, a diverse range of data has been analyzed, thus it seemed not reasonable to attempt proper execution of such a complex process from a feasibility standpoint.

After the models have been fit, convergence diagnostics have been generated in the form of trace plots as well as Gelman-Rubin [193] tables, where the latter provides a numerical summary for convergence. If convergence failed, iterations of the model have been increased until a maximum of 12000 iterations. If the model still failed to converge, it has been discarded in favor of the other models it was being compared to. If all models for a specific data type would have failed to converge, the model structure would have been simplified by removing the random effects for participants for instance.

Next, the fit of the models has been graphically compared via predictive posterior check plots. From each model's posterior distribution, 100 samples have been generated and compared to the observed data. This allowed to visually inspect to what extent the posterior sufficiently simulates the observed data.

Additionally, bar plots of the 95% credibility intervals of the posterior distribution and violin plots of the residuals have been created for every coefficient of the fixed effects. The bar plots allow to easily compare center estimates and the certainty of the models, whereas the violin plots allow to evaluate how well the model fits the data (i.e., how large is the residual error) and observe whether the inherent bounds of the data are reflected in the skewness of the residual distribution. Here, residuals have been calculated by subtracting the center estimates from the observed data. A residual error close to zero indicates thus a good fit to the observed data.

Finally, leave-one-out cross-validation (LOO-CV) scores [194] have been used to compare the models. Where as the informal model comparison in the previous step is only based on graphical evaluations, LOO-CV is a numerical comparison between models that provides a score that can be interpreted in the same manner as common information criteria such as the Akaike's Information Criterion (AIC) [195] or the Watanabe-Akaike Information Criterion [196] where smaller values represent better forecasting accuracy and goodness-of-fit, as well as lower model complexity [129, 130]. Thus, for evaluating the data in the next section, the model with the lowest LOO-CV has been chosen.

Final Model Inference

For the final model chosen, center estimates and 95 % credibility intervals have been reported in form of a Center, Lower, Upper (CLU) estimates table. This table summarizes the effect of the individual coefficient (e.g. embodiment group) onto the dependent variable (e.g. proprioceptive drift).

Based on the posterior samples from the final model, hypothesis testing has been conducted in the form of parameter estimation. For instance, take the hypothesis that the supportive embodiment group is faster as compared to the suppressive embodiment, and a very simple GLM where we compare the time per successfully inserted peg just between the conditions, i.e. `Time ~ Condition`, which would be described in mathematical terms as

$$Y_i = \beta_0 + \beta_1 x_{1i}, \tag{3.5}$$

where Y is the predicted variable time for participant i, β_0 is the coefficient for the reference level or intercept (supportive embodiment) and β_1 is the coefficient for the suppressive em-

bodiment group. x is a so called dummy coded variable, which is 0 in case of the supportive group and 1 in case of the suppressive group. This means that the effect of the suppressive group is added to the intercept and depending on the sign of β_1, this may either decrease or increase the predicted value Y as compared to the intercept alone. Using these terms, the hypothesis that the supportive group performs faster can be formulated as $\beta_0 < \beta_0 + \beta_1$, which can be summarized to $\beta_1 > 0$ by subtracting β_0 on both sides. Thus, to compute the posterior probability that the supportive embodiment group is faster than the suppressive embodiment group, we simply have to take the average of the coefficient distribution of β_0, where $\beta_1 > 0$. We can also obtain the evidence ratio, by dividing the posterior probabilities of this hypothesis with the null hypothesis ($\beta_1 \leq 0$) to infer how significant this posterior probability is in comparison to the null effect. In [139] an evidence ratio scale is suggested (see table 3.5), indicating different levels of evidence ratios in favor or against a hypothesis, which is being used in the present study to evaluate the hypotheses based on the evidence ratios.

Evidence Ratio	**Evidence**
> 100	Extreme
30	Very strong
10	Strong
3	Moderate
> 1	Anecdotal
< 1	Anecdotal
0.33	Moderate
0.01	Strong
0.003	Very strong
< 0.0001	Extreme

Table 3.5: Evidence ratio scale as suggested by Kelter et al. [139]. The dashed line separates evidence in favor of the null hypothesis (lower part) from the alternative hypothesis (upper part).

Finally, the difference δ of center and lower & upper 95 % HDI limits between two coefficients (e.g. the intercept and β_1) is being calculated and transformed to the original scale (e.g. seconds for the time). With this difference and posterior probability and evidence ratios, we can then infer how likely our hypothesis is, assess how plausible it is compared to the null effect and quantify the difference in data between the embodiment groups. As an (hypothetical) example, imagine we get a difference of 20 [2, 45] seconds, a posterior probability of 90 % and an evidence ratio of 20. We could state then that there is a difference of 20 [2, 45] seconds with a 90 % probability that this difference is greater than zero and that there is strong evidence to support this hypothesis in favor of the null hypothesis.

4 Results

In this section the results of the user study are being presented. Analysis of the surveys revealed medium differences in SoO, medium differences in SoA, and large differences in SoS. The IPA analysis of the interviews shows that ... The position judgement task yielded no differences in proprioceptive drift between the embodiment groups. Regarding the task performance, no learning could be observed for either of the task performance measures count and time. A comparison in absolute performance per trial showed that 1), the supportive embodiment group was 14 % more effective in inserting pegs and 2), "the difference in time needed to insert pegs was only relevant in the first trial (x seconds), whereas the in the succeeding trials, there was no significant difference".

4.1 Embodiment Manipulation Check

This section is dedicated to the manipulation check of embodiment, consisting out of explicit (survey, interview) and implicit measures (proprioceptive drift).

4.1.1 Survey

Likert scale surveys on the embodiment sub-components SoO, SoA and SoS have been analyzed in the present study using Ordinal Logistic Regression (OLR). In the Bayesian framework, multiple Likert items can be elegantly estimated at the same time by adding them as a random factor into the GLM. To do so, these Likert items should be internally consistent, therefore the consistency between the individual SoE question and control question per SoE sub-component has been analyzed using Cronbach's alpha [197]. For this purpose, the scores of the control question have been inverted. Calculating Cronbach's alpha for the SoE sub-component items of the survey yielded the following results:

SoE Component	Cronbach's alpha	95 % CI
SoO	0.895	[0.827, 0.943]
SoA	0.825	[0.714, 0.905]
SoS	0.871	[0.789, 0.930]

Table 4.1: Cronbach's alpha for the SoE sub-component items with a 95 % CI.

For all three SoE sub-components, the individual items yield a Cronbach's alpha value above 0.7, which is generally considered as an acceptable value for internal consistency [188, 189]. Therefore, for each SoE sub-component, both Likert items have been added as random factors to the GLM, where the scores of the control questions have been inverted. As discussed in section 2.2.3, for each SoE sub-component, a set of various OLR models has been fit to account for category-specific effects, unequal variances between conditions or items. The best model has been chosen based on the LOO-CV score and model parsimony. For sake of readability, only the best models are being reported in the following sub-sections, see appendices C.1, C.2, and C.3 for full model comparisons and convergence diagnostics.

SoO

Data Exploration
To explore the observed data visually, a rain cloud plot has been generated for each of the Likert items, showing the distribution, summary statistics as well as the individual data points. Here, the white circle in the box plot indicates the mean whereas the black line in the box plot indicates the median.

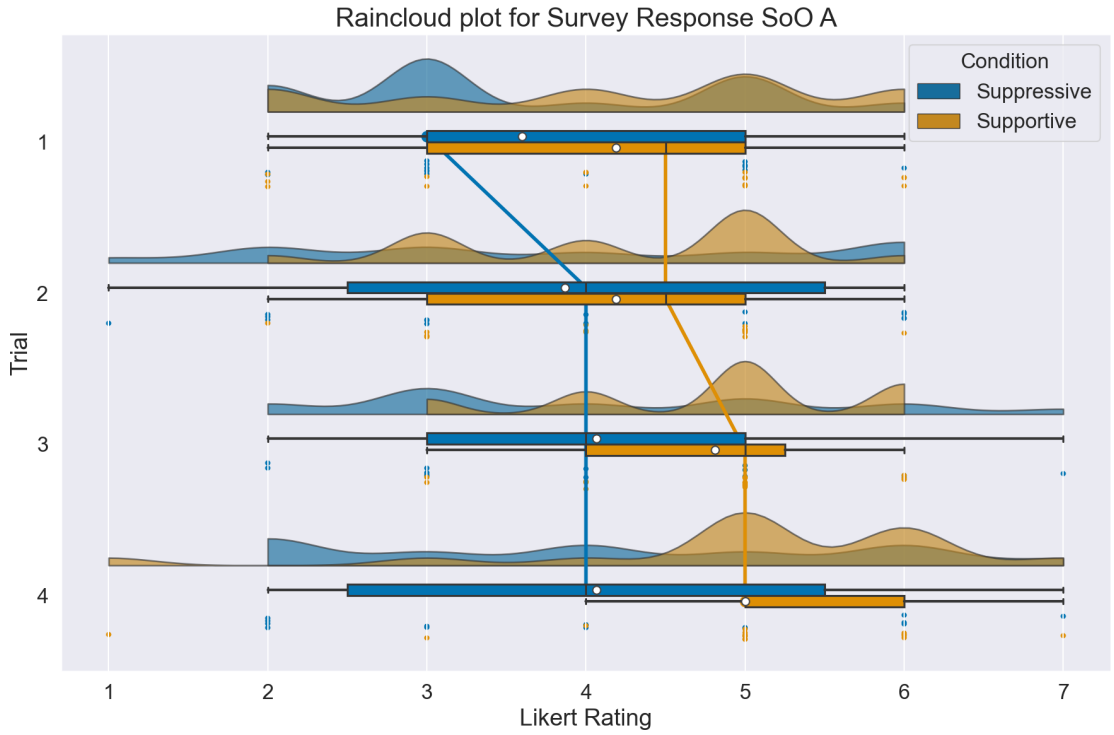


Figure 4.1: Rain cloud plot for SoO item A, "It felt like the robot hand was my own hand".

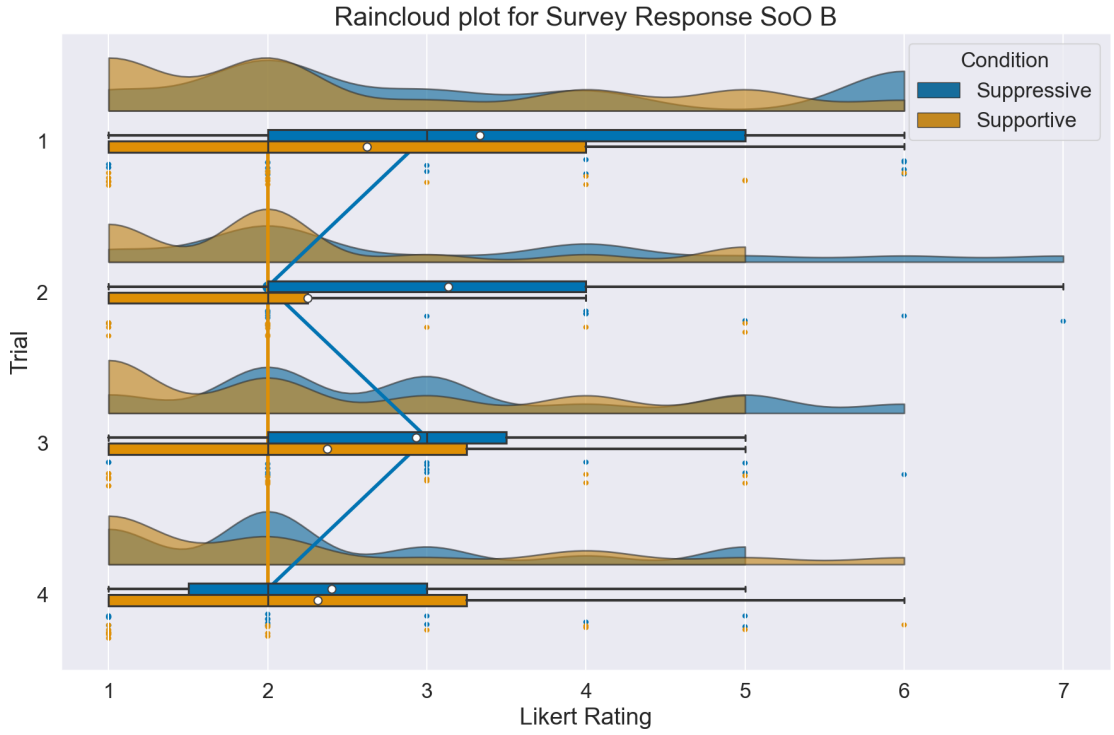


Figure 4.2: Rain cloud plot for SoO item B, "It seemed as if I had more than one right hand".

For item A, *"It felt like the robot hand was my own hand"*, it can be observed that there is a large amount of variance in both groups, however the variance of the supportive embodiment group seems to decrease in trials three and four. Furthermore, based on the medians, modes and means it can be observed that the supportive embodiment group is more centered towards

	Supportive					Suppressive				
Trial	N	Mean	Median	sd	IQR	N	Mean	Median	sd	IQR
1	16	4.19	4.5	1.42	2	15	3.6	3	1.30	2
2	16	4.19	4.5	1.11	2	15	3.87	4	1.73	3
3	16	4.81	5	0.981	1.25	15	4.07	4	1.53	2
4	16	5	5	1.41	1	15	4.07	4	1.71	3

Table 4.2: Summary statistics per trial and condition for SoO item A.

	Supportive					Suppressive				
Trial	N	Mean	Median	sd	IQR	N	Mean	Median	sd	IQR
1	16	5.38	6	1.67	3	15	4.67	5	1.88	3
2	16	5.75	6	1.34	1.25	15	4.87	6	1.81	2
3	16	5.62	6	1.45	2.25	15	5.07	5	1.49	1.5
4	16	5.69	6	1.62	2.25	15	5.6	6	1.35	1.5

Table 4.3: Summary statistics per trial and condition for SoO item B.

the right side than the suppressive embodiment group. Additionally, the modes are more pronounced of the supportive embodiment group, whereas the suppressive embodiment group has generally a very flat distribution (with exception of trial one). Overall it seems that based on this Likert item, the SoO did not significantly change after the second trial for the suppressive embodiment group whereas it seems to have increased trial by trial for the supportive group.

For item B, *"It seemed as if I had more than one right hand"*, the plot shows that both groups have a large amount of variance and that generally there is a large overlap between the distributions of the two embodiment groups, indicating that they are relatively similar. Additionally, it can be seen that the median of the supportive embodiment group remains the same throughout all of the trials, whereas the median of the suppressive embodiment group switches back and forth between the trials, thus staying the same on average. That said, it appears that the supportive embodiment is generally more located to the left side of the rating spectrum, indicating higher ownership.

Overall, it can be observed for both items that the supportive embodiment group perceived a higher SoO than the suppressive embodiment group, however it has to be acknowledged that there is a large amount of variance in both groups.

Model Comparison

To analyze this data, several Bayesian OLR models have been fit, of which the parameter tables and convergence diagnostics can be found in appendices C.1.1 and C.1.2 respectively. To analyze their fit, predictive posterior plots, center & 95 % HDI bar plots, residual violin plots and LOO-CV scores have been generated, which can be found in appendix C.1.3. Summarizing the main findings, the following can be observed: In the predictive posterior check plots, it can be seen that all models perform more or less equally well, where we can see that the center estimates are relatively close to the observed data and uncertainty limits that are of similar in range. This is also reflected in the CLU bar plots of the individual parameters. The residual violin plots show that the residual errors of the models are all very similar. It can be noted that the models have generally less errors when estimating the SoO scores for the supportive embodiment group than compared for the suppressive embodiment group, for which the error distribution ranges mostly between $[-1, +1]$. Overall, the models appear to perform very similar and fit the data relatively well.

To numerically evaluate the models, LOO-CV scores have been calculated for all four models, which can be found below in table 4.4. As can be seen, the OLR model that models unequal

variance between the two Likert items obtains the best LOO-CV score, as compared to the other models. As all of the models appear to fit the data relatively similarly, this model has been chosen based on the LOO-CV score for further analysis.

Model	LOO	SE
Unequal var. between items	707.954	23.061
Unequal var. between conditions	717.445	21.309
Cumulative	717.951	21.290
Category-adjacent	730.430	23.094

Table 4.4: LOO-CV and SE scores of OLR models of the SoO survey responses.

Final Model Inference
Below in table 4.5, center and 95 % HDI's estimates of the fixed effects of the unequal variance between items model can be found:

Parameter	Center	Lower	Upper
Suppressive	-0.5	-1.4	0.3
Trial 2	0.1	-0.3	0.6
Trial 3	0.4	-0.1	0.8
Trial 4	0.5	0.1	1.0
Suppressive:Trial 2	0.1	-0.6	0.7
Suppressive:Trial 3	0.0	-0.6	0.6
Suppressive:Trial 4	0.1	-0.5	0.7

Table 4.5: CLU table for the fixed effects of posterior distribution of the OLR model for SoO on the probit scale.

Note that the thresholds for the individual categories are not being reported, as only the parameter of the embodiment condition and the parameters for the trials are of interest to estimate the level of embodiment between groups over time. Note that the reference level (the intercept) is the supportive embodiment group in trial one.

Finally, using the chosen model, parameter estimation based hypothesis testing has been conducted for each of the four trials. It was expected that the suppressive embodiment group has a lower perceived SoO than the supportive embodiment group. Center, upper and lower estimates, evidence ratio and posterior probabilities of the difference between the SoO ratings between the supportive and the suppressive group can be found in table 4.6.

Hypothesis	Center	Lower	Upper	Evid. Ratio	Post. Prob.
$\beta_{Suppressive} < 0$	-0.51	-1.22	0.18	8.19	0.89
$\beta_{Suppressive} + \beta_{Suppressive:Trial_2} < 0$	-0.45	-1.14	0.24	6.21	0.86
$\beta_{Suppressive} + \beta_{Suppressive:Trial_3} < 0$	-0.53	-1.22	0.17	8.58	0.9
$\beta_{Suppressive} + \beta_{Suppressive:Trial_4} < 0$	-0.4	-1.11	0.29	4.87	0.83

Table 4.6: Summary statistics of the hypotheses testing for each of the four trials for the SoO Likert scale.

For OLR models, differences between the embodiment groups can be interpreted analogous to standard effect sizes like Cohen's *d* as discussed in section 2.2.3. For trial one, it can be observed that the difference δ between the suppressive and supportive group is $-0.51[-1.22, 0.18]SD$, with P($\delta > 0$) = 0.89 and a moderate evidence ratio of 8.19. In other words, given the data, the selected priors and model, the suppressive group rated their perceived SoO with standard difference of -0.51 lower as compared to the supportive group. For trial two, the observed difference δ between the suppressive and supportive group is $-0.45[-1.14, 0.24]SD$, with P($\delta > 0$)

= 0.86 and a moderate evidence ratio of 6.21. For trial three, the observed difference δ between the suppressive and supportive group is $-0.53[-1.22, 0.17]SD$, with P($\delta > 0$) = 0.90 and a moderate evidence ratio of 8.58. In the last trial, it can be observed that the difference δ between the suppressive and supportive group is $-0.4[-1.11, 0.29]SD$, with P($\delta > 0$) = 0.83 and a moderate evidence ratio of 4.87.

Overall, it can be observed that there is moderate evidence that in all four trials the suppressive group had approximately 0.47 SD lower SoO ratings than the supportive embodiment group, which corresponds to a medium effect size in terms of Cohen's d [198]. Additionally, it can be observed that the lowest difference can be found in the last trial, with also the lowest evidence ratio and posterior probability, indicating the perceived SoO converges towards the last trial. However it has to be noted that the value zero is included in all of the four 95 % HDIs, implying that we cannot rule out the null effect as it is still under the most 95 % probable values.

SoA

Data Exploration

Below in figures 4.3 and 4.4, the rain cloud plots for items A and B of the SoA questions can be found.

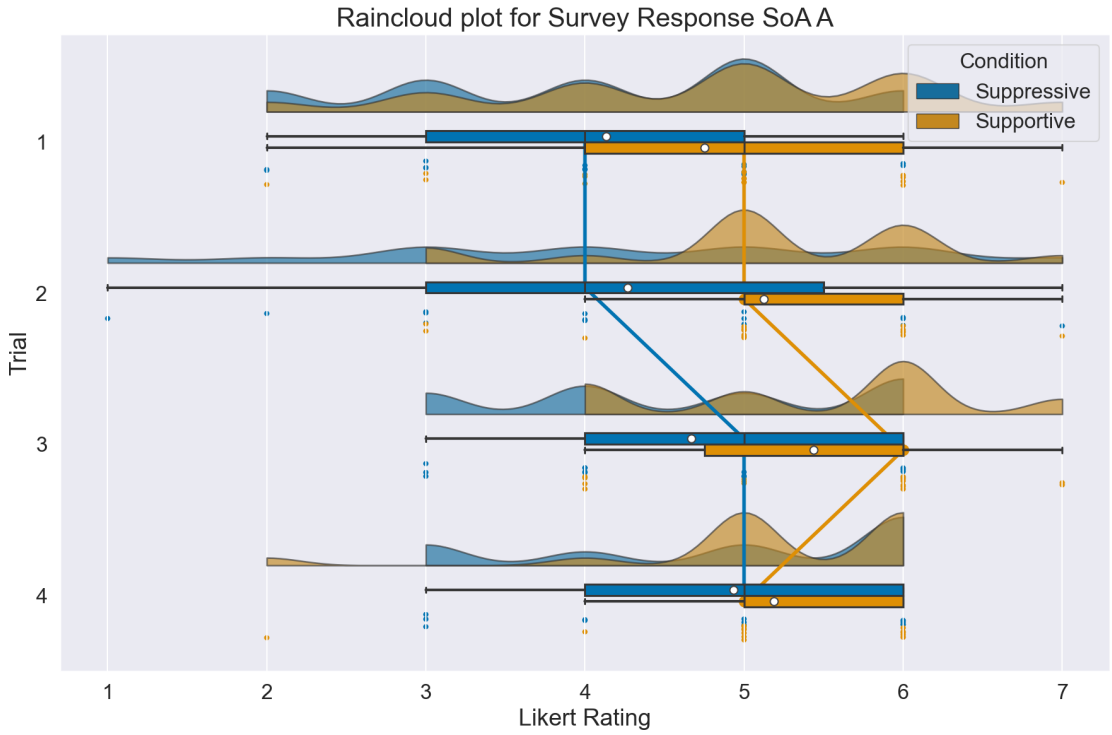


Figure 4.3: Rain cloud plot for SoA item A, "The robot hand moved like I wanted it to, like it obeyed my will".

	Supportive					Suppressive				
Trial	N	Mean	Median	sd	IQR	N	Mean	Median	sd	IQR
1	16	4.75	5	1.34	2	15	4.13	4	1.30	2
2	16	5.12	5	1.09	1	15	4.27	4	1.67	2.5
3	16	5.44	6	1.03	1.25	15	4.67	5	1.18	2
4	16	5.19	5	1.05	1	15	4.93	5	1.22	2

Table 4.7: Summary statistics per trial and condition for SoA item A.

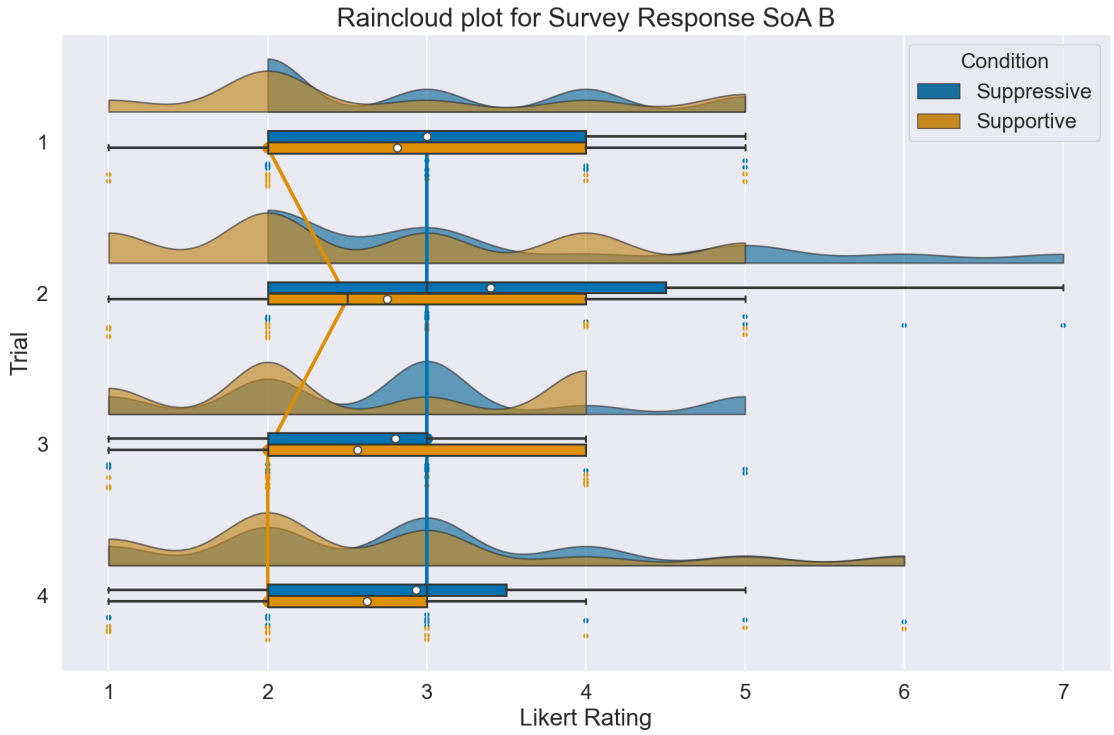


Figure 4.4: Rain cloud plot for SoA item B, "It seemed as if the robot hand had a will of its own".

	Supportive					Suppressive				
Trial	N	Mean	Median	sd	IQR	N	Mean	Median	sd	IQR
1	16	5.19	6	1.38	2	15	5	5	1.13	2
2	16	5.25	5.5	1.34	2	15	4.6	5	1.64	2.5
3	16	5.44	6	1.15	2	15	5.2	5	1.21	1
4	16	5.38	6	1.41	1	15	5.07	5	1.39	1.5

Table 4.8: Summary statistics per trial and condition for SoA item B.

For item A, *"The robot hand moved like I wanted it to, like it obeyed my will"*, a large overlap between the embodiment groups can be observed in the first trial where there supportive group is slightly more distributed to the right side than the suppressive group. The difference between the embodiment groups becomes more pronounced in trials two and three, however in the last trial the groups seem to converge in SoA ratings. Generally it appears that the supportive group has less variance compared to the suppressive group, which is especially noticeable in trial two, where the distribution of the suppressive group is almost flat. Overall, for item A it seems that both embodiment groups have a similar rating in the beginning and end of the experiment, however in trials two and three, the supportive group seems to have a higher perceived SoA than the suppressive group.

For item B, *"It seemed as if the robot hand had a will of its own"*, the distribution of the supportive group is in all trials more located to the left side as compared to the suppressive group. The medians of both embodiment groups do not change, except for trial two for the supportive group. Note that the distribution of the suppressive group moves successively more towards the left side in trials three and four, where we can see a large overlap of the distributions in the last trial. Overall. it appears that for item B, the supportive is generally more prominent on the left side of the Likert rating than the suppressive embodiment group, however the groups converge in the last trial.

Overall, for both items it appears that the supportive embodiment group has a higher perceived SoA rating than the suppressive group, however the groups converge in the last trial.

Model Comparison
To analyze this data, several Bayesian OLR models have been fit, of which the parameter tables and convergence diagnostics can be found in appendices **??** and C.2.2 respectively. To analyze their fit, predictive posterior plots, center & 95 % HDI bar plots, residual violin plots and LOO-CV scores have been generated, which can be found in appendix C.2.3. Summarizing the main findings, the following can be observed: In the predictive posterior plot it appears that all models have a very similar fit compared to each other and that they predict the observed data well. The similar fit between the models is also reflected in the CLU bar plots of the individual parameters, with exception of the categorical model which estimates slightly smaller upper and lower limits for the `trial` parameter. The residual violin plots show that the errors of the models are also nearly identical, which are mostly centered around 0, ranging between $[-1,+1]$. Overall, the models appear to perform very similar and fit the data relatively well.

To numerically evaluate the models, LOO-CV scores have been calculated for all four models, which can be found below in table 4.4:

Model	**LOO**	**SE**
Cumulative	724.855	23.988
Unequal var. between conditions	724.855	23.988
Unequal var. between items	731.777	6.921
Category-adjacent	737.996	26.805

Table 4.9: Loo-CV and SE scores of OLR models of the SoA survey responses.

As can be seen, the base cumulative OLR model obtains the best LOO-CV score, as compared to the other models. As all of the models appear to fit the data relatively similarly, this model has been chosen based on the LOO-CV score for further analysis.

Final Model Inference
Below in table 4.10, center and 95 % HDI's of the fixed effects of the unequal variance between items model can be found, which are on the probit scale.

Parameter	**Center**	**Lower**	**Upper**
Suppressive	-0.35	-1.05	0.32
Trial 2	0.17	-0.25	0.61
Trial 3	0.40	-0.03	0.87
Trial 4	0.25	-0.19	0.67
Suppressive:Trial 2	-0.27	-0.90	0.33
Suppressive:Trial 3	-0.12	-0.74	0.48
Suppressive:Trial 4	0.14	-0.47	0.75

Table 4.10: CLU table for the fixed effects of posterior distribution of the OLR model for SoA on the probit scale.

For trial one it can be observed that the difference δ between the suppressive and supportive group is $-0.35[-0.93, 0.22]SD$, with P($\delta > 0$) = 0.85 and a moderate evidence ratio of 5.17. For trial two, the observed difference δ between the suppressive and supportive group is $-0.63[-1.21, -0.06]SD$, with P($\delta > 0$) = 0.97 and a strong evidence ratio of 29.26. For trial three, the observed difference δ between the suppressive and supportive group is $-0.48[-1.06, 0.1]SD$, with P($\delta > 0$) = 0.91 and a strong evidence ratio of 10.71. In the last trial, it can be observed that the difference δ between the suppressive and supportive group is $-0.21[-0.78, 0.35]SD$, with P($\delta > 0$) = 0.74 and a anecdotal evidence ratio of 2.83.

Hypothesis	Center	Lower	Upper	Evid. Ratio	Post. Prob.
$\beta_{Suppressive} < 0$	-0.35	-0.93	0.22	5.71	0.85
$\beta_{Suppressive} + \beta_{Suppressive:Trial_2} < 0$	-0.63	-1.21	-0.06	29.26	0.97
$\beta_{Suppressive} + \beta_{Suppressive:Trial_3} < 0$	-0.48	-1.06	0.1	10.71	0.91
$\beta_{Suppressive} + \beta_{Suppressive:Trial_4} < 0$	-0.21	-0.78	0.35	2.83	0.74

Table 4.11: Summary statistics of the hypotheses testing for each of the four trials for the SoA Likert scale.

Overall, it seems that there is a small to medium effect size difference in trial one, which then increases in trial two to a medium to large difference. This difference decreases then in trial three and four, where ultimately only a small difference can be observed. These differences also correspond with the findings from the rain cloud plots, where we could observe large overlap in distributions in the first trial, higher differences in distributions in trials two and three and in the last trial a decrease in differences, indicating that the groups converged in agency.

Note however that value zero is included in all HDI's besides the second trial, thus the null effect being still under the 95 % most probable values. However, based on the evidence ratios and posterior probabilities it seems that there is good evidence that in trial two and three there is indeed a difference between the suppressive and supportive groups.

SoS

Data Exploration
Below in figures 4.5 and 4.6, the rain cloud plots for items A and B of the SoS questions can be found.

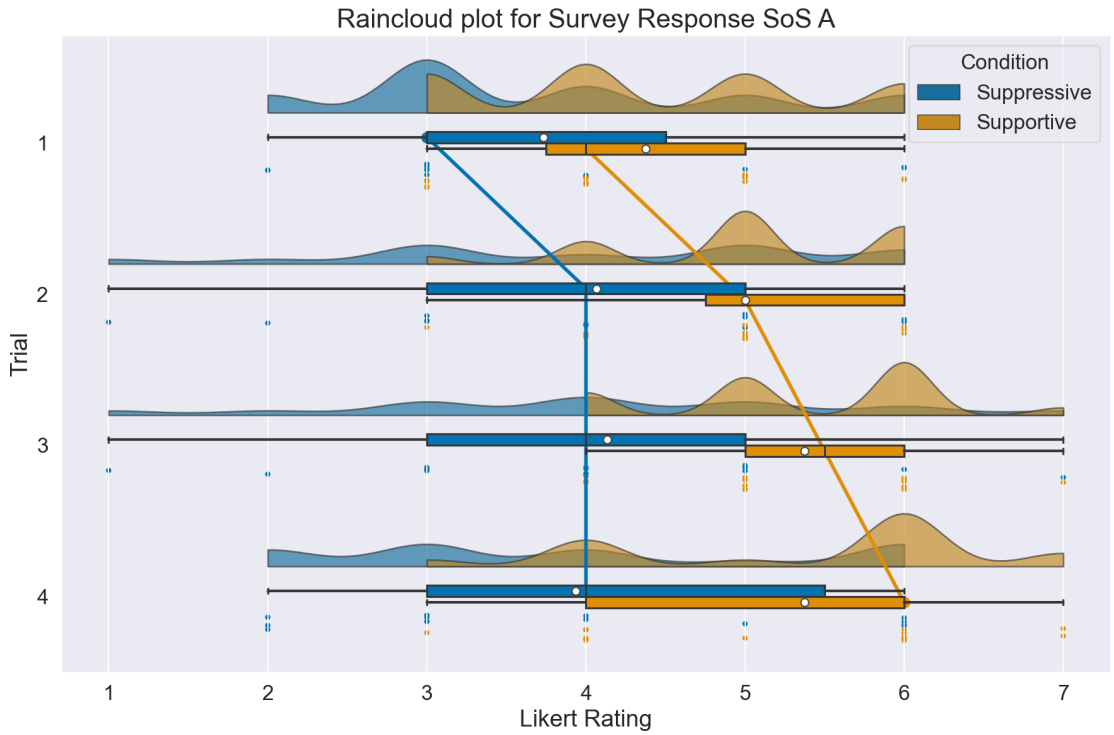


Figure 4.5: Rain cloud plot for SoS item A, "It seemed like my hand was in the location where the robot hand was".

For item A, *"It seemed like my hand was in the location where the robot hand was"*, it can be observed that the median of the supportive group increases trial by trial whereas the median

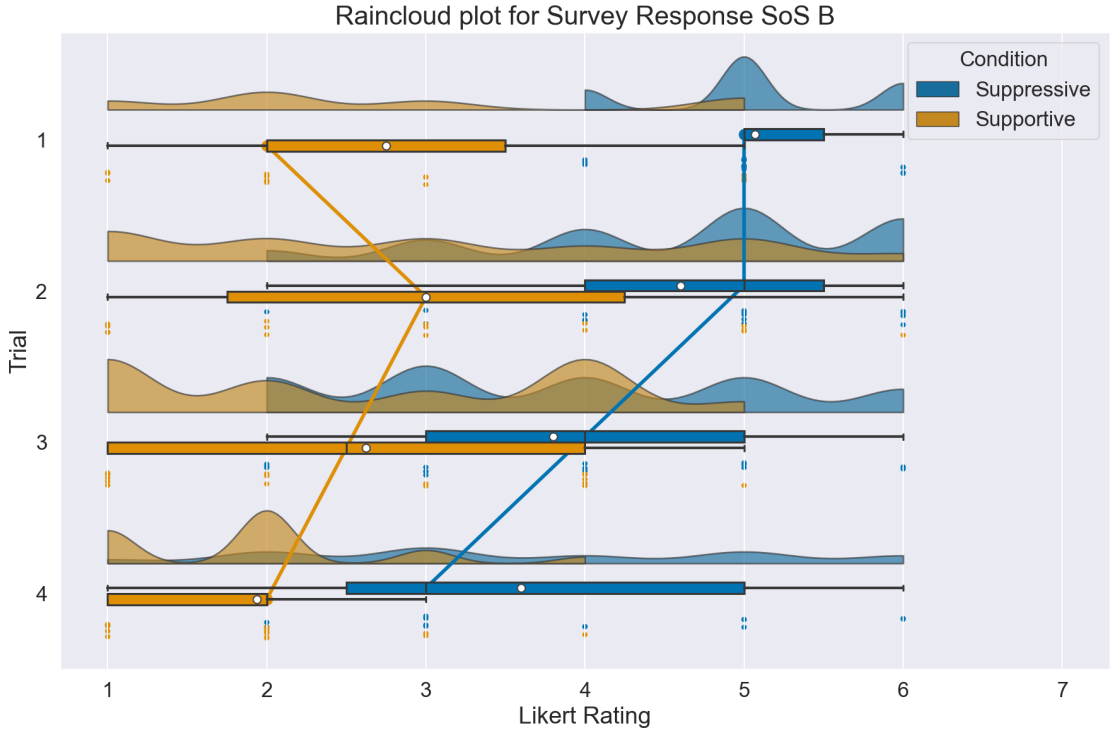


Figure 4.6: Rain cloud plot for SoS item B, "I had difficulties locating my hand in the remote environment".

	Supportive					Suppressive				
Trial	N	Mean	Median	sd	IQR	N	Mean	Median	sd	IQR
1	16	4.38	4	1.09	1.25	15	3.73	3	1.28	1.5
2	16	5	5	0.894	1.25	15	4.07	4	1.53	2
3	16	5.38	5.5	0.885	1	15	4.13	4	1.60	2
4	16	5.38	6	1.20	2	15	3.93	4	1.53	2.5

Table 4.12: Summary statistics per trial and condition for SoS item A.

	Supportive					Suppressive				
Trial	N	Mean	Median	sd	IQR	N	Mean	Median	sd	IQR
1	16	5.25	6	1.48	1.5	15	2.93	3	0.704	0.5
2	16	5	5	1.67	2.5	15	3.4	3	1.24	1.5
3	16	5.38	5.5	1.41	3	15	4.2	4	1.37	2
4	16	6.06	6	0.854	1	15	4.4	5	1.55	2.5

Table 4.13: Summary statistics per trial and condition for SoS item B.

of the suppressive group only increases from the first to the second trial and then remains at the same location. Additionally, a large amount of variance can be found in the suppressive embodiment group, which is especially pronounced in trials two and three where the distribution is very flat. Generally, the supportive group seems to have better SoS ratings than the suppressive group for this Likert item.

For item B, *""I had difficulties locating my hand in the remote environment"*, it can be observed that the suppressive group has very low variance in the first trial, while being located on the far right side of the plot compared to supportive group. The distribution of the suppressive group becomes increasingly flat with every trial, with the mean and median decreasing. The support-

ive group is very flat in the first two trials and becomes then more pronounced towards the left side in trial three and four. Generally, the supportive group has better ratings as compared to the suppressive group.

Overall, for both items there can be clear difference found in the two groups, suggesting that the manipulation of SoS was effective.

Model Comparison
To analyze the observed data, several Bayesian OLR models have been fit, of which the parameter tables and convergence diagnostics can be found in appendices **??** and C.3.2 respectively. To analyze their fit, predictive posterior plots, center & 95 % HDI bar plots, residual violin plots and LOO-CV scores have been generated, which can be found in appendix C.3.3. Summarizing the main findings, the following can be observed: In the predictive posterior plot it is shown that all models have a very similar fit and predict the observed data well. Also in the bar plot it can be observed that all models have a very similar fit in terms of center and upper and lower 95 % HDI's with the exception of the categorical model which seems to have slightly smaller lower and upper limits as well as a lower center estimates for the `trial` parameter. This similarity between the models is also reflected in the residual violin plot, in which the models are nearly identical. The residual distributions are centered around zero and range mostly $[-1, +1]$.

To numerically evaluate the models, LOO-CV scores have been calculated for all four models, which can be found below in table 4.14:

Model	**LOO**	**SE**
Unequal var. between items	747.491	24.09
Cumulative	756.278	22.813
Unequal var. between conditions	756.278	22.813
Category-adjacent	768.373	25.77

Table 4.14: Loo-CV and SE scores of OLR models of the SoS survey responses.

As can be seen, the OLR model that models unequal variance between the two Likert items obtained the best LOO-CV score, as compared to the other models. As all of the models appear to fit the data relatively similarly, this model has been chosen based on the LOO-CV score for further analysis.

Final Model Inference
Below in table 4.15, center and 95 % HDI's of the fixed effects of the unequal variance between items model can be found, which are on the probit scale.

Parameter	**Center**	**Lower**	**Upper**
Suppressive	–1.16	-1.81	-0.49
Trial 2	0.13	-0.30	0.54
Trial 3	0.43	0.00	0.88
Trial 4	0.75	0.31	1.17
Suppressive:Trial 2	0.14	-0.46	0.76
Suppressive:Trial 3	0.21	-0.39	0.83
Suppressive:Trial 4	-0.09	-0.72	0.52

Table 4.15: CLU table for the fixed effects of posterior distribution of the OLR model for SoS on the probit scale.

For all trials, we can observe a posterior probability of 1 and extreme evidence ratios in favor of the respective hypotheses. Moreover, we can observe large differences in effect sizes for all trials. Overall, there is significant evidence that there is a large difference in SoS ratings between the embodiment groups.

Hypothesis	Center	Lower	Upper	Evid. Ratio	Post. Prob.
$\beta_{Suppressive} < 0$	-1.15	-1.71	-0.6	2665.67	1
$\beta_{Suppressive} + \beta_{Suppressive:Trial_2} < 0$	-1.01	-1.57	-0.47	887.89	1
$\beta_{Suppressive} + \beta_{Suppressive:Trial_3} < 0$	-0.94	-1.49	-0.4	532.33	1
$\beta_{Suppressive} + \beta_{Suppressive:Trial_4} < 0$	-1.25	-1.8	-0.71	2999	1

Table 4.16: Summary statistics of the hypotheses testing for each of the four trials for the SoS Likert scale.

4.1.2 Interview

Semi structured interviews has been conducted to inquire about the perceived SoO, SoA, & SoS level of the participants. The interviews have been audio recorded and subsequently summarized. The transcribed data has then been analyzed via IPA [48], similar as in the study by [49]. Next to themes of perceived SoO, SoA and SoS, the following reoccurring themes emerged from the IPA analysis: "Interplay between SoO and SoO", "Expectations and adjustments to constant & sudden mismatches", and "Effects of delay on SoO and SoA". These and a number of corresponding sub-themes are being discussed below.

Perceived SoO

Regarding SoO, 10 participants of the supportive group stated that the robot hand felt like their own hand whereas only two participants from the suppressive group reported about ownership of the hand.

> "I never felt like I was operating a robot but it felt like my actual hand. So yes, it felt pretty close to being my own actual hand." (P21, Supportive)

> "For the hand it was pretty good, because it did what I asked it to do." (P23, Suppressive)

In addition, one participant of the supportive group actively said the hand did not feel like their own hand whereas five participants of the suppressive group stated that the hand did not feel like their own.

> "It definitely felt like I was puppeteering something rather than the arm or the hand feeling as part of my own body." (P. 13 , Supportive)

> "Didn't feel like I had a lot of ownership of the hand, had to squeeze my fingers really hard to grab the peg so it didn't feel like my own hand." (P. 24, Suppressive)

When being asked about the robot arm instead of the robot hand, eight participants of the supportive group and three participants of the suppressive embodiment group reported a feeling of ownership. Moreover, four participants of each embodiment group stated felt the opposite way, feeling no ownership over the arm.

> "So I wasn't really thinking about this, I was using a robot arm just as my own arm, until the last trial when the arm was doing weird stuff, that made it obvious I am attached to something. Especially by the 3rd trial it felt really like my own arm and hand." (P. 28, Supportive)

> "First it didn't feel like my arm in the position judgement task, but once I was moving it myself it did start to feel like my own arm and hand, although the movements

> were a little bit slower than what I would expect in your life. So it still feels like it's my own hand, but as if there are some modifications to my own hand." (P. 6, Suppressive)

> "The arm I had troubles with because it didn't really track the angle of my elbow, so sometimes the robot arm was different than my own arm pose and it was also bumping into stuff sometimes which was a little bit frustrating. Sometimes my wrist didn't turn all the way I wanted it to. Multiple times the peg was at an angle and but no matter how much I turned my wrist it didn't seem I could actually turn the wrist of the robot arm, that was pretty frustrating and that made me feel disconnected." (P. 26, Supportive)

> "When I did things slowly I felt like it was my hand. But when I moved very quickly or moved very far, I quickly lost that sensation, because there is a slight delay, so I think you have to do things very slowly or you will mess it up. But I just did everything slowly and could compensate in for the delay in mind. But for moving the arm, it felt like the arm was way bigger than me, I felt like I moved my arm way more than the robot arm actually moved which caused kind of a disconnect because I felt like should be easily able to reach one of the pegs but I couldn't." (P29, Supportive)

Additionally, there were some aspects relating to the vision of the participant that impeded SoO. For instance, one participant of the supportive group noted that the camera perspective was not feeling like the natural human 1PP, suggesting that the camera perspective was perhaps too high in relation to the position of the arm:

> "If I bring the hand to my face, it's not going actually in front of the camera but more to my "chest", that also might add to disconnect of ownership." (P. 22, Suppressive)

Similarly, one participant also mentioned that he had less feeling of ownership because of lacking depth perception:

> "I think is important that I missed the depth. To feel like if it's your own hand, you have to have a good orientation and depth." (P. 15, Suppressive)

While statements are are rather anecdotal, they suggest a link between the SoO and SoS.

Perceived SoA

For the SoA, 11 participants of the supportive group stated that the robot hand moved as they wanted it to whereas six participants from the suppressive group also felt this way. One participant of the supportive and four participants of the suppressive group reported that the hand did not move as they wanted it to.

> "The hands, I think it was really intuitive. They did mostly what I wanted them to do." (P10. Supportive)

> "Had the feeling like it was a glove without fingers. I could not control fingers separately, it seemed to work only properly if I closed the entire glove. During the experiment I got more used to this, I noticed that it didn't really matter if I moved a finger a little bit, the hand would not do anything then. At first I felt like this is working better, but in the 3rd and 4th trial I felt like, no, it doesn't work better, I am just working better with it, the system still sucks." (P 20, Suppressive)

> "I found it hard to use the hand, because I feel that the hand didn't really do what I wanted it to do. It wasn't really diligent when I wanted to close the hands just a little bit. Not always I had the feeling that it will listen to the way that I closed hands. I had to exaggerate a lot." (P. 15, Suppressive)

When asked about the haptic feedback in the hand, 12 out 16 participants stated that the haptic feedback helped them to better grasp the pegs. Specifically, the common theme here was knowing how much force is needed to get a proper grip.

> "Yes, the feedback helped me a lot. Because with the resolution of the HMD not that being detailed, there's a difference between holding the peg with a firm grab or just having the hand loosely around the peg." (P. 9, Supportive)

> "I felt it was really that it would have the shape of the object." (P. 10, Supportive)

> "The haptic feedback really gave me the sense that okay, fine, I'm actually touching it. Even the first bit, which gives me confidence that I can go ahead and completely grasp it with the haptic feedback. So that was something that I was constantly trying to feel on my hands." (P. 12, Supportive)

> "It didn't really correspond to my hand I feel like. It felt really like I had to squeeze very tight to make the robot hand grab the peg. But in reality I shouldn't fully close my fingers for that." (P. 24, Suppressive)

Interestingly, also some participants of the suppressive group reported that they felt haptic feedback in their hand even though there was no haptic feedback activated:

> "Sometimes I felt forces in my hand when was trying to grasp things, it seemed to be pushing back a bit." (P. 30, Suppressive)

For the robot arm, four participants of the supportive group mentioned that they had good control over the robot arm whereas only participant of the suppressive group also mentioned this.

> "The system was very intuitive to use, did not feel like I needed to adjust to anything. As soon as you turn on the camera and turn on the machine, everything is kind of self explanatory for me. I do know where my arm is supposed to be as well, as like, really well. I can instantly tell how to control the thing. Like, I don't have any issues piloting that thing." (P. 13, Supportive)

On the other side, four participants of the supportive group and six participants of the suppressive group reported that they had lacking control over the robot arm. Reasons for reduced agency due to the arm include not being able to control the elbow or wrists of the robot hand, sudden movements of the robot arm for certain movements, the arm having more joints than a human arm, allowing unnatural arm poses, and the general damping that is present in the haptic feedback of the Virtuose:

> "I did not feel like I could control the elbow, it was really annoying when I wanted to position my arm in another way to better grasp the pegs but I couldn't manage to do so. Sometimes I wanted to turn my wrist but you cant and you have to turn your whole arm." (P. 27, Supportive)

> "I was surprised that sometimes moving it requires quite a lot of energy. Sometimes I really needed to push. I was in doubt whether it was because part of the equipment was pushing against something or if it was because I just needed to, like push harder (against the haptic feedback)." (P. 18, Supportive)

> "Most of the time the arm was doing what I wanted it to do, but there were a few moments when it didn't because the arm has the different joints and the arm moves differently than a normal human would. So I could move the robot arm differently than I could move my own arm. The first time it was really strange, was really confused at first. On the other hand it was also helpful because you had more degrees of freedom and range than a normal arm would have. So, most of the time the arm did what I wanted it to do, but either slower or in a different way, so that the arm was in a different position than I thought it would be." (P. 19, Supportive)

> "It takes some time until you get where you want to go, and then you overshoot and have to go back. I thought more like I was sort of fighting sometimes against the force feedback than with the force feedback." (P. 1, Suppressive)

> "I had a lot of time the feeling it had its own will, it was hard to control because there was a lot of force resisting for certain movements, which made it difficult to maneuver between the pegs." (P 14, Suppressive)

> "Because of the resistance and the delay, I had to move my hand in sections to not overshoot, and then check again. So you go forward, like 10 centimeters, wait, go forwards, and then adjust." (P. 15, Suppressive)

Perceived SoS

Regarding the SoS, two participants of the supportive group and 14 participants of the suppressive group reported that depth perception was an issue for them and they had troubles locating their hand in the remote space.

> "My spatial awareness was really good, depth perception was also good." (P. 9, Supportive)

> "My spatial awareness was pretty good. I think it would mostly be limited by the feeling that there still a small lack of depth perception. Apart from that, I felt like I had a good idea of where I was and how we should get from A to B." (P 17, Supportive)

> "It was definitely more difficult because of the 2D view, I used shadows to determine where I was, otherwise I could not make out the depth because the colours of the mounts were black and orange which was so bright that I could not see any depth in it. And also with the disconnect with your arm it was a little bit harder to guess the depth, because I had to move further with my own arm than the robot arm." (P. 29, Supportive)

Some participants of the suppressive group mentioned that the perspective was decreasing their spatial awareness:

> "I think I would have performed way better if it was just like the same perspective as I was sitting, just like with my own eyes. It also gives you more of a realistic way of controlling your own arm and moving stuff around, from your own perspective." (P. 3, Suppressive)

> "I was really confused when I saw the top view of the environment, because I didn't expect it. It made me feel disoriented and confused." (P. 6, Suppressive)

> "I think I got the feeling for the X and Y direction pretty quickly but especially the Z-axis was very hard for me to get a sense of. I think it was because of the top down view of the camera. I didn't really have a sense of depth." (P. 30, Suppressive)

Others also reported the urge to look around:

> "I also had the natural urge to move my head around but that didn't work of course. I think it was doable but I do think it would benefit from either being 3D or moving your head around." (P. 29, Supportive)

> "I feel like the HMD gives you a false sense of spatial awareness, because in real life I would have moved my head to look around the pegs but here I cannot do that. I feel like that made it more difficult for moving my hand further away from here because I could not adjust my vision to it." (P. 22, Supportive)

Interplay between SoO and SoA

There seems to be a connection between the SoO and the SoA, affecting each other. For instance,here are some cases were a good SoA lead to a good SoO:

> "If I wouldn't have been able to place myself there as if it were my own arm, I don't think I would have been able to control it as smoothly as I did. I think the feeling of 'This is my arm' would be lower then." (P. 17, Supportive)

> "For the hand it (SoO) was pretty good, because it did what I asked it to do." (P23, Suppressive)

> "When I tried to lift the peg from the top, that's when I really felt like I was you know, controlling the arm directly like my own." (P. 17, Supportive)

> "I got a lot better, it felt more like my own arm and that made it feel more natural and more easier." (P. 27, Supportive)

Similarly, when the robot is not behaving as intended and agency is reduced, it reduced SoO:

> "It didn't feel like it was attached to me, sometimes I stopped thinking about that there was a second hand controlling the robot hand, but that was mostly when everything was working out, when it did what I wanted it to do. But every time the hand did not do what I wanted it to do it became more apparent that this is not my hand." (P. 20, Suppressive)

> "The arm I had troubles with because it didn't really track the angle of my elbow, so sometimes the robot arm was different than my own arm pose and it was also bumping into stuff sometimes which was a little bit frustrating. Sometimes my wrist didn't turn all the way I wanted it to. Multiple times the peg was at an angle and but no matter how much I turned my wrist it didn't seem I could actually turn the wrist of the robot arm, that was pretty frustrating and that made me feel disconnected." (P. 26, Supportive)

> "It didn't feel like my own hand because I didn't really do things that I wanted to do." (P. 15, Suppressive)

> "The arm gave force feedback, but here was nothing to push against, it just felt weird and that kind of got me out of it (the SoO illusion)." (P. 2, Suppressive)

In another case, a participant was struggling with controlling the wrist and elbow movements of the robot, and was adjusting with both hands the orientation of the Virtuose end-effector to create the desired movements on the remote side:

> "For the arm I felt a lot less ownership than the hand, because I used both my hands to turn the wrist and elbow and that really didn't feel like my own arm." (P. 23, Suppressive)

> "It became more part of my body during the different trials. But in the beginning, I was really figuring out like, it took me a while to find out that moving your elbow also has a great effect on the robot, because I was mainly focused on the hands and all the static movements So that realization came over time and then it felt more naturally afterwards." (P. 8, Supportive)

> "In the first trial it (SoO) was not really smooth. But as I was progressing, I think at the time it was third or fourth trial, it was way more smoother. I knew that I was easily located. And my movements were much more fluid. So yes, at that point of time, I did feel like it was very much in synchrony and it felt like my own body." (P.12, Supportive)

However, some participants reported that while they had a good SoA, they did not feel a good SoO but rather that they were controlling a character in a video game:

> "I think especially due to the delay that hand felt a bit more separate from my body, it feels very much that I control the hand, so I have a full sense of control, but I have a lower sense of ownership. It's a bit playing a video game where you have full control over a character but you know that you are not the character." (P. 22, Supportive)

> "Generally I had good control, but it felt more like third person game, than my own body." (P 31, Suppressive)

This shows that while a good SoA might help increasing the SoO, SoA and SoO can be perceived separately. Related to this, some participants also mentioned that the robot arm/hand felt more like a tool or controller to them, rather than their own body part:

> "I didn't think of controlling my hand but that my hand was the controller. So I wasn't thinking about what do my own hand movements do but how can I control the movement of the other hand. I used my hand more as a tool to control the other hand." (P. 22, Supportive)

> "The hand was more like a trigger for me, to open and close it, instead of a feeling of a real hand. Would say it felt rather not like my own body, it was more like playing a game and trying to get it to do something, that didn't really feel like it was my own hand." (P. 3, Suppressive)

> "I felt like my movements were way more extreme that the hand's, I had to exaggerate. I was a bit frustrated because the hand was not doing what I was doing in the real world I ended up using my hand more as a remote, as a tool." (P. 14, Suppressive)

Expectations and adjustments to constant & sudden mismatches

Generally, it appears that participants could adjust to constant mismatches. For instanc,e some participants were at the beginning confused about how to move the fingers of the hand:

> "Once I figured out I should pay more attention to the grasp than the individual fingers, that worked consistently pretty well. Overall, I could consistently just use it as if it was my own hand." (P 26, Supportive)

Another constant mismatch that participants got used to, was the delay that was present in the overall system:

> "The delay wasn't that much, only a few milliseconds. You got used to it." (P 28, Supportive)

However, although depth perception was constant throughout the experiment, participants of the suppressive embodiment group could not adjust to it. Instead, they tried to deal with this issue by using monocular cues to infer about depth:

> "I didn't feel like that got better, I just got more of an idea how to deal with the depth. For instance the little finger that goes on top of the peg rather than around it, then I knew it just had to move the hand a little bit more down to grasp it. So I tried to use some tricks to work around the depth perception issue, here using the lowest finger as indicator." (P. 31, Suppressive)

> "Didn't really adjust to the depth perception but changed my strategies to deal with it." (P. 24, Suppressive)

> "I compensated for it (lacking depth perception) by using angled pegs to better see the holes, so I dealt with it but it didn't change over time." (P26, Supportive)

> "I tried to use shadows to estimate how close the pegs were, but the was very soft so that was also not that useful." (P. 7, Suppressive)

> "You have to compensate that with some orientation, maybe shadow or size, but that was difficult because there weren't a lot of orientation things present in the scene to verify on which height you are. So I had to determine height via the haptic feedback with touching the mounting plate but that was really sometimes hard to notice where you touch the mounting plate." (P. 15, Suppressive)

> "I think the delay was still difficult to adjust to til the end of the game. But at in the 2nd, 3rd trial, you know that there is delay, so you can at least anticipate it." (P.16, Suppressive)

Finally, sudden mismatches would typically disrupt the illusion of SoE, e.g. when a participants needs to be clutched out and in again in order to extend the workspace.

> "Especially when you are outside of the workspace and need to be clutched in again, you loose the feeling of ownership" (P. 1, Suppressive)

> "Just sometimes the remote space is larger than your own workspace and you can't reach as far, that really made it less intuitive, felt like a cheat code the clutching in and out" (P. 27, Supportive)

> "It was really annoying, the feeling that you should be able to reach the upper most left peg but that you really can't and that I need to ask you to clutch me in and out" (P. 31, Suppressive)

Effects of delay on SoO and SoA

Both groups mentioned that for precise movements they felt lower SoO due to the delay present in the overall system:

> "I think I have a low SoO because of the delay. I feel like have to wait for my hand to see if it has reacted and that disconnects my body from the hand." (P. 22, Supportive)

> "Because of the delay you immediately get the feeling this is not my own hand." (P. 15, Suppressive)

However, some participants said that even though the delay reduces the SoO, it still feels like their own body part:

> "Well, this delay kind of makes the sense less real, that it's your own hand. But it didn't feel like a separate hand because it did save movements as I did." (P. 16, Suppressive)

For others, the delay was easy to adjust to:

> "The delay wasn't that much, only a few milliseconds. You got used to it." (P 28, Supportive)

4.1.3 Proprioceptive Drift

Proprioceptive drift has been measured as the absolute difference in position between the initial and end position of the position judgement task, as described in section 3.4.1. This has been done before the exploration and the peg & hole task (trial one) and afterwards (trial two).

Data Exploration
Below, the rain cloud plots of the resulting proprioceptive drift measurements can be seen for the X, Y and Z axis respectively:

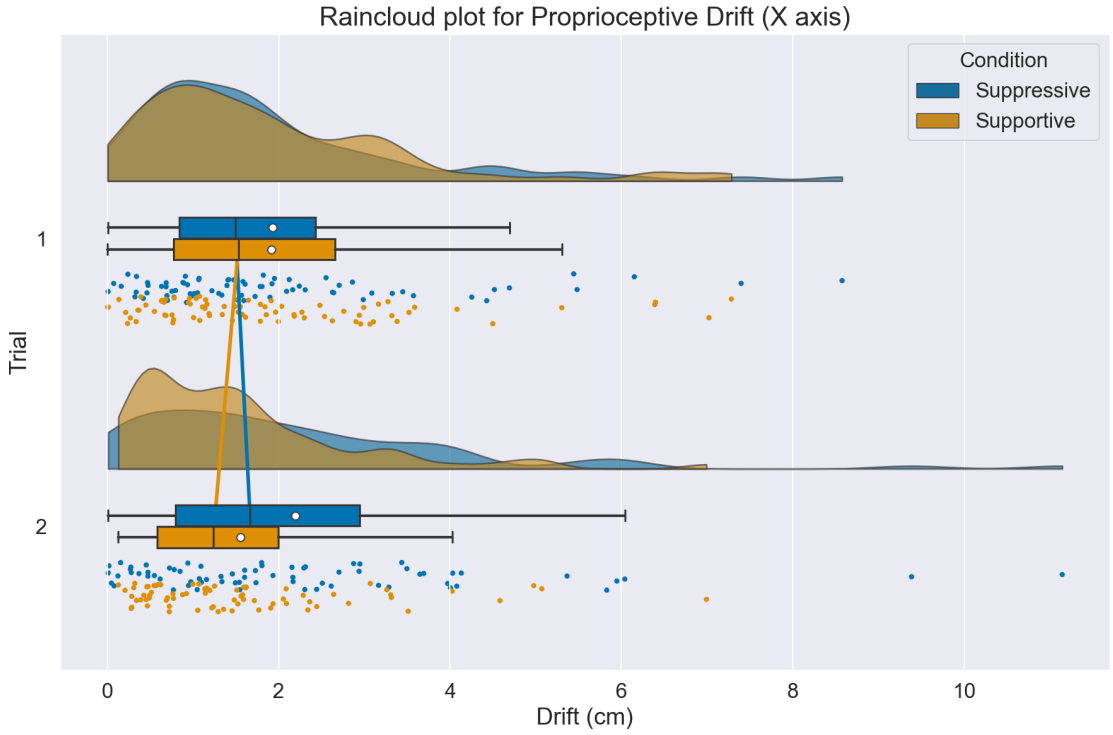


Figure 4.7: Rain cloud plot for proprioceptive drift on the X axis (Forwards).

		Supportive					Suppressive				
Axis	Trial	N	Mean	Median	sd	IQR	N	Mean	Median	sd	IQR
X	1	80	1.91	1.53	1.58	1.89	75	1.93	1.49	1.69	1.59
	2	75	1.55	1.24	1.30	1.42	70	2.20	1.67	2.06	2.15
y	1	80	1.89	1.31	1.62	1.65	75	1.60	1.27	1.34	1.81
	2	75	1.54	1.26	1.25	1.42	70	1.70	1.34	1.30	1.52
Z	1	80	2.39	1.80	1.90	2.20	75	2.90	2.48	2.04	2.78
	2	75	2.65	1.84	2.36	2.81	70	3.32	2.31	3.24	3.39

Table 4.17: Summary statistics of the proprioceptive drift measurements.

Summarizing the main findings from the plots, the following can be observed: As expected, all distributions are skewed to the right, caused by taking the absolute value of the position differences. When inspecting the first trial (before any visuo-proprioceptive stimulation), the suppressive and supportive embodiment group have nearly the same median on the X and Y axis and a difference in medians on the Z axis for the first trial, as expected. Here, the difference on the Z axis was expected based on the the visual manipulations of perspective and depth cues to create a different levels of SoS. In the second trial, the supportive embodiment group increases in drift towards the hand for both the X and Y axis, and stays the same for the Z axis.

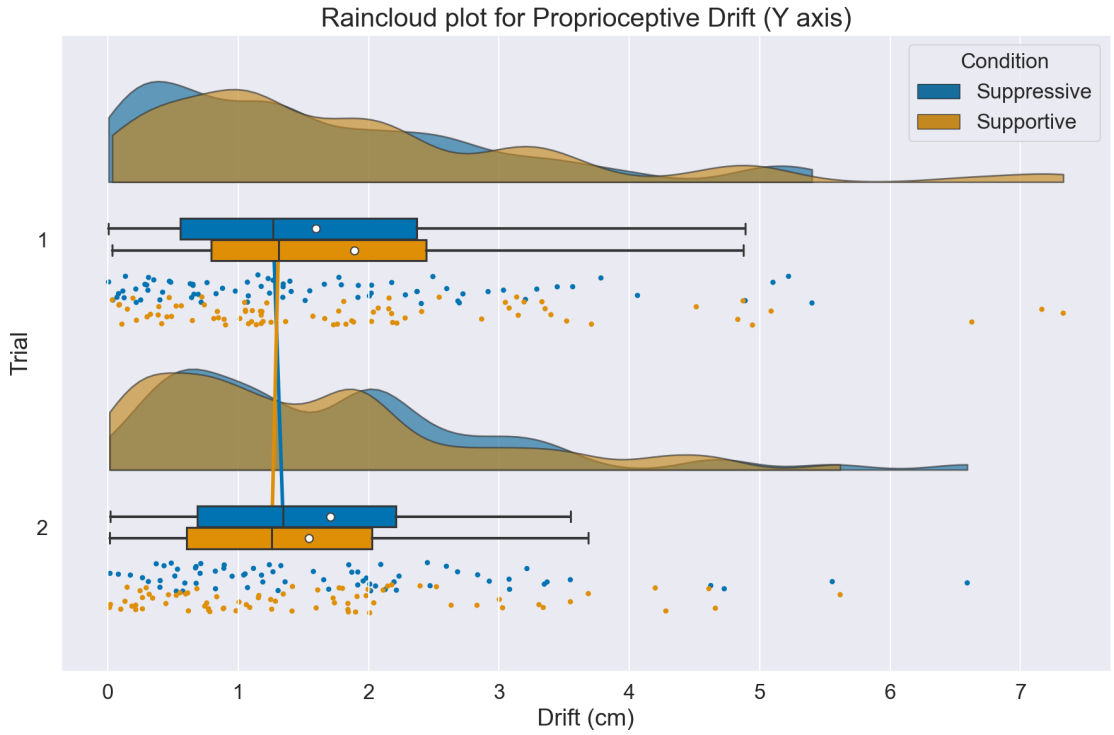


Figure 4.8: Rain cloud plot for proprioceptive drift on the Y axis (Left).

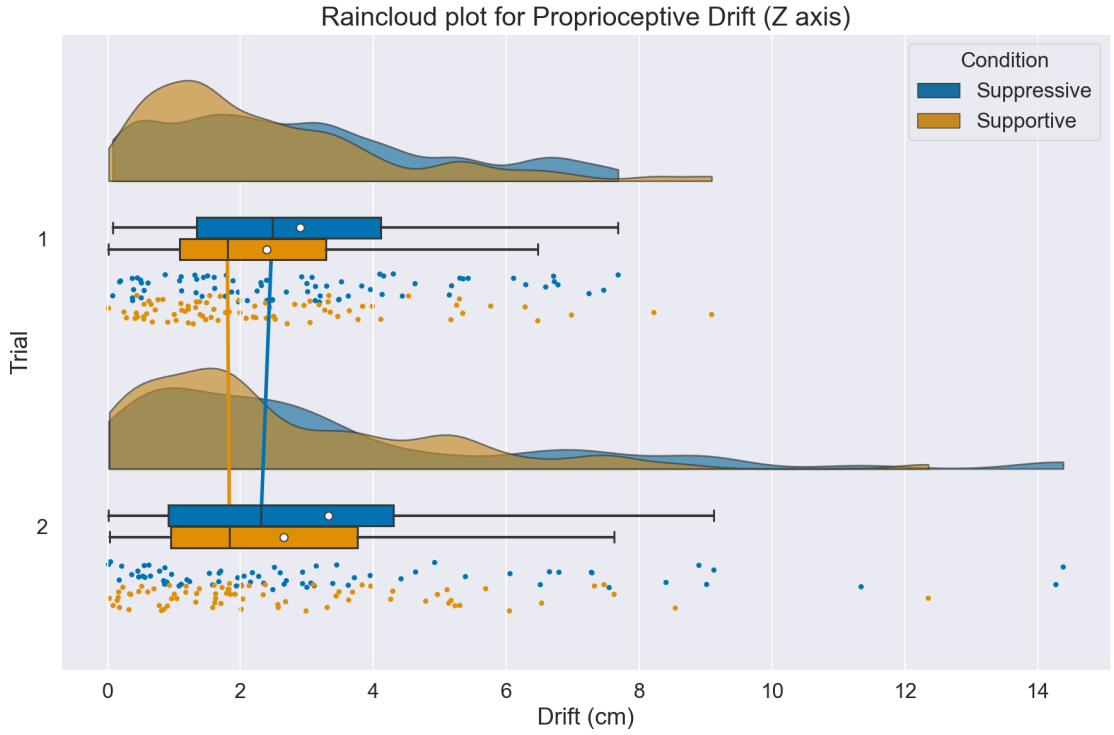


Figure 4.9: Rain cloud plot for proprioceptive drift on the Z axis (Upwards).

The suppressive group on the other hand decreases in drift on the X and Y axis, and increases in drift on the Z axis. Note that large overlaps between the distributions of the embodiment groups can be observed for all axes in both trials.

Model Comparison
To analyze this data, several Bayesian OLR models have been fit, of which the parameter tables and convergence diagnostics can be found in appendices C.4, C.5 and C.6 for the different axes respectively. To analyze their fit, predictive posterior plots, center & 95 % HDI bar plots, residual violin plots and LOO-CV scores have been generated for each axis. Summarizing the main findings, the following can be observed:

For all three axes, very similar looking posterior predictive check plots have been obtained, in which the Gaussian model fails to capture the skewness of the distribution of the observed data, whereas the Gamma and exponentially modified Gaussian models provide a comparatively similar, good fit to the observed data. When comparing the plots to each other, more variance can be observed on the Z axis as compared to the X and Y axis.

In the bar plots, it can be observed that for all three axes, the Gaussian and Gamma models produce very similar looking center and upper and lower HDI's, whereas the exponentially modified Gaussian model predicts different center estimates as well as smaller HDI's limits, indicating more certainty in its predictions.

To investigate this difference in certainty, residual violin plots have been generated for all three models on all three axes. In all of these plots it can be observed that the residual distributions look very similar. This indicates that the difference in center estimates in the bar plots is of rather smaller magnitude compared to the the observed data. Additionally it can be observed that the Gaussian model is consistently shifted lower on the Y axis of the residual plots, reflecting its inability to capture the lower limit of the data.

To numerically determine the best fitting model, LOO-CV scores of all three models have been calculated for each of the axes, as shown in tables 4.18. For all three axes, the Gamma model

Axis	**Model**	**LOO**	**SE**
X	Gamma	973.943	32.092
	Exp. Gaussian	987.914	32.699
	Gaussian	1167.551	49.990
Y	Gamma	910.997	28.434
	Exp. Gaussian	919.097	29.515
	Gaussian	1055.249	33.993
Z	Gamma	1210.646	30.013
	Exp. Gaussian	1225.235	31.317
	Gaussian	1389.133	40.963

Table 4.18: Loo-CV and SE scores of the proprioceptive drift models.

obtained the lowest LOO-CV score. As all models produced similar results in the graphical analysis of the models (i.e. predictive posterior checks, bar plots and violin plots), the Gamma model has been chosen for further analysis for all three proprioceptive drift axes.

Final Model Inference
Below in table 4.19, the CLU table of the fixed effects of the Gamma model for the X axis can be found, where the intercept is the supportive group in trial one. Using the chosen model, parameter estimation based hypothesis testing has been conducted for each of the two trials. It was expected that the suppressive embodiment group has less proprioceptive drift than the supportive embodiment group. Center, upper and lower estimates, evidence ratio and posterior probabilities of the differences in proprioceptive drift between the supportive and the suppressive group can be found in table 4.20.

For trial one, a difference δ of 0.04 [-0.38, 0.88] centimeters can be observed with P($\delta > 0$) = 0.54 and an anecdotal evidence ratio of 1.2. For trial two, a difference δ of 0.64 [0.05, 4.54] cen-

Parameter	Center	Lower	Upper
Intercept	0.64	0.43	0.85
Suppressive	0.02	-0.28	0.32
Trial 2	-0.21	-0.48	0.06
Suppressive:Trial 2	0.33	-0.06	0.72

Table 4.19: CLU table for the fixed effects of posterior distribution of the proprioceptive drift models (X axis) on the log scale.

Hypothesis	Center	Lower	Upper	Evid. Ratio	Post. Prob.
$\beta_{Suppressive} > 0$	0.04	-0.38	0.88	1.2	0.54
$\beta_{Suppressive} + \beta_{Suppressive:Trial_2} > 0$	0.64	0.05	4.54	54.94	0.98

Table 4.20: Summary statistics of the hypotheses testing for each of the two trials for the proprioceptive drift measurement (X axis). Differences have been exponentiated from the log scale to the original scale.

timeters can be observed with P($\delta > 0$) = 0.93 and a very strong evidence ratio of 54.94. These estimates correspond with the findings from the rain cloud plot in figure 4.7. As expected, there is nearly no difference in proprioceptive drift before any visuo-motor stimulation (trial one) and difference in drift in the second trial.

Below in table 4.21, the CLU table for the proprioceptive drift on the Y axis can be found. The corresponding hypothesis table can be found in table 4.22.

Parameter	Center	Lower	Upper
Intercept	0.63	0.43	0.84
Suppressive	-0.16	-0.46	0.15
Trial 2	-0.2	-0.47	0.08
Suppressive:Trial 2	0.26	-0.14	0.66

Table 4.21: CLU table for the fixed effects of posterior distribution of the proprioceptive drift models (Y axis) on the log scale.

Hypothesis	Center	Lower	Upper	Evid. Ratio	Post. Prob.
$\beta_{Suppressive} > 0$	-0.27	-0.57	0.37	0.18	0.15
$\beta_{Suppressive} + \beta_{Suppressive:Trial_2} > 0$	0.16	-0.16	3.13	2.79	0.74

Table 4.22: Summary statistics of the hypotheses testing for each of the two trials for the proprioceptive drift measurement (Y axis). Differences have been exponentiated from the log scale to the original scale.

For trial one, a difference δ of -0.27 [-0.57, 0.37] centimeters can be observed with P($\delta > 0$) = 0.15 and an anecdotal evidence ratio of 0.15 (in favor of the null hypothesis). For trial two, a difference δ of 0.16 [-0.16, 3.13] centimeters can be observed with P($\delta > 0$) = 0.74 and an anecdotal evidence ratio of 2.79. These estimates correspond with the findings from the rain cloud plot in figure 4.8. As expected, there is nearly no difference in proprioceptive drift before any visuo-motor stimulation (trial one). Against our expectation, there seems to be barely any evidence in trial two for a difference in proprioceptive drift between the embodiment groups.

Below in table 4.23, the CLU table for the proprioceptive drift on the Z axis can be found. The corresponding hypothesis table can be found in table 4.24.

For trial one, a difference δ of 0.52 [-0.18, 1.91] centimeters can be observed with P($\delta > 0$) = 0.91 and a strong evidence ratio of 10.16. For trial two, a difference δ of 0.68 [-0.61, 6.6] centimeters can be observed with P($\delta > 0$) = 0.93 and a strong evidence ratio of 14.09. Additionally,

Parameter	Center	Lower	Upper
Intercept	0.87	0.66	1.08
Suppressive	0.2	-0.1	0.5
Trial 2	0.1	-0.18	0.38
Suppressive:Trial 2	0.03	-0.37	0.43

Table 4.23: CLU table for the fixed effects of posterior distribution of the proprioceptive drift models (Z axis) on the log scale.

Hypothesis	Center	Lower	Upper	Evid. Ratio	Post. Prob.
$\beta_{Suppressive} > 0$	0.52	-0.18	1.91	10.16	0.91
$\beta_{Suppressive:Trial_2} > 0$	0.07	-0.60	1.58	1.29	0.56
$\beta_{Suppressive} + \beta_{Suppressive:Trial_2} > 0$	0.68	-0.61	6.60	14.09	0.93

Table 4.24: Summary statistics of the hypotheses testing for each of the two trials for the proprioceptive drift measurement (Z axis). Additionally, only the interaction effect between trial and embodiment condition is looked at, due to the expected bias of the visual manipulations between embodiment groups (camera perspective / depth perception). Differences have been exponentiated from the log scale to the original scale.

the interaction effect between trial and embodiment condition, $\beta_{Suppressive:Trial_2}$, has been investigated to check whether the change between trial one and two is different between the embodiment conditions. This has been done due to the potential bias of the camera perspective and the manipulation of depth perception. Considering only $\beta_{Suppressive:Trial_2}$ for the hypothesis test yielded a difference δ of 0.07 [-0.6, 1.58] centimeters with P($\delta > 0$) = 0.56 and an anecdotal evidence ratio of 1.29. These estimates correspond with the findings from the rain cloud plot in figure 4.9. As expected, there is a difference in proprioceptive drift along the Z axis in the first trial, likely caused by the visual bias between the embodiment groups. When looking at $\beta_{Suppressive:Trial_2}$, thus not including the potential offset caused by visual bias, there seems to be no difference in the change of proprioceptive drift between trials, when comparing the two embodiment groups. As there is not interaction effect in change, one could consider the difference in trial one as an offset caused by visual bias as induced via the chosen manipulations for SoS.

4.2 Task Performance

To assess the task performance and learning behavior of the participants, effectiveness (number of successfully inserted pegs) and efficiency (time needed per successfully inserted peg) have been analyzed.

Visualizing the task performance over trials (see appendix C.7.2 and C.7.1) shows that no significant learning behavior can be observed, neither on the population, nor on the participant level. In context of the tweak finder model this means that either the amplitude (pool of undiscovered tweaks) is very small or that the learning rate (catch rate) is so low, that no visible learning can be observed [129]. For completeness sake, both variables have been fit with nonlinear learning models, of which the estimated learning curves can also be found in appendix C.7.2 and C.7.1. See section 5.2 for a more detailed discussion on the lack of learning effects.

For the remainder analysis of task performance, a standard GLM as in the previous sections has been used. In this model effectiveness and efficiency has been evaluated with interchangeable trials which means that we can only make inferences about absolute task performance per trial and that we cannot quantify individual parameters of learning as in the TFM paradigm.

4.2.1 Effectiveness

Effectiveness has been measured via the number of successfully inserted pegs per trial.

Data Exploration
To explore the data of this effectiveness measure, a rain cloud plot has been created, which can be found in figure 4.10:

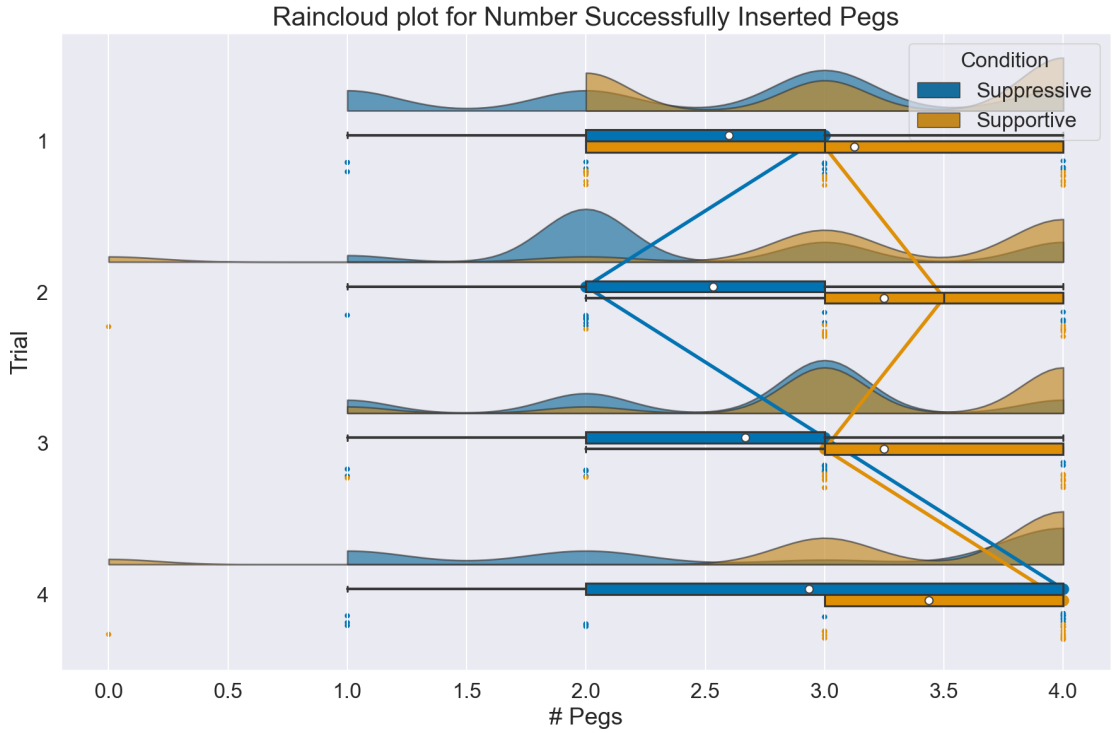


Figure 4.10: Rain cloud plot for for number of successfully inserted pegs.

	Supportive					Suppressive				
Trial	N	Mean	Median	sd	IQR	N	Mean	Median	sd	IQR
1	16	3.12	3	0.885	2	12	3	3	0.739	0.5
2	15	3.47	4	0.640	1	14	2.64	2	0.842	1
3	15	3.4	3	0.632	1	13	2.92	3	0.641	0
4	15	3.67	4	0.488	1	12	3.42	4	0.900	1.25

Table 4.25: Summary statistics per trial and condition for number of successfully inserted pegs.

For the supportive embodiment group it can be observed that the number of pegs inserted improves slightly over the course of the trials as the mean and median increases in every trial (except for the third trial). Additionally, it can observed that the variance decreases over trials when inspecting the spread of the violin plot as well as the whiskers of the box plot. The suppressive group on the other hand has first a drop in performance in the second trial which can be observed in the decreased median and mode. In the succeeding trials however, the performance of the suppressive group improves progressively as the mode, median and mean increase with every trial. Overall, both groups improve on average when comparing the initial trial to the last trials. Comparing the embodiment groups with each other, it appears that the supportive group performance consistently better and is also characterized by less variance as compared to the suppressive group.

Model Comparison
To analyze the observed data, a Poisson, a Binomial and a Beta-Binomial model have been fit, of which the parameter tables and convergence diagnostics can be found in appendices C.8.1 and C.8.2 respectively. To analyze their fit, predictive posterior plots, center & 95 % HDI bar plots, residual violin plots and LOO-CV scores have been generated, which can be found in appendix C.8.3. Summarizing the main findings, the following can be observed:

In the predictive posterior plot, it can be seen that the Poisson model does not capture the upper limit of four pegs per trial and even predicts up two 12 pegs per trial. In contrast, the Binomial and Beta-Binomial model capture the upper limit, albeit with an underestimation of $P(X = 2)$ and $P(X = 3)$, and a slight overestimation of $P(X = 4)$. Here, both models are very similar in their center estimates and HDI limits. Overall, the Poisson model fails to capture the distribution of the observed data whereas the Binomial and Beta-Binomial model seem to predict the data sufficiently.

The bar plot of the individual model parameters also reflects the similarity between the Binomial and Beta-Binomial model as the center and 95 % HDI's are almost the same. In contrast, the Poisson model has for all parameters different center estimates as well as smaller upper and lower HDI limits.

The residual violin plot shows very similar distributions for the Binomial and Beta-Binomial model except for trial one and four for the suppressive group, where the distribution of the Binomial model has a higher density around the zero value and therefore lower prediction errors. In contrast, the residual distribution of the Poisson model is located more on the positive spectrum, indicating that this model generally underestimates the success rate, as the residual distributions are calculated by subtracting the center estimates from the observed data.

To numerically evaluate the models, LOO-CV scores have been calculated for all three models, as shown in table 4.26. As can be seen, the binomial model has the lowest LOO-CV score and has therefore been chosen as the model with the best fit, for further analysis.

Model	**LOO**	**SE**
Binomial	242.298	10.908
Beta-Binomial	244.118	10.826
Poisson	368.285	2.728

Table 4.26: LOO-CV scores for Poisson, binomial and beta-binomial models.

Final Model Inference
Below in table 4.27, the CLU table of the fixed effects can be found, where the intercept contains supportive embodiment and trial one as the reference level.

Parameter	**Center**	**Lower**	**Upper**
Intercept	1.51	0.70	2.37
Suppressive	-0.78	-1.93	0.30
Trial 2	0.23	-0.71	1.16
Trial 3	0.23	-0.67	1.18
Trial 4	0.62	-0.33	1.67
Suppressive:Trial 2	-0.31	-1.53	0.88
Suppressive:Trial 3	-0.15	-1.37	1.06
Suppressive:Trial 4	-0.18	-1.56	1.10

Table 4.27: CLU table for the fixed effects of posterior distribution of the binomial model.

Using the chosen model, parameter estimation based hypothesis testing has been conducted for each of the four trials. It was expected that the suppressive embodiment group is less effective in inserting pegs than the supportive embodiment group. Center, upper and lower estimates, evidence ratio and posterior probabilities of the differences between the supportive and the suppressive group can be found in table 4.35.

Hypothesis	Center	Lower	Upper	Evid. Ratio	Post. Prob.
$\beta_{Suppressive} < 0$	-0.1443	0.4420	0.0207	11.54	0.92
$\beta_{Suppressive} + \beta_{Suppressive:Trial_2} < 0$	-0.1937	-0.4673	0.0195	35.7	0.97
$\beta_{Suppressive} + \beta_{Suppressive:Trial_3} < 0$	-0.1586	-0.4709	0.0206	16.62	0.94
$\beta_{Suppressive} + \beta_{Suppressive:Trial_4} < 0$	-0.1306	-0.5492	0.0130	16.54	0.94

Table 4.28: Summary statistics of the hypotheses testing for each of the four trials for the number of successfully inserted pegs. Differences are have been converted using the inverse-logit function and are thus on the odds ratio scale.

The Binomial model has been estimated using the logit link function. To generate the CLU differences for the hypothesis table, the inverse-logit function has been applied to the parameters of table 4.27:

$$logit^{-1} = \frac{e^x}{1+e^x} \tag{4.1}$$

By applying the inverse-logit function, the differences between the embodiment groups are on the odds ratio scale and reflect the chance of inserting one peg successfully in a given trial. For instance, the difference in chance for the first trial is 14.43 [1.95, 46.73] %, which means that the suppressive group is 14.43 % less likely to insert a peg successfully than the supportive group. In trials two, three and four, the model predicts an odds ratio of -19.37 [-46.73, 1.95], -15.86 [-47.09, 2.06] and -13.06 [-54.92, 1.3] respectively. For all trials, the posterior probability is higher than 90 %, additionally it can be observed that for all trials there is strong evidence against the null effect, except for trial 2, where we observe a very strong evidence ratio. This suggests that there is compelling evidence in favor of the hypothesis that the supportive embodiment group was more effective than the suppressive embodiment group.

To provide further evidence, the odds ratios that have been previously obtained via the inverse-logit function have been inserted into the Binomial probability formula in equation 4.2, in order to estimate the probability of inserting X pegs for a given trial:

$$P(x) = \binom{n}{x} p^x q^{n-x} \tag{4.2}$$

where $P(x)$ is the probability of x successes, n is the number of trials[1], p is the probability of success for a single trial and q is the probability of failure for a single trial. Using the center estimates, the resulting probabilities can be found in tables 4.29 and 4.30.

Trial	P(X)	P(X=0)	P(X=1)	P(X=2)	P(X=3)	P(X=4)
1	0.81906	0.00107	0.01940	0.13180	0.39770	0.45210
2	0.85069	0.00050	0.01133	0.09680	0.36768	0.52370
3	0.85069	0.00050	0.01133	0.09680	0.36768	0.52370
4	0.89379	0.00013	0.00428	0.05407	0.30335	0.63817

Table 4.29: Binomial probabilities for the supportive group of inserting X pegs per trial.

[1] "Trial" refers here to the number of trials in a peg & hole trial, which corresponds to the number of pegs, which is always 4.

Trial	P(X)	P(X=0)	P(X=1)	P(X=2)	P(X=3)	P(X=4)
1	0.67476	0.01186	0.09630	0.29330	0.39700	0.20010
2	0.65699	0.01384	0.10606	0.30471	0.38908	0.18631
3	0.69209	0.00899	0.08082	0.27248	0.40829	0.22943
4	0.76319	0.00315	0.04054	0.19599	0.42107	0.33925

Table 4.30: Binomial probabilities for the suppressive group of inserting X pegs per trial.

It can be observed that the supportive group is very effective in inserting the pegs as most of their success rate is distributed between $P(X = 3)$ and $P(X = 4)$, with $P(X = 4)$ having the highest probability through out all trials. The suppressive group on the other hand has most of its success rate concentrated between $P(X = 2)$ and $P(X = 4)$, with $P(X = 3)$ having the highest rate. Additionally, the differences in binomial possibilities between the groups has been summarized in table 4.31:

Trial	P(X)	P(X=0)	P(X=1)	P(X=2)	P(X=3)	P(X=4)
1	-0.1443	0.01079	0.07690	0.16150	-0.0007	-0.25200
2	-0.1937	0.01334	0.09473	0.20791	0.02140	-0.33739
3	-0.1586	0.00849	0.06949	0.17568	0.04061	-0.29427
4	-0.1306	0.00302	0.03626	0.14192	0.11772	-0.29892

Table 4.31: Differences in binomial probability for inserting X between the suppressive and supportive group.

Here, it becomes evident that the supportive group was much more successful in inserting all four pegs as compared to the suppressive group (around 25-30 %) whereas the probability for inserting three pegs successfully is relatively similar (except for the last trial).

Overall, the estimated differences from the hypothesis tests and the binomial probabilities correspond with the rain cloud plot in figure 4.10, in which the supportive group consistently performances better than the suppressive embodiment group. Overall, it can be stated that there is strong evidence that the suppressive embodiment group was more effective in the peg & hole task than the suppressive embodiment group.

4.2.2 Efficiency

To compare the task performance in terms of efficiency between the groups and within trials, the time per successfully inserted peg has been measured.

Data Exploration
To explore the data visually, a rain cloud plot has been generated, showing the distribution, summary statistics as well as the individual data points.

	Supportive					Suppressive				
Trial	N	Mean	Median	sd	IQR	N	Mean	Median	sd	IQR
1	50	106	69.5	96.3	89.8	39	110	100	64.4	84
2	52	99	91.2	46.5	54.7	38	134	103	97.5	55.6
3	52	90.8	77.4	55.8	49	40	103	77.1	63.5	68.8
4	55	100	66.4	75.6	86.2	44	98.3	81.9	60.6	74.6

Table 4.32: Summary statistics per trial and condition for efficiency measure.

From the distributions in figure 4.11, it can be observed that for both groups the data is skewed in all trials to the right, which is expected as the variable time has a positive lower bound. Overall, it can be observed that there is a large overlap in the distributions of the two embodiment

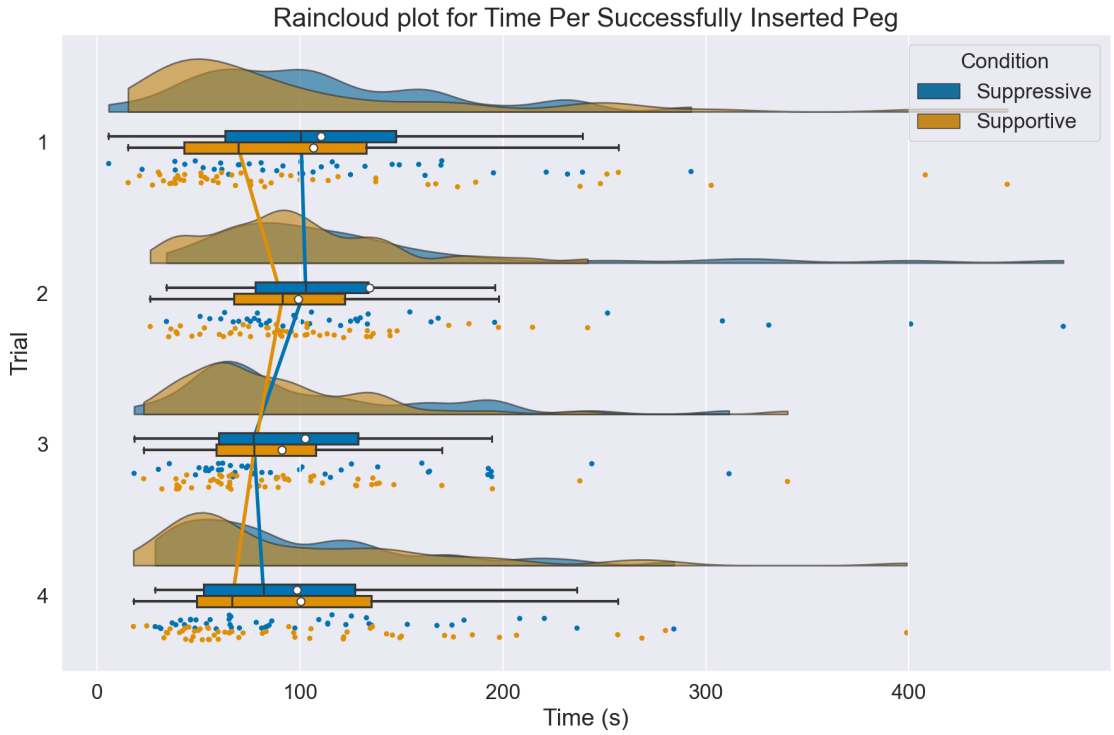


Figure 4.11: Rain cloud plot for time needed per successfully inserted peg.

groups, where the supportive group has a central tendency slightly more towards the left side than the suppressive group, based on the medians. Over time, the supportive group first increases in the time needed to insert the pegs, but then decreases in trial three and four. For the suppressive group, the central tendency remains approximately the same in the second trial, decreases in the third trial and then slightly increases again in the fourth trial. In the end, both groups have the same mean and almost identical distribution, albeit slightly different medians.

Model Comparison

To analyze this data, several GLMs have been fit, including a Gaussian, Gamma and exponentially modified Gaussian model. The parameter tables and convergence diagnostics can be found in appendices C.1.1 and C.1.2 respectively. To analyze their fit, center & 95 % HDI bar plots, residual violin plots and LOO-CV scores have been generated, which can be found in appendix C.1.3. Summarizing the main findings, the following can be observed:

The predictive posterior checks show that the Gaussian model predicts implausible (negative) values and also predicts on average more time needed per peg than can actually be observed in the data. The Gamma and the exponential modified Gaussian model on the other hand both capture the lower bound of the data and the predicted the samples follow nicely the observed data, thus suggesting a good model fit on first sight.

The coefficients in the bar plots show that the Gaussian and Gamma model are very similar in their estimation of the coefficients whereas the exponentially modified Gaussian model mostly predicts lower center estimates, with more certainty.

If this model does indeed predict a better fit to the data, this should be reflected in the shape of the residuals. Generating violin plots of the residuals show that generally the modes of all models appear to be slightly below zero, thus overestimating the time needed to insert the pegs. Moreover, the first and fourth trial are less certain than the second and third trial which is also reflected in the box plots of the rain cloud plot in figure 4.11. Overall, the residual errors of

the models are relatively similar, where the exponentially modified Gaussian model seems to predict the values slightly better, especially for higher time values.

Finally, to numerically determine the best fitting model, LOO-CV scores have been calculated for all three models, as can be seen in table 4.33:

Model	**LOO**	**SE**
Exp. Gaussian	3638.517	32.799
Gamma	3649.572	31.011
Gaussian	3739.082	33.010

Table 4.33: LOO-CV scores for Gaussian, exp. Gaussian, and Gamma models.

As can be seen, the exponentially modified Gaussian model has the lowest LOO-CV score. Together with the findings from the predictive posterior checks, bar plots and residual violin plots, this supports the hypothesis that the exponential modified Gaussian model fits the best. Therefore, this model has been chosen as the model with the best fit, for further analysis.

Final Model Inference
Below in table 4.34, the CLU table of the fixed effects can be found, where the intercept contains the supportive embodiment group in trial one as reference level.

Parameter	**Center**	**Lower**	**Upper**
Intercept	4.36	4.21	4.49
Suppressive	0.17	-0.02	0.35
Trial 2	0.24	0.09	0.39
Trial 3	0.13	-0.01	0.29
Trial 4	0.07	-0.08	0.22
Suppressive:Trial 2	-0.07	-0.28	0.14
Suppressive:Trial 3	-0.15	-0.37	0.07
Suppressive:Trial 4	-0.10	-0.33	0.11

Table 4.34: CLU table for the fixed effects of posterior distribution of the exponentially modified Gaussian model on the log scale.

Using the chosen model, parameter estimation based hypothesis testing has been conducted for each of the four trials. It was expected that the suppressive embodiment group requires more time to insert the pegs than the supportive embodiment group. Center, upper and lower estimates, evidence ratio and posterior probabilities of the differences in time between the supportive and the suppressive group can be found in table 4.35.

Hypothesis	**Center**	**Lower**	**Upper**	**Evid. Ratio**	**Post. Prob.**
$\beta_{Suppressive} > 0$	14.09	-1.35	37.31	27.37	0.96
$\beta_{Suppressive} + \beta_{Suppressive:Trial_2} > 0$	9.28	-19.68	82.67	7.24	0.88
$\beta_{Suppressive} + \beta_{Suppressive:Trial_3} > 0$	1.32	-21.64	61.37	1.45	0.59
$\beta_{Suppressive} + \beta_{Suppressive:Trial_4} > 0$	5.19	-18.28	65.56	3.49	0.78

Table 4.35: Summary statistics of the hypotheses testing for each of the four trials for the time needed per successfully inserted peg. Differences have been exponentiated from the log scale to the original scale (in seconds).

For trial one, a difference δ of 14.09 [-1.35, 37.31] seconds can be observed with P($\delta > 0$) = 0.96 and a strong evidence ratio of 27.37. In trial two, a difference δ of 9.28 [-19.28, 82.67] seconds is being estimated with P($\delta > 0$) = 0.88 and a moderate evidence ratio of 7.24. For trial three, a difference δ of 1.32 [-21.64, 61.37] seconds can be observed with P($\delta > 0$) = 0.59 and an

anecdotal evidence ratio of 1.45. For the last trial, the model predicts a a difference δ of 5.19 [-18.28, 65.56] seconds with P($\delta > 0$) = 0.78 and a moderate evidence ratio of 3.49.

Overall, it appears that there is strong evidence for a difference of approximately 14 seconds per peg in the first trial, however, in the other trials this difference fades and becomes insignificant (e.g. a difference of 1 second is not relevant for a mean of about 99 seconds). This also corresponds to the observation from the rain cloud plots as both groups seem to converge towards the last trial.

5 Discussion

This section evaluates how effective the manipulation of SoE in the present study was and discusses the hypotheses that have been formulated at the beginning of the study, namely supportive embodiment leads to better dexterous task performance (H1) and to an increased learning rate (H2) for dexterous tasks when compared to a suppressive embodiment.

5.1 Embodiment Manipulation Check

Surveys & Interviews

The survey analysis in section 4.1.1 yielded medium, small to medium and large effect size differences in SoO, SoA and SoS, with moderate, mixed (moderate, strong, anecdotal) and extreme evidence ratios, respectively. These findings are supported by the introspective interview analysis in which 1), participants of the supportive embodiment group reported feeling of ownership more often for the hand and the arm; 2), the supportive embodiment group reported higher agency for the hand as well as for the arm; and 3), all participants of the suppressive group reported to have difficulties estimating their position in the remote environment whereas only a few participants of the supportive group had troubles with self-localization.

Based on the survey and interview findings, it appears that the manipulation for SoS worked very well. Moreover, participants of the suppressive group consistently reported that they tried to estimate the height with cues such as positioning the peg at an angle, using shadows in the remote scene as reference or positioning the thumb of the remote hand over the peg and moving the arm down until it touches the peg. These cues correspond to monocular cues as discussed in section 3.2.2, which provide information about depth when looking at the scene with one eye only [30], indicating that the lacking depth perception is not only reflected in the perceived SoS but also in the behaviour of the participants. An additional cue to compensate for a lack in depth perception was using sound such as the "clunk" sound when a peg was successfully inserted into a hole, as to confirm the proper insertion. The use of such cues highlights the need for additional perceptual information to properly assess distances in the remote environment and indicates that the combination of a 3PP with a lack of depth perception was very effective in creating differences in SoS between the two embodiment groups. Note that depth perception has not been used as SoS manipulation in other studies. As the present study only compared the joined contribution of perspective and depth perception, it would be therefore interesting to compare the individual contribution of perspective (1PP vs 3PP) and depth perception (binocular vs monocular depth cues) in a 2x2 factorial study.

In contrast to SoS, the manipulations for SoO and SoA seemed to have a weaker effect, which is reflected in smaller differences in the survey responses and less reported sensations of agency and ownership in the interviews. One reason that could explain this behaviour is that for both embodiment groups the same setup has been used to control the remote arm. This might have had two effects, namely 1), visuo-proprioceptive stimulation (caused by the arm) was the same for both groups, outweighing other perceptual cues and 2), technical limitations of the arm deteriorating SoA, which in turn decreased SoO.

The first possible explanation is that both groups received the same level of visuo-proprioceptive stimulation which might have had a significantly greater impact compared to perspective and haptic feedback manipulations for SoO and SoA, thus outweighing these perceptual cues and creating two similar levels of SoO and SoA. Congruent visuo-proprioceptive stimulation has been proven as a very effective manipulation of SoO and SoA, as discussed in section 2.1.2. Moreover, in [4] various perceptual cues have been manipulated to create a supportive and suppressive embodiment, including visuo-proprioceptive stimulation which

was found to have the highest impact on SoO and SoA. This is supported by the fact that several participants of the suppressive embodiment group reported to have felt haptic feedback in their hands although there was no feedback present. This resembles the effect of the RHI, indicating higher ownership levels

The second possible explanation is the impact of technical limitations decreasing SoA, which in turn may have affected SoO. The most prominent theme from the interviews was the impact of SoA on SoO, that is, whenever the robot was doing something unexpected or when the participant could not control the robot as desired, it pulled them out of the embodiment illusion. Examples for such instances included participants not being able to control the wrist or the elbow of the remote arm properly, the arm doing unexpected, sudden movements or participants not being able to reach the upper left peg and thus needing to be clutched out and in again to extend their workspace. See section 7.1 for a detailed list of identified technical limitations and proposed solutions. In the literature, there is a strong correlation reported between SoO and SoA [86, 199–201]. Although they can be disassociated [202–205], literature suggests that they can strengthen each other if they co-occur [34, 206, 207]. In context of the predictive coding theory [155], not only perception arises but also self-awareness [205, 208]. Here, the brain creates hierarchical generative models about both the causal structures of the outside world as well as about the "the most likely to be 'me'" [208]. To minimize the prediction error, the brain may employ perceptual inference or active inference. The former describes the process by which the brain calculates the probability whether its current multisensory input has one common cause (i.e., one self-owned hand) or two causes (i.e, one self-owned hand and an artificial hand), given the current sensory state and the prior probability of a common cause [205]. If the estimated probability is higher for a common cause, then the brain merges the two disparate limb sensations together and the RHI occurs. If it is higher for two common causes, the sensations of the limbs remain separate and no RHI occurs. In other words, if the predicted input matches the actual sensory input, the predicted model is confirmed. In case of a mismatch, a prediction error occurs and depending of the magnitude of the error, the predicted model needs to be updated [155, 205]. The latter case, active inference, describes the process by an action is performed to create a new sensory state which is tested against the predictions of the brain (e.g., reaching with an arm to estimate distance to object) [155]. If the new sensory state matches the prediction of the brain, SoA arises and the predicted model stays valid. Therefore, via active inference also the model of one's self may be updated, highlighting the connection between SoO and SoA. To investigate the correlation between SoO and SoA, Shepherd's Pi correlations [209] have been calculated for SoE sub-components (see appendix E.1). Indeed, a significant, positive, moderate correlation ($r = 0.58$) could be found for SoO and SoA, supporting this hypothesis. That said, the lack in differences of SoO and SoA is most likely due to a combination of the same level of visuo-proprioceptive stimulation and technical limitations disrupting SoO and SOA.

Proprioceptive drift

Proprioceptive drift as measured in the position judgment task showed a difference in drift between the supportive and suppressive group of 0.64 cm on the X axis and 0.16 cm on the Y axis in the second trial. As expected, there was an offset on the Z axis in the first trial (0.52 cm) between the supportive and suppressive embodiment group, which was likely caused by manipulations of point of view and depth perception. In the second trial, a difference of 0.68 cm could be found, where the interaction effect between embodiment condition and trial contributed only 0.07 cm. This negligible interaction effect suggests that both groups approximately changed by the same rate in drift from trial one to trial two. As no visuo-proprioceptive stimulation has taken place before trial one, this suggests that one could consider the difference in trial one as an offset caused by visual bias. Thus, when simply subtracting this offset in the first

trial from the second trial, a difference of 0.16 cm is obtained for the second trial, which is the same as on the Y axis.

In the study by Lengenhagger et al. [110], a position judgement task was conducted in VR with the entire of the body of the participant, here proprioceptive drift was measured on the X and Y axis, where they found a significant difference only on the X axis. The study by Hamasaki et al. [210], similar to the one by Lengenhagger et al., measured proprioceptive drift on multiple axes (X, Y and Z). Also here, the authors found significant differences only on the X axis. The predominant effect of proprioceptive drift on the X axis may relate to the fact that distances along the fronto-parallel (forwards) axis are perceived as shorter as they actually are [32, 211, 212]. Therefore, participants may have judged their "end" position closer to their initial position, creating larger distance errors as compared to the other axes. Note, that because of the top/-down view, the fronto-parallel plane lies actually on the Z axis for the suppressive group: if they move their real arm down, it will move forwards (i.e., towards the mounting plate) on the rendered screen, whereas if they move their arm forwards, it will move upwards on the HMD stream. Therefore, if we consider proprioceptive drift only along the fronto-parallel plane, we should compare the change in drift from trial one to trial two for the supportive group on the X axis and for the suppressive group on the Z axis. By calculating the median difference in percentages for the respective axes, we obtain a drift of 18.95 % for the supportive group and a drift of 6.85 % for the suppressive group. To compare these values to other studies, the difference of relative drift between the X axis from the supportive group and the Z axis from the suppressive group has been calculated, rendering a difference in proprioceptive drift (from the first to the second trial) of 12.1 %. This has been done as proprioceptive drift studies often miss baseline measurements [27], therefore not allowing us to analyze the change in drift before and after stimulation. Other studies measuring proprioceptive drift via robot hands reported a difference in drift of 29.03 % (1.55 ± 4.37 cm (sync) and 2.00 ± 2.91 cm (async)) [35] and 7.83 % (3.96 ± 2.21 cm (sync) and 3.65 ± 2.06 cm (async)) [74] (experiment 2). Note that the SoE induction method was visuo-tactile stimulation in [35], whereas it was visuo-proprioceptive stimulation in [74]. To this end, the difference in proprioceptive drift along the front-parallel axis between trial one and two as measured in the position judgement task seems comparable to the aforementioned studies, however these are just two sample points with different setups and different induction methods.

5.2 Embodiment & Learning Behaviour

In section 3.1 the hypothesis has been formulated that a higher level of embodiment leads to be faster learning rates as compared to a lower level of embodiment. The results in appendix C.7 show that there was no significant learning behaviour on the participant level for either of the effectiveness nor the efficiency measure.

For the effectiveness measure, it could have been that the (trainable part of the) task was so simple, that the amplitude was very small, meaning that learners were already very close to their asymptote from the beginning of the experiment. For instance, for first trial it could be observed that the supportive group had a mean of 3.2 (last trial 3.5), which is already very close to the maximum score of four pegs.

Concerning the efficiency measure, technical limitations imposed by the Virtuose setup (e.g. needing to clutch in and out to extend the workspace etc.) could have distorted the actual time needed to learn the task.

Additionally, fatigue may have affected the efficiency of the participants as a consequence of the duration of the experiment in combination with wearing the VR headset. Nausea, dizziness, disorientation, fatigue, and instability (referred to as VR induced symptoms and effects (VRISE)) are modulated by the time spent in VR [213]. Kourtesis et al. [214] suggest that a maximum duration between 55 and 70 minutes for VR immersion. This time may have very well

been exceeded, as participants generally spent at least an hour in VR by the end of the peg & hole task. Moreover, the haptic feedback provided by the Virtuose may have caused fatigue in the right arm as the gravity compensation accounts only for the weight of the manipulator itself and the attached HGlove, but not for the weight of the participant's arm. Therefore, participants were still required to lift the weight of their own arm for a long period of time, which was often mentioned in the interview. Such fatigue effects could have worked against the asymptote, causing the learning curve to curl up towards the last trial. This could also explain the convergence between embodiment groups over time for the efficiency measure. Another reason could have been the chosen order to insert the pegs into the respective holes. To increase the complexity of the peg & hole task, it was not allowed to remove a peg once it had been inserted into one of the target holes. This required to carefully choose the order in which pegs are inserted as certain sequences make the task significantly harder to execute as more maneuvering around the pegs is required. Whereas some participants stuck to the same order, some participants tried a wide range of orders, where some orders were more efficient than others. Therefore, too much freedom in the execution may have caused a significant amount of variation in the time spend to execute the task, masking the effects of embodiment on efficiency.

Similarly, the amount of trials in the experiment could have not been enough to find the optimal order of inserting the pegs, which would indicate that the amount of trials was rather testing the efficiency of participants than training it [129].

5.3 Embodiment & Task Performance

In section 3.1, the hypothesis has been formulated that a higher level of embodiment leads to be better dexterous task performance as compared to a lower level of embodiment. This hypothesis is based on the free energy principle and the predictive coding theory [155], in which the brain creates a model of the world that generates explanations and predictions about its sensations, against which sensory samples are tested to update beliefs about their causes. In context of the predictive coding theory, a higher level of SoE would be associated with less prediction errors over body image and schema as sensory states match prior beliefs about their causes. In the present study, task performance was assessed in terms of efficiency (time) and effectiveness (count) within the peg & hole task. Next to the Bayesian analysis in section 4.2, an additional correlation analysis between embodiment components and task performance measures has been performed in appendix E.3.

5.3.1 Efficiency

For the efficiency measure, the analysis in section 4.2.2 showed that in the first trial there was strong evidence that the supportive group was approximately 14 seconds faster per peg as compared to suppressive embodiment group. However, in the subsequent trials the evidence in favor of the suppressive group becomes moderate to anecdotal and the difference in time fades to an insignificant amount (i.e., a difference of 1 second is not relevant compared to a mean of 99 seconds). Overall, it appears that there was only a small difference in efficiency between the embodiment group, which converged over time. The lack of difference in the efficiency measure is probably due to the same reasons as stated in section 5.2 (e.g. technical limitations, fatigue etc.). This may be supported by the correlation analysis in appendix E.3 in which none of the SoE components showed a significant correlation with the time measure.

5.3.2 Effectiveness

For the effectiveness measure, the analysis in section 4.2.1 showed that for all trials there is strong evidence that the supportive group was more effective, that is, they inserted more pegs successfully. Results showed that the supportive group was on average 16 % more likely to insert a peg successfully as compared to the suppressive embodiment group. Moreover, when

calculating the Binomial probability of successfully inserting a given number of pegs per trial, it was shown that the supportive group had on average a success rate of approximately 45 to 64 % for inserting all four pegs per trial, which was around 25 to 30 % higher than the suppressive group.

Based on the interviews and surveys, it was concluded that medium, small and large differences could be found for SoO, SoA, and SoS respectively between the embodiment groups. To investigate the contribution of the individual embodiment sub-components on task performance, a correlation analysis between embodiment responses and task performance measures has been calculated in appendix E.3. Here, a significant correlation could only be found with SoS, which has a weak, positive correlation ($r = 0.32$, $p < 0.0001$) with the count measure. SoO showed no significant correlations ($r = -0.014$, $p = 0.878$) while SoA was nearly significant with a positive, weak correlation ($r = 0.16$, $p = 0.084$), however both correlations had low power.

Taken together, this suggests that participants of the supportive group perceived significantly higher SoS as a consequence of the first person perspective and the stereopsis created by the stereo camera. As such, they may have experienced better spatial awareness in the remote environment which enabled them to perform more effectively in the peg & hole task as compared to the suppressive embodiment group. This finding is in line with the study by Medeiros et al. [165] where a 1PP lead to better spatial awareness and increased task performance. Similarly, in the study by Mroczkowski et al. [124], stereopsis lead to better predictive control for a grasping task as compared to no stereopsis. In context of the predictive coding theory, the improved spatial awareness could have led to lower visual prediction errors, and thus to a more accurate model of the world (i.e. the remote environment) as well as a more accurate model of the body schema of the surrogate, allowing for better eye–hand coordination and distance estimation in the remote space, resulting in better effectiveness in the peg & hole task.

However, the question arises why SoO and SoA were not significantly correlated with task performance. As for SoA nearly significant correlations have been observed, it may be that correlations have been masked by technical limitations such as the inability of controlling the wrist or elbows of the surrogate arm, introducing variation into the data and thus distorting the underlying linear relationship between SoA and effectiveness. The same may hold for SoO. The aim of the following subsections is to illuminate the cause of the lacking correlations with task performance, and explain to what extent embodiment can enhance dexterous task performance. First, contrasting studies on ownership and task performance are reviewed, showing that the effect of ownership on task performance depends on whether the cues that have been used to elicit ownership are task relevant. Based on motor imagery, we show how egocentric and allocentric perceptual cues that give rise to SoO can impact the inverse model and the forward model in motor learning. We then relate each of the SoE sub-components to egocentric and allocentric cognition and explain when and how embodiment can impact dexterous task performance.

Previous Studies on Ownership and Task Performance

In Grechuta et al. [19], ownership was modulated with respect to the (a)synchrony of visual and tactile stimuli. Here it was observed that higher ownership was positively correlated with task performance. The authors interpret this observation as a consequence of multimodal integration, in which reduced perceptual ambiguity correlates with ownership and task performance. Their explanation is challenged by Rosa et al. [164], which postulate that the statement by Grechuta et al. simply suggests that there may be a causality of multimodal integration to evoking ownership and a causality of multimodal integration to motor control, however it does not follow that ownership strictly increases task performance. Rosa et al. support their hypothesis with the results of their study, which show that agency was correlated with task performance while ownership was not correlated with task performance. They propose that a causality from

ownership to task performance only exists if the chosen perceptual cue to manipulate ownership is also related to motor control. Indeed, in the study by Rosa et al. connectedness was used to alter embodiment, however it did not provide any proprioceptive information due to flaws in their implementation, which made the arm movements incongruent to the hand movements, leading to reduced ownership, agency and task performance. In contrast, Grechuta et al. used congruent tactile and visual cues that provided valuable sensorimotor information, increasing ownership and task performance simultaneously.

To support the notion that ownership does not strictly modulate task performance, one may consider a recent study by Grechuta et al. [215], which shows that the plasticity of body ownership also depends on the consistency of body- and action-independent sensory cues. In their study, they manipulated the congruency of task-irrelevant distal (visual and auditory) cues (e.g. time of the day) for a goal-oriented air hockey task in a VR environment. Their results show that ownership was modulated by task-irrelevant, distal cues, while task performance was not. In contrast, a previous study by Grechuta et al. [161] showed that within the same task, action-dependent distal cues lead to increased ownership and task performance levels. These cues included a temporal cue where the time of the cue was synchronized with the time of the hit of the hockey puck; a spatial cue which originated from the location of the hit; and a semantic cue, reflecting performance of hitting the hockey goal in a binary fashion (i.e., success or failure). Here, "the directionality of the error indicated by the spatial distribution of the sound, the speed of the puck indicated by the temporal characteristics of the sound, as well as the knowledge of results all constituted error signals which could supervise corrective motor commands" [161].

Motor Imagery & Ownership in Motor Learning

Note how all of the cues in the study by Grechuta et al. [161] have been provided after hitting the puck. One could suggest that they have contributed to motor imagery, that is, mentally simulating the task: After finishing a trial, the participant reflects back on the provided cues, reliving the last trial and imagines how they would execute the hit in the next trial. Indeed, motor imaginary has been shown in various studies to significantly improve the effectiveness and efficiency of motor performance (see [216,217] for reviews). From a theoretical perspective, this might be explained by the concept of forward models and inverse models [218].

Forward models are thought to be computational units in the brain, encoding the dynamics of our body parts as they interact with the environment to anticipate the outcome of voluntary movements [218–220]. When a voluntary, goal-oriented action is performed, a copy of the motor commands sent from the motor cortices to the muscles is also sent to an internal forward model (efference copy) that predicts the sensory consequences of the action (corollary discharge) [221]. The Central Nervous System (CNS) is then informed about how closely the anticipated movement matched its actual counterpart by comparing the corollary discharge with the real sensory patterns (reafference). In other words, the forward model predicts the state of our body after the upcoming movement as well as the sensory consequences that the movement is going to cause. Such predictions are said to then provide a more reliable estimate of the state of the body compared to actual sensory input, which suffers from noise and delay [220, 222]. The study by Kilteni et al. [220] supports the notion that the sensory predictions generated by motor imagery engages the forward model in the brain the same way as physically executing the imagined movement does. In the words of Kilteni et al. [220], "when we imagine moving one hand to touch the other, the forward models simulate the imagined action based on the efference copy and predict the end states of the limbs after the imagined action as well as their sensory consequences, just as they do for real movements. If the predicted positions of the limbs after the imagined movement indicate contact between two body parts, then the forward models generate tactile predictions".

In contrast, inverse models invert the forward model by predicting the necessary motor command to cause a desired change of a body state [218]. For instance, for a desired arm trajectory, the inverse model generates the motor command to change the position of the arm towards the designated target. In [223], Tanaka describes how motor imagery can be used to generate predictions within the inverse model using the body image. Tanaka describes this relation via the neuropathological case of Waterman [50], a patient that lost his sense of touch and proprioception and thus lost control over his motor movements, leading to a significant deficit in his body schema. Surprisingly, through his own efforts he regained (to some extent) control over his body during two years of rehabilitation. Through vision and motor imagery, Waterman managed to execute basic skills again, such as standing up or eating. According to Tanaka, this was possible due to the body image: By imaging himself from an allocentric perspective, he was able to recognize and grasp the motor imagery of potential actions that stand out from the perceived situation. An allocentric perspective describes the location and orientation of objects in the scene [224] and is responsible for building and storing spatial representations in memory [225] based on visual information [226]. Tanaka [223] argues that in order to learn basic skills, one needs to understand basic procedures by transposing the observed movement into one's own. By imagining his body image from the allocentric perspective, Waterman could simulate and observe movements that he could learn from. Such simulation of movements correspond to the construction of the inverse model, where a desired state is transformed into a motor action that produce this state [218, 223]. Thus, motor imagery of the body image may have facilitated the learning of new skills for Waterman. Note that Waterman had to actively focus his mind for executing simple movements and could thus not transition from the cognitive phase of motor learning to the associative phase. Tanaka suggests that Waterman was missing the sensory feedback (e.g. proprioception) to update his body schema and refine his motor skills.

Interestingly, whereas Tanaka [223] relates motor imagery to the inverse model of motor functions, it is linked by Kilteni et al. [220] to the forward model. The difference between the two internal models is reflected in the goal and the reference frame of the imagination task: Whereas in Tanaka's essay motor imagery is performed from an allocentric point of view to simulate movement, it is done from an egocentric point of view to predict sensory consequences of movement in Kilteni et al.'s study. Note how ownership may be linked to both motor imagery methods of Tanaka and Kileni et al.. The allocentric motor imagery method is dependent on the body image, which is comprised of perception, beliefs and emotional attitudes towards our own body [227]. As such, it is the subjective mental image of one's own body and inherently requires ownership as a prerequisite. For instance, it has been shown that ownership of a rubber hand modulates the perceived size of the own hand [228]. The egocentric motor imagery method may also be linked to ownership as both appear to recruit the same multi-sensory signals as ownership [229, 230]. For instance, ownership and motor imagery both have shown to lead to attenuation of somatosensory signals via the forward model [160, 220].

The question arises why there was no correlation between SoO and task performance, but a correlation between SoS and task performance for the present study. The answer may entail that not just the body image plays an important role for simulating movements to engage the inverse model, but also the allocentric perception itself and consequently the imagination thereof. In [226] it is suggested that SoS can be thought of as the blending of two parallel representations: the egocentric mapping of somatosensory sensations into the external space, which is primarily associated with peripersonal space, and the abstract allocentric coding of the position occupied in the environment, which is primarily associated with visual perspective. As such it may have been that the increased depth perception of the supportive group lead to better spatial estimations in the allocentric view, allowing to better estimate how the arm of the surrogate should be moved in order to insert the pegs into the holes. In the next section, we will try to answer our initial research question by relating each embodiment component

to egocentric and allocentric cognition and explain when and how embodiment may impact dexterous task performance.

Synthesis

In this section, we will try to answer our initial research question in which we wanted to investigate to what extent embodiment affects dexterous task performance.

SoO appears to not directly modulate task performance, as shown by the studies of Grechuta et al. [161, 215], Rosa et al. [164] as well as the results of the present study. Instead, it seems that the contribution of SoO on task performance depends on the multi-sensory integration of perceptual cues that give rise to the sensation of ownership and how relevant these cues are to the task at hand. Specifically, egocentric and allocentric perception may be used in motor imagery to update internal models of motor of control, as suggested in [223] and [220]. Whereas the body image is objectified and projected into the allocentric perspective to simulate movements and engage the inverse model, egocentric cues are used to make predictions of sensory consequences to update the forward model and generate more reliable body state estimates and refine motor control. Note that both allocentric and egocentric cognition are related to SoO: whereas the body image as mental, subjective image of one's self is intrinsically linked to ownership, egocentric perceptual cues, such as proprioception, tactile and vision, are the same cues that are said to generate a feeling of ownership as consequence of multi-sensory integration. Ownership may thus play a role in training operators via motor imagery and may even influence task performance via unconscious motor imagery [231].

Sos may be viewed as the blending of two parallel representations: the egocentric mapping of somatosensory sensations into the external space, which is primarily associated with peripersonal space, and the abstract allocentric coding of the position occupied in the environment, which is primarily associated with visual perspective [226]. The former relates to the usually used definition in RHI studies (the feeling of being located in a volume of space [10]) whereas the latter describes the spatial presence in an environment [28, 29]. Similarly to SoO, SoS may not always contribute to task performance. In the studies of Gorisse et al. [20] and Medeiros et al. [165] a 1PP lead to higher SoS and task performance (e.g. navigation times) as compared to a 3PP. Similarly, in the present study, a 1PP with binocular depth cues lead to an increase in SoS and effectiveness as compared to a 3PP with monocular depth cues. As such, it is proposed that the increased level of SoS in these studies lead to better spatial awareness, i.e., the allocentric coding of the position occupied in the remote environment, providing the inverse model with more affordances to generate action possibilities in the remote environment. For instance, the 1PP view in the studies by Gorisse et al. [20] and Medeiros et al. [165] may have resulted in less collisions, because the inverse model was more constrained in which motor movements were allowed (i.e. would avoid collision) based on the increased spatial awareness. However, this may not generalize to all cases as shown in the study by Debarba et al. [41], in which a 1PP lead to a significant increase in SoO and SoS as compared to a 3PP, but did not result in any significant differences in task performance (i.e., time to reach a target). Thus, although participants felt more spatially aware, this feeling did not translate into better task performance. It may have been that in this particular study, the 3PP did not provide significantly more allocentric (or egocentric) information as compared to the 1PP, as both perspectives appear to have given a similar field of view when comparing them side by side.

Several studies have suggested that SoA arises from a match of predicted and actual sensory feedback in the forward model [222, 232, 233]. As such, SoA is dependent on egocentric cues that provide the necessary sensory information to the forward model in order to make these predictions. The study by Falcone et al. [4], investigated the effect of various perceptual cues on embodiment as well as on task performance. The results show a significant (but weak) correlation between SoE (i.e., SoO and SoA) and task performance. The perceptual cue that af-

fected SoO and SoA the most, was visuo-proprioceptive synchronicity which is an egocentric cue. As such, this cue may have been used in the forward model to generate more accurate predictions of the body state (i.e., hand position), resulting in more accurate motor control and task performance. Similarly, in the study by Rosa et al. [164] a significant, moderate correlation was found between SoA and task performance when manipulating connectedness for a tracing task in VR. Here, an increased level of connectedness (a fully connected arm) lead to decreased SoA and task performance as compared to a lower level of connectedness (floating hand). The authors explain this result by the technical limitations of their study, as the fully connected arm was not accurately following the movements of the participants, but would instead produce unexpected movements during the tracing task. This suggests that the floating hand was providing higher visuo-proprioceptive synchronicity, generating more accurate predictions in the forward model about the hand position. However, an increased level of SoA does not always have to result in an increased task performance as illustrated in the present study. When inspecting the SoA survey responses, one can see that, while there were little differences between the suppressive and supportive embodiment group, participants reported generally positive SoA scores. However, when inspecting the correlation analysis, no significant correlation could be found for either the effectiveness nor the efficiency measure. It appears thus that, although participants felt generally in control of the surrogate, SoA did not affect task performance. SoA, as stated above, arises from a match of predicted and actual egocentric cues, however, it may have been that allocentric cues were more critical to executing the peg & hole task than egocentric ones. In other words, it could have been more important to estimate your body position relative to the environment (i.e., distance from the peg to the hole) than being able to control the movement of the robot arm itself. This may be supported by the positive correlation between SoS (i.e., allocentric perception) and effectiveness in the present study. This suggests that a feeling of agency is not enough to affect task performance, it also depends to what extent the task at hand is sensitive to allocentric and egocentric perception.

To conclude, we could show for each SoE sub-component examples in which higher levels of embodiment did not affect task performance. Instead, it seems that the contribution of embodiment to task performance is dependent on the task at hand and to what extent egocentric and allocentric perception generate motor commands and sensory predictions that are critical to executing the task.

Finally, let us come back to the paper by Toet et al. [6] in which the authors propose embodiment as a tool to increase transparency in telerobotics. Telerobotics, unlike prosthetics and VR, is inherently subjected to the presence of time delay, deteriorating transparency in form of delayed control signals and multi-sensory (e.g., visual, auditory, haptic) feedback. Thus although it has been shown in the present study that embodiment does not directly impact task performance, it still may be of significant when time delays are present. For instance, SoO and SoA have been shown to mitigate the susceptibility to delay [234]. Once established, these embodiment components can cause people to neglect asynchronies in visuo-tactile stimulation that would be otherwise detected [77, 160]. Additionally, from a predictive coding account one could propose that higher embodiment leads to less prediction errors (e.g. about the body image and schema), which may lead to less cognitive load, which has been correlated with free energy [156]. Note that the present study did not consider delay (i.e., different levels of delay), as there was not enough time for a pilot study to explore suitable delay values. Future research should thus investigate the effect of embodiment on task performance in the presence of time delays.

6 Conclusion

This study was set out to investigate the effect of SoE on task performance in telerobotics. Telerobotics is the remote control of a robot and requires high transparency for natural and intuitive interaction. Application areas include rescue & search missions, remote surgeries, bomb disposals or nuclear waste decommissioning [2]. It has been proposed that increased SoE leads to higher transparency and in turn increases task performance [6]. Embodiment has been shown to correlate with brain regions mediating motor control [11–13] and has been shown to directly correlate with task performance in domains such as prosthetics [14–18] and VR [4,19,20,161], however it has not been explicitly investigated yet in the domain of telerobotics. Therefore, this study sought out to answer the research question:

> To what extent does the level of Sense of Embodiment affect the level of dexterous task performance in telerobotics?

To answer this question, a user study has been conducted in which two groups of embodiment levels have been created (supportive & suppressive embodiment), which had to perform a peg & hole task using a telerobotic setup over multiple trials. To create the two levels of embodiment, a set of perceptual cues consisting out of haptic feedback, point of view and depth perception has been manipulated. Embodiment has been analyzed via self reported questionnaires and interviews as well as proprioceptive drift. Task performance has been analyzed in terms of effectiveness (number successfully inserted pegs) and efficiency (time per successfully inserted pegs).

Results show that the chosen perceptual manipulations created medium and small differences in SoO and SoA, and very strong differences in SoS between the suppressive and supportive group. Results of the task performance measures show that 1), no learning behaviour could be observed on the participant level for either effectiveness nor efficiency, 2) there was no relevant difference in the time needed to insert the pegs between the embodiment groups and 3), the supportive embodiment group was 16 % more likely to insert a peg successfully, and 25 to 30 % more likely to insert all pegs in one trial as compared to the suppressive group, thus being generally more task effective. To analyze the relation between task performance measures and SoE sub-components, a correlation analysis has been performed, in which only SoS correlates significantly with effectiveness ($r = 0.32$, $p < 0.0001$).

Based on the results of the present study and other related studies in the field of embodiment and task performance, it has been shown that neither SoO [164, 215], nor SoA, nor SoS [41] strictly modulate task performance. Instead, it is proposed that the contribution of embodiment to task performance is dependent on the task at hand and to what extent egocentric [220] and allocentric [223] perception generate motor commands and sensory predictions in the internal models of the brain [218] that are critical to executing the task.

The present study contributes to the understanding of when and how embodiment can increase task performance in the domain of telerobotics. As such, the results of the present study may guide the design of more effective and efficient telerobotic systems, which may benefit to a plethora of complex and challenging domains, such as tele-surgery or remote rescue & search missions. Additionally, the present study investigated depth perception as a novel perceptual cue to manipulate SoS, which led to significant differences between the embodiment groups in perceived SoS. Further research should be conducted to formally validate depth perception as manipulation tool for SoS. To this extent, recommendations for further research are given in chapter 7.

7 Recommendations

This chapter discussed technical limitations of the used system, lessons learned throughout the user study, proposes how the user study could be improved if it were repeated again and concludes with a set of questions for future research.

7.1 Technical Limitations & Proposed Solutions

This section outlines technical limitations of the used setup, effects thereof on SoE and solutions to alleviate these issues.The following technical limitations of the system have been identified and may be considered in the context of SoE, as shown in table 7.1.

7.1.1 Virtuose & Franka

The setup of the Virtuose and Franka included multiple limitations such as difficulties in controlling the wrist and the elbow of the Franka arm, implausible joint states, sth sth dynamics feedback Virtuose, inadequate scaling between Franka and Virtuose movements as well as insufficient gravity compensation.

Wrist

Multiple participants reported that they had issues turning the wrist of the robot arm. This was caused by a mistake in the Virtuose API in which the joint of the Virtuose end-effector was set to the second to last joint rather than the actual last joint. Thus, when trying to rotate the wrist, the remote arm will rotate in an unexpected way, potentially decreasing agency.

Elbow

Additionally, participants reported that they had no control over the "elbow" joint of the remote arm which was frustrating when they wanted to maneuver around the pegs. Indeed, participants could not actively control this joint, instead they only had control over the end-effector, whereas the other joints moved following the principle of inverse kinematics to guide the end-effector to its desired location and orientation. To alleviate this problem, the current setup could be extended with a wireless motion tracking system of the human arm as done in [235] and and impose kinematic constraints to the control architecture such that the user can also control the elbow of the arm. Moreover, this setup could also be used for alleviating the next problem:

Implausible Joints

Participants reported that sometimes the robot arm would execute movements with joint configurations that are unnatural, that is, they are implausible for the anatomy of the human arm. Whenever this occurred, it would make it apparent that the remote arm is not a human arm, breaking the illusion of ownership. While multiple studies have shown that body representation is flexible and that ownership can be experienced for anatomically implausible [236, 237] or even non-bodily objects [238–240], these studies did not suddenly change the appearance or plausibility of the hand during the experiment, instead first induce embodiment typically via synchronous visuo-tacile or visuo-proprioceptive stimulation [85] and then measured ownership[1]. A key difference to present study however is that the joint configuration of the robot arm suddenly changed during the experiment and thus broke the illusion of embodiment. In context of the predictive coding frame work, this corresponds to a large predictive error that cannot be easily explained away and thus updates the body-schema to a new model in which the remote is not perceived as one's own arm. This implies that for scenarios where ownership

[1] Moreover, ownership ratings were generally higher the more they resembled the human anatomy and likeliness.

System Component	Limitation	Affected SoE components	Proposed Solution
Virtuose & Franka	Inaccurate control of wrists	SoA, SoO	Correct implementation
	No control of Elbow	SoA, SoO	Position tacker
	Implausible joint positions	SoO	Kinematic constraints based anthropomorphic possible configurations
	Virtuose haptic feedback in free space	SoO, SoA	Force-Torque sensor
	Limited workspace caused by insufficient scaling between local and remote arm	SoA, SoO	Adjust scaling
QB hand & HGlove	No solid force feedback of QB hand	SoO	Increase gain & Tune dead band controller or use different hardware
	QB hand grasping capabilities	SoA, SoO	Different hardware
	Loose brackets of HGlove	SoO	Different brackets or different gloves
Other	No head movement control	SoA, SoS	Predict and map head movements to camera
	Audio cues	SoA, SoO	Noise cancelling headphones

Table 7.1: Technical limitations of the used setup, effects thereof on SoE and proposed solutions.

is desired over a robotic arm (e.g. remote social interaction via a humanoid robot) that the remote arm should not suddenly be driven into anatomically implausible configurations. To avoid such instances, one could modify the controller of the arm such that it not only drives the end-effector to a desired pose but also imposes constrained based on anthropomorphic plausible movements. Efforts for such controller schemes have been made in [241] and [242], where intra-arm and inter-arm coordination characteristics have been replicated for a specific set of tasks. However, these studies did not evaluate as how natural these movements were perceived by participants. It would thus be interesting to investigate the impact of anthropomorphic plausible arm movements on SoO versus non-anthropomorphic plausible arm movements. Additionally, one could also consider the impact of velocity and acceleration of the remote arm movements [243] on SoO.

Dynamics

Another theme that emerged from the IPA analysis was that the force feedback of the Virtuose felt stronger for participants when moving the arm to the right side as compared to the left side. This behaviour of the arm was caused by the haptic feedback controller which reflects the dynamics of the robot arm back to the user. Specifically, the internal friction of the joints of the remote arm got reflected when moving in free space, thus rendering haptic feedback although the user is not interacting with any object. In the present study, a very specific robot pose was used to create the illusion of a connected arm from either a 1PP or 3PP. This meant that when moving the arm inwards (left), less joints are being moved and thus less internal friction is caused as compared to moving the arm outwards (right), for which significantly more forces may be rendered. Several participants reported that it felt generally very difficult to move the arm to the right side when grasping for pegs or putting a peg back into its starting position in order to grasp another peg. Some participants mentioned that for very large forces their SoO got disrupted as they were pushing against a force although they could see the robot arm wasn't touching any objects in the remote environment. This visuo-motor conflict thus may have decreased the overall ownership and agency of both embodiment groups and may present a general limitation of such passivity based controllers in terms of transparency [178].

Recently, Lenz and Behnke [244] used a force-torque sensor to measure the forces and torques applied to the wrists of robot arm in order to generate a weightless feeling when moving the arm in free space. However, note that they do not consider the effect of time delays on the stability of the robot, which becomes as such unstable for latencies higher than 50 ms [245]. An alternative solution would be to use passivity layers as in [8,177] and use the forces and torques measured by the sensor as force input to the passivity layer rather than the forces measured via joints of the remote robot[2].

Scaling between Virtuose and Franka Movements

Another limiting factor was that sometimes the participants were not able to reach the upper left peg and needed to be clutched out and in again. Porssut et al. [247] analysed the impact of articular limits on SoO when performing a reaching movement and showed that reaching an articular limit while the avatar's arm is not fully extended leads to a visuo-proprioceptive conflict which disrupts the illusion of embodiment and thus decreases SoO. A similar effect could be observed in the present study as several participants reported decreased agency and ownership levels when they needed to be clutched out and in again. For future studies attention should be paid to proper calibration between the scaling of remote and local movements.

[2]This method is currently being tested by i-Botics team in the ANA Avatar XPRIZE challenge [246] and under active development.

7.1.2 HGlove & QB Softhand

Another set of drawbacks impacting SoE can be found in the implementation of the QB Softhand and HGlove, namely the force feedback and the design of the finger brackets of the HGlove.

Force Feedback

Participants reported that the force feedback that was rendered via the HGlove was not perceived as a solid force as they could grasp through the force feedback if they wanted to. This may have lead to a visuo-proprioceptive conflict as they could close their hand further than the QB softhand in the remote scene. For instance, they could fully enclose a peg with the remote hand but they could close their own hand even further. As a consequence, participants could feel their own fingers touching when clenching their fist. This effect may have had a similar impact on SoO as the reaching limits described by Porssut et al. [247].

The force feedback of this setup may have been not strong enough due to technical limitations. The control signal for the external torque that is rendered to the HGlove is dependent on the current that is drawn by the motor as well as the grasping velocity of the QB Softhand [173]. The latter is obtained by taking the derivative of the position signal of the Softhand's motor and smoothing the resulting velocity data with a moving average filter to reduce the noise of the data. The velocity is used to determine whether the QB Softhand is opening or closing, as different friction parameters have to be used for the respective movement. Here, the controller simply checks whether the velocity is bigger or greater than zero. A noisy signal velocity signal when grasping an object creates a "switching" effect, where the local hand switches rapidly between applying forces for opening and closing, as the velocity is very low when an object is being grasping (near zero). This presents a challenge as one can choose to increase the gain of the force feedback to render a stronger, more solid force, however this also amplifies noise present in the velocity signal, creating stronger switching effects and lead to rapid changes in force feedback. For the present study, the maximum amount of filtering and gain has been used, to create the largest possible forces while maintaining an acceptable level of delay for the user. Additionally, a dead band controller has been added to system as described in [24], to deal with the switching behaviour at low velocities. While this controller significantly reduced switching behaviour, it appears that it not fully resolved the problem and that velocity limits of the dead band controller should be adjusted. If done properly, this should allow for increasing the gain to render stronger haptic feedback to the operator.

Finger Brackets

Another limitation surrounding the setup of the HGlove and the QB Softhand were the brackets of the finger mounts of the HGlove. These brackets would come loose sometimes during the trials and needed to be adjusted by the participants themselves or the experimenter, thus interrupting the experiment and possibly breaking the illusion of SoE. For the supportive group (force feedback on), the haptic feedback of the hand would sometimes become unstable and thus requiring to adjust the brackets again, which is disadvantage compared to the suppressive group (force feedback off), as they would have less interruptions. Furthermore, the brackets of the HGlove pull on your fingers, rather than them pushing against your fingers; some participants mentioned that this felt unnatural (thus potentially decreasing SoO as compared to suppressive group).

7.1.3 Other

This section contains other aspect of the experimental setup that may have impacted the results of the study. Specifically, this includes the inability to control the position and orientation

of the remote camera as well as auditory cues that could have been used to determine proper insertion of the pegs.

Head Movement Control

In the present study it was not possible to control the movements the camera with your head, as the position of the camera was fixed. Ventre-Dominey et al. [248] compared amongst others the effect of congruent and static head movement control over a robot on the perceived embodiment. They found that SoS and SoA significantly increased in the congruent condition as compared to the static condition; no significant changes could be observed for ownership (enfacement). Here, the sense of agency was stronger than the sense of localization. The increase in agency is in line with previous studies in the RHI paradigm and their VR related counterparts where congruent movements between real and artificial body-parts induce a sense of agency [**?**, 34, 54]. While location in their study referred to the sense of feeling relocated in another body, the spatial awareness of the users might have also been increased due to motion parallax caused by moving the head, thus improving the depth perception of the user. In a follow up study by Farizon et al. [249] compared visuo-motor (a)synchrony of head movements as induction method for SoE. Their results show that congruent head movements lead to significant sensation of agency and location as compared to incongruent head movements. Similar to [248] Ownership (enfacement) increased as well during the congruent visuo-motor stimulation however not significantly. Taken together, the results from [248] and [249] suggest that head movement control increases SoA and SoS, but not SoO.

Some participants reported that they felt the urge to move their head. An urge to move your head might indicate a lack in SoA or SoS. An open question is how strong the contribution of head movement control to SoO and SoA is as compared to other perceptual cues? An additional question could be how to reduce the delay between tracking of head movements (see e.g. [250, 251]) and actuating the remote camera in order to avoid motion sickness. One possible solution may be to predict the head movements of the user based on salient points in the scene as done in [252] and forward the predicted head poses to the head movement controller.

Audio Cues

In the present study, participants were able to use auditory cues from the remote environment to determine whether a peg was properly inserted into its respective hole. This may have benefited participants of the suppressive group which were limited to monocular depth cues and generally reported difficulties for estimating visually determing whether the peg was properly inserted or not. Next to a possible effect on task performance, this may also have affected embodiment. Radziun and Ehrsson [253] investigated the effect of auditory cues on ownership and proprioceptive drift in four different conditions in the RHI: (a)synchronous touches without auditory cues and synchronous touches with (a)synchronous auditory cues. The authors observed that synchronous auditory cues created a stronger embodiment illusion which was reflected in higher ownership ratings as well as higher proprioceptive drift. These results support the findings of Darnai et al. [254] in which auditory cues increased proprioceptive drift as compared to no auditory cues. Radziun and Ehrsson postulate that auditory cues are "used in the formation of the coherent multisensory representation of one's own body". Taken together, the results of Radziun and Ehrsson and Darnai et al. suggest that auditory cues may have increased the ownership levels of both embodiment groups in the present study. While the contribution of this effect is not known, one may choose for future studies to give the suppressive group noise cancelling headphones and the supportive groups no headphones (or headphones with audio depending on the experimental setup), to create larger differences in embodiment and remove potential bias on task performance.

7.2 Lessons Learned

Several experimental flaws could be observed in the present study, specifically regarding experimental design affecting embodiment and methods used to measure embodiment (and cognitive workload). This section discusses the lessons learned and proposes how the present study could be redone.

7.2.1 Manipulating Embodiment

While visuo-proprioceptive stimulation is a powerful perceptual cue to manipulate embodiment (as discussed in section 2.1.2), it has not been included in the present study in form of delay as there was not enough time for a pilot study to explore suitable delay values that would allow for a stable system while not biasing task performance. When adding delay to a real telemanipulation system it should be considered how control architectures that are meant to keep the system stable, will affect transparency and task performance. In case of the passivity layer based approaches [8, 177] as used in in the present study, the excess energy introduced by the delay is being dissipated by inserting additional damping into the system which is also reflected in the haptic feedback of the master device, requiring the operator to use more force to move the haptic device. In that case, it would not be clear if adding delay reduces task performance as consequence of lower embodiment or because participants need to physically use more effort to control the master device. Additionally, delay may affect task performance when considering time based measures. For instance, in the current study participants reported that due to delay they had to stop and wait for the hand to rest before making a new movement as to prevent overshooting movements.

However, in retrospect, it appears that both embodiment groups having the same level of visuo-proprioceptive stimulation contributed to a large part to the lacking differences in SoO and SoA. Waltemate et al. [255] investigated the effect of delay on motor performance, SoO, SoA and perceived delay for a complex full body task in VR. In their literature review, the authors highlight that there is no golden rule for specifying the highest acceptable delay for a desired amunt of motor and perceptual performance, as several studies involving different setups and sensorimotor tasks draw different conclusions for the perceptual threshold of latency detection. The authors observe that delays above 75 ms significantly reduced motor performance whereas perceived agency and ownership significantly declined from delay values between 120 ms and 210 ms. Additionally, their findings suggest that motor performance accuracy may predict perceptual judgments over ownership and agency better than delay: For a high number of performance errors, smaller delays may be perceived as non-simultaneous whereas for a low number of errors, higher delays might go unnoticed. The authors postulate that the same delay value may or may not deteriorate perceptual judgments depending on the complexity of the task. For simple tasks the delay might go unnoticed, while for more complex tasks leading to higher chances of performance errors, delays might exert a stronger impact on perceived agency, ownership, and simultaneity.

Another experimental flaw was that participants could use auditory cues from the remote environment to determine whether a peg was properly inserted or not. This may have benefited especially the suppressive embodiment group as they had to rely on monocular depth cues and reported difficulties to visually determine the proper insertion. Additionally, auditory cues may have also increased the ownership levels of the suppressive embodiment group [253, 254]. This could have been easily avoided by using noise cancelling headphones. Moreover, one could choose to provide auditory cues only to the supportive embodiment group in order to increase their ownership levels even further. That said, the contribution of auditory effects on task performance and ownership is unknown for the present study.

7.2.2 Measuring embodiment and cognitive load

Another set of flaws could be observed for measuring embodiment and cognitive load. Physiological measures (pupil dilation, heart rate and skin conductance) have been measured in the present study as indicator for 1) embodiment and 2) cognitive load (as predictor for learning free energy), as shown in appendix D. Ultimately these measures have not been further used for analysis due to flawed methodology as 1) no base line measures have been created, 2) no subjective ratings for workload have been collected and 3) no threat has been inserted into the experiment to create visible spikes in the data as measure of embodiment.

As no base line measurement has been collected for any of the physiological measures before the experiment, it is not clear how significant the observed values are as compared to the resting state. Even if there would have been differences observed between groups, it would have been unclear whether this difference is relevant.

Furthermore, physiological data has been collected as index for learning, however to measure learning behaviour as a function of physiological measurements, subjective measures (e.g. the NASA TLX survey [256]) should be employed as reference measurement. This is important, as physiological measurements may not be a suitable method for measuring learning behaviour [257] as 1) even though increased skills reduced the required amount of effort to fulfill the task, participants may chose to increase task performance even further, rather than reduce cognitive load; 2) cognitive workload as induced by the peg and hole task could have been much stronger than the effect on cognitive workload as induced by learning and 3) the effect of SoE manipulations is different than the effect of learning on cognitive workload.

To measure SoE via physiological measurements, visible spikes have to be created in these measurements in order compare the different levels of SoE [6]. This is typically done by inserting a threat towards the end of the experiment, a classic example would be the RHI, where a knife is stabbed into the rubber hand. In the present study no such threat has been used, instead just continuous measurements have been used, which revealed no significant changes over time. It is not clear whether the lack in differences is due to a small difference in embodiment or because of the absence of the threat.

Another flaw in measuring embodiment may have been the methodology of the proprioceptive drift measurement. In the position judgment task, the arm of the participants has been "randomly" displaced by the experimenter, however it is questionable to what extent this was purely random. This could be alleviated in future studies by displacing the arm via the Virtuose rather than the experimenter. Here, the Virtuose could move the participant's arm randomly to a set of predefined positions in order to avoid any bias caused by experimenter.

7.2.3 Reiteration Proposal

If I would try to answer my initial research question again I would propose to do it in three phases:

1. Fix relevant technical issues and if feasible, extend the current setup,
2. Conduct a pilot with the current setup to verify the contribution of perceptual manipulations on embodiment,
3. Select the best perceptual cues from phase 2 and repeat the current study with the improved setup.

First, the most relevant technical issues of the current setup should be fixed. This includes for instance proper scaling between the Virtuose and Franka movements to cover the complete remote workspace and avoid disrupting the embodiment illusion by clutching participants out and in. Moreover, the force sensor that is currently used by the i-Botics team in the Xprize com-

petition could be used to allow for weightless operation in free space and remove unnecessary haptic feedback. This is feasible as the present study has been executed as part of the i-Botics teams, thus the force sensor and the accompanying software stack is readily available. Furthermore, the maximum velocity and acceleration of the remote arm may be limited based on [243] in order to avoid disruption of embodiment due to sudden movements of the remote arm. Additionally, the system could be improved by implementing the arm tracking and extending the controller to constrain movements along anatomically plausible poses as in [241, 242]. Furthermore, the current hand control could be improved by replacing the current hardware on the operator side as well on the remote side: The HGlove could be replaced with an exoskeleton that allows for tactile feedback such as the SenseGlove Nova [258] or the Haptx DK2 [259]. These devices have additionally the advantage that they can provide force feedback to more than just three fingers and they avoid the problem of finger brackets becoming loose as both devices are gloves that enclose the entire hand. In order to make use of the tactile feedback, the QB Softhand could be extended with tactile sensors on its fingertips as done in [260] or it could be replaced with a robotic hand that can sense pressure such as the Psyonic Hand [261] which additionally allows to control all fingers individually rather than just closing and opening of the hand. The added tactile feedback could then increase perceived ownership and agency [35, 74]. Another addition to the system that would be interesting to investigate is adding the ability to control the pose of the remote camera by tracking the head movements of the operator and mapping them to the camera in the remote scene as done in [250, 251]. To alleviate the effects of delay on motion sickness, the movements of the operator could be predicted by the method of Lee et al. [252] and forwarded to the camera control system. Here, one could compare synchronous to asynchronous or static camera movement to create different levels of SoO and SoA, as shown in [248, 249]. That said, such alterations to the system would take a significant amount of time to properly implement and might be out of scope for the actual research question. In case the HGlove and QB Softhand are not replaced by other hardware, the brackets of the HGlove should be redesigned such that they don't slip off and the dead band controller should be tuned to avoid the switching behaviour for low velocities. Finally, noise cancelling headphones should be used to remove any auditory cues from the remove environment. During the experiment, the participants could be in a Skype call (or similar) with the experimenter, to allow for communication in case of emergency.

Second, a pilot study should be conducted that explores in a factorial within-subjects design the individual contribution of the perceptual cues used to manipulate SoO, SoA and SoS, similar to [4]. As this pilot solely focuses on embodiment and not on task performance or learning effect, it can be conducted in a within design, allowing to investigate a wider range of perceptual cues while reducing idiosyncratic effects. Next to the cues used in the present study (Hand force feedback, perspective, depth perception), also the effect of (a)synchronous visuoproprioceptive stimulation (i.e., different delay values) should be investigated. Additionally, the cues could be extended (with (a)synchronous) arm tracking & anthropomorphic constraints, tactile feedback and head tracking, depending on the made improvements of the first phase. Proprioceptive drift could be measured again with the position judgement task as done in the present study, however the arm of the participant should be displaced rather by the Virtuose than by experimenter, to allow for more random displacements. Next to surveys, interviews and proprioceptive drift, physiological measures should be included to measure SoE such as skin conductance, heart rate and pupil dilation. In this case, base line measurements at rest state should be created before and after the experiment, subjective workload questionnaires should be administered in order to correlate them to the physiological measurements, and finally, a threat should be inserted at the end of the experiment to measure relevant spikes in the selected measure. Based on the results, the best combination of perceptual cues can be used to create a greater difference in SoE for the supportive and suppressive embodiment group.

Finally, in the last phase, the current study should be repeated with the improved setup and the selected perceptual cues. Here, the peg & hole task should be altered to obtain a more valid efficiency measure. For instance, one could use a similar setup as in [61], where a larger number of pegs can be inserted per trial and participants should be instructed to insert as many pegs as possible for a given time.

7.3 Further Research

This section proposes further future studies alongside their associated research questions.

7.3.1 QB Softhand Pose Tracking

A follow up study on the pose tracking controller that was implemented in [173] could be conducted and the effect it has on embodiment and task performance. Results of the interviews suggest that the pose tracker subconsciously helped the user in controlling of the QB Softhand. It would be interesting to formally validate this in a within-subjects design study where various degrees of stiffness in the feedback are used, that is, the amount of force that is being used to guide the user along the first synergy of the QB Softhand.

Moreover, it would be interesting to investigate to what extent the pose tracking feedback limits the agency over the hand as too strong feedback would require the operator to use considerable amount of effort to push against the mounts of the HGlove, thus reducing transparency. On the other hand, it has been shown that assisted help (e.g. haptic feedback) can increase task performance by ignoring erroneous commands of the user, and lead to higher perceived agency although actual agency is reduced over the device [262]. Possible research question could be thus "How does the stiffness of the pose tracking feedback affect the intuitiveness of the system" and "How does the stiffness of the pose tracking feedback affect the agency over the system". To answer these questions, a user study could be employed in which a wide range of objects, has to be grasped. For instance, a box and place task with different objects could be used or puzzle box task with different kinds of puzzle pieces. To measure usability, one could use the usability scale of the USUS survey [263]. To measure agency, one could use the Likert items used in the present study or use embodiment surveys of related robotic hand studies [35, 36, 74]. The surveys should be administered after every condition in order to prevent recall bias. Additionally, one could use proprioceptive drift or skin conductance as an implicit measure. The findings could be of significance for designing control architectures of underactuated robotic hands to increase intuitiveness and transparency. Moreover, the effect of the pose tracking feedback on intuitiveness and agency may transfer to other robotic hands that are not underactuated but also potentially limited in the movement capabilities they provide.

7.3.2 Depth Perception as Embodiment Manipulation Tool

In the present study, depth perception has been used a novel perceptual cue to manipulate SoS by either streaming both images or just one image of a stereo camera to an HMD. Based on results of the current study, depth perception may be a promising perceptual cue to manipulate SoS. As the present study only compared the joined contribution of perspective and depth perception, it would be therefore interesting to compare the individual contribution of perspective (1PP vs 3PP) and depth perception (binocular vs monocular depth cues) in a 2x2 factorial study. The research question for this study could be formulated as "To what extent does perspective affect SoS as compared to depth perception"? To answer this question, the same setup and task as in the present study could be used. To measure the effect of the perceptual cues, one could use the SoS questions from the current study or the questions from [41]. Additionally, one could use proprioceptive drift or skin conductance as an implicit measure. Depending on the contribution of depth perception on the perceived SoS, it may present an effective manipulation for embodiment that is straightforward to implement in VR and telerobotics.

A Appendix: Ethical Documents

This study has been approved by the ethics committee Computer and Information Science of the University of Twente, under the reference number RP 2021-224. Below you can find the information brochure and the consent form that has been given to the participants.

A.1 Information Brochure

Information Brochure

Introduction:

Dear participant, thank you for being interested in partaking in this study. Tele-robotics is the remote control of robotic devices that aims to combine the physical strengths of robots and the cognitive abilities of humans over an arbitrary distance to a remote environment that is difficult to reach or dangerous for humans. Use cases can be found in e.g. rescue & search missions, surgeries, social interactions or inspection & maintenance.

In this study you will operate a tele-robotic setup where you will control a robotic arm & hand via a hand exoskeleton in a remote environment in order to execute a set of user tasks.
The exoskeleton (Hglove) will allow you to feel haptic feedback based on grasping motions of a robot hand that is attached to the robot arm. For instance, if you would grasp for a bottle in the remote environment, you would feel similar forces exerted by the exoskeleton.

Additionally, the exoskeleton is attached to a haptic 6 degrees of freedom device (Virtuose 6D), which allows you to feel the dynamics of the robot arm. For example, if you would grasp the bottle mentioned above and slam it against a table, your arm movement would be stopped mid-air, once the robot arm reaches the table.

During the experiment you will be wearing a Head Mounted Headset (HMD) to view the remote environment. Additionally you will be wearing an Empatica E4 wristband which will measure physiological data such as heart rate or skin conductance. You can see the setup that is going to be used below in figures 1 and 2.

The expected total duration of the experiment is approximately between 1 and 1.5 hours.

The rest of this brochure is structured as follows:

1. Requirements for participation are being listed;
2. The general procedure of the experiment is outlined;
3. Possible risks and safety measures are being explained;
4. Data collection and storage is discussed;
5. Compensation for your participation;
6. Contact details.

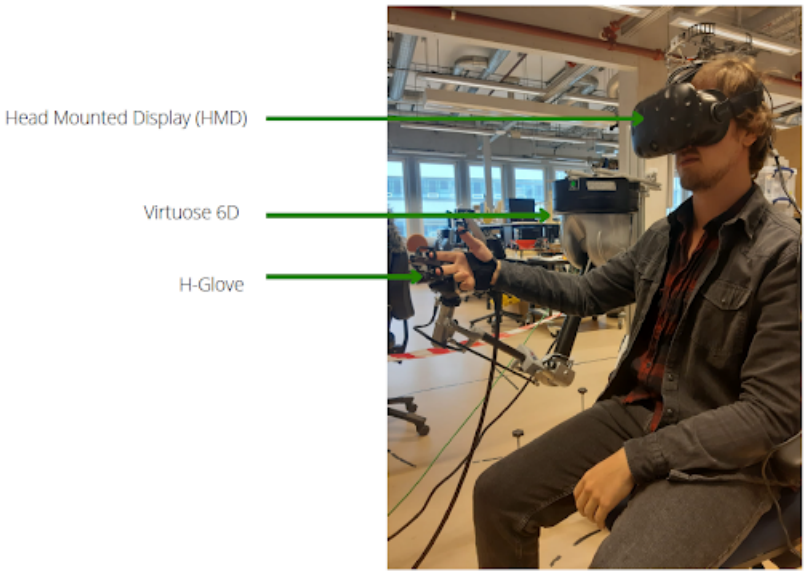


Figure 1: Operator wearing haptic devices and HMD.

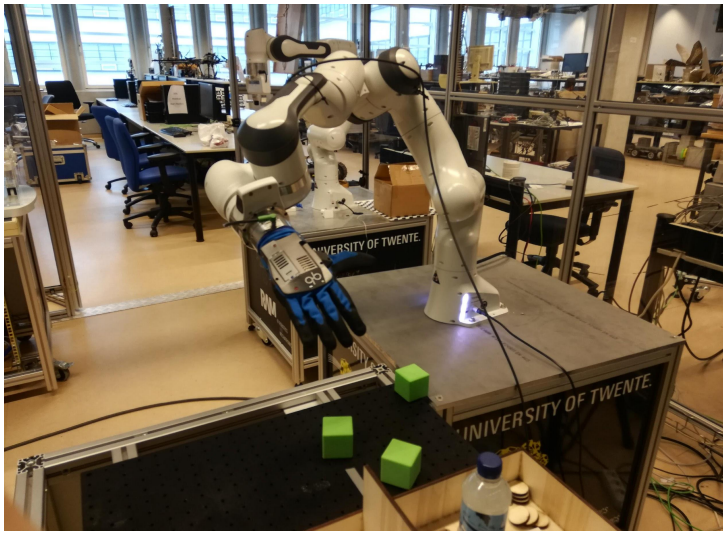


Figure 2: Robot arm with robot hand.

Participant Requirements:

In order to participate in this study you must full fill the following conditions:

- At least 18 years old
- Speak fluently English
- Right handed
- No upper body injuries
- Normal to normative vision
- You have not participated in a pilot of this study
- Not susceptible to motion sickness in VR

What you will do:

1. You are asked to fill in a demographics questionnaire;
2. At the beginning of the experiment, there will be a hand calibration procedure which requires you to open and close your right hand when wearing the exoskeleton.
3. Additionally there will be a short eye calibration procedure, which will map your pupil dilation to various brightness levels when wearing the HMD.
4. Next, there will be a proprioceptive drift measurement. Your hand will be displaced from an experimenter. You are then asked to move your hand back to its initial location. This task will be repeated 5 times.
5. Afterwards, the main task of the experiment will be conducted, which is a Peg & Hole task. This task requires you to grab a set of pegs from one mounting plate and place them into the holes of another mounting plate. This task will be repeated for four trials. We will explain the tasks in detail before and during the experiment.
6. After each trial, you will be asked to verbally respond to a short questionnaire about your experience. You will be able to see the survey in the HMD.
7. Afterwards all Peg & Hole trials are completed, step 4 is repeated.
8. Finally a brief interview will be conducted, asking you questions about experience.

What we will do:

1. We will provide you the instructions and information needed to accomplish the tasks;
2. We will set the equipment on you (Exoskeleton, HMD, E4 wristband);
3. We will instruct you on how to use the tele-robotic setup;
4. We will answer to all your questions;
5. We will take care that the experimental session will be carried out in safety and with respect to the COVID-19 measures.

Possible risks:

- Unexpected motions, force jumps, instability
  - Caused by the haptic feedback device attached to your hand
  - Caused by the haptic feedback device attached to your arm
- Nausea caused by the HMD

Safety Measures:

- Unexpected motions, force jumps, instability of the haptic feedback devices
  - There is a central emergency button for both haptic feedback devices as well as an emergency button for the robot arm
  - You can only move the robot arm and experience the corresponding haptic feedback when the experimenter activates this functionality. Next to the central emergency button mentioned above, the experimenter can disable haptic feedback and movement capabilities of the system via a GUI.
  - There are force limits in the haptic feedback devices to ensure safe operation at all times.
  - There are 2 researchers present at all times.
- Nausea caused by the HMD
  - The camera view which will be rendered inside the HMD is static, that is, the camera itself is not moving, but you can move your head freely. As a result, the scene itself is not changing significantly in terms of position or orientation, mitigating the chance of becoming motion sick.
  - You can take a break at any point needed and then continue if you want to.

What data will we collect:

We collect your questionnaire responses, trajectory data of the end-effector, pupil dilation, data from an empatica E4 wristband (e.g. heart pulse rate, skin conductance), task performance measures (e.g. completion time) and audio recordings of the interviews. All of the data will be anonymised, such that no personally identifiable data will be stored.

Note that pupil dilation data and data from the E4 wristband can allow accidental diagnostic findings. For instance, unusual changes in pupil dilation might be a symptom of an injury to the brain from physical trauma or a stroke, but may also be related to other factors such as drug use. Additionally, data from the empatica E4 can allow accidental diagnostic findings for e.g. arrhythmia. Note that the researcher is not a trained medical professional who can interpret such data but if unusual changes in the physiological data is being detected, you will be informed about this.

Compensation:

You will receive a 10 Euro bol.com voucher as compensation for your participation. You have the right to withdraw from the experiment at any point without having to give a reason and also then you will receive the voucher.

Feel free to contact me or my supervisor for any additional questions.

Principal Researcher:
Nils Rublein
+49 15774507457
n.rublein@student.utwente.nl

Supervisor:
Dr. Douwe Dresscher
+31 534892047
d.dresscher@utwente.nl

You can email the ethics committee for complaints or independent advice.
ethicscommittee-cis@utwente.nl

A.2 Consent Form

Consent Form

The research concerning this consent form involves the operation of a tele-robotic setup where the participant is asked to perform various user tasks by manipulating a robotic arm via a hand exoskeleton that is attached to a haptic feedback device. Additionally, the participant is wearing a Head Mounted Display (HMD) to view the remote environment of the robot arm.

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the study information in the brochure or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	□	□
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	□	□
I understand that taking part in the study involves: 1. Executing a teleoperation task that requires me to wear a HMD and a set of haptic devices; 2. I will verbally complete survey questionnaires and will take part in interviews, and that my answers will be collected and analyzed; 3. End-effector trajectory data will be collected and analyzed; 4. Pupil dilation, heart rate and skin conductance will be collected and analyzed; 5. Audio recordings (interview) will be collected and analyzed 6. Any personalized data will be anonymized and used for research and will be publicly available 7. I have the right to withdraw this consent without the need to give any reason within 48 hours of signing the consent.	□	□
Risks		
I am aware of the following risks and methods used to mitigate them (see information brochure) ● Unexpected motions, force jumps, instability caused by the haptic feedback devices ● Nausea caused by the HMD	□	□
Use of the data		
I understand that data collected from the study may lead to accidental medical diagnostics (e.g. arrhythmia).	□	□
I understand that personal information collected about me that can identify me [e.g. my name], will not be shared beyond the study team.	□	□
I understand that all physiological data will be destroyed once it has been analyzed.	□	□

Consent to be Audio Recorded

I agree to be audio recorded during the interview. □ □

Future use and reuse of the information by others [Optional]

I give permission for the questionnaire and interview answers as well as the end-effector trajectory data that I provide to be archived so it can be used for future research and learning. □ □

I agree that my information may be shared with other researchers for future research studies that may be similar to this study or may be completely different. The information shared with other researchers will not include any information that can directly identify me. Researchers will not contact me for additional permission to use this information. □ □

Signatures

_____	_____	_____
Participant name	Signature	Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

Nils Rublein_____	_____	_____
Researcher Name	Signature	Date

If I request further information about the research, now or in the future, I may contact:

Principal Researcher:
Nils Rublein
+49 15774507457
n.rublein@student.utwente.nl

Supervisor:
Dr. Douwe Dresscher
+31 534892047
d.dresscher@utwente.nl

If you have any complaints about this research, please direct them to the secretary of the Ethics Committee of the Faculty of Electrical Engineering, Mathematics and Computer Science at the University of Twente, P.O. Box 217, 7500 AE Enschede (NL), email: ethics-comm-ewi@utwente.nl).

B Appendix: User Materials

This appendix includes materials for the user, that is the interviews and surveys that have been used.

B.1 Survey Questions

B.1.1 Demographics Survey

What is your age?

- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65 and older

What is your gender?

- Male
- Female
- Other

Did you have in past or currently have any upper body injuries?

- Yes
- No

How much do play video games on a weekly basis?

- I don't play video games at all
- Less than on hour
- Between 1 and 3 hours
- Between 3 and 5 hours
- More than 5 hours

Have you experienced virtual reality with a head mounted display?

- Never
- Once
- Sometimes
- Regularly

Have you become nauseous in VR?

- No, as I have never experienced VR before
- Never
- Once
- Sometimes
- Regularly

Have you experienced a telerobotic interaction before?

- Never
- Once
- Sometimes
- Regularly

B.1.2 SoE Survey

The following questions will be asked verbally to the participant on a seven level likert scale (1 = strongly disagree, 7 = strongly agree). There are two questions per SoE component, where one of the questions is a control question. These questions will be administered in situ after each trial, where the order of the questions is randomized.

- SoO: It seemed like the robot hand was my hand
- SoO: It felt as if the touch I felt was not caused by the robot hand
- SoA: The robot hand moved like I wanted it to, like it obeyed my will / as if it was my real/own hand
- SoA: It seemed as if the robot hand had a will of its own
- SoS: It seemed like my hand was in the location where the robot hand was
- SoS: I had difficulties locating my hand in the remote environment

B.2 Interview Questions

The interview is semi-structured and follows a 5 phase structure [45]. Below each phase of the interview can be found with the goal of the respective phase and its corresponding questions.

Phase 1: Ice Breaker
Get participant talking, put participants at ease, create rapport.

- You are finally done, how do you feel?

Phase 2: Introduction
Bring up topic, shift focus toward research questions. Start with easy, non-threatening questions.

- How would you describe the overall interaction/experience
- How did it feel when you first entered the system (e.g. put on HMD, moved with the robot arm, etc.)

Phase 3: Key Questions
Consider key questions within a broader perspective w.r.t to the main research question.

- SoO: When you were controlling the hand, did it feel as part of your body?
- SoA: How would you describe the movements of the hand? / the control you had over the hand?
- SoA: Did you feel the arm/hand moved as you wanted it to?
- SoS: Did you feel you had a good perception of the space in that the hand was moving?
- SoE & Learning Curve: In the beginning of the interview you said, you felt *XYZ* when you entered the system. How did you feel over time / how would you describe your adjustment to the system over time?
- Task Performance & Learning Curve: How would you describe your task performance over time?

Phase 4: Cooling off
"Your interview may have been intense with very detailed questions. At this point, you may want to pull back and ask more general questions or summarize the interview. Ask any follow-up questions in light of the entire interview."

- What was the most difficult/annoying aspect? What would you improve on the system for a more intuitive interaction?
- Is there any sensation or anything else that you experienced which we did not ask you during the experiment that you would like to share with us?

Phase 5: Wrap Up
Bring closure to the discussion, debrief on study purposes.

- Explain concept of SoE, conditions, relation to task performance and learning
- Ask for any questions, comments, feedback

C Appendix: Statistical Analysis

For readability, the OLR model names have been enumerated as follows:

Number	Model
M_1	Cumulative base model
M_2	Category-adjacent
M_3	Unequal var. between conditions
M_4	Unequal var. between items

Table C.1: OLR model numbers.

C.1 SoO Survey

C.1.1 Parameter tables

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept[1]	1.0	1843	-2.9	0.8	-4.4	-3.0	-1.2
b_Intercept[2]	1.0	1636	-1.4	0.7	-2.7	-1.5	0.2
b_Intercept[3]	1.0	1630	-0.8	0.7	-2.1	-0.8	0.8
b_Intercept[4]	1.0	1614	-0.5	0.7	-1.8	-0.6	1.1
b_Intercept[5]	1.0	1715	0.4	0.7	-0.9	0.4	2.0
b_Intercept[6]	1.0	1708	1.7	0.7	0.4	1.7	3.3
b_ConditionSuppressive	1.0	3490	-0.5	0.4	-1.3	-0.5	0.3
b_Trial2	1.0	3150	0.2	0.2	-0.3	0.2	0.6
b_Trial3	1.0	617	0.3	0.2	-0.1	0.3	0.8
b_Trial4	1.0	1427	0.5	0.2	0.0	0.5	0.9
b_ConditionSuppressive:Trial2	1.0	2907	0.0	0.3	-0.6	0.0	0.6
b_ConditionSuppressive:Trial3	1.0	324	0.0	0.3	-0.6	0.0	0.7
b_ConditionSuppressive:Trial4	1.0	270	0.2	0.3	-0.4	0.2	0.9
sd_Item__Intercept	1.0	179	1.4	0.9	0.4	1.2	3.7
sd_Participant__Intercept	1.0	795	0.9	0.2	0.7	0.9	1.3
disc		24000	1.0	0.0	1.0	1.0	1.0
r_Item[Q1,Intercept]	1.0	1662	-0.1	0.7	-1.3	-0.2	1.3
r_Item[Q2,Intercept]	1.0	1696	0.9	0.7	-0.3	0.8	2.4

Table C.2: Parameter table for the cumulative base model.

C.1.2 Convergence Diagnostics

C.1.3 Fitted Response Analysis

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept[1]	1.0	1916	-2.4	1.2	-4.9	-2.4	0.1
b_Intercept[2]	1.0	1509	-1.4	1.0	-3.7	-1.4	0.9
b_Intercept[3]	1.0	1457	-0.4	1.0	-2.7	-0.4	1.9
b_Intercept[4]	1.0	1424	-0.3	1.0	-2.6	-0.3	1.9
b_Intercept[5]	1.0	1431	0.3	1.0	-1.9	0.3	2.6
b_Intercept[6]	1.0	1407	0.9	1.0	-1.3	0.9	3.2
b_Trial2	1.0	5896	0.1	0.2	-0.2	0.1	0.4
b_Trial3	1.0	6270	0.3	0.2	-0.0	0.3	0.6
b_Trial4	1.0	5923	0.4	0.2	0.1	0.4	0.7
b_Trial1:ConditionSuppressive	1.0	897	-1.4	2.9	-7.2	-1.4	4.7
b_Trial2:ConditionSuppressive	1.0	898	-1.4	2.9	-7.1	-1.4	4.7
b_Trial3:ConditionSuppressive	1.0	899	-1.5	2.9	-7.2	-1.5	4.6
b_Trial4:ConditionSuppressive	1.0	899	-1.4	2.9	-7.1	-1.4	4.7
bcs_ConditionSuppressive[1]	1.0	937	1.1	3.0	-5.1	1.2	7.0
bcs_ConditionSuppressive[2]	1.0	905	0.5	2.9	-5.5	0.6	6.2
bcs_ConditionSuppressive[3]	1.0	899	1.0	2.9	-5.0	1.0	6.7
bcs_ConditionSuppressive[4]	1.0	903	1.1	2.9	-4.9	1.2	6.9
bcs_ConditionSuppressive[5]	1.0	894	1.5	2.9	-4.5	1.6	7.3
bcs_ConditionSuppressive[6]	1.0	903	0.7	2.9	-5.3	0.8	6.5
sd_Item__Intercept	1.0	2948	1.3	1.0	0.3	1.0	4.2
sd_Participant__Intercept	1.0	3016	0.7	0.1	0.5	0.7	1.0
disc		14000	1.0	0.0	1.0	1.0	1.0
r_Item[Q1,Intercept]	1.0	1404	-0.3	1.0	-2.6	-0.3	1.9
r_Item[Q2,Intercept]	1.0	1405	0.3	1.0	-1.9	0.3	2.6

Table C.3: Parameter table for the category-adjacent model.

C.2 SoA Survey

C.2.1 Parameter tables

C.2.2 Convergence Diagnostics

C.2.3 Fitted Response Analysis

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept[1]	1.0	816	-3.0	0.7	-4.4	-3.0	-1.4
b_Intercept[2]	1.0	806	-1.5	0.7	-2.8	-1.5	-0.0
b_Intercept[3]	1.0	989	-0.8	0.7	-2.2	-0.9	0.6
b_Intercept[4]	1.0	1190	-0.6	0.7	-1.9	-0.6	0.9
b_Intercept[5]	1.0	1181	0.4	0.7	-1.0	0.3	1.8
b_Intercept[6]	1.0	1246	1.7	0.7	0.3	1.7	3.1
b_ConditionSuppressive	1.0	1369	-0.5	0.4	-1.3	-0.5	0.2
b_Trial2	1.0	474	0.1	0.2	-0.3	0.1	0.6
b_Trial3	1.0	3898	0.3	0.2	-0.1	0.3	0.8
b_Trial4	1.0	3328	0.5	0.2	0.0	0.5	0.9
b_ConditionSuppressive:Trial2	1.0	687	0.0	0.3	-0.6	0.0	0.6
b_ConditionSuppressive:Trial3	1.0	4244	0.0	0.3	-0.6	0.0	0.6
b_ConditionSuppressive:Trial4	1.0	3606	0.2	0.3	-0.4	0.2	0.8
sd_Item__Intercept	1.0	141	1.4	0.9	0.4	1.1	3.6
sd_Participant__Intercept	1.0	164	1.0	0.2	0.7	0.9	1.3
disc		12000	1.0	0.0	1.0	1.0	1.0
r_Item[Q1,Intercept]	1.0	1121	-0.1	0.6	-1.4	-0.1	1.2
r_Item[Q2,Intercept]	1.0	1110	0.9	0.6	-0.4	0.8	2.3

Table C.4: Parameter table for the unequal variances model.

C.3 SoS Survey

C.3.1 Parameter tables

C.3.2 Convergence Diagnostics

C.3.3 Fitted Response Analysis

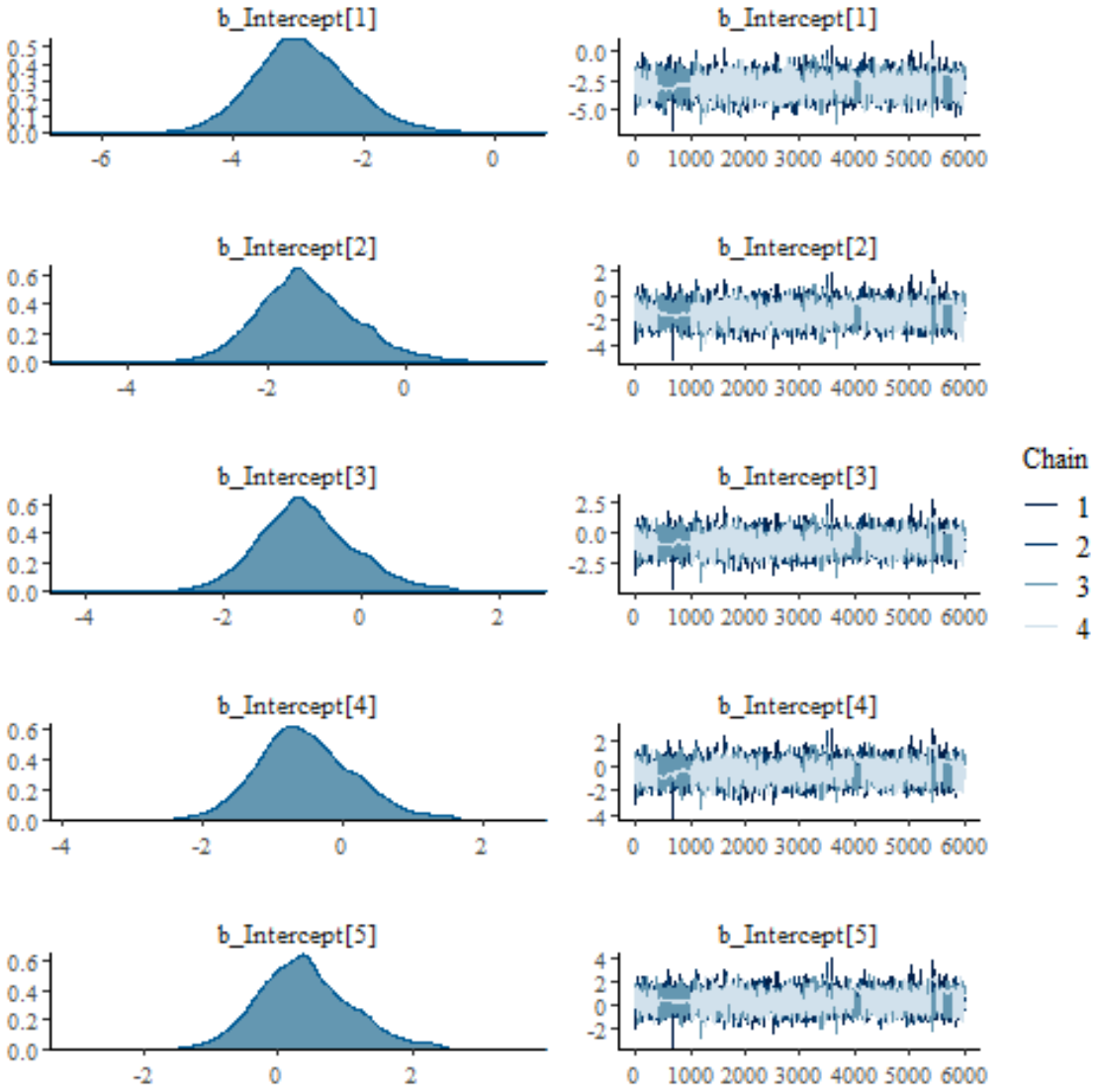


Figure C.1: Density & trace plots for the cumulative model.

C.4 Proprioceptive Drift X Axis

C.4.1 Parameter tables

C.4.2 Convergence Diagnostics

C.4.3 Fitted Response Analysis

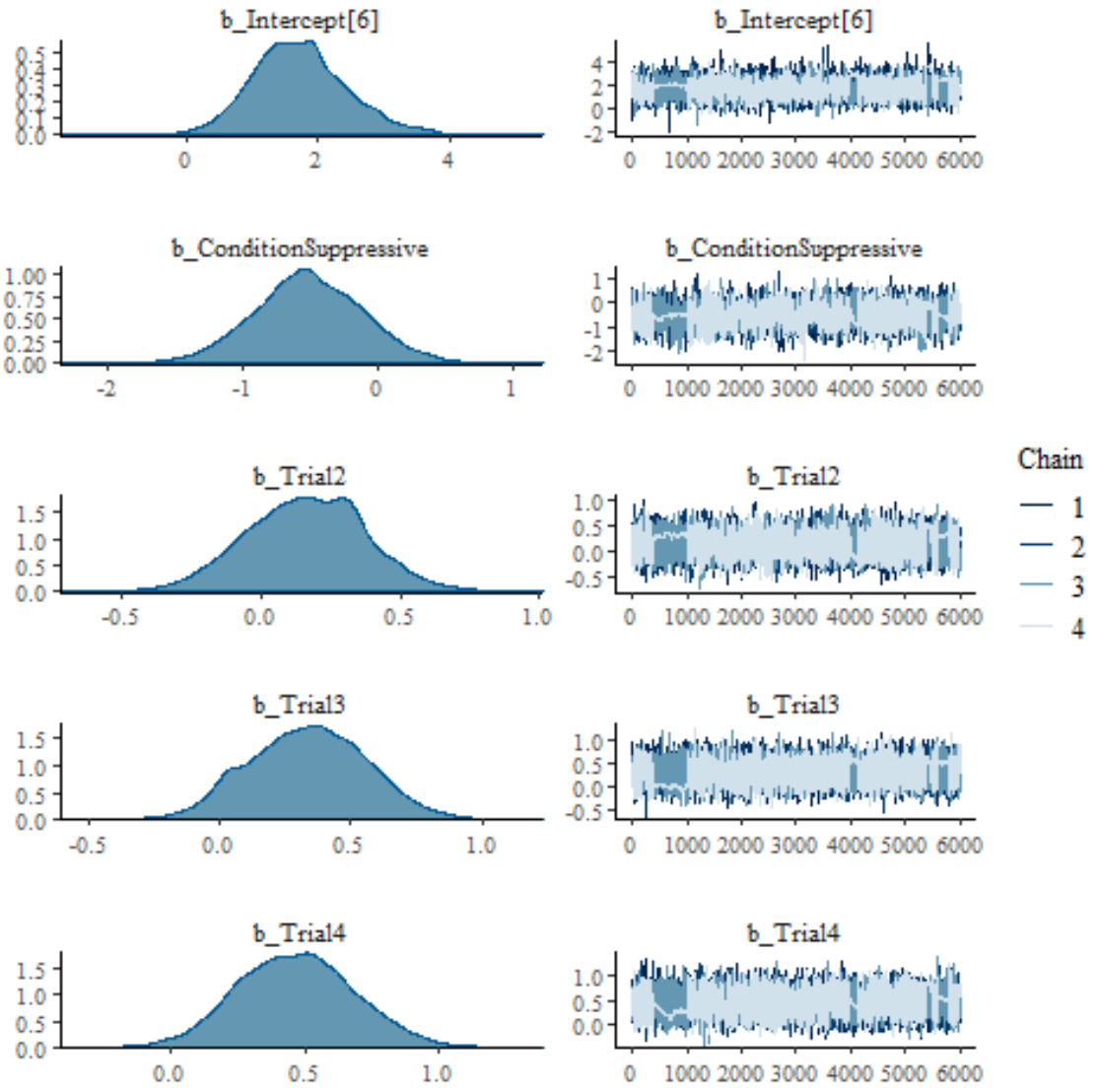


Figure C.2: Density & trace plots for the cumulative model.

C.5 Proprioceptive Drift Y Axis

C.5.1 Parameter tables

C.5.2 Convergence Diagnostics

C.5.3 Fitted Response Analysis

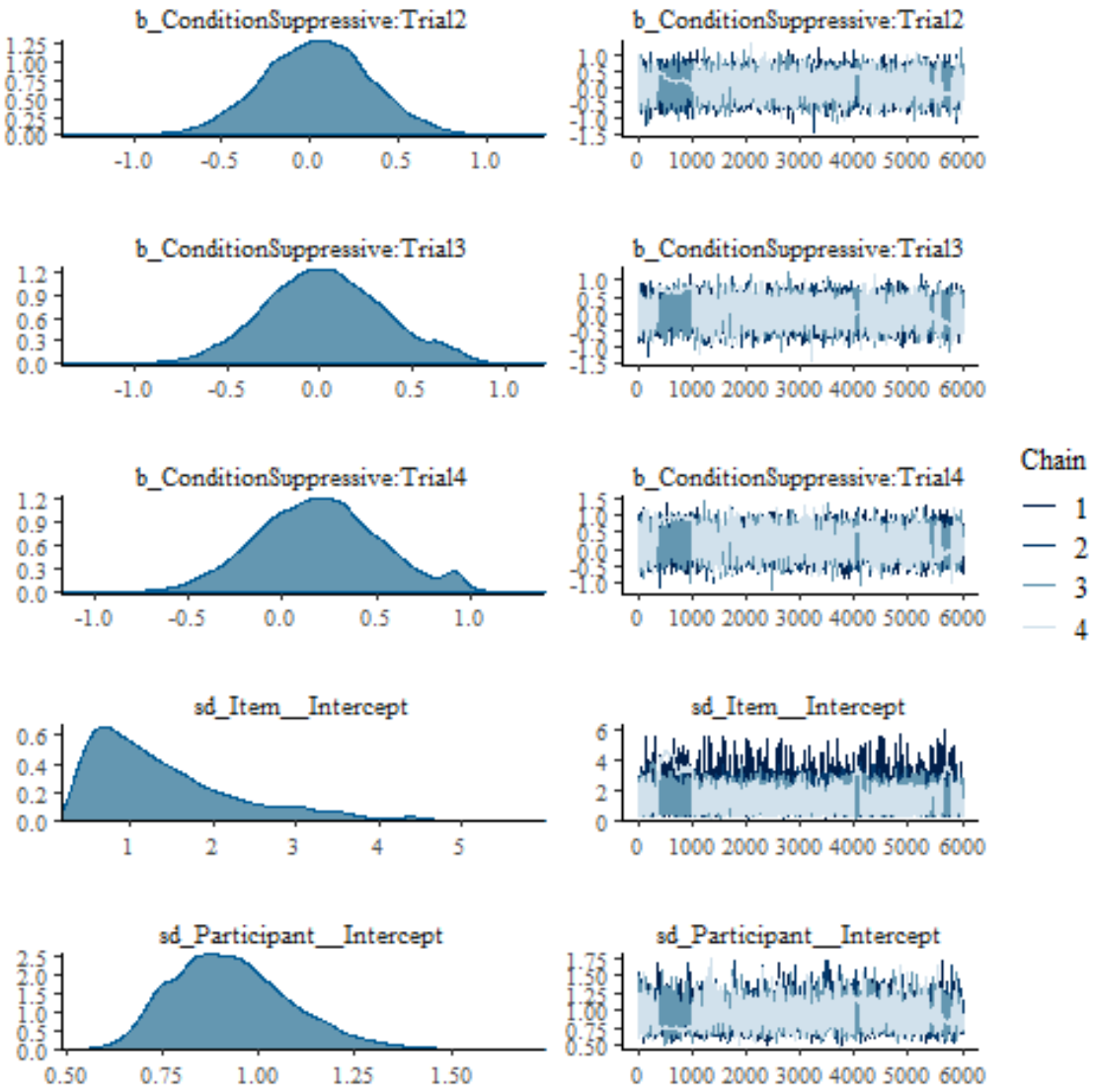


Figure C.3: Density & trace plots for the cumulative model.

C.6 Proprioceptive Drift Z Axis

C.6.1 Parameter tables

C.6.2 Convergence Diagnostics

C.6.3 Fitted Response Analysis

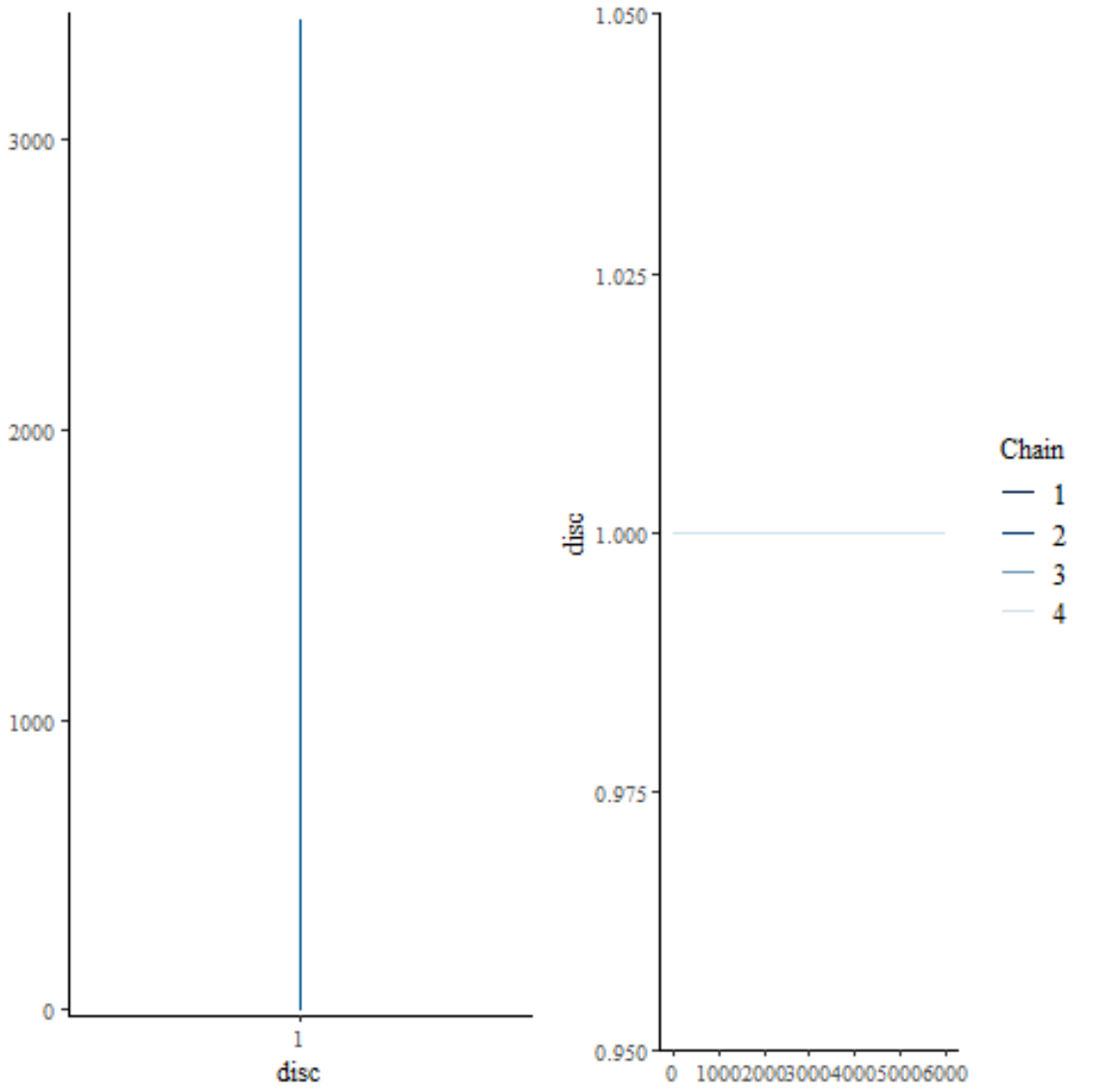


Figure C.4: Density & trace plots for the cumulative model.

C.7 Task Performance: Modelling Learning Behavior

To model learning behavior, the non-linear TFM as discussed in section 2.3.2 is being used. For both task performance measures, number of inserted pegs (discrete) and time per peg (continuous), the same model structure as well as the same conditional effects on the priors has been used, only the specification of the prior distributions themselves differed.

The model structure that has been used is specified as follows:

```
Data ~ exp(Asymp) + exp(Ampl)*inv_logit^-1(Chrt^Trial),
```
(C.1)

where `Asymp` describes the maximum performance (Asymptote), `Ampl` describes the learning potential (Amplitude) and `Chrt` describes the learning rate (catch rate). This model structure has been used for both data types as both the family distributions for discrete (Poisson, binomial, beta-binomial) and continuous (Gaussian, Gamma) data are compatible with the log-link function, and thus the same linearization can be used within the model structure. This in turn allows us to fit the respective family distribution via the identity-link function, which facilitates the model fitting as the Markov chains of the model will easier converge.

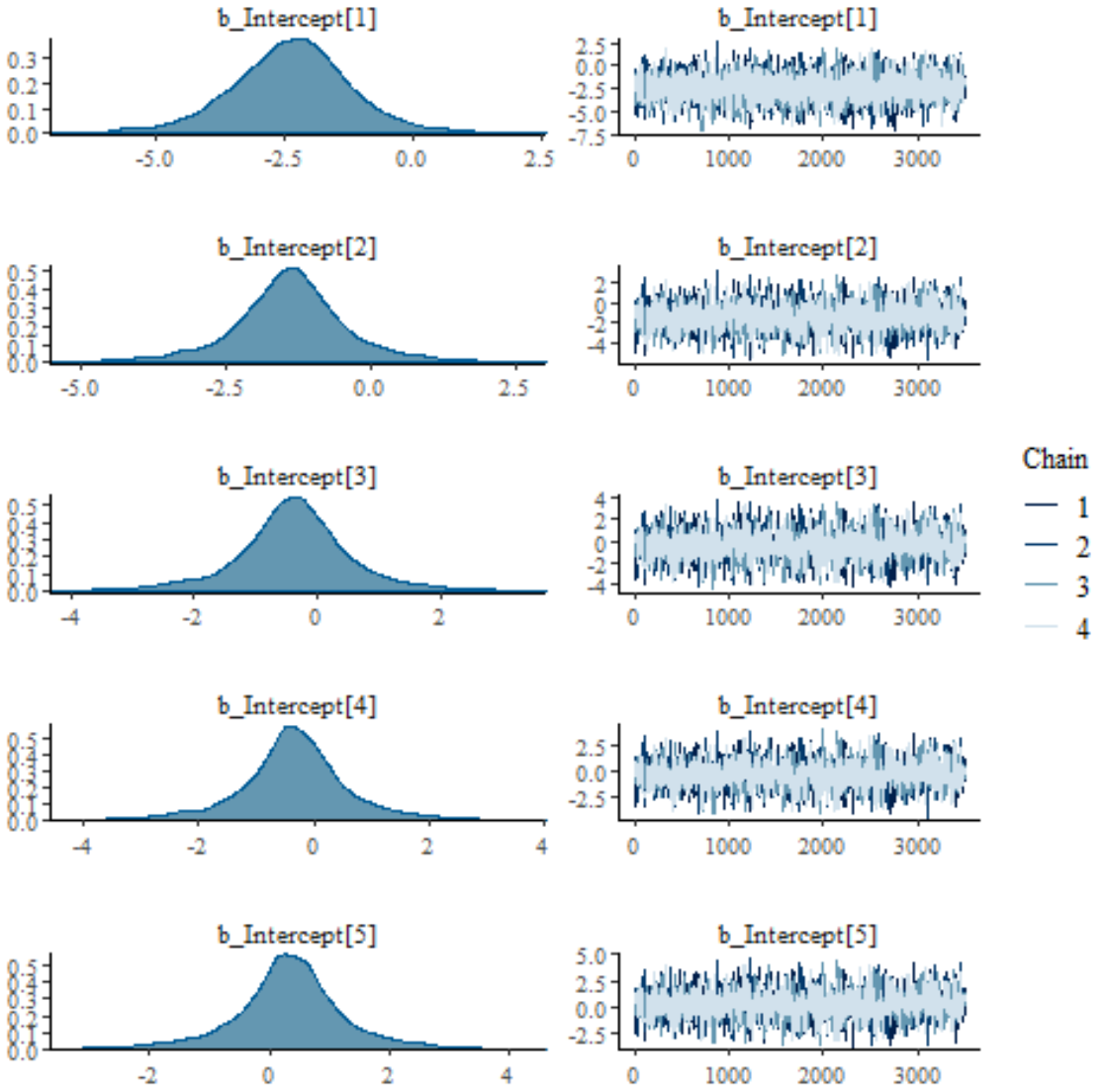


Figure C.5: Density & trace plots for the categorical model.

Moreover, the same conditional effects on the priors have been used for both data types,. Specifically the priors in equation C.2 model 1) differences in learning potential, learning rate and maximum performance between the embodiment groups by adding the term `Condition` to the priors `Chrt`, `Ampl` and `Asymp` respectively; 2), fatigue by adding the term `Trial` to `Asymp`; and 3) individual differences between participants in all three variables, `amplitude`, `asymptote`, and `catch rate`, by adding random effect the `(1|Participant)`:

$$\begin{aligned} \texttt{Ampl} &\sim \texttt{Condition} + (1|\texttt{Participant}), \\ \texttt{Chrt} &\sim \texttt{Condition} + (1|\texttt{Participant}), \\ \texttt{Asymp} &\sim \texttt{Condition} + \texttt{Trial} + (1|\texttt{Participant}) \end{aligned} \tag{C.2}$$

Finally, the only missing part is the specification of the prior distributions themselves, as non-linear models are not compatible with default priors. Thus, weakly informative priors are being specified as follows: First, for both data types it has been assumed that will they have the same learning rate. Based on [129], learning curves of motor tasks can drop very fast, with survival rates as lows as 20 %, but may also have very high survival rates, like 99 % for instance. Below,

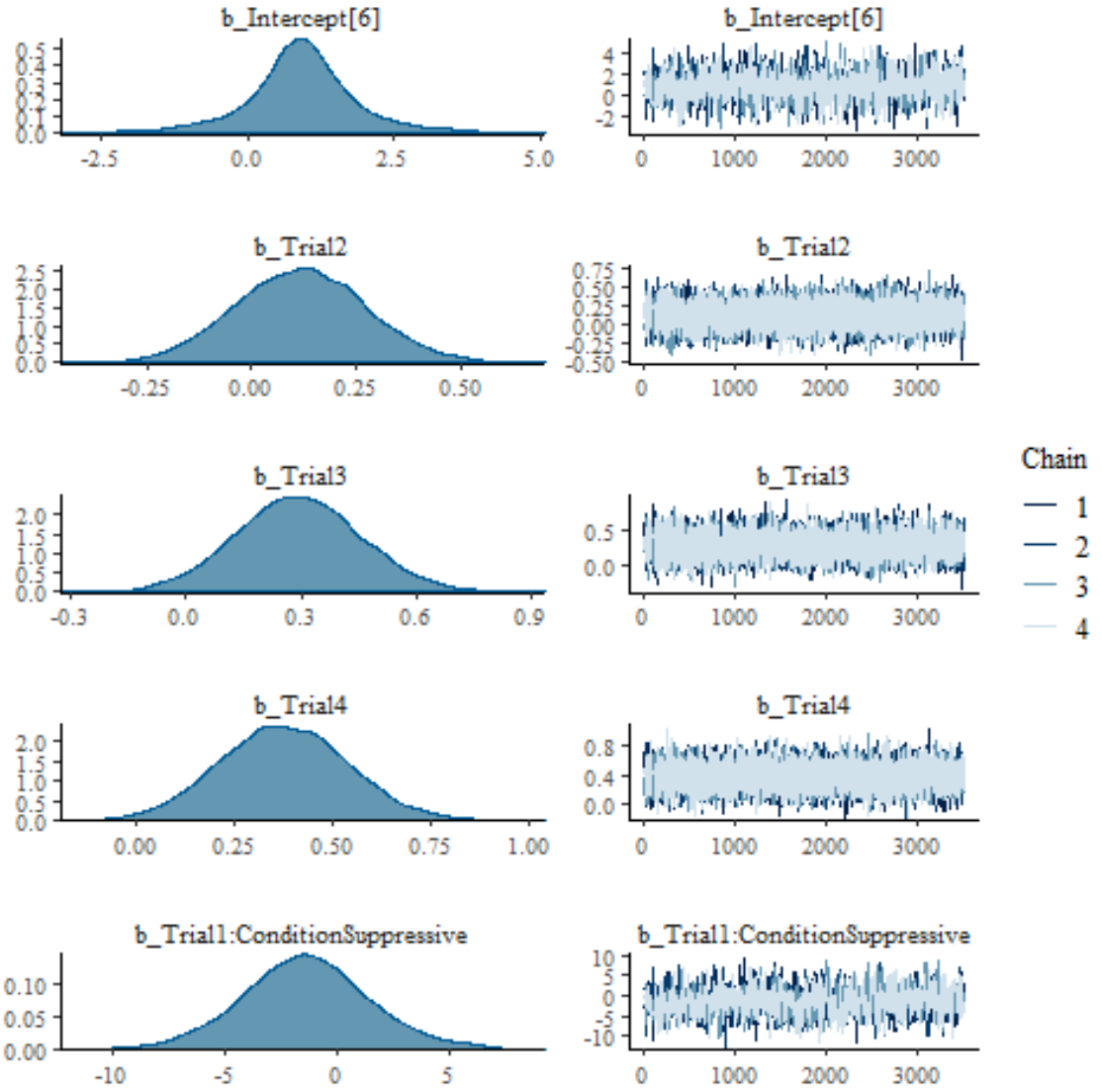


Figure C.6: Density & trace plots for the categorical model.

the corresponding prior on the normal scale and the inverse-logit scale can be found in figure C.73:

Next, the priors for the asymptote and the amplitude have been specified separately for the number of pegs and time per peg, due to the different upper and lower bounds. For the former it is relatively straight forward, as the lowest possible score is 0 and the maximum possible score is 4. However, since the prior needs to be transformed to the log scale and $log(0)$ is undefined, +1 has been added to the observed data and the bounds for the priors, therefore changing the lowest and highest score to 1 and 5 respectively. Thus, for the asymptote, the maximum possible score, it seems reasonable to have a lower and upper bound of [3,5]. For the amplitude, the learning potential (or the performance in the first trial), it seems plausible to have lower and upper bounds of [1,3]. By using these bounds to create normal distributions and the corresponding distributions on the log scale, the resulting priors can be seen in figure C.74 and C.75:

Similarly, the priors for the asymptote and the amplitude have been constructed based on expected upper and lower bounds. For the asymptote, it seems reasonable to have a maximum performance between 20 and 80 seconds. For the amplitude, initial times in the first trial

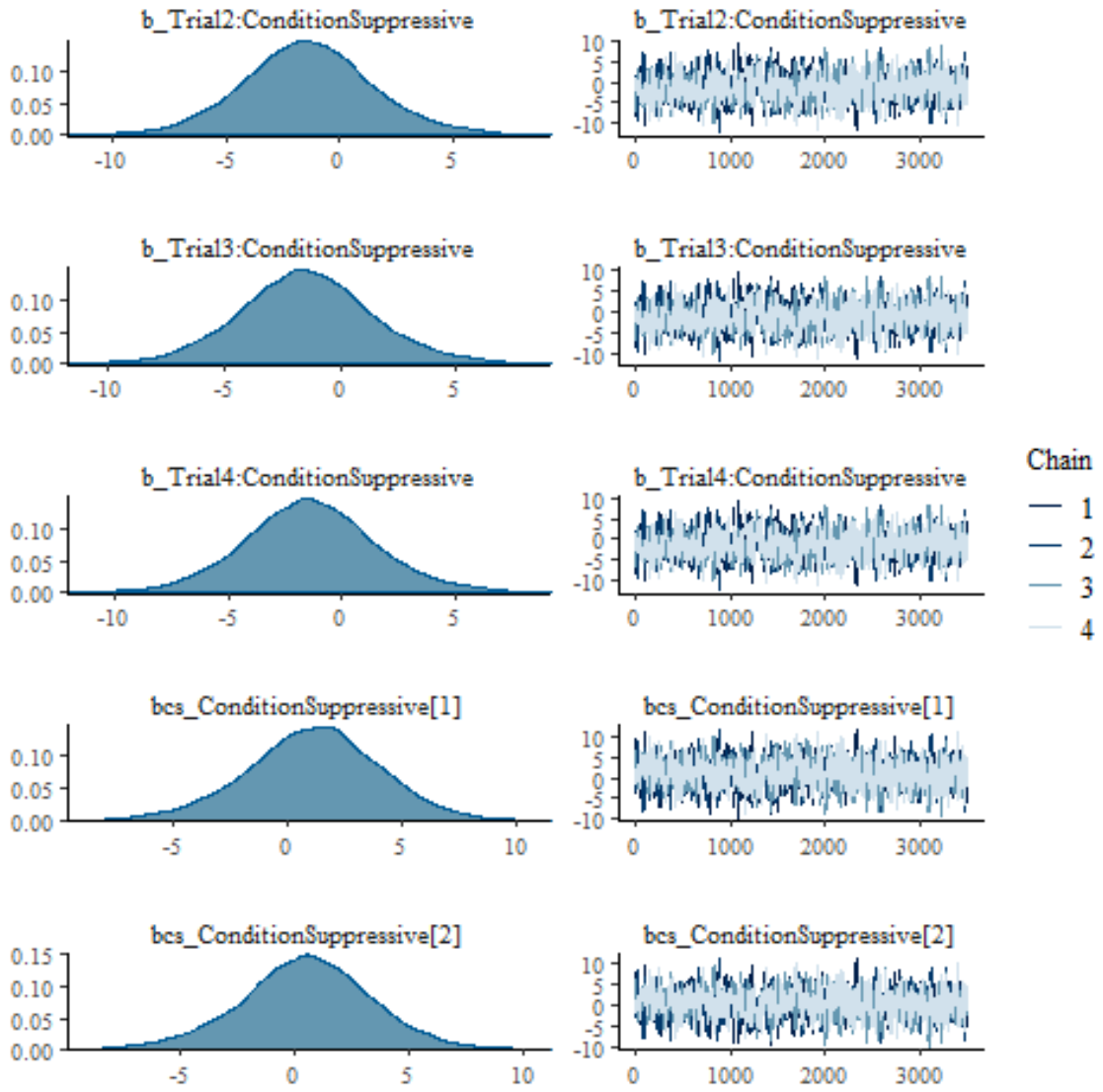


Figure C.7: Density & trace plots for the categorical model.

between 60 and 180 seconds seem legitimate. By constructing the corresponding priors on the normal and log scale, the figure C.76 and C.77 have been generated:

C.7.1 Learning Behavior: Count

As can be seen, not significant learning behaviour can be found in terms of effectiveness, i.e., the number of pegs successfully inserted. While on the population level, a slight increase with every trial is visible, there is for a large part no clear learning behaviour visible on the participant level. For completeness sake, a non-linear learning model with a binomial distribution has been fit. Note, that here the data has been modified by adding plus one peg to each measurement, as discussed in the previous section. As can be seen, the learning model also does not predict any learning for either of the two embodiment groups.

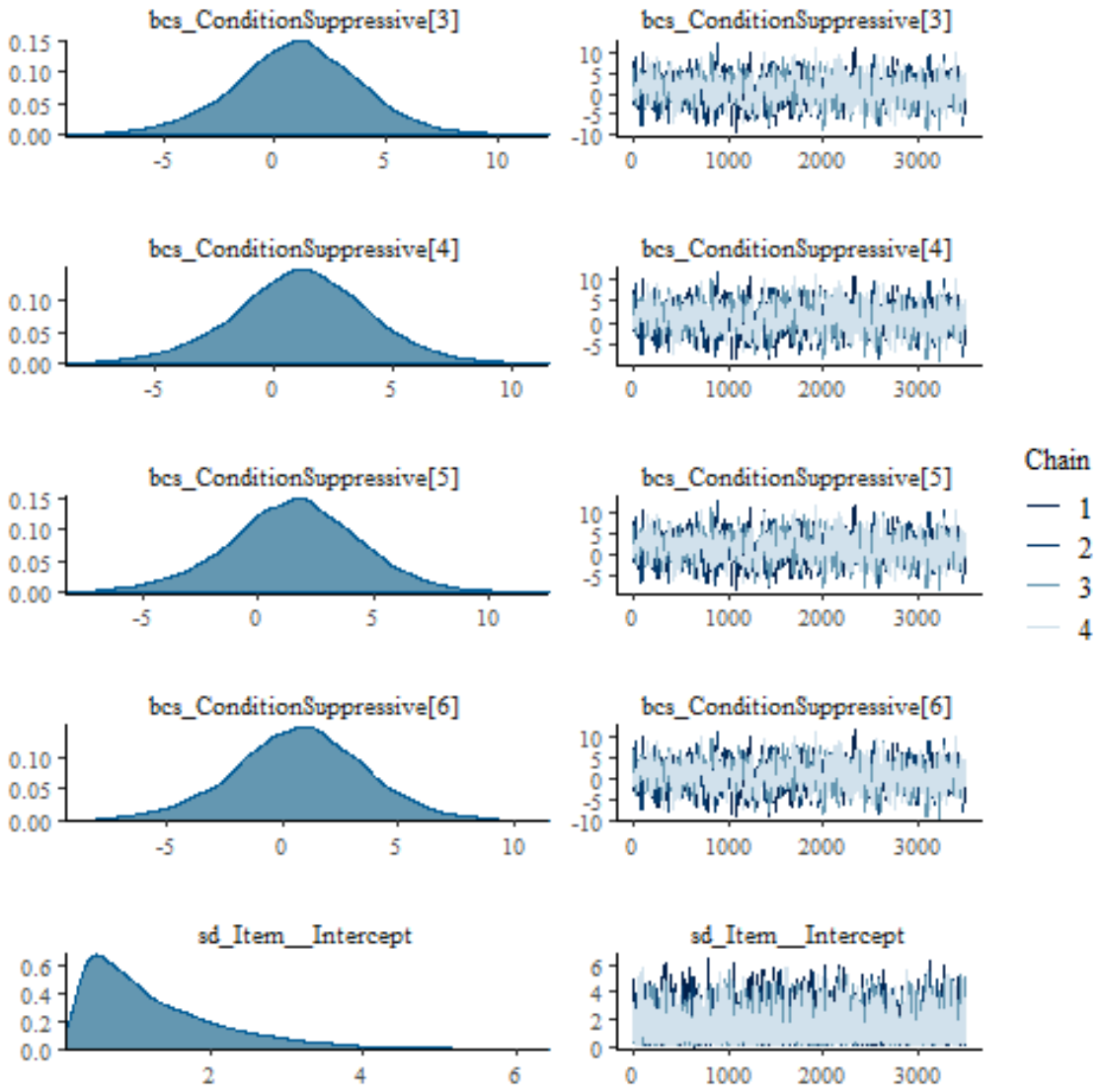


Figure C.8: Density & trace plots for the categorical model.

C.7.2 Learning Behavior: Time

As can be seen, not significant learning behaviour can be found in terms of efficiency, i.e., the time needed per successfully inserted peg. For completeness sake, a non-linear learning model with a Gamma distribution has been fit. While the learning model does predict learning, this is most likely due to the chosen priors, and does not reflect the actual performance of participants.

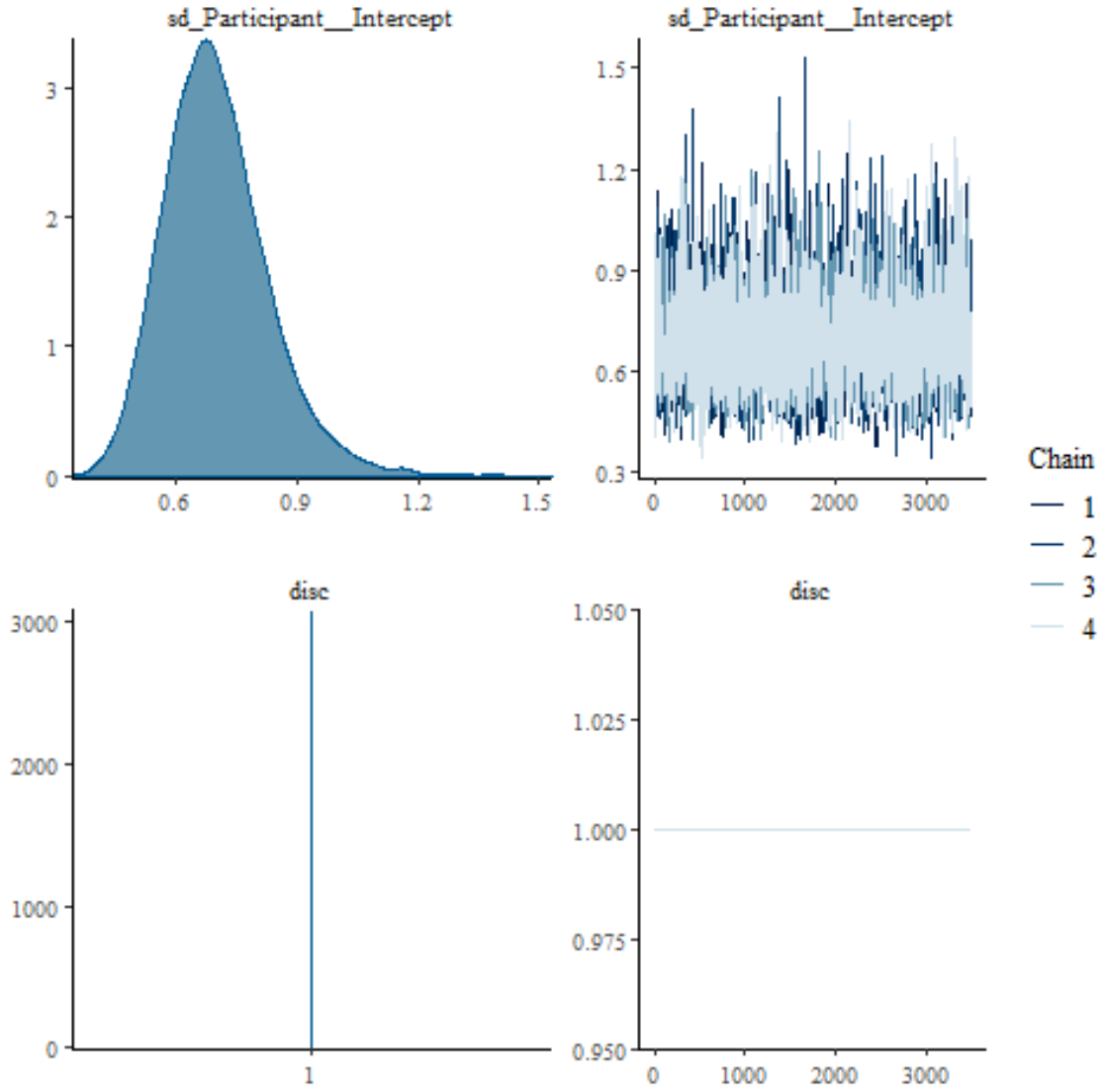


Figure C.9: Density & trace plots for the categorical model.

C.8 Task Performance: Number Pegs

For every model, we used the default priors calculated by BRMS as can be seen in table C.24:

C.8.1 Parameter tables

C.8.2 Convergence Diagnostics

C.8.3 Fitted Response Analysis

Predictive posterior checks with 100 samples per model have been generated as can be seen below in figure C.96:

Bar plots of the 95 % HDI's for each of the coefficients have been generated in figure C.97:

Next, violin plots of the residuals from the posterior distributions of all three models have been generated:

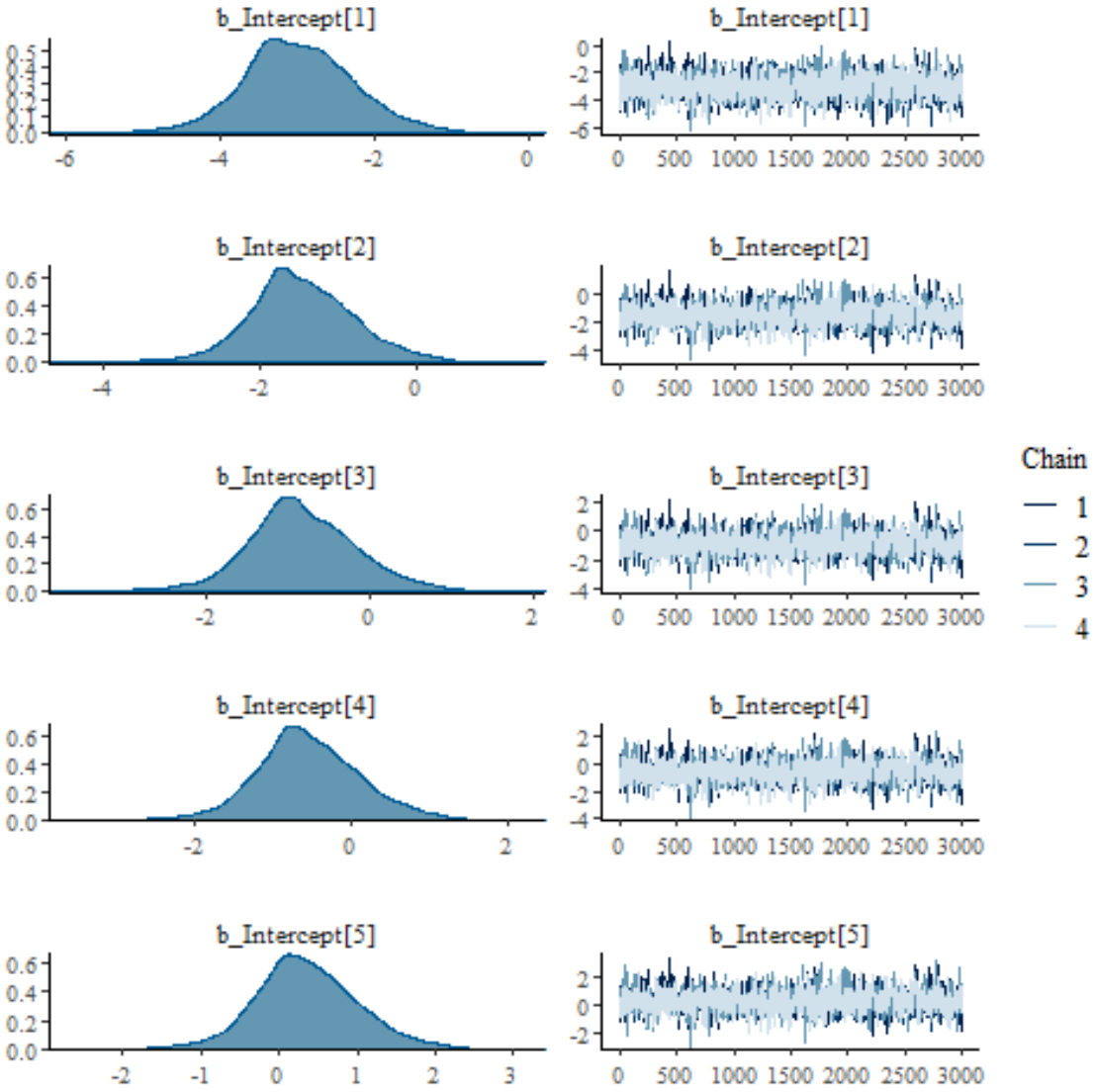


Figure C.10: Density & trace plots for the unequal variance between conditions model.

C.9 Task Performance: Time

To analyze this data, three Bayesian GLMMs have been fit with `Time` as a function of dummy-coded factors `Condition` and `Trial` and the two-way interaction thereof. Additionally, random effects in the participants have been included to account for individual differences, resulting in the overall model

```
Time ~ Condition*Trial + (1|Participant).
```
(C.3)

For each of the three models a different family distribution has been used, namely a Gaussian, an exponentially modified Gaussian and a Gamma distribution. As time is continuous data with a positive lower bound, it is expected that the exponentially modified Gaussian produces the best fit, as this distribution has an additional parameter, rate β, which can adjust for the skewness of the lower bounded data.

For every model, the default priors calculated by `BRMS` have been used, where the priors for the intercept, sd and sigma are all shared by the three models, beta is used in the exponentially

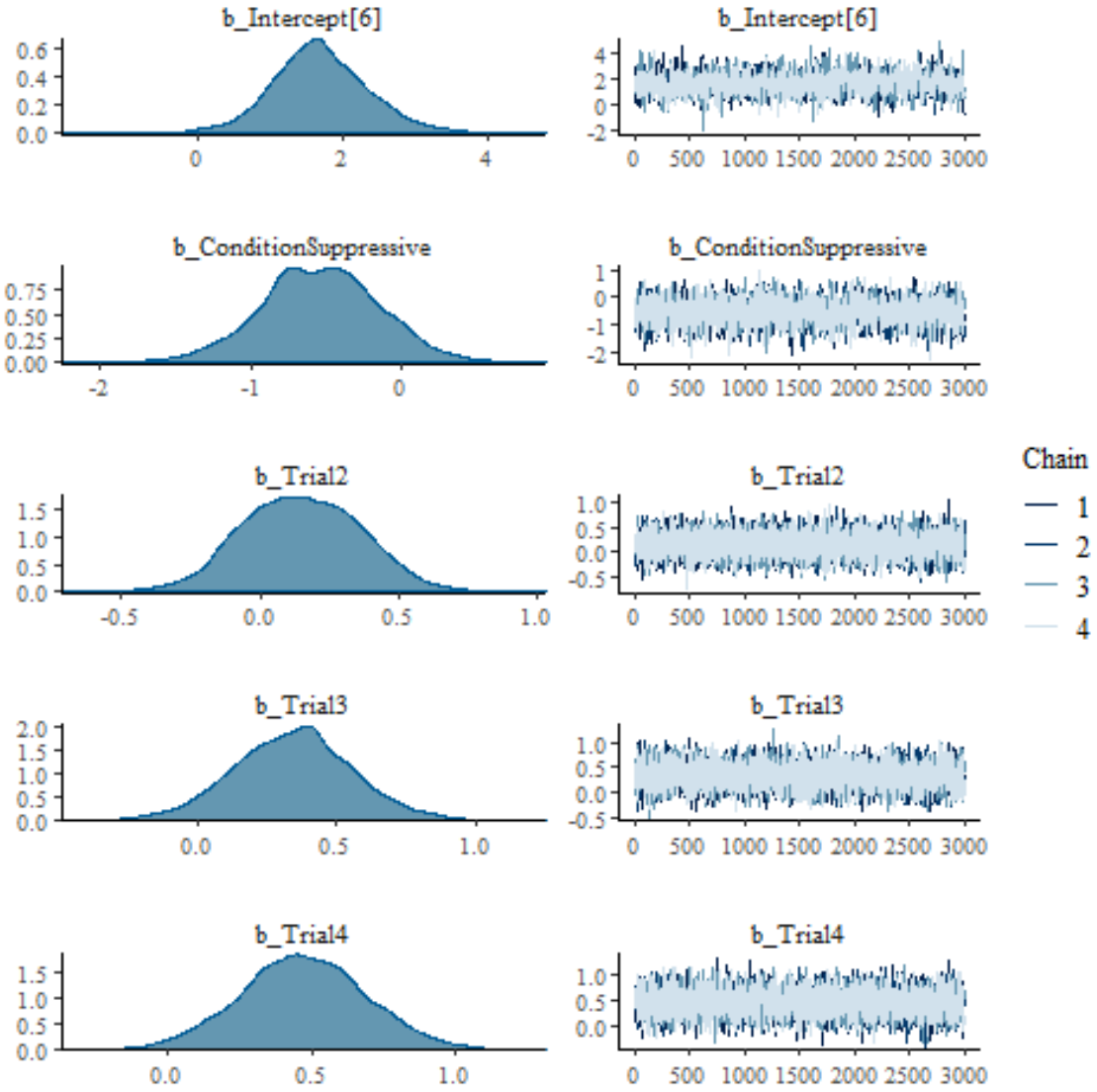


Figure C.11: Density & trace plots for the unequal variance between conditions model.

modified Gaussian model, and shape for the Gamma model. These priors can be found below in table C.28:

All three models have been run with four chains and 4000 iterations (2000 warm-up and 2000 sampling). The log-function has been used as a link function for each of the models. All three models convergence without any divergence problems.

C.9.1 Parameter tables

C.9.2 Convergence Diagnostics

C.9.3 Fitted Response Analysis

Predictive posterior checks with 100 samples per model have been generated as can be seen below in figure C.107:

As can be seen, the Gaussian model predicts negative time values, which are impossible. Furthermore, the mode of the predictions is located to far on the right, thus predicting on average more time needed per peg than can actually be observed in the data. Lastly, when considering the average of the samples, the respective mode is too low as compared to the observed data.

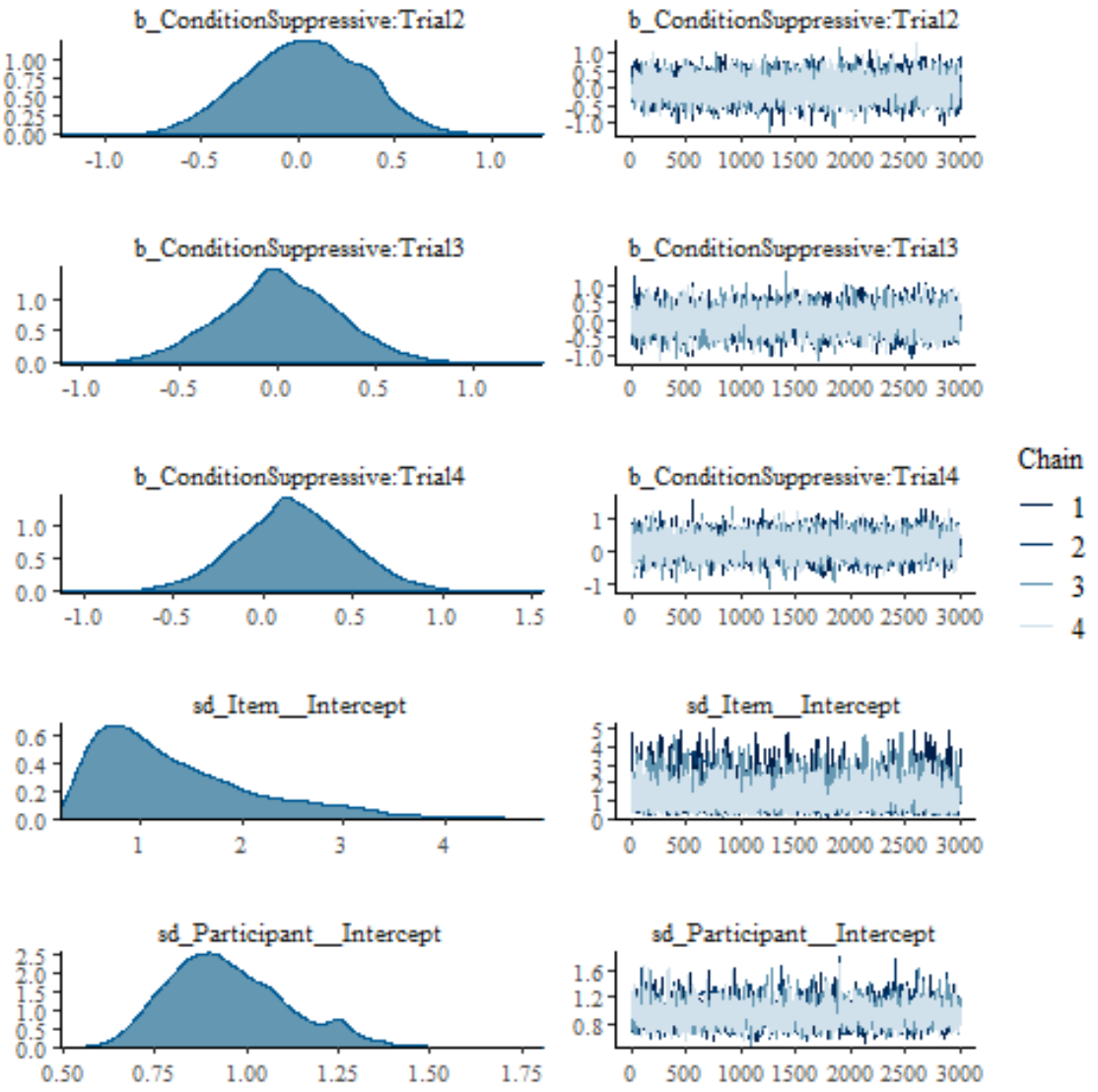


Figure C.12: Density & trace plots for the unequal variance between conditions model.

Next, the exponentially modified Gaussian model predicts the observed data well on the left side, however it seems more uncertain on the right tail of the distribution. The location of the mode seems to be predicted well as well, albeit a little too certain, i.e. the mode of the predictive posterior samples seems a little too high. Lastly, the gamma model produces a similar fit compared to the exponentially modified Gaussian model. Here, the mode seems to be slightly better in terms of the probability it assigns. Both exponential modified Gaussian and Gamma model are more uncertain on the predictions on the right tail of the distributions.

Note that both Gamma and Exponential modified Gaussian model both also seem to predict negative (implausible) values. However, it should be noted that the observed data also appears negative on the left tail. This might be due to the smoothing that is internally done in the `pp_check()` function that generates the posterior predictive check. To verify this theory, we can sample the predictive posterior distribution for each model and examine the proportion of predictions, that get a negative 2.5 % credibility (left tail) limit assigned:

As can be seen in table C.32, the Gaussian model assigns 75.1 % of the predictions to lower 2.5 % quantile whereas neither the exponentially modified Gaussian nor the Gamma model produce implausible negative predictions in the lower 2.5 % credibility limit.

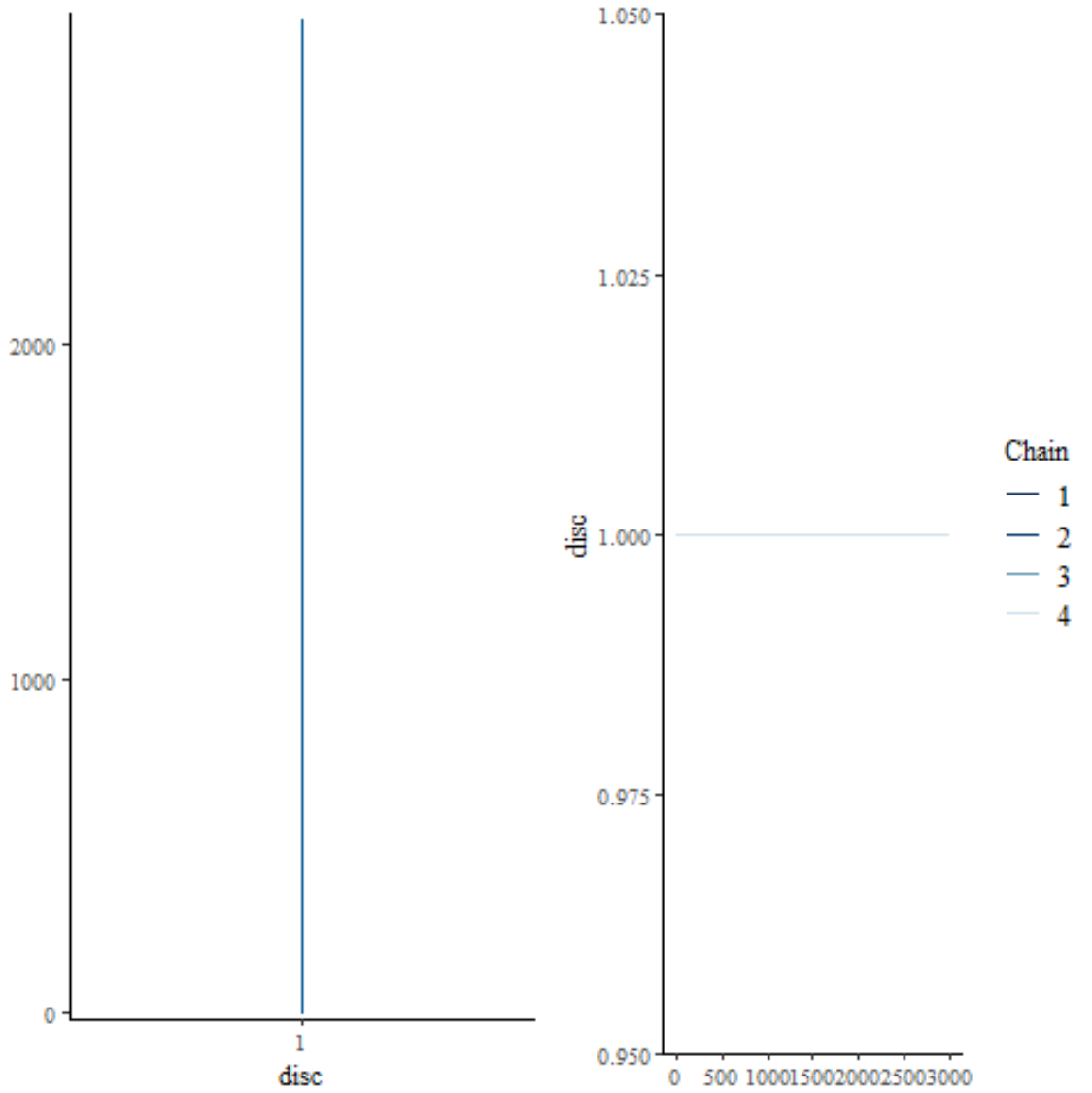


Figure C.13: Density & trace plots for the unequal variance between conditions model.

For every model, bar plots of the 95% credibility intervals of the posterior distribution and violin plots of the residuals of the predictive posterior distribution will be reported.

First, the bar plots of the 95 % HDI's for each of the coefficients is depicted in figure C.108: Note that the coefficients have been exponential from the log scale for interpretability. The intercept (Supportive embodiment condition, trial one) is therefore on the original scale of the data, and the other coefficients are multiplicative.

Starting with the intercept it can be observed that the Gamma and the Gaussian model are very similar in terms of center estimates and credibility limits, whereas the exponentially modified Gaussian model expects a lower center estimate and is also more certain in its predictions, which is reflected in a smaller 95 % credibility interval.

For the coefficient of suppressive condition it can be seen that all models agree on the same center estimate but vary slightly in their range of credibility intervals. Note that all the center estimates and most of the credibility intervals have a value above one, thus indicating a higher time needed to insert the pegs for the suppressive embodiment condition (as this coefficient is multiplied with the intercept).

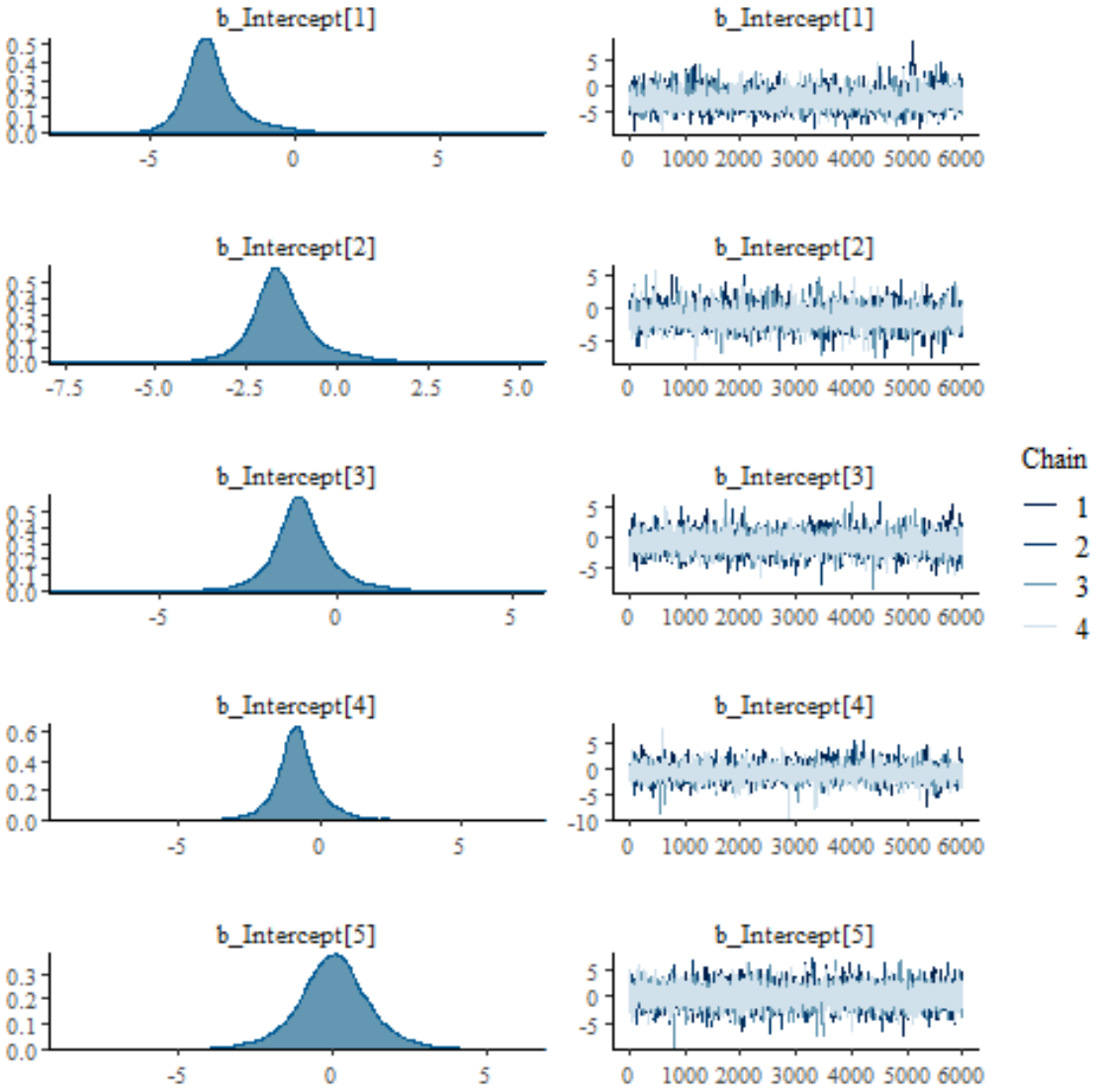


Figure C.14: Density & trace plots for the unequal variance between items model.

For the coefficients of the trials it can be observed that the Gamma and Gaussian model are very similar to each other where as the exponentially modified Gaussian model predicts higher center estimates. Except for the last trial, the range of the credibility intervals are relatively similar for all three models.

Lastly, for the coefficient of the interaction between suppressive and trial, it can be seen that the Gamma and Gaussian model are again very similar. In contrast, the exponentially modified Gaussian model generally is generally more certain, which is reflected in its smaller credibility interval range. Furthermore, it can be seen that this model also predicts different center estimates as compared to Gamma and Gaussian model. Specifically, trials two and four have slightly higher center estimates whereas the trial 3 has a lower center estimate.

In general it can be seen that the Gaussian and Gamma model are very similar in their estimation of the coefficients whereas the exponentially modified Gaussian model mostly predicts different center estimates, with more certainty.

If this model does indeed predict a better fit to the data, this will be reflected in the shape of the residuals. Hence, the violin plots of the residuals from the posterior distributions can be seen in figure C.109, grouped by condition and trial:

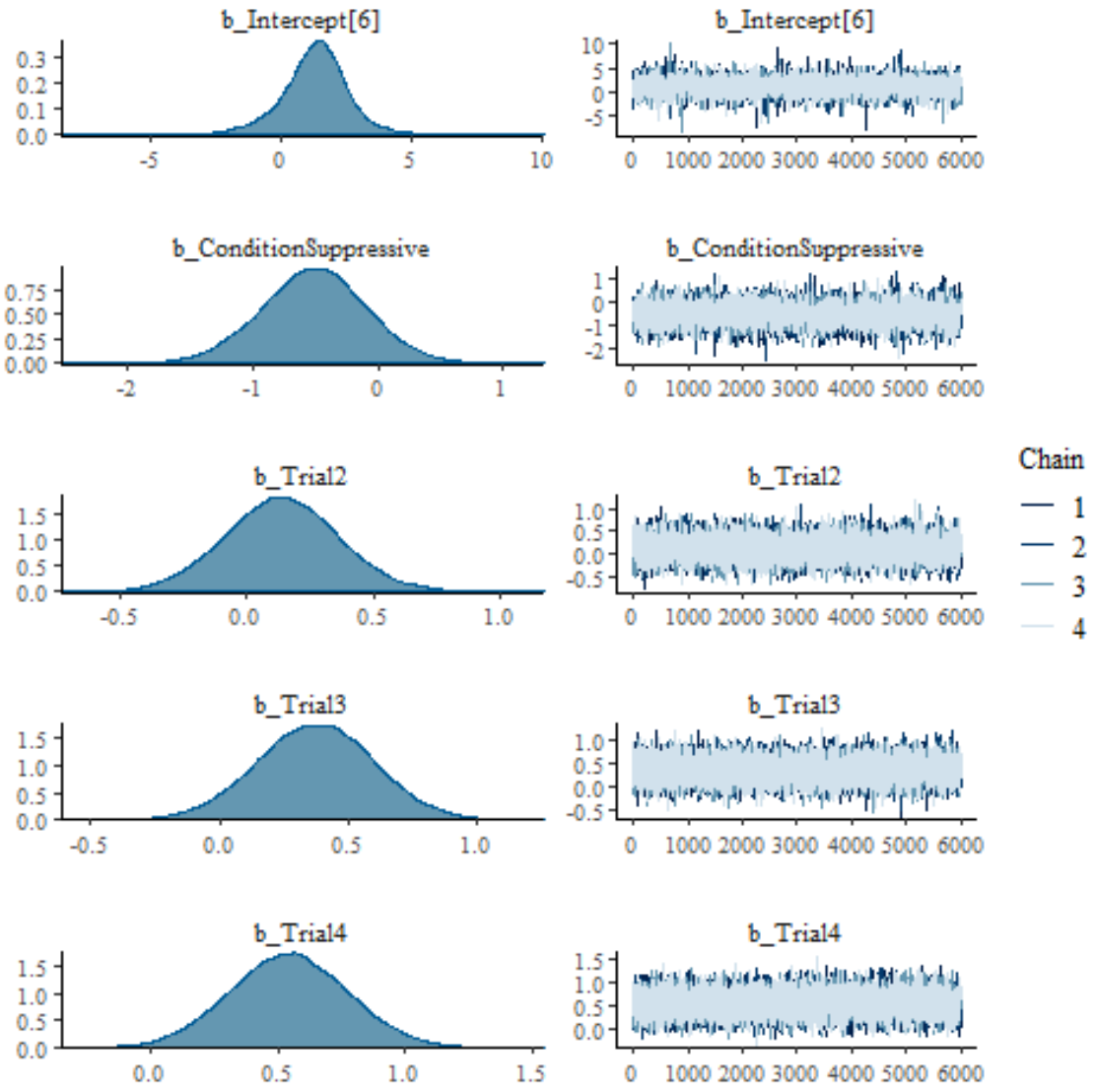


Figure C.15: Density & trace plots for the unequal variance between items model.

Here, the residuals are calculated by subtracting the predicted center estimate from the observed data, thus negative residual values indicate a higher time predicted and positive residual values less time. A residual value of zero would indicate perfect prediction.

As can be seen, for the first trial in the supportive embodiment group, the mode of the exponentially modified Gaussian model is closer to 0, thus indicating less errors, followed successively by the Gamma and Gaussian model. Here, the shape of the residual distributions is very similar between all three of the models. For the first trial in the supportive suppressive group, the modes of the three models are very similar, however the exponentially modified Gaussian model is cut off at the bottom, thus indicating less errors on for higher time predictions.

In the second trial, the residual distributions are very similar for both of the supportive and suppressive group respectively.

In the third trial, all three models are very similar to each other for the supportive group. For the suppressive group, the exponentially modified Gaussian model shows some more errors on the top of the violin plot, indicating more errors for smaller time predictions.

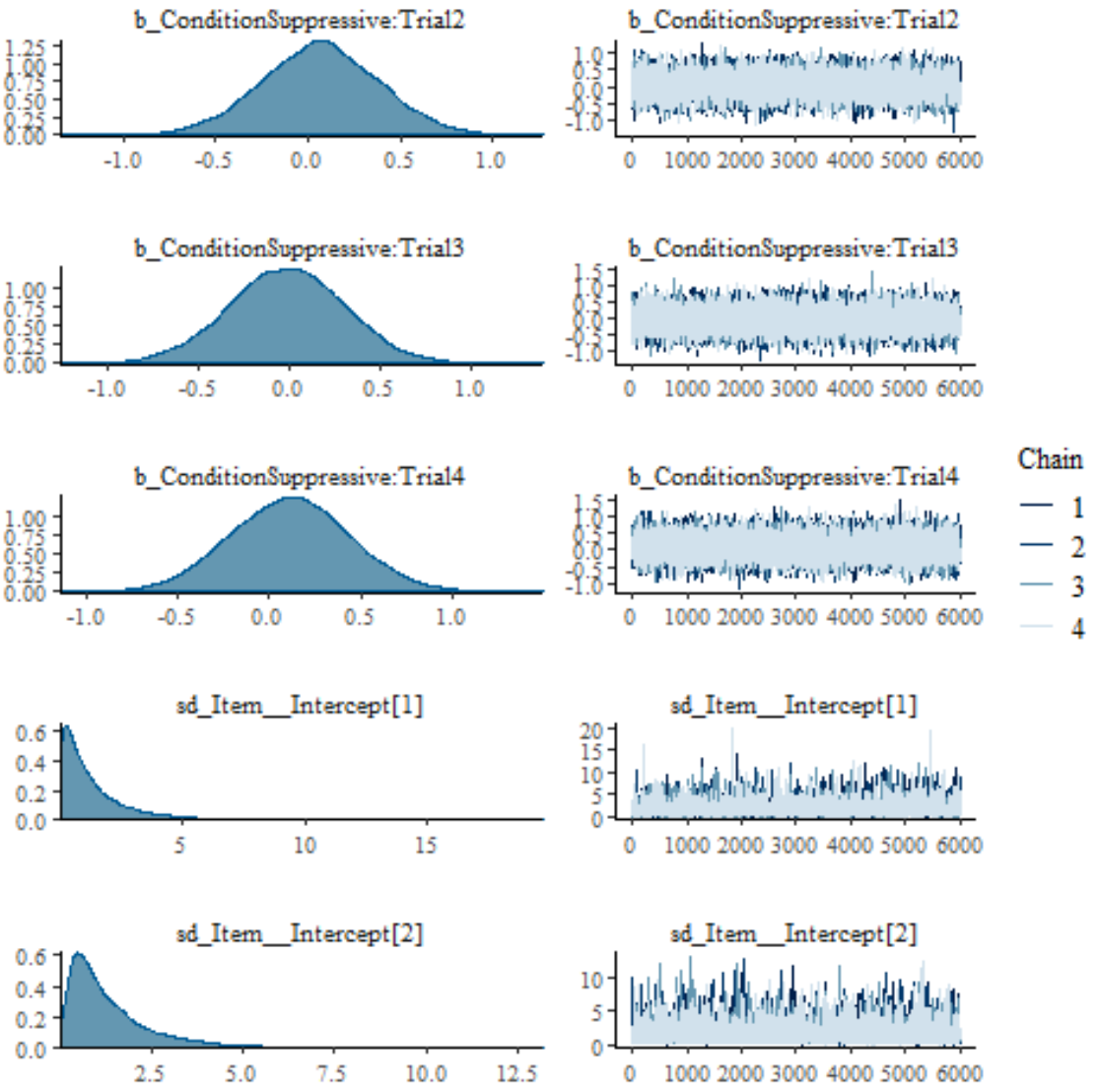


Figure C.16: Density & trace plots for the unequal variance between items model.

In the last trial, the shapes of the residual distributions are very similar for both of the supportive and suppressive group respectively. For both conditions, the mode of the exponentially modified Gaussian model is slightly closer to zero as compared to the Gamma and Gaussian model. In addition, the bottom of the violin plot is more cut off than of the other models, indicating less error for predictions of higher time values.

Generally, the modes of all models appear to be slightly below zero, thus overestimating the time needed to insert the pegs. Moreover, the first and fourth trial are less certain than second and third trial which is also reflected for instance the box plots of the rain cloud plot in figure 4.11. Overall, the residual errors of the models are relatively similar, where the exponentially modified Gaussian model seems to predict the values slightly better, especially for higher time values.

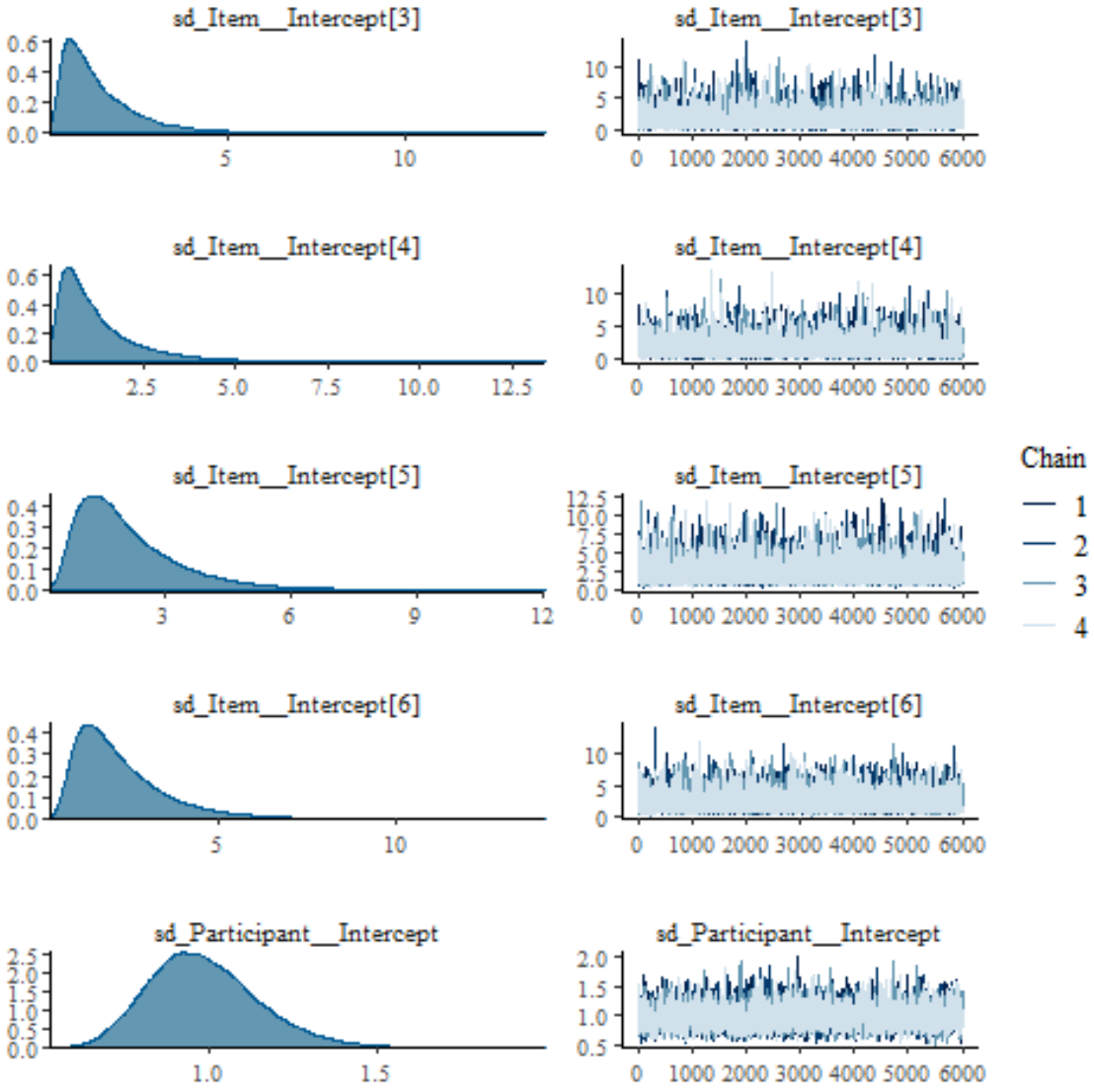


Figure C.17: Density & trace plots for the unequal variance between items model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept[1]	1.0	2665	-2.8	1.2	-4.7	-3.0	0.2
b_Intercept[2]	1.0	6773	-1.5	1.0	-3.4	-1.6	0.9
b_Intercept[3]	1.0	5998	-1.0	1.0	-3.0	-1.0	1.3
b_Intercept[4]	1.0	4577	-0.8	1.0	-2.7	-0.8	1.5
b_Intercept[5]	1.0	9779	0.1	1.3	-2.6	0.1	2.7
b_Intercept[6]	1.0	7935	1.3	1.4	-1.7	1.3	3.9
b_ConditionSuppressive	1.0	3722	-0.5	0.4	-1.4	-0.5	0.3
b_Trial2	1.0	10527	0.1	0.2	-0.3	0.1	0.6
b_Trial3	1.0	11419	0.4	0.2	-0.1	0.4	0.8
b_Trial4	1.0	8875	0.5	0.2	0.1	0.5	1.0
b_ConditionSuppressive:Trial2	1.0	10784	0.1	0.3	-0.6	0.1	0.7
b_ConditionSuppressive:Trial3	1.0	11038	-0.0	0.3	-0.6	-0.0	0.6
b_ConditionSuppressive:Trial4	1.0	9733	0.1	0.3	-0.5	0.1	0.7
sd_Item__Intercept[1]	1.0	6949	1.3	1.4	0.0	0.9	5.2
sd_Item__Intercept[2]	1.0	10892	1.4	1.3	0.1	1.0	4.9
sd_Item__Intercept[3]	1.0	10320	1.4	1.2	0.2	1.1	4.7
sd_Item__Intercept[4]	1.0	4880	1.3	1.3	0.1	0.9	4.8
sd_Item__Intercept[5]	1.0	5640	2.2	1.4	0.7	1.8	6.0
sd_Item__Intercept[6]	1.0	11204	2.2	1.4	0.6	1.8	5.9
sd_Participant__Intercept	1.0	5988	1.0	0.2	0.7	1.0	1.4
cor_Item__Intercept[1]__Intercept[2]	1.0	24000	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[1]__Intercept[3]	1.0	24000	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[2]__Intercept[3]	1.0	17114	0.1	0.4	-0.7	0.1	0.8
cor_Item__Intercept[1]__Intercept[4]	1.0	20232	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[2]__Intercept[4]	1.0	18220	0.1	0.4	-0.7	0.1	0.8
cor_Item__Intercept[3]__Intercept[4]	1.0	14858	0.1	0.4	-0.7	0.1	0.7
cor_Item__Intercept[1]__Intercept[5]	1.0	18954	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[2]__Intercept[5]	1.0	18528	0.1	0.4	-0.6	0.1	0.7
cor_Item__Intercept[3]__Intercept[5]	1.0	18437	0.1	0.4	-0.6	0.1	0.8
cor_Item__Intercept[4]__Intercept[5]	1.0	14196	0.1	0.4	-0.6	0.1	0.8
cor_Item__Intercept[1]__Intercept[6]	1.0	18967	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[2]__Intercept[6]	1.0	18215	0.1	0.4	-0.6	0.1	0.7
cor_Item__Intercept[3]__Intercept[6]	1.0	14850	0.1	0.4	-0.6	0.1	0.7
cor_Item__Intercept[4]__Intercept[6]	1.0	15634	0.1	0.4	-0.6	0.1	0.7
cor_Item__Intercept[5]__Intercept[6]	1.0	14665	0.1	0.4	-0.6	0.1	0.8
disc		24000	1.0	0.0	1.0	1.0	1.0
r_Item[Q1,Intercept[1]]	1.0	2629	0.3	1.2	-1.5	0.0	3.3
r_Item[Q2,Intercept[1]]	1.0	2660	0.5	1.2	-1.1	0.2	3.7
r_Item[Q1,Intercept[2]]	1.0	6887	-0.1	1.0	-1.9	-0.1	2.3
r_Item[Q2,Intercept[2]]	1.0	6895	0.5	1.0	-1.2	0.4	3.0
r_Item[Q1,Intercept[3]]	1.0	6139	-0.2	1.0	-2.1	-0.2	2.1
r_Item[Q2,Intercept[3]]	1.0	6403	0.5	1.0	-1.4	0.4	2.8
r_Item[Q1,Intercept[4]]	1.0	4378	-0.2	1.0	-2.1	-0.2	2.1
r_Item[Q2,Intercept[4]]	1.0	4674	0.4	1.0	-1.5	0.3	2.7
r_Item[Q1,Intercept[5]]	1.0	10051	-0.9	1.3	-3.6	-0.9	1.7
r_Item[Q2,Intercept[5]]	1.0	10109	0.9	1.3	-1.8	0.9	3.6
r_Item[Q1,Intercept[6]]	1.0	8197	-1.2	1.4	-4.3	-1.1	1.3
r_Item[Q2,Intercept[6]]	1.0	8109	0.5	1.3	-2.4	0.6	3.1

Table C.5: Parameter table for the unequal item variances model.

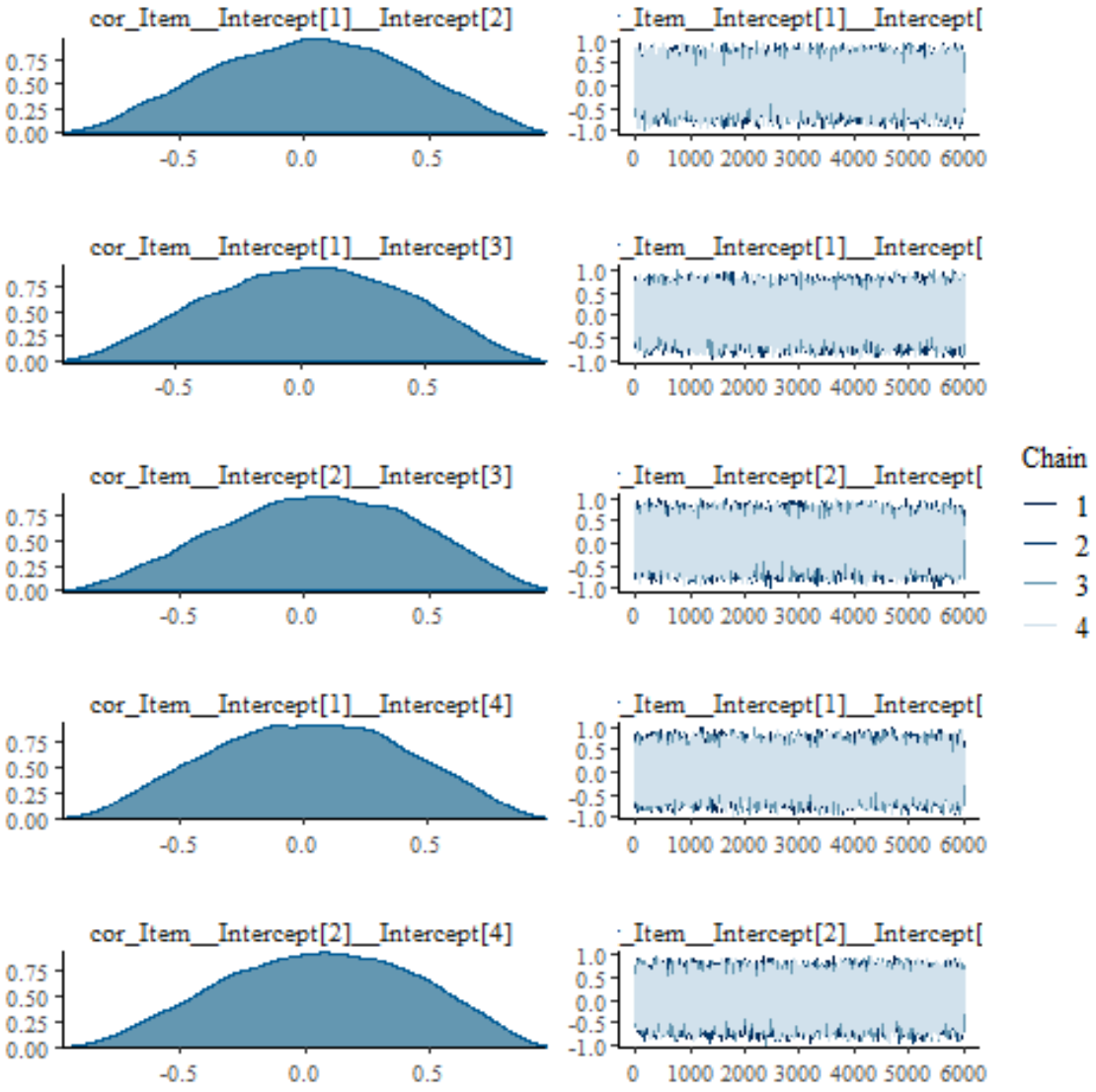


Figure C.18: Density & trace plots for the unequal variance between items model.

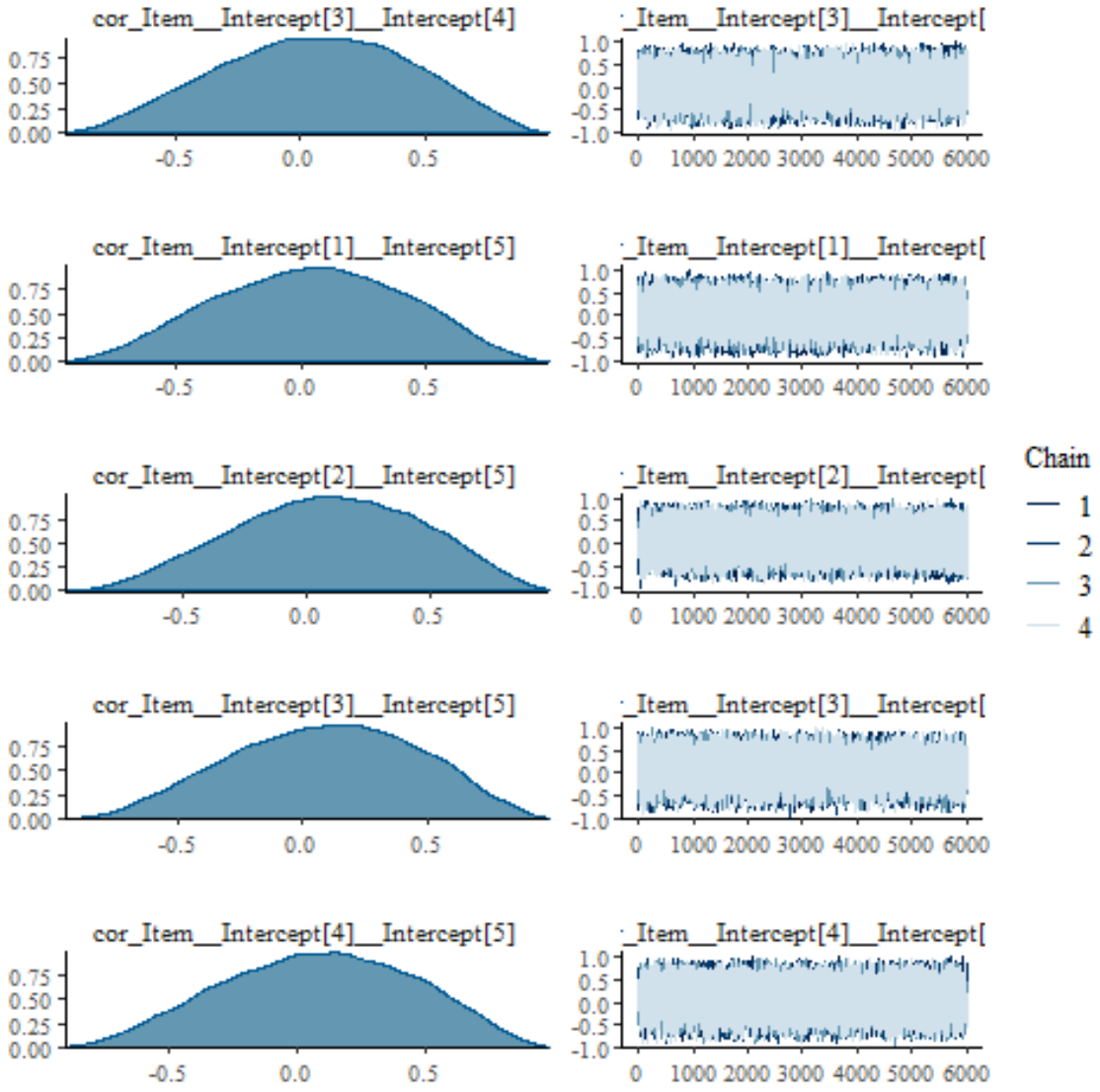


Figure C.19: Density & trace plots for the unequal variance between items model.

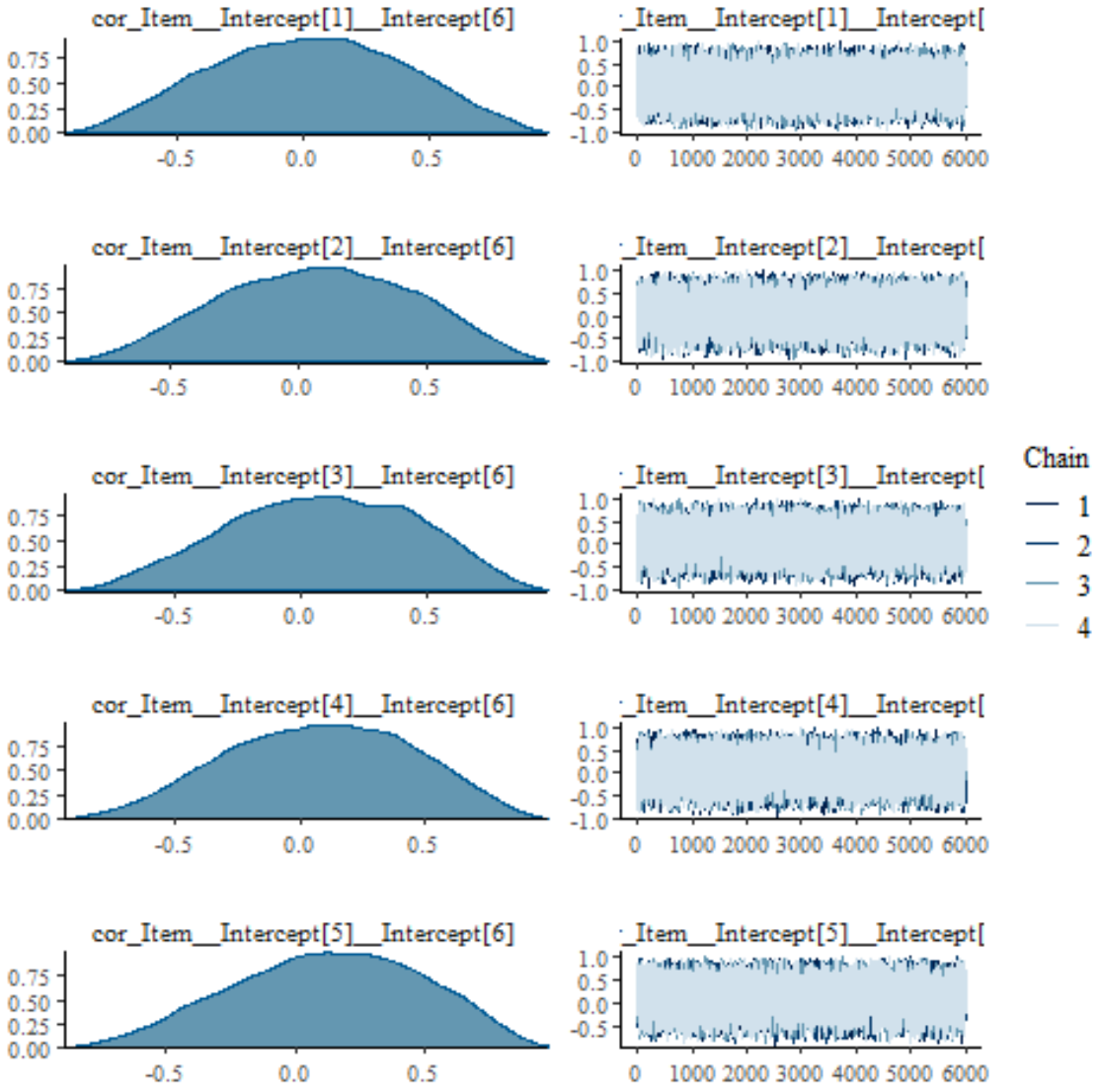


Figure C.20: Density & trace plots for the unequal variance between items model.

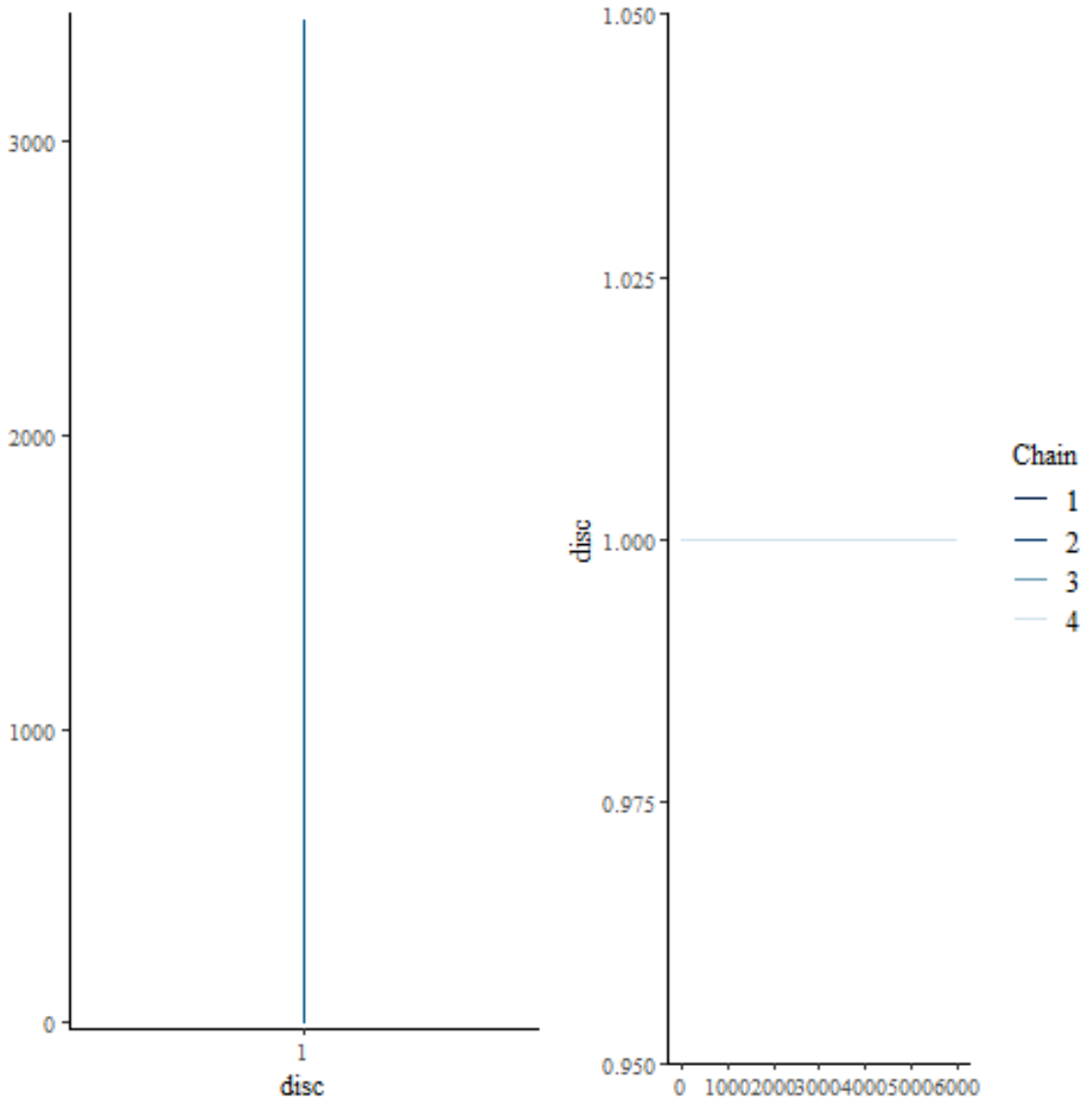


Figure C.21: Density & trace plots for the unequal variance between items model.

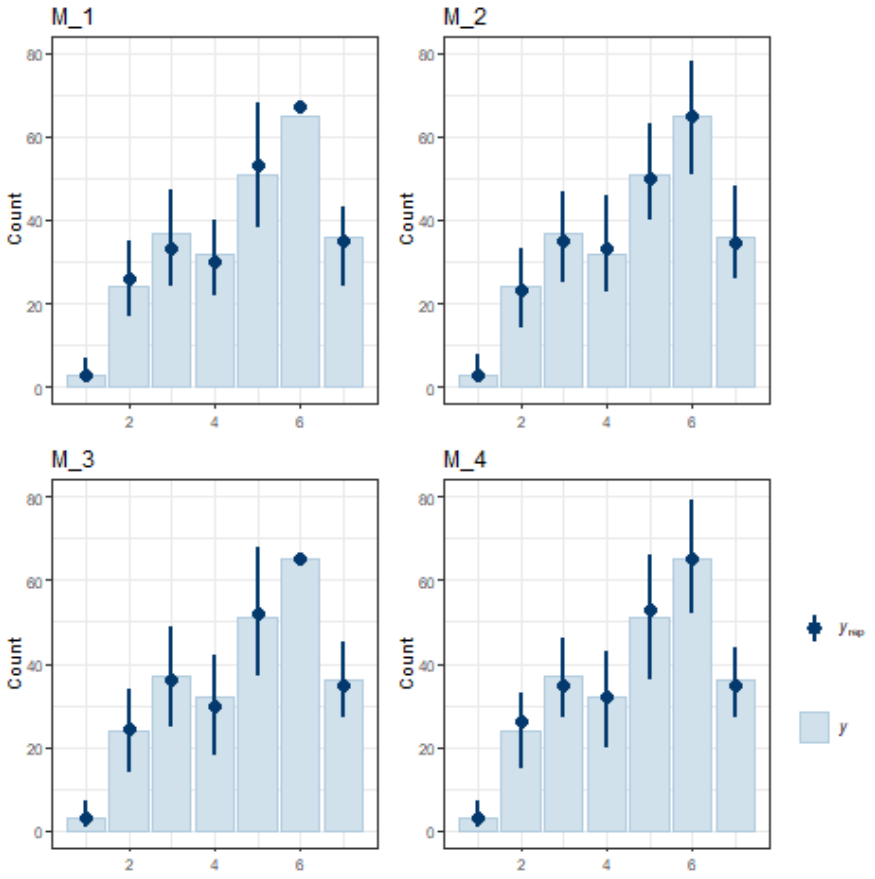


Figure C.22: Predictive posterior checks of the SoO OLR models.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept[1]	1.0	577	-3.0	0.6	-4.3	-3.0	-1.8
b_Intercept[2]	1.0	642	-2.1	0.5	-3.1	-2.1	-1.0
b_Intercept[3]	1.0	598	-1.2	0.5	-2.2	-1.3	-0.1
b_Intercept[4]	1.0	553	-0.8	0.5	-1.7	-0.8	0.4
b_Intercept[5]	1.0	528	0.1	0.5	-0.9	0.0	1.2
b_Intercept[6]	1.0	536	1.7	0.5	0.7	1.6	2.8
b_ConditionSuppressive	1.0	2856	-0.4	0.3	-1.0	-0.4	0.3
b_Trial2	1.0	4191	0.2	0.2	-0.3	0.2	0.6
b_Trial3	1.0	3815	0.4	0.2	-0.0	0.4	0.9
b_Trial4	1.0	4666	0.2	0.2	-0.2	0.2	0.7
b_ConditionSuppressive:Trial2	1.0	4680	-0.3	0.3	-0.9	-0.3	0.3
b_ConditionSuppressive:Trial3	1.0	4723	-0.1	0.3	-0.7	-0.1	0.5
b_ConditionSuppressive:Trial4	1.0	5497	0.1	0.3	-0.5	0.1	0.8
sd_Item__Intercept	1.0	546	0.8	0.7	0.1	0.6	2.7
sd_Participant__Intercept	1.0	3547	0.7	0.1	0.5	0.7	1.0
disc		24000	1.0	0.0	1.0	1.0	1.0
r_Item[Q1,Intercept]	1.0	473	0.0	0.5	-0.9	-0.0	1.1
r_Item[Q2,Intercept]	1.0	472	0.3	0.5	-0.6	0.2	1.4
lprior	1.0	377	-15.9	0.5	-17.3	-15.8	-15.3
log-posterior	1.0	3357	-403.3	6.4	-416.7	-403.0	-391.7

Table C.6: Parameter table for the cumulative base model.

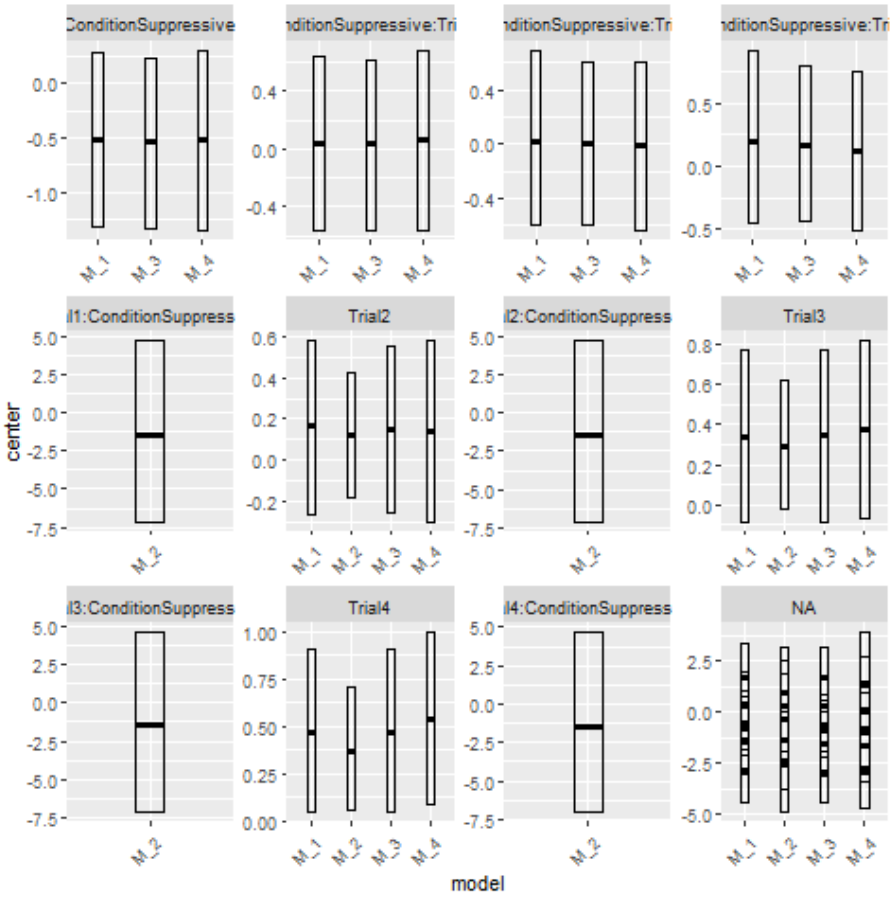


Figure C.23: Center & 95 % HDI bar plots of the SoO OLR models.

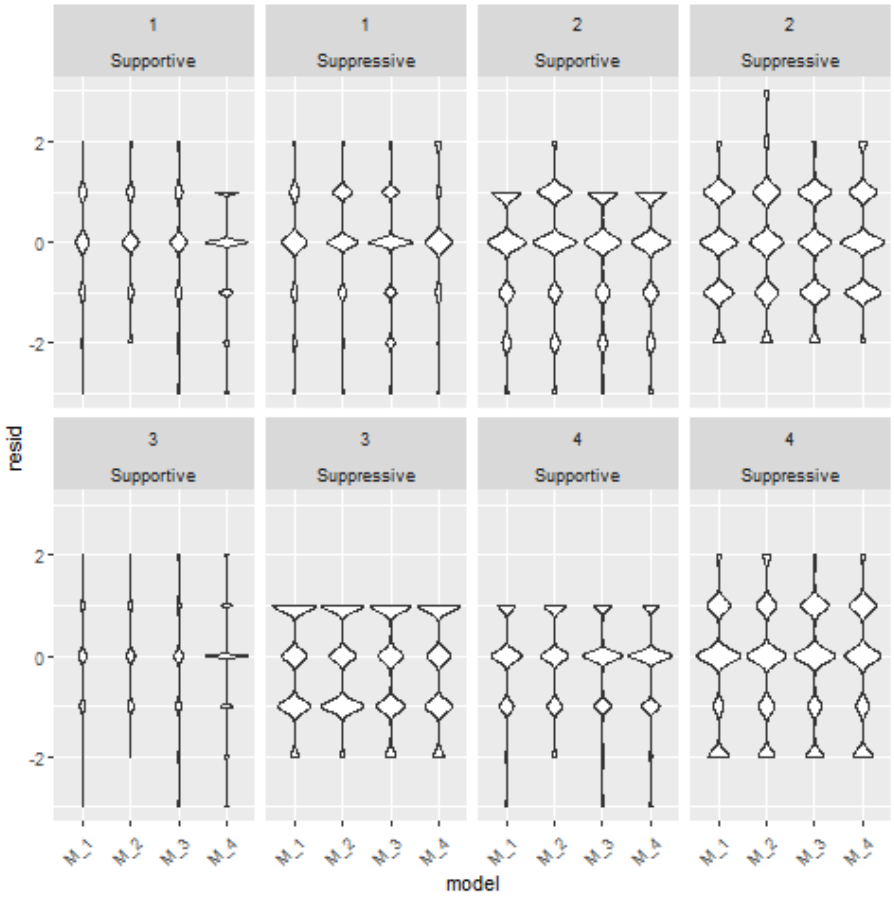


Figure C.24: Violin plots of the residuals of the SoO OLR models.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept[1]	1.0	1383	-4.1	3.1	-11.8	-3.5	-0.4
b_Intercept[2]	1.0	1622	-1.4	0.8	-3.1	-1.4	0.5
b_Intercept[3]	1.0	1311	-0.7	0.7	-2.4	-0.8	1.0
b_Intercept[4]	1.0	1289	-0.3	0.7	-2.0	-0.4	1.4
b_Intercept[5]	1.0	1121	0.0	0.7	-1.6	0.0	1.8
b_Intercept[6]	1.0	1280	1.3	0.7	-0.3	1.3	3.0
b_Trial2	1.0	9634	0.2	0.2	-0.2	0.1	0.5
b_Trial3	1.0	8905	0.3	0.2	0.0	0.3	0.7
b_Trial4	1.0	9606	0.2	0.2	-0.1	0.2	0.5
b_Trial1:ConditionSuppressive	1.0	969	-1.6	2.6	-6.7	-1.5	3.7
b_Trial2:ConditionSuppressive	1.0	904	-1.8	2.6	-7.0	-1.7	3.5
b_Trial3:ConditionSuppressive	1.0	976	-1.7	2.6	-6.9	-1.7	3.5
b_Trial4:ConditionSuppressive	1.0	977	-1.5	2.6	-6.8	-1.5	3.7
bcs_ConditionSuppressive[1]	1.0	1225	-1.3	3.9	-10.2	-0.8	5.0
bcs_ConditionSuppressive[2]	1.0	995	1.3	2.6	-4.0	1.3	6.5
bcs_ConditionSuppressive[3]	1.0	961	0.9	2.6	-4.3	0.9	6.1
bcs_ConditionSuppressive[4]	1.0	893	1.5	2.6	-3.8	1.4	6.6
bcs_ConditionSuppressive[5]	1.0	988	1.5	2.6	-3.6	1.4	6.6
bcs_ConditionSuppressive[6]	1.0	975	1.2	2.6	-4.1	1.2	6.4
sd_Item__Intercept	1.0	1862	0.9	0.9	0.1	0.5	3.3
sd_Participant__Intercept	1.0	3665	0.5	0.1	0.3	0.5	0.7
disc		14000	1.0	0.0	1.0	1.0	1.0
r_Item[Q1,Intercept]	1.0	1133	-0.1	0.7	-1.6	-0.1	1.6
r_Item[Q2,Intercept]	1.0	1129	0.1	0.7	-1.4	0.1	1.8

Table C.7: Parameter table for the category-adjacent model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept[1]	1.0	577	-3.0	0.6	-4.3	-3.0	-1.8
b_Intercept[2]	1.0	642	-2.1	0.5	-3.1	-2.1	-1.0
b_Intercept[3]	1.0	598	-1.2	0.5	-2.2	-1.3	-0.1
b_Intercept[4]	1.0	553	-0.8	0.5	-1.7	-0.8	0.4
b_Intercept[5]	1.0	528	0.1	0.5	-0.9	0.0	1.2
b_Intercept[6]	1.0	536	1.7	0.5	0.7	1.6	2.8
b_ConditionSuppressive	1.0	2856	-0.4	0.3	-1.0	-0.4	0.3
b_Trial2	1.0	4191	0.2	0.2	-0.3	0.2	0.6
b_Trial3	1.0	3815	0.4	0.2	-0.0	0.4	0.9
b_Trial4	1.0	4666	0.2	0.2	-0.2	0.2	0.7
b_ConditionSuppressive:Trial2	1.0	4680	-0.3	0.3	-0.9	-0.3	0.3
b_ConditionSuppressive:Trial3	1.0	4723	-0.1	0.3	-0.7	-0.1	0.5
b_ConditionSuppressive:Trial4	1.0	5497	0.1	0.3	-0.5	0.1	0.8
sd_Item__Intercept	1.0	546	0.8	0.7	0.1	0.6	2.7
sd_Participant__Intercept	1.0	3547	0.7	0.1	0.5	0.7	1.0
disc		24000	1.0	0.0	1.0	1.0	1.0
r_Item[Q1,Intercept]	1.0	473	0.0	0.5	-0.9	-0.0	1.1
r_Item[Q2,Intercept]	1.0	472	0.3	0.5	-0.6	0.2	1.4

Table C.8: Parameter table for the unequal variances model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept[1]	1.0	2316	-2.8	1.3	-4.8	-3.0	0.4
b_Intercept[2]	1.0	3101	-2.0	0.9	-3.6	-2.1	0.3
b_Intercept[3]	1.0	4976	-1.2	0.8	-2.7	-1.3	0.8
b_Intercept[4]	1.0	2390	-0.9	0.8	-2.4	-0.9	0.9
b_Intercept[5]	1.0	860	-0.1	0.9	-1.9	-0.1	2.0
b_Intercept[6]	1.0	5388	1.3	1.1	-1.2	1.4	3.3
b_ConditionSuppressive	1.0	3881	-0.4	0.3	-1.1	-0.4	0.3
b_Trial2	1.0	7052	0.2	0.2	-0.3	0.2	0.6
b_Trial3	1.0	3886	0.4	0.2	-0.0	0.4	0.9
b_Trial4	1.0	7939	0.2	0.2	-0.2	0.2	0.7
b_ConditionSuppressive:Trial2	1.0	7528	-0.3	0.3	-0.9	-0.3	0.3
b_ConditionSuppressive:Trial3	1.0	4831	-0.1	0.3	-0.7	-0.1	0.5
b_ConditionSuppressive:Trial4	1.0	8603	0.2	0.3	-0.5	0.2	0.8
sd_Item__Intercept[1]	1.0	6344	1.4	1.5	0.0	0.9	5.4
sd_Item__Intercept[2]	1.0	3901	1.1	1.3	0.0	0.7	4.6
sd_Item__Intercept[3]	1.0	3725	1.0	1.1	0.0	0.6	4.1
sd_Item__Intercept[4]	1.0	2052	0.9	1.1	0.0	0.5	4.0
sd_Item__Intercept[5]	1.0	1547	1.1	1.1	0.1	0.8	4.1
sd_Item__Intercept[6]	1.0	2183	1.5	1.4	0.1	1.1	5.1
sd_Participant__Intercept	1.0	2403	0.7	0.1	0.5	0.7	1.0
cor_Item__Intercept[1]__Intercept[2]	1.0	24000	0.0	0.4	-0.7	-0.0	0.7
cor_Item__Intercept[1]__Intercept[3]	1.0	12877	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[2]__Intercept[3]	1.0	2686	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[1]__Intercept[4]	1.0	24000	-0.0	0.4	-0.7	-0.0	0.7
cor_Item__Intercept[2]__Intercept[4]	1.0	4172	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[3]__Intercept[4]	1.0	3418	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[1]__Intercept[5]	1.0	5018	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[2]__Intercept[5]	1.0	12920	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[3]__Intercept[5]	1.0	9715	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[4]__Intercept[5]	1.0	5829	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[1]__Intercept[6]	1.0	4099	-0.0	0.4	-0.7	-0.0	0.7
cor_Item__Intercept[2]__Intercept[6]	1.0	4784	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[3]__Intercept[6]	1.0	3729	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[4]__Intercept[6]	1.0	2334	0.0	0.4	-0.7	0.0	0.8
cor_Item__Intercept[5]__Intercept[6]	1.0	4562	0.0	0.4	-0.7	0.0	0.7
disc		24000	1.0	0.0	1.0	1.0	1.0
r_Item[Q1,Intercept[1]]	1.0	2531	0.4	1.3	-1.5	0.1	3.7
r_Item[Q2,Intercept[1]]	1.0	2355	0.4	1.3	-1.4	0.1	3.6
r_Item[Q1,Intercept[2]]	1.0	2896	0.2	0.9	-1.4	0.0	2.5
r_Item[Q2,Intercept[2]]	1.0	3096	0.3	0.9	-1.2	0.1	2.7
r_Item[Q1,Intercept[3]]	1.0	4871	0.0	0.8	-1.4	-0.0	2.0
r_Item[Q2,Intercept[3]]	1.0	4856	0.2	0.8	-1.2	0.1	2.3
r_Item[Q1,Intercept[4]]	1.0	2320	-0.0	0.7	-1.5	-0.0	1.7
r_Item[Q2,Intercept[4]]	1.0	2334	0.1	0.7	-1.4	0.0	1.8
r_Item[Q1,Intercept[5]]	1.0	743	-0.1	0.9	-2.0	-0.1	1.9
r_Item[Q2,Intercept[5]]	1.0	728	0.2	0.9	-1.7	0.1	2.3
r_Item[Q1,Intercept[6]]	1.0	5540	-0.6	1.1	-3.2	-0.4	1.3
r_Item[Q2,Intercept[6]]	1.0	5344	0.0	1.0	-2.5	0.1	2.0

Table C.9: Parameter table for the unequal item variances model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept[1]	1.0	483	-3.4	0.5	-4.4	-3.4	-2.4
b_Intercept[2]	1.0	291	-2.0	0.4	-2.7	-2.0	-1.0
b_Intercept[3]	1.0	311	-1.0	0.4	-1.7	-1.1	-0.2
b_Intercept[4]	1.0	310	-0.7	0.4	-1.4	-0.7	0.2
b_Intercept[5]	1.0	315	-0.1	0.4	-0.8	-0.2	0.7
b_Intercept[6]	1.0	392	1.3	0.4	0.6	1.3	2.2
b_ConditionSuppressive	1.0	2762	-1.2	0.3	-1.9	-1.2	-0.6
b_Trial2	1.0	2028	0.1	0.2	-0.3	0.1	0.5
b_Trial3	1.0	2001	0.4	0.2	-0.1	0.4	0.8
b_Trial4	1.0	2535	0.7	0.2	0.3	0.7	1.1
b_ConditionSuppressive:Trial2	1.0	2507	0.2	0.3	-0.4	0.1	0.8
b_ConditionSuppressive:Trial3	1.0	1970	0.3	0.3	-0.3	0.3	0.9
b_ConditionSuppressive:Trial4	1.0	1833	-0.0	0.3	-0.7	-0.0	0.6
sd_Item__Intercept	1.0	240	0.5	0.6	0.0	0.3	2.3
sd_Participant__Intercept	1.0	1294	0.7	0.1	0.5	0.7	0.9
disc		24000	1.0	0.0	1.0	1.0	1.0
r_Item[Q1,Intercept]	1.0	280	0.0	0.3	-0.5	-0.0	0.7
r_Item[Q2,Intercept]	1.0	281	0.1	0.3	-0.4	0.1	0.8

Table C.10: Parameter table for the cumulative base model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept[1]	1.0	483	-3.4	0.5	-4.4	-3.4	-2.4
b_Intercept[2]	1.0	291	-2.0	0.4	-2.7	-2.0	-1.0
b_Intercept[3]	1.0	311	-1.0	0.4	-1.7	-1.1	-0.2
b_Intercept[4]	1.0	310	-0.7	0.4	-1.4	-0.7	0.2
b_Intercept[5]	1.0	315	-0.1	0.4	-0.8	-0.2	0.7
b_Intercept[6]	1.0	392	1.3	0.4	0.6	1.3	2.2
b_ConditionSuppressive	1.0	2762	-1.2	0.3	-1.9	-1.2	-0.6
b_Trial2	1.0	2028	0.1	0.2	-0.3	0.1	0.5
b_Trial3	1.0	2001	0.4	0.2	-0.1	0.4	0.8
b_Trial4	1.0	2535	0.7	0.2	0.3	0.7	1.1
b_ConditionSuppressive:Trial2	1.0	2507	0.2	0.3	-0.4	0.1	0.8
b_ConditionSuppressive:Trial3	1.0	1970	0.3	0.3	-0.3	0.3	0.9
b_ConditionSuppressive:Trial4	1.0	1833	-0.0	0.3	-0.7	-0.0	0.6
sd_Item__Intercept	1.0	240	0.5	0.6	0.0	0.3	2.3
sd_Participant__Intercept	1.0	1294	0.7	0.1	0.5	0.7	0.9
disc		24000	1.0	0.0	1.0	1.0	1.0
r_Item[Q1,Intercept]	1.0	280	0.0	0.3	-0.5	-0.0	0.7
r_Item[Q2,Intercept]	1.0	281	0.1	0.3	-0.4	0.1	0.8

Table C.11: Parameter table for the cumulative base model.

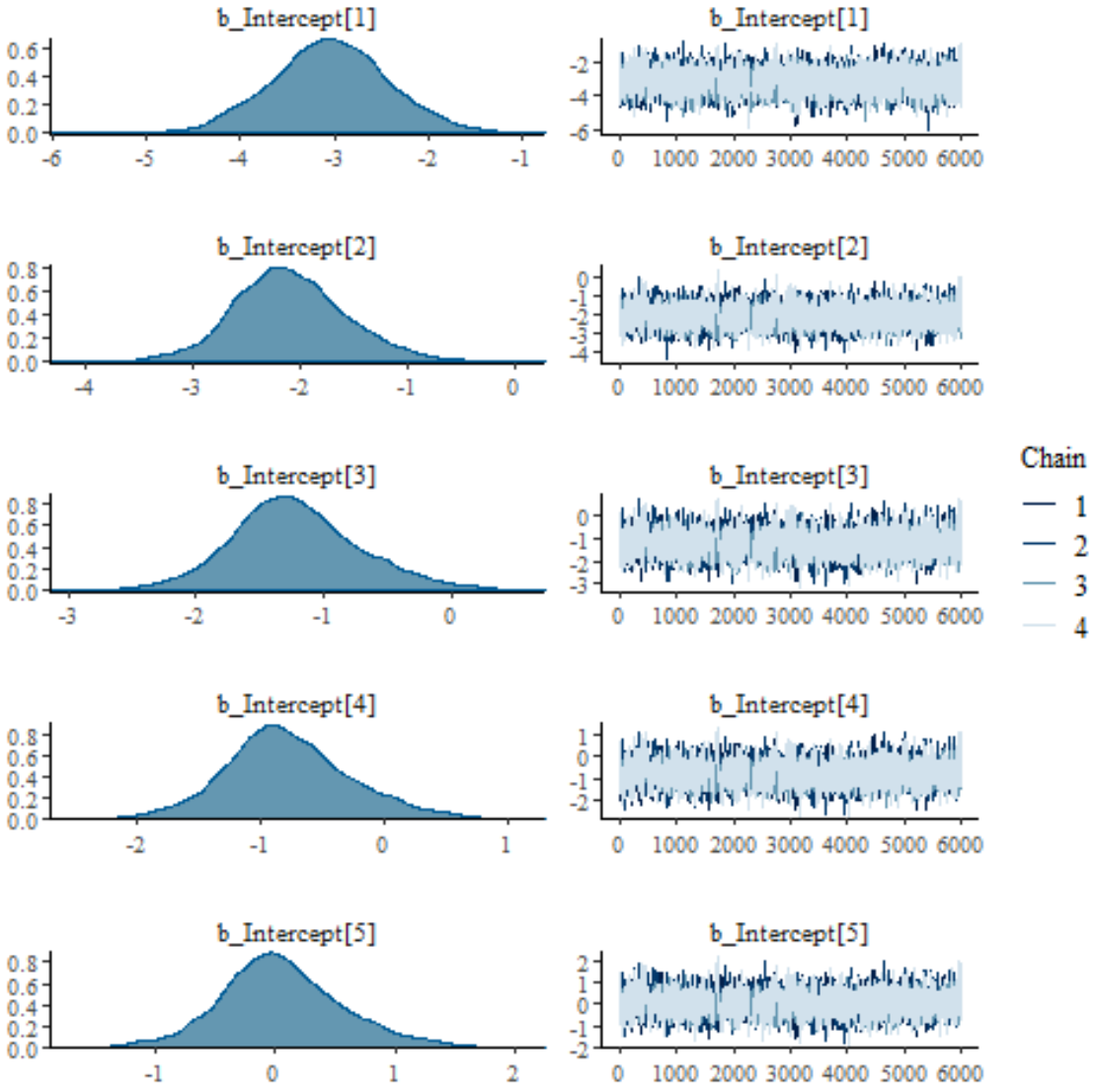


Figure C.25: Density & trace plots for the cumulative model.

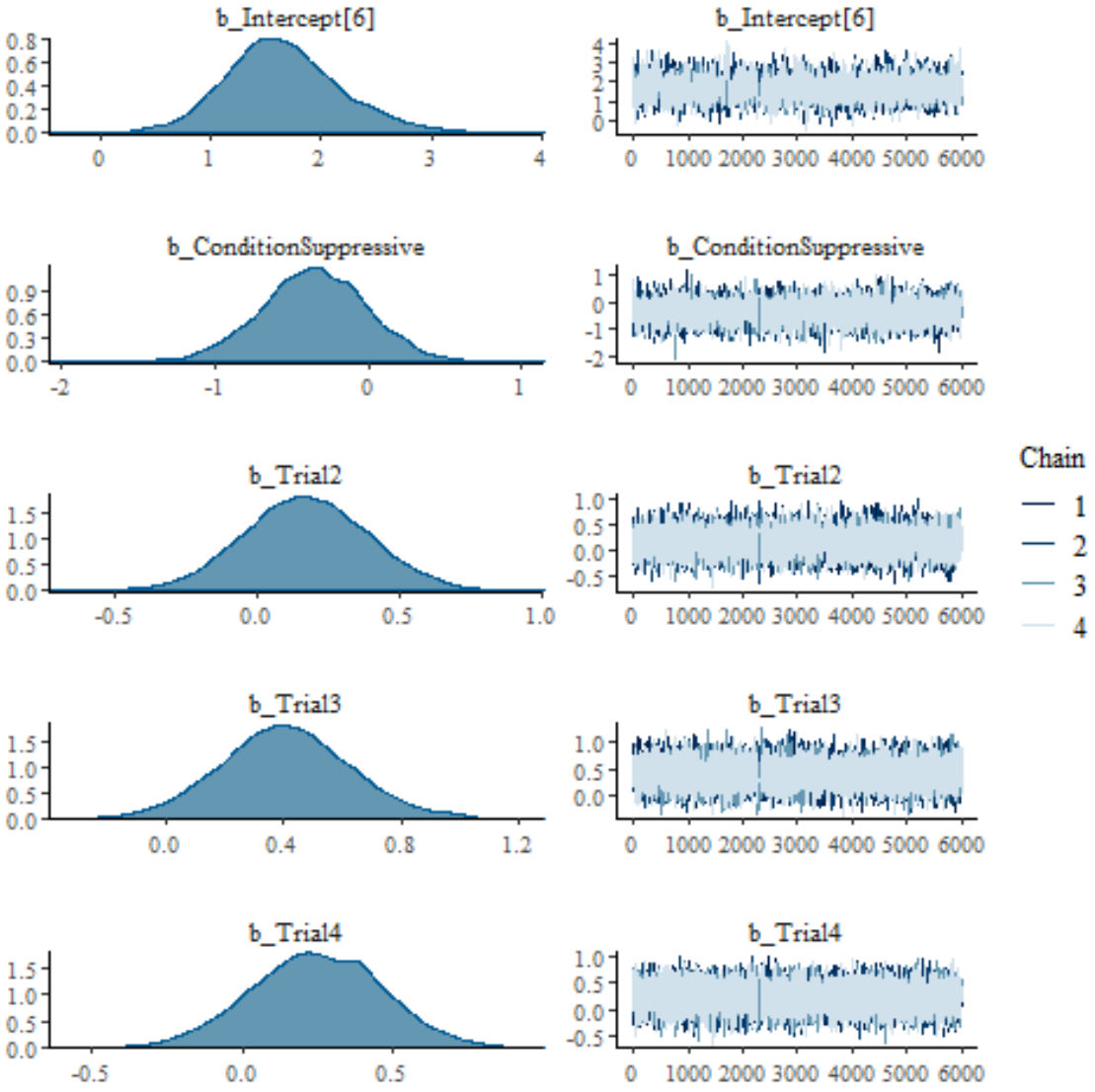


Figure C.26: Density & trace plots for the cumulative model.

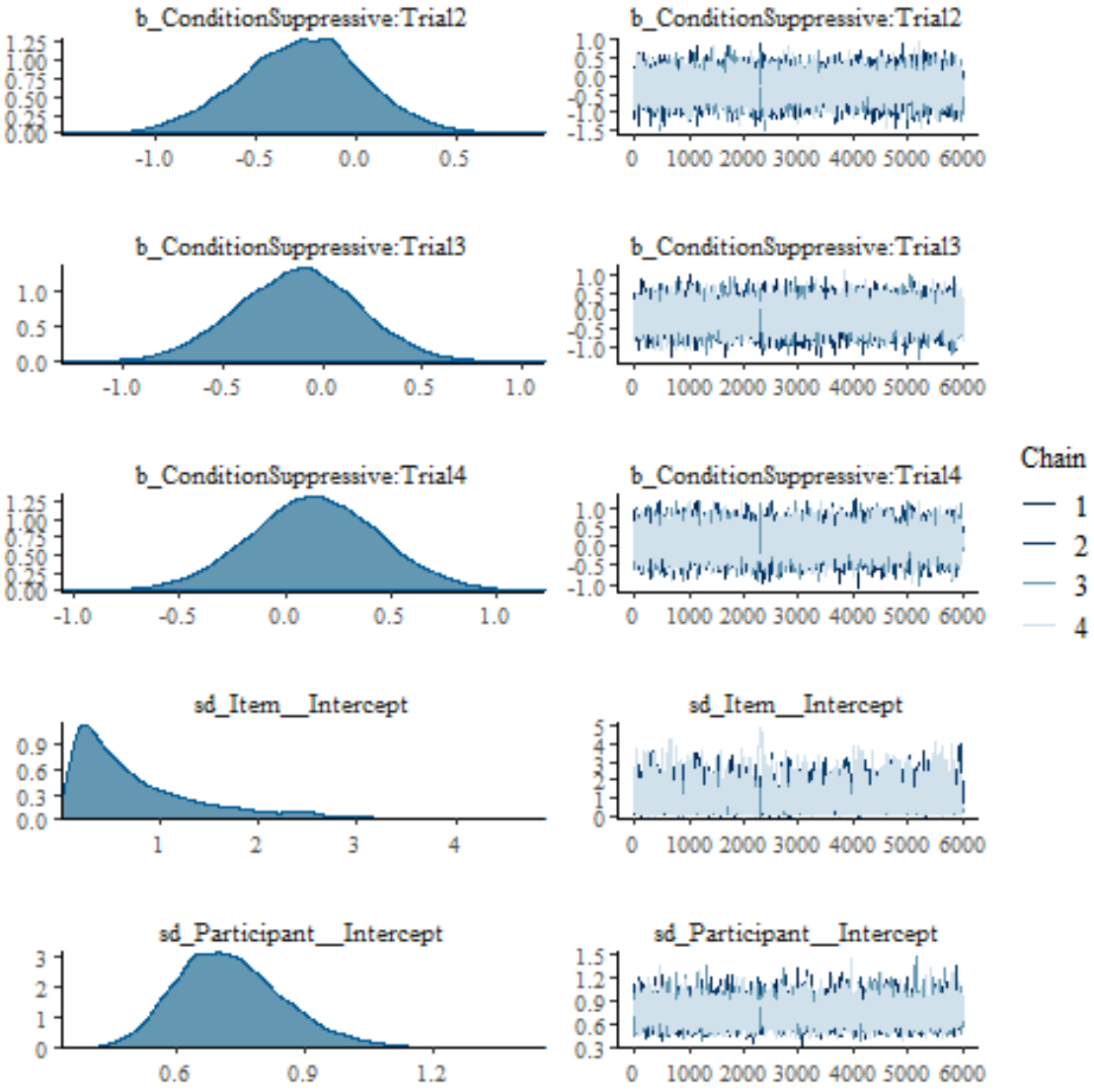


Figure C.27: Density & trace plots for the cumulative model.

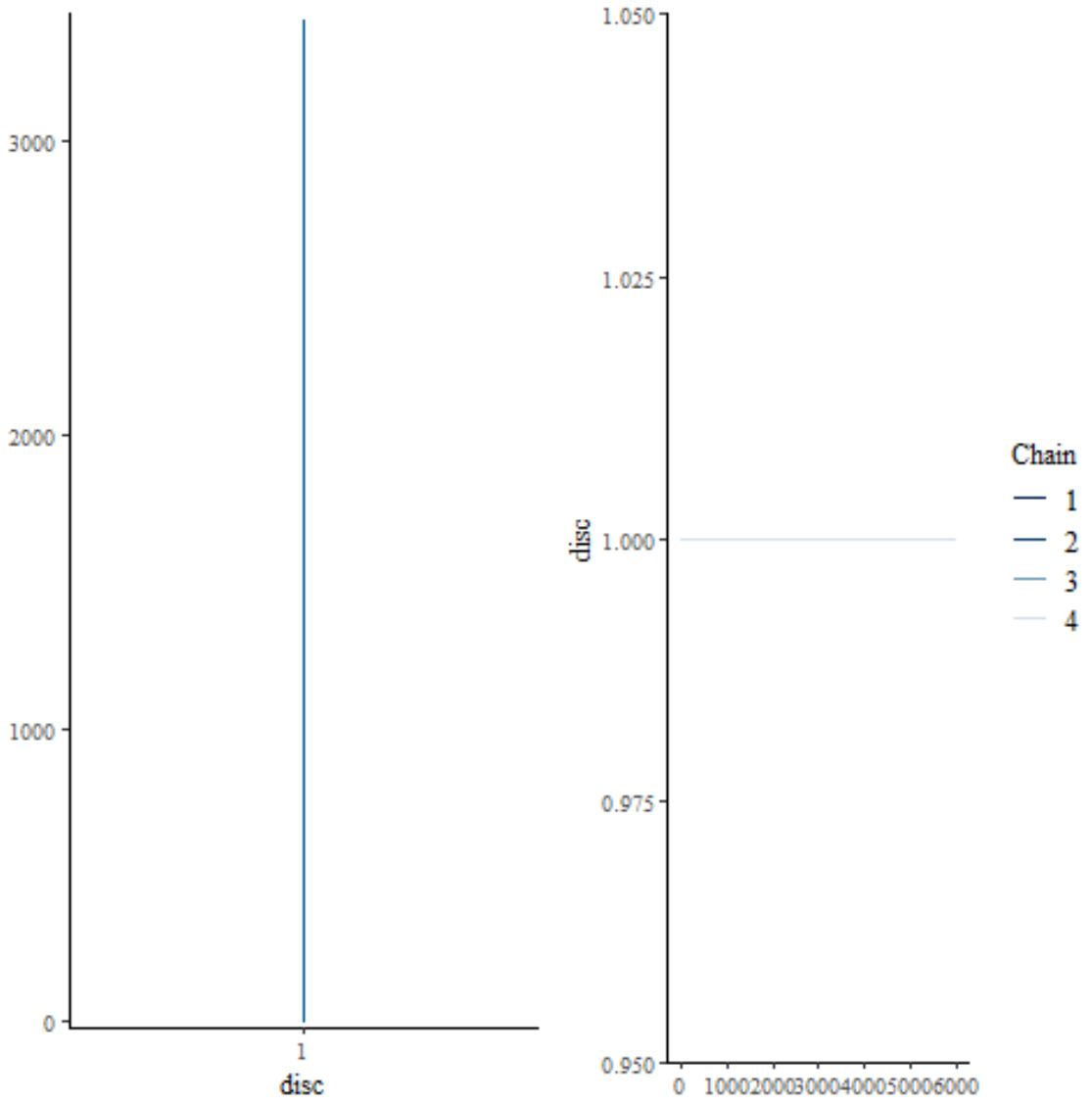


Figure C.28: Density & trace plots for the cumulative model.

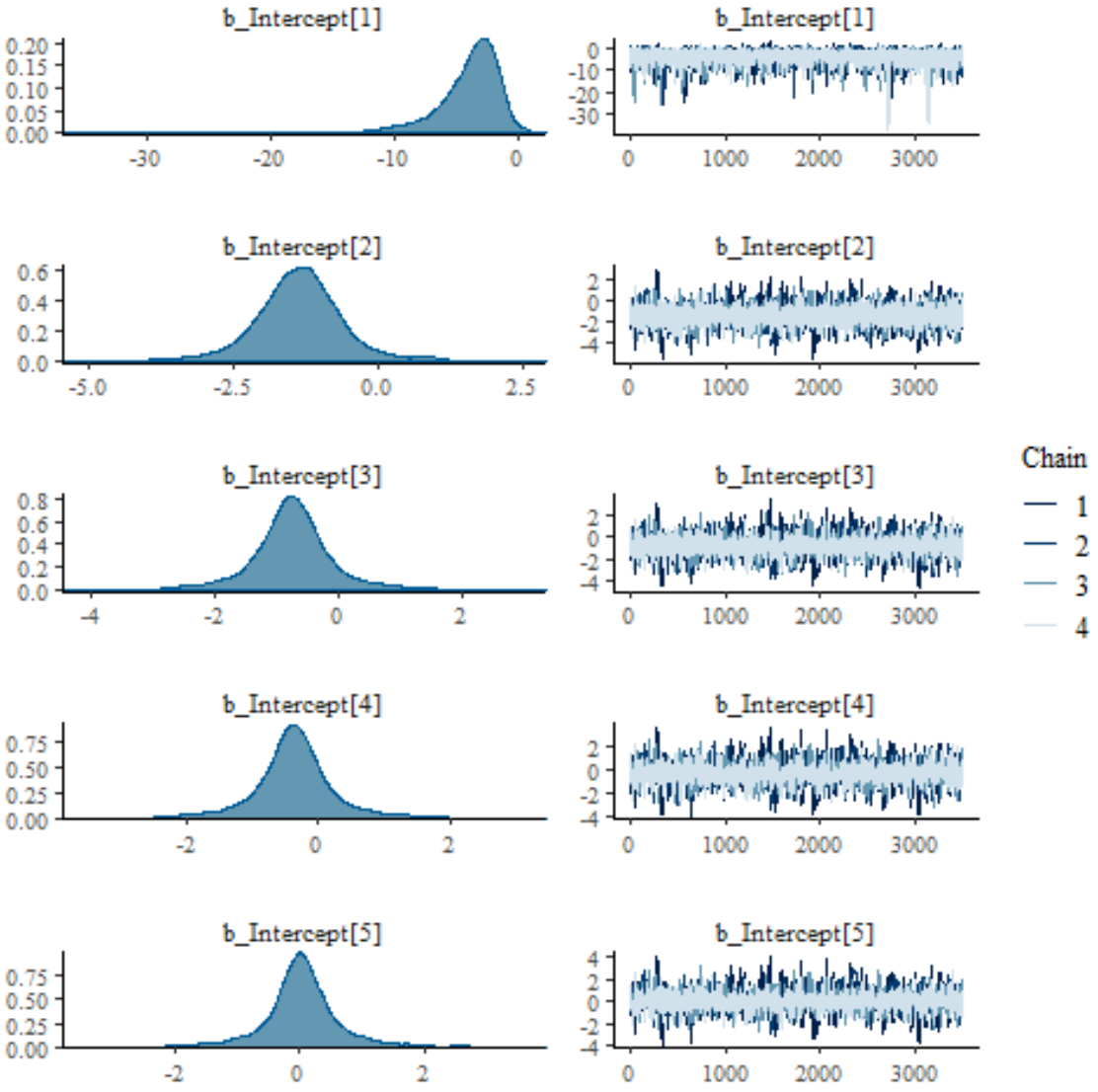


Figure C.29: Density & trace plots for the categorical model.

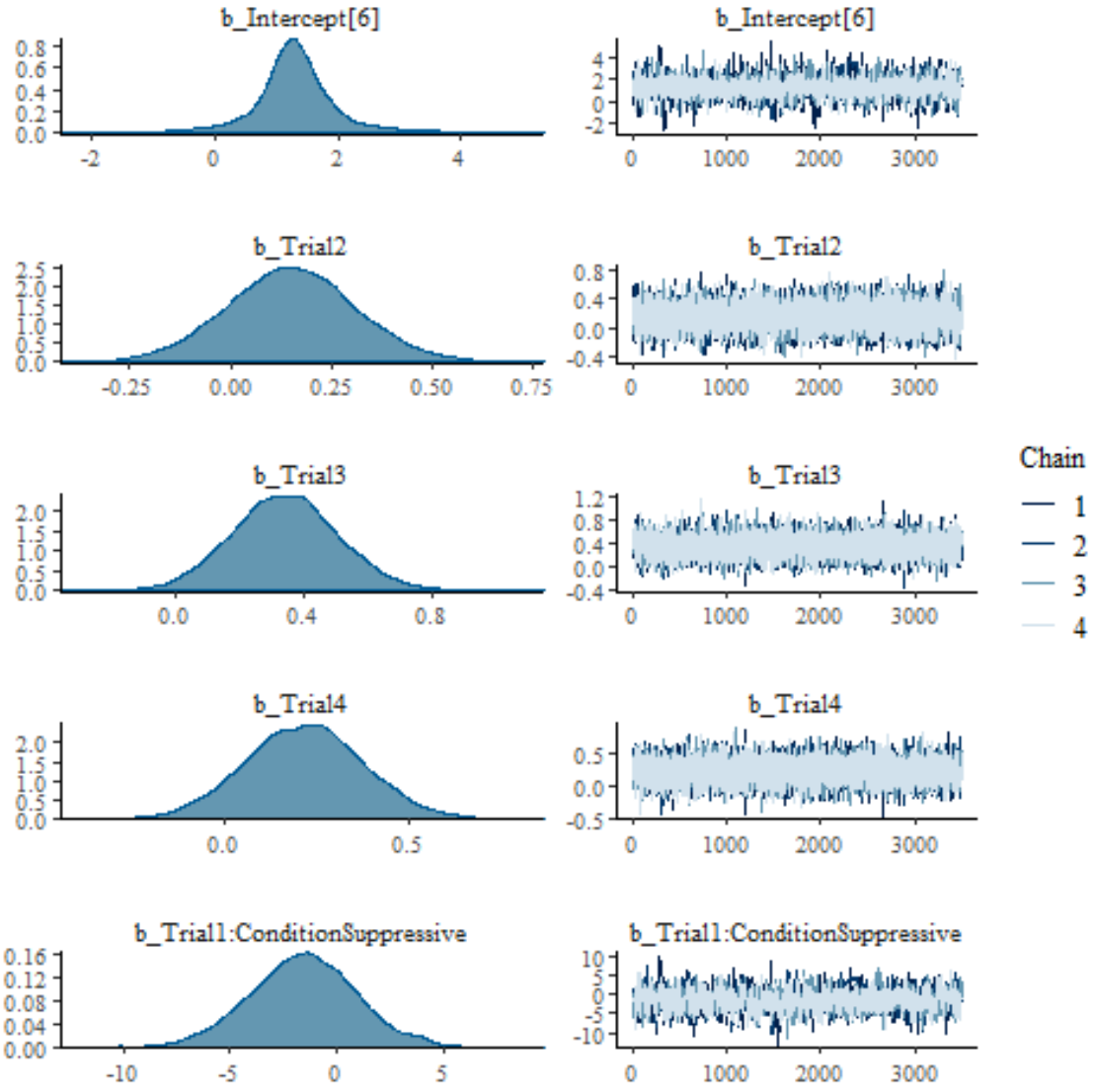


Figure C.30: Density & trace plots for the categorical model.

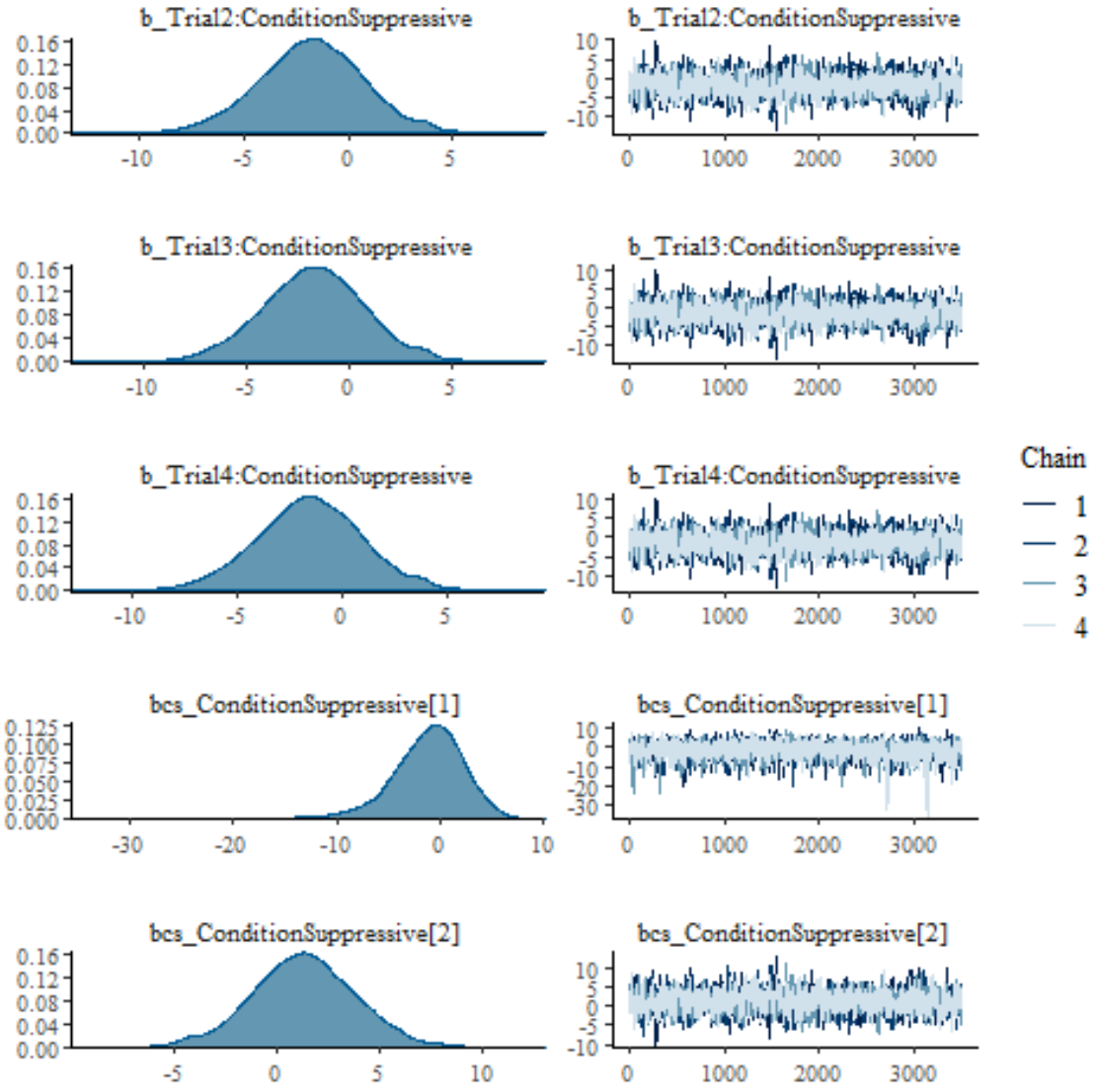


Figure C.31: Density & trace plots for the categorical model.

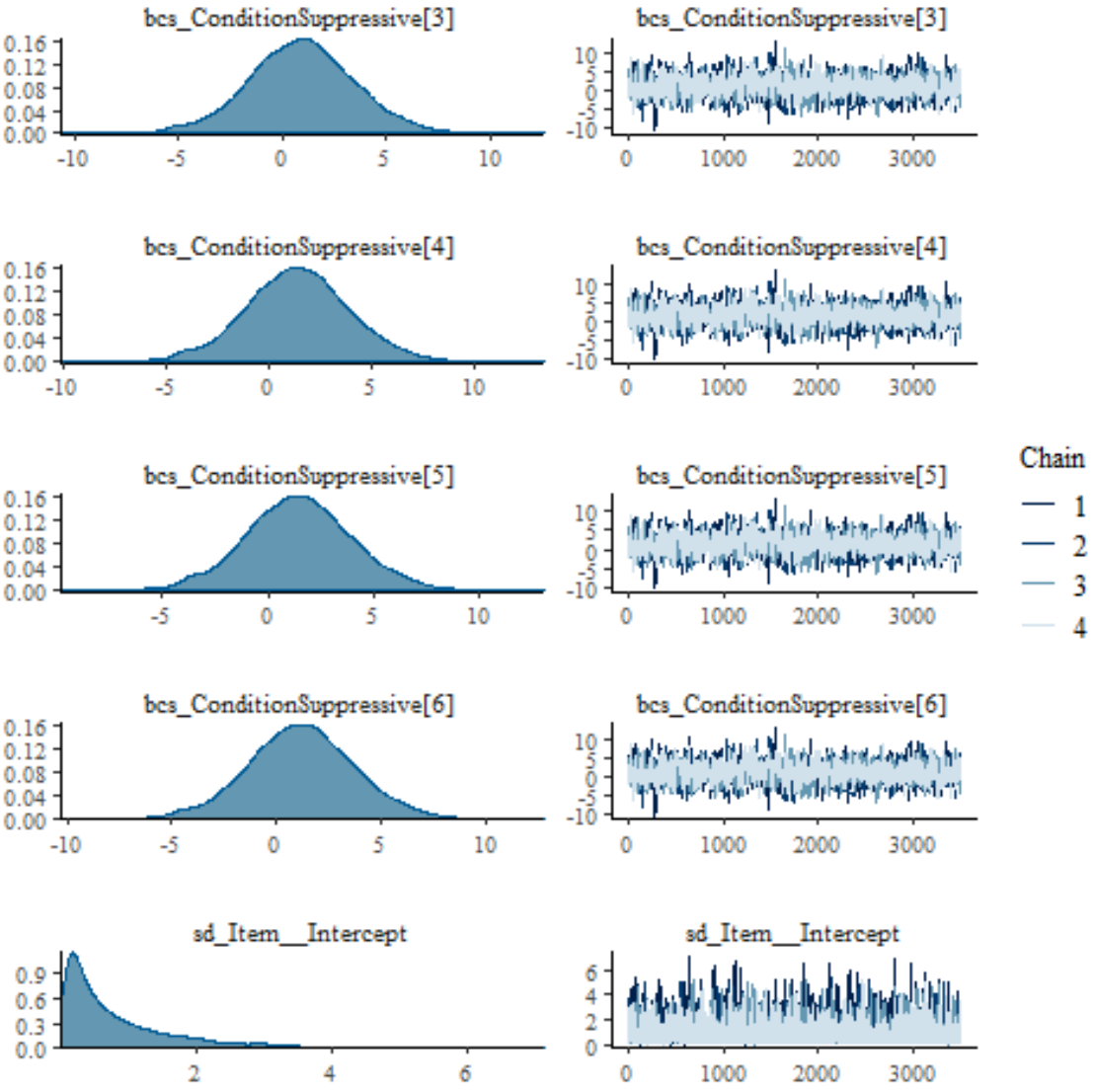


Figure C.32: Density & trace plots for the categorical model.

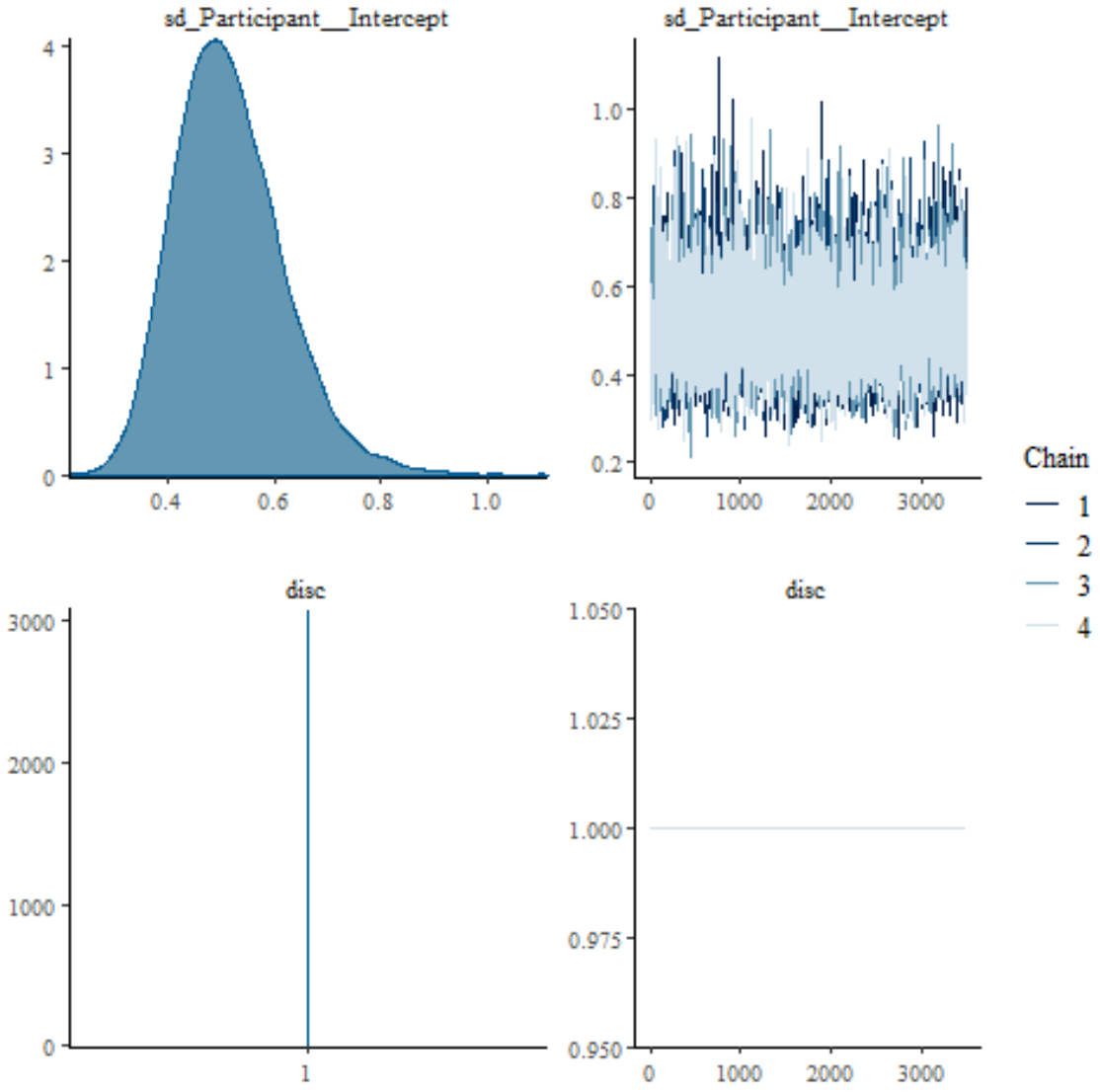


Figure C.33: Density & trace plots for the categorical model.

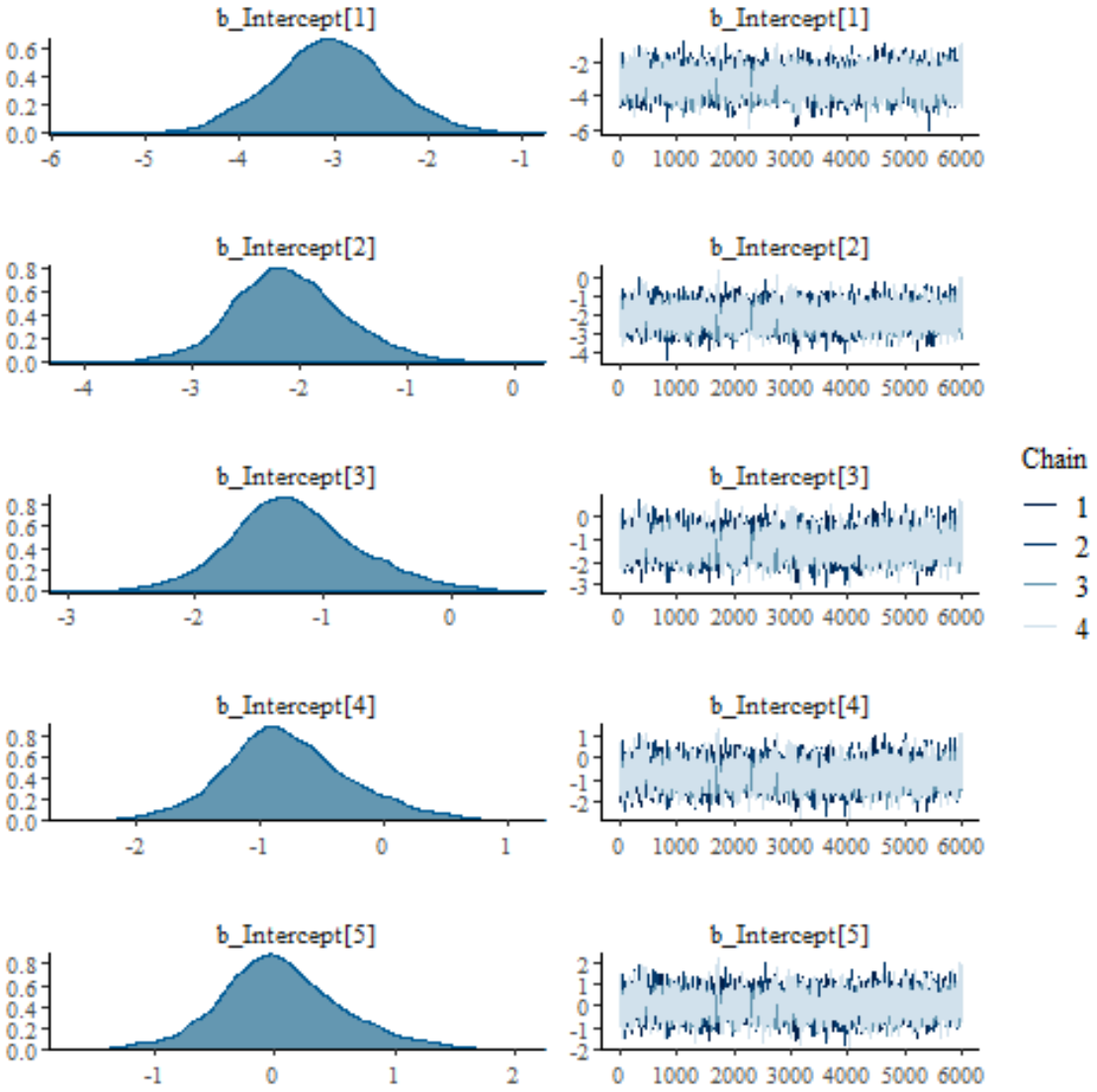


Figure C.34: Density & trace plots for the unequal variance between conditions model.

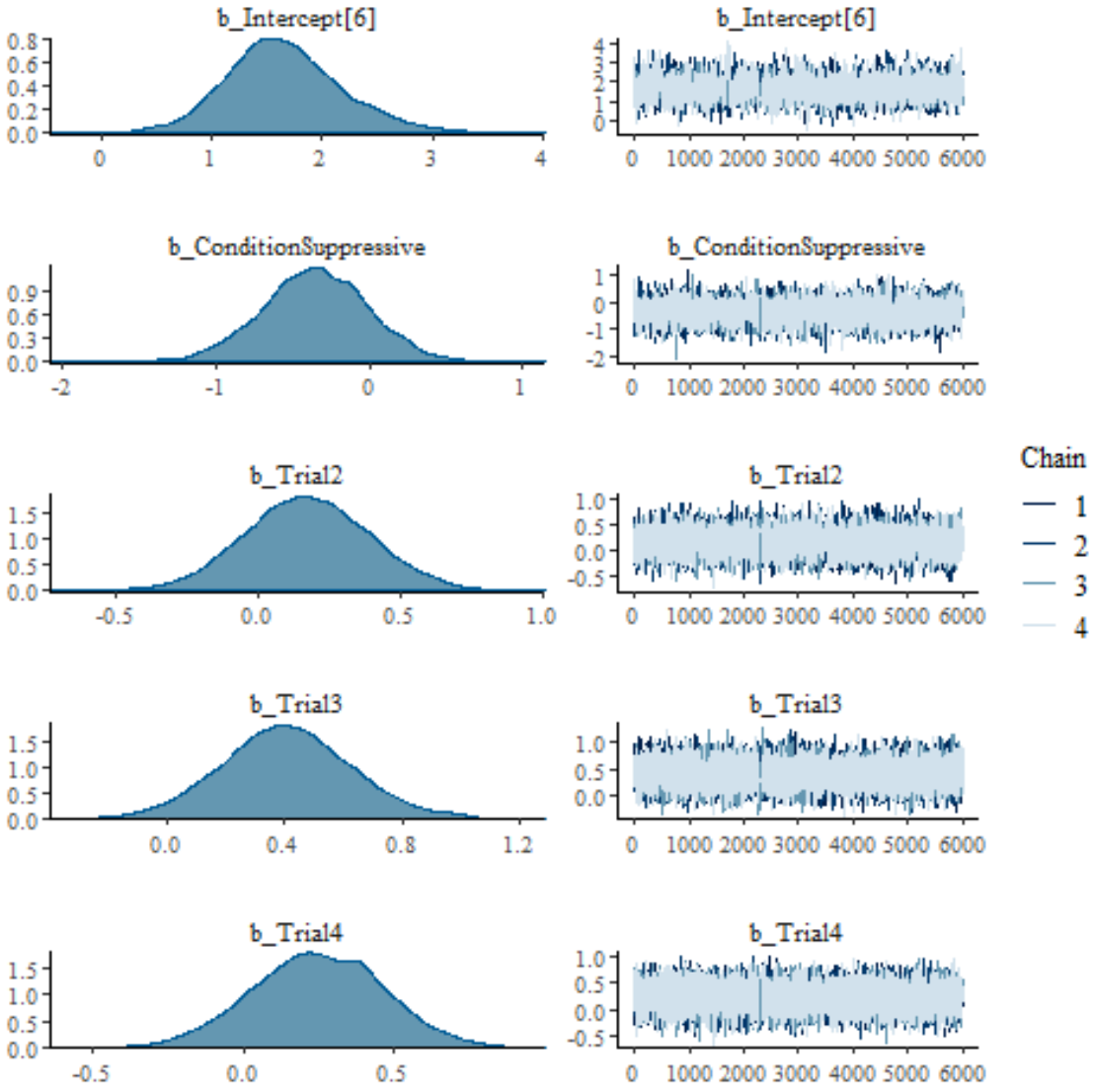


Figure C.35: Density & trace plots for the unequal variance between conditions model.

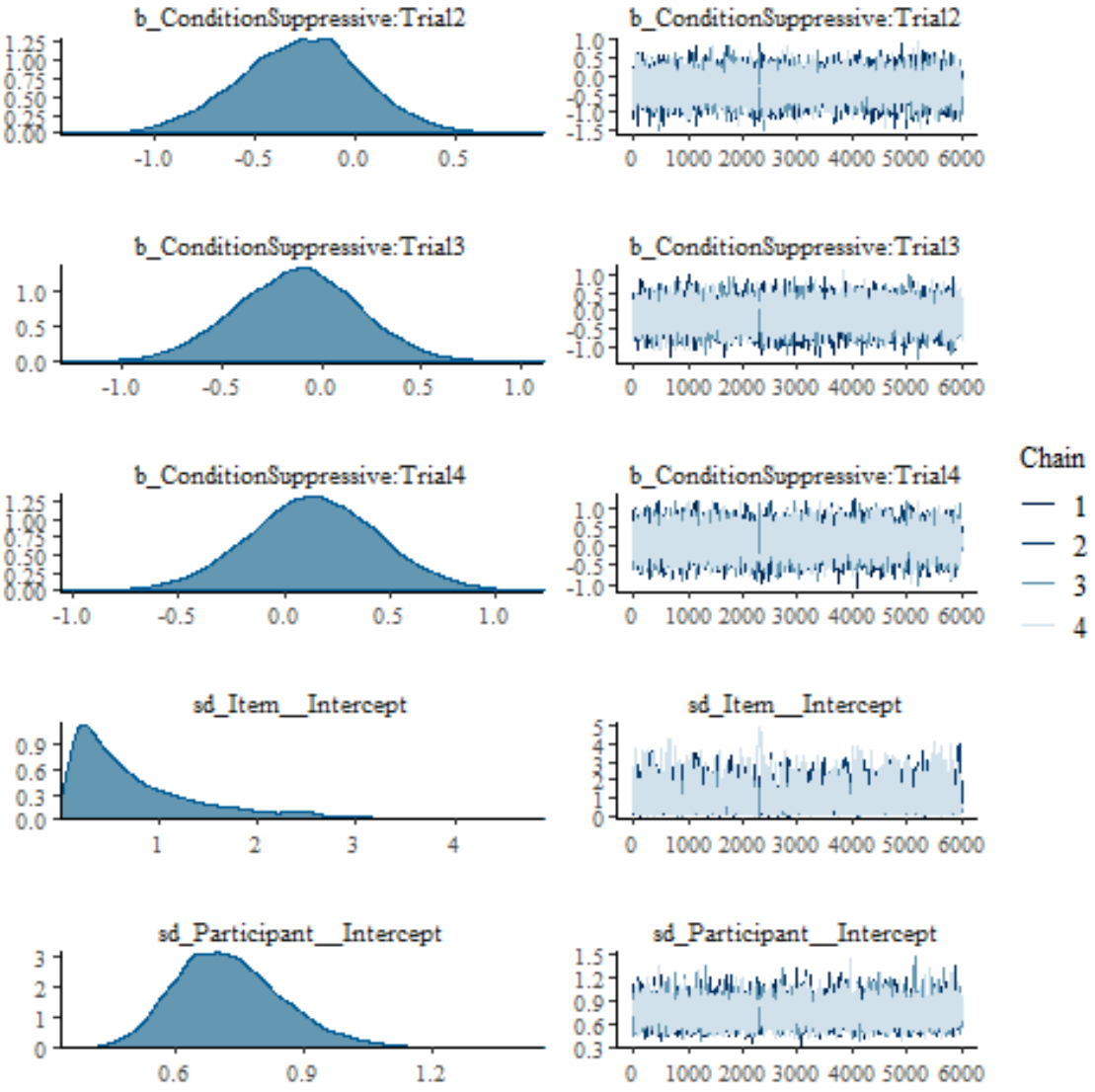


Figure C.36: Density & trace plots for the unequal variance between conditions model.

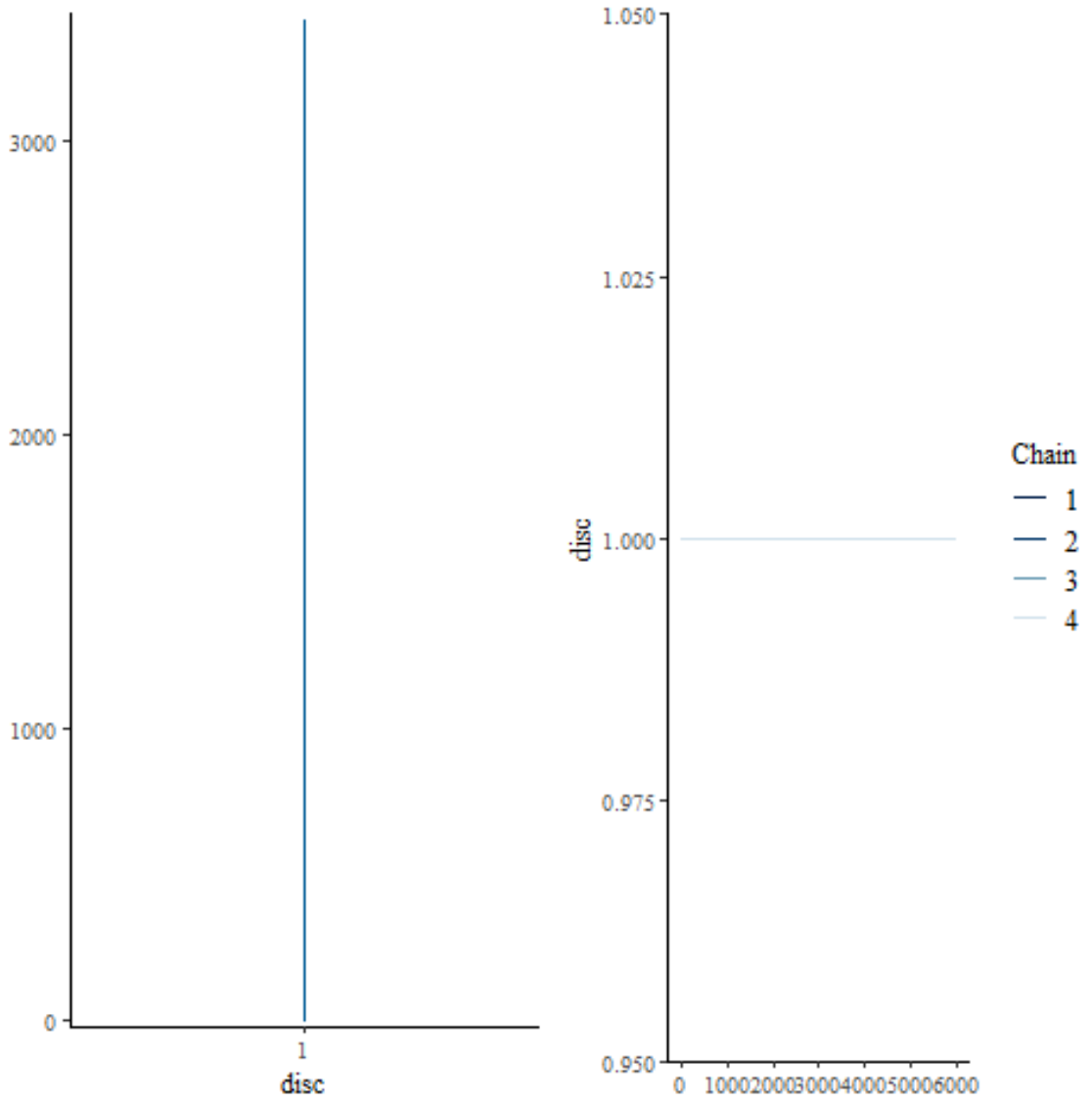


Figure C.37: Density & trace plots for the unequal variance between conditions model.

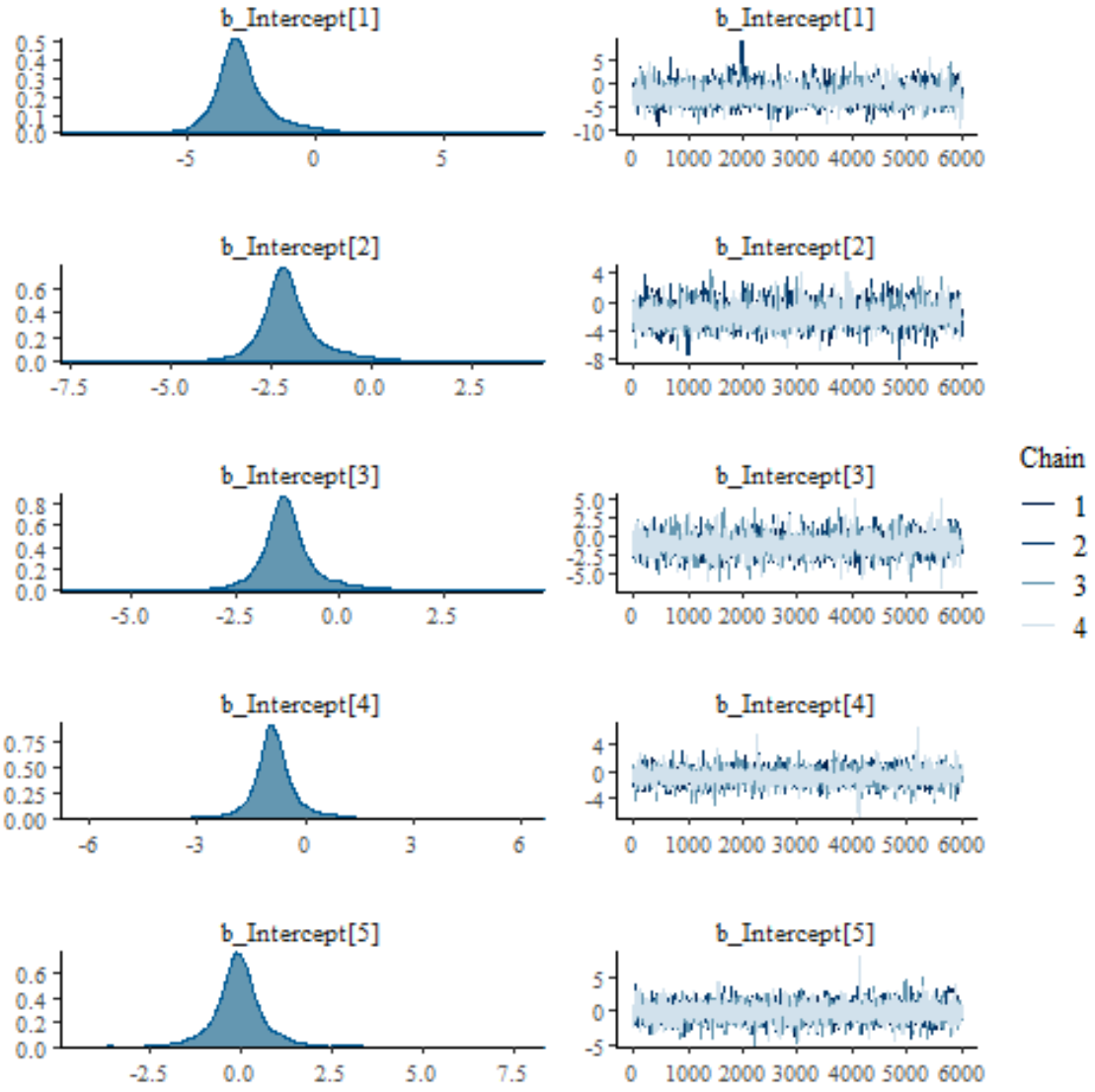


Figure C.38: Density & trace plots for the unequal variance between items model.

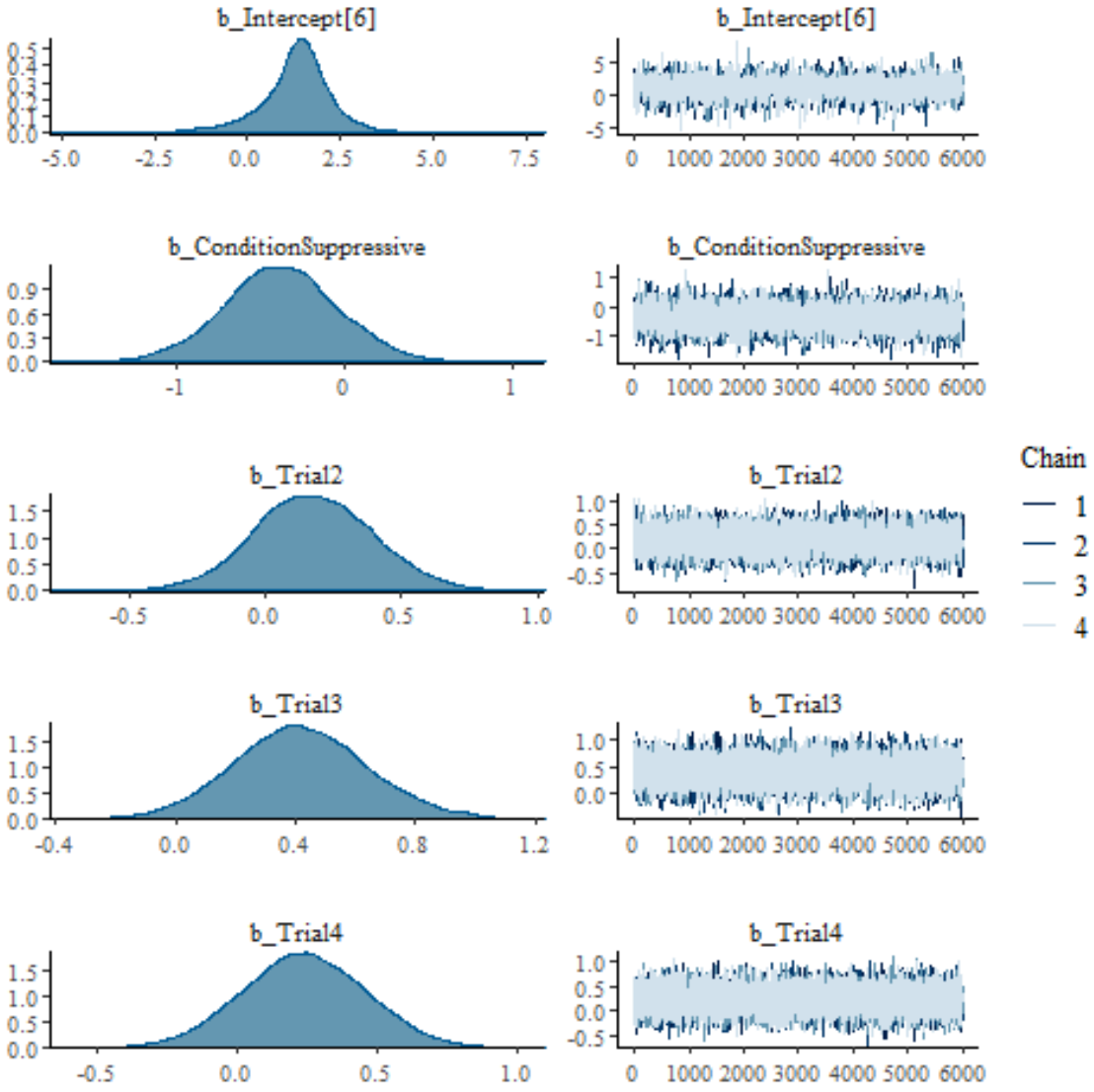


Figure C.39: Density & trace plots for the unequal variance between items model.

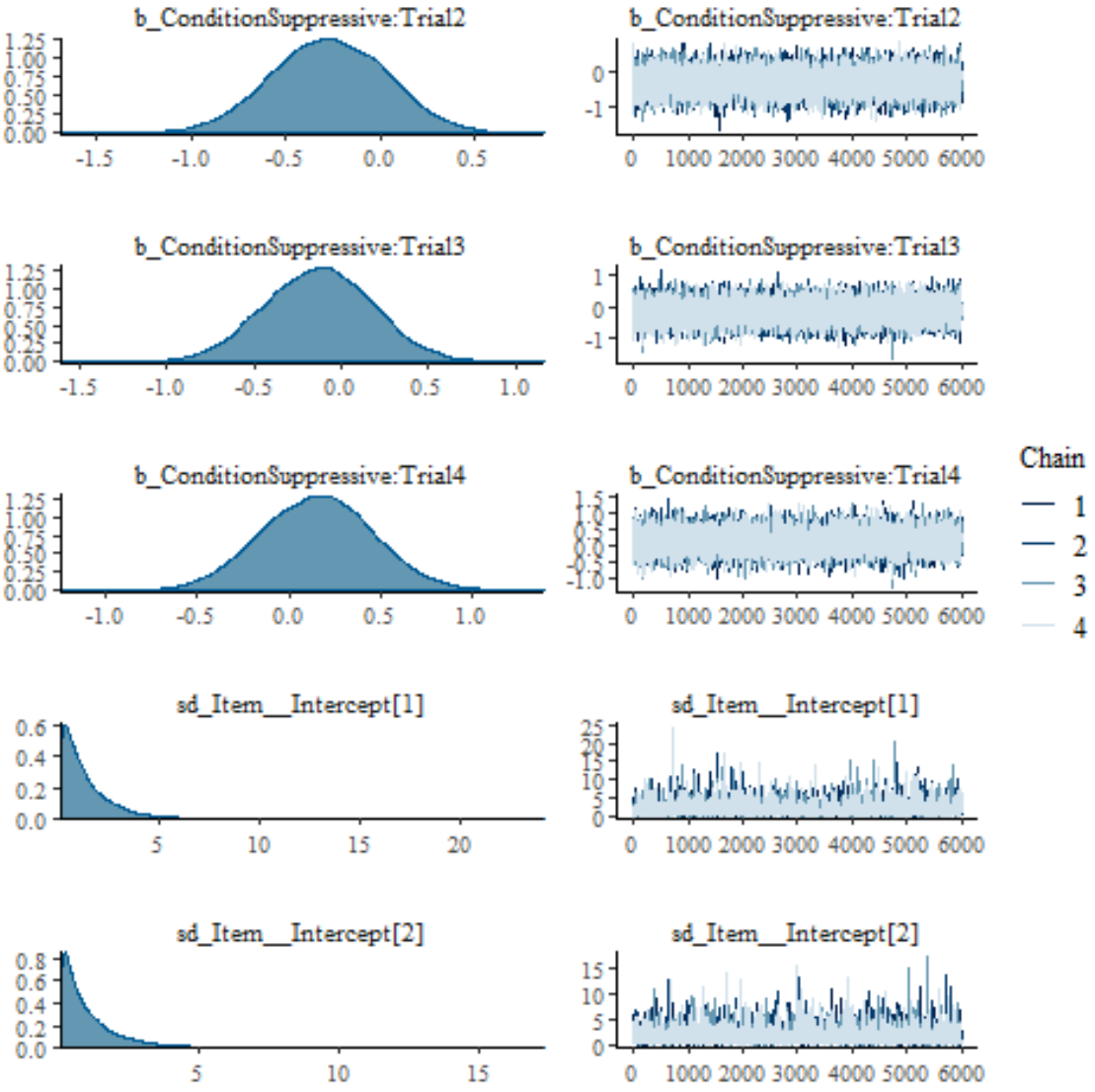


Figure C.40: Density & trace plots for the unequal variance between items model.

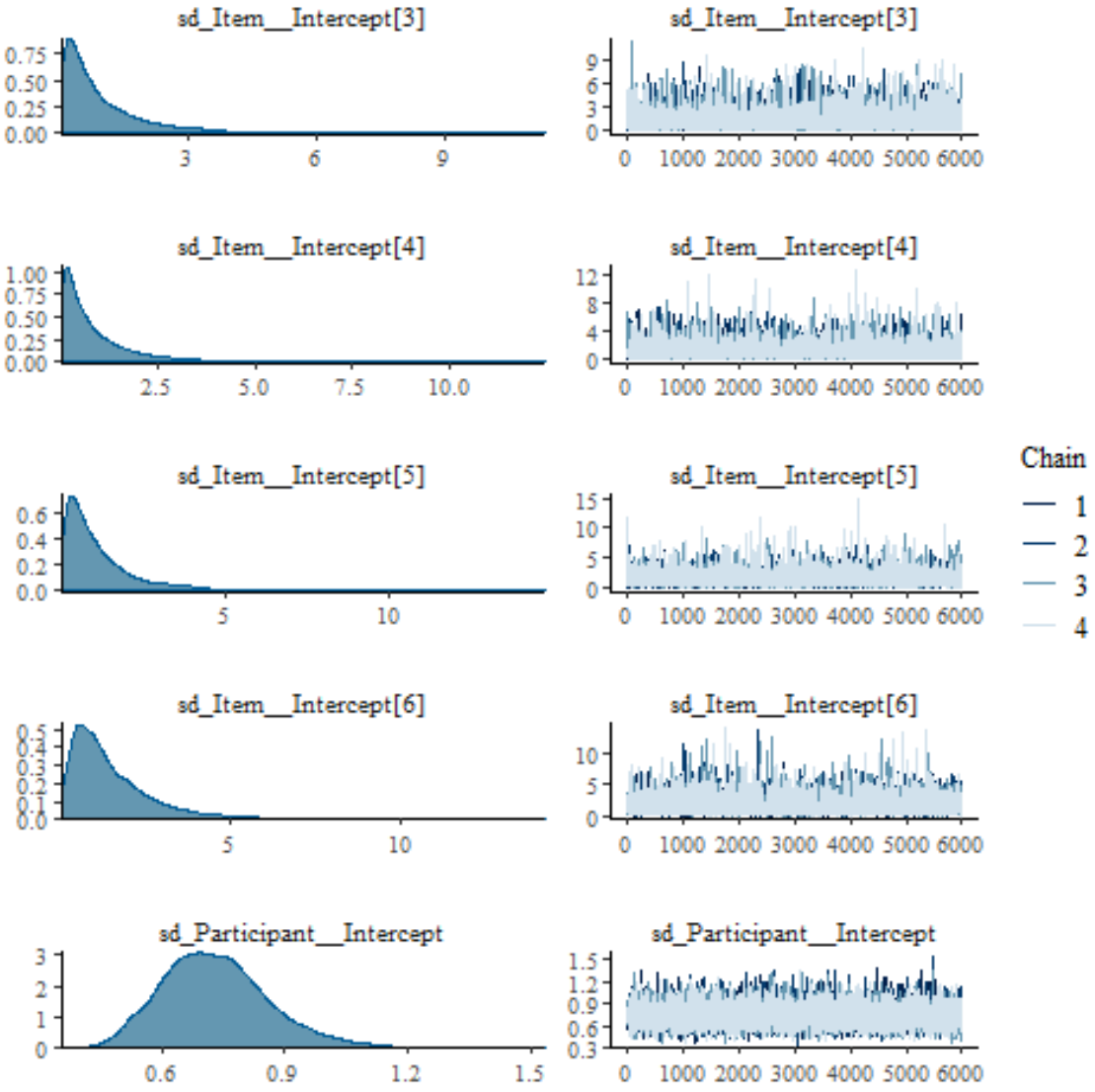


Figure C.41: Density & trace plots for the unequal variance between items model.

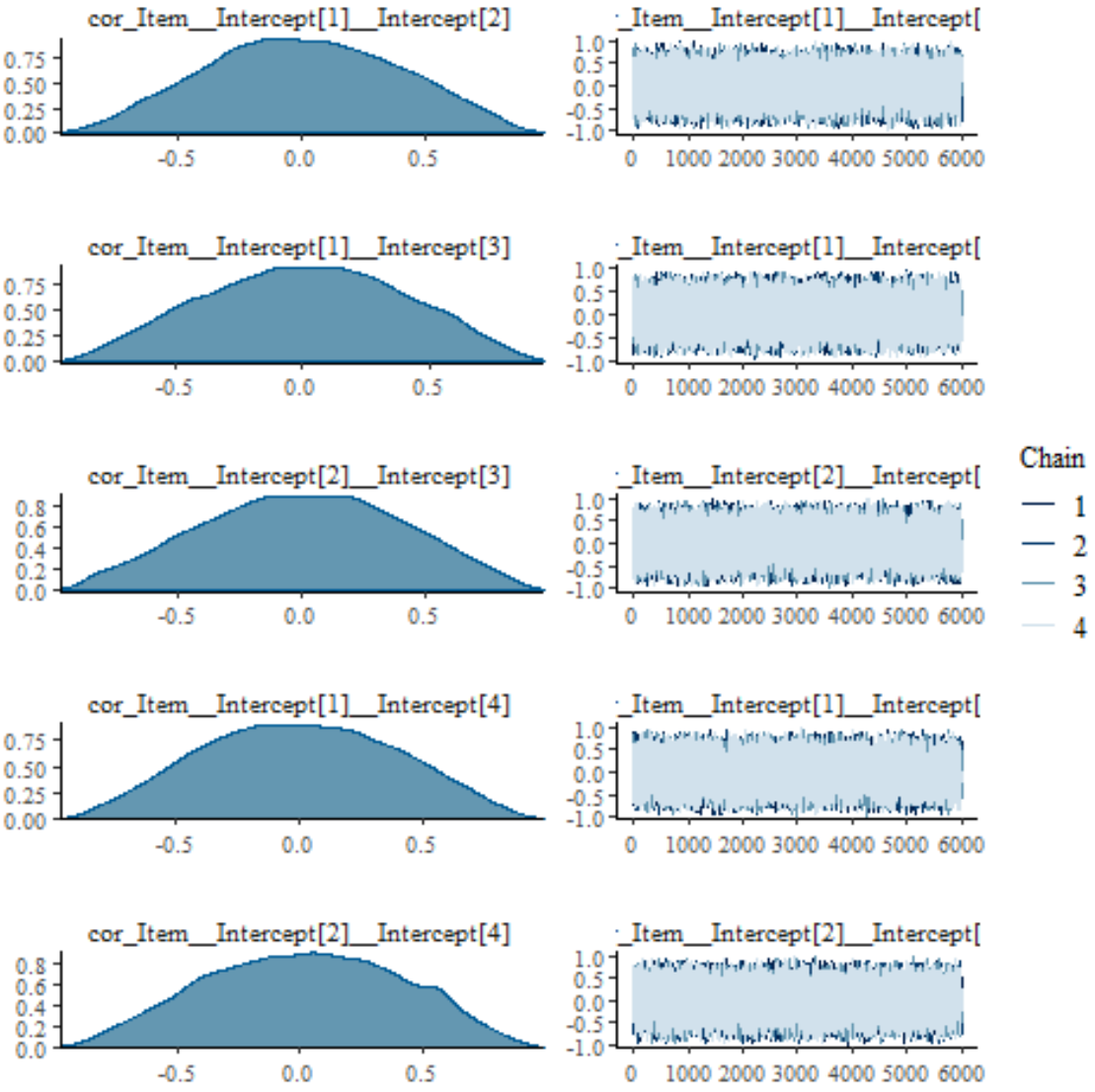


Figure C.42: Density & trace plots for the unequal variance between items model.

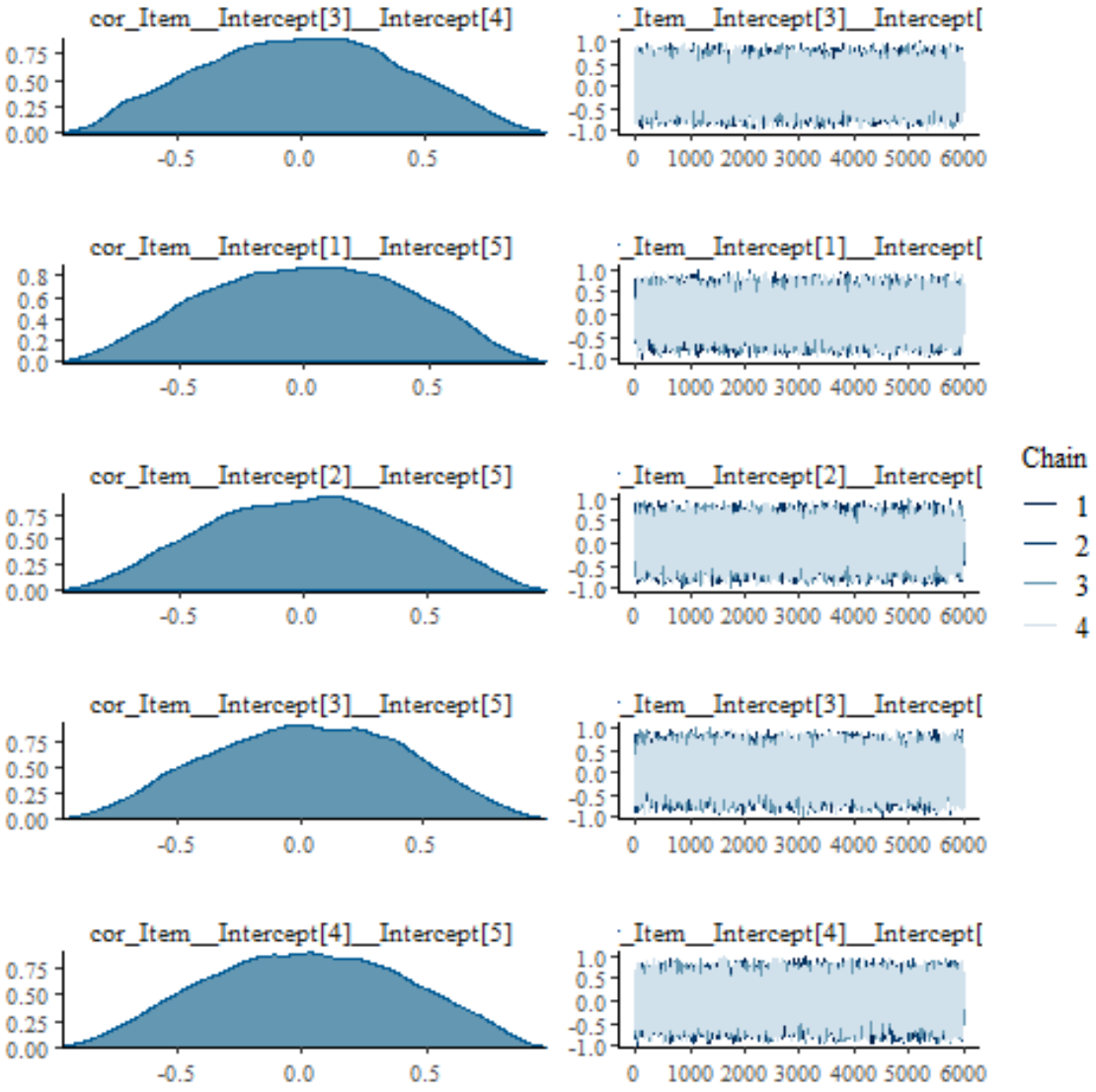


Figure C.43: Density & trace plots for the unequal variance between items model.

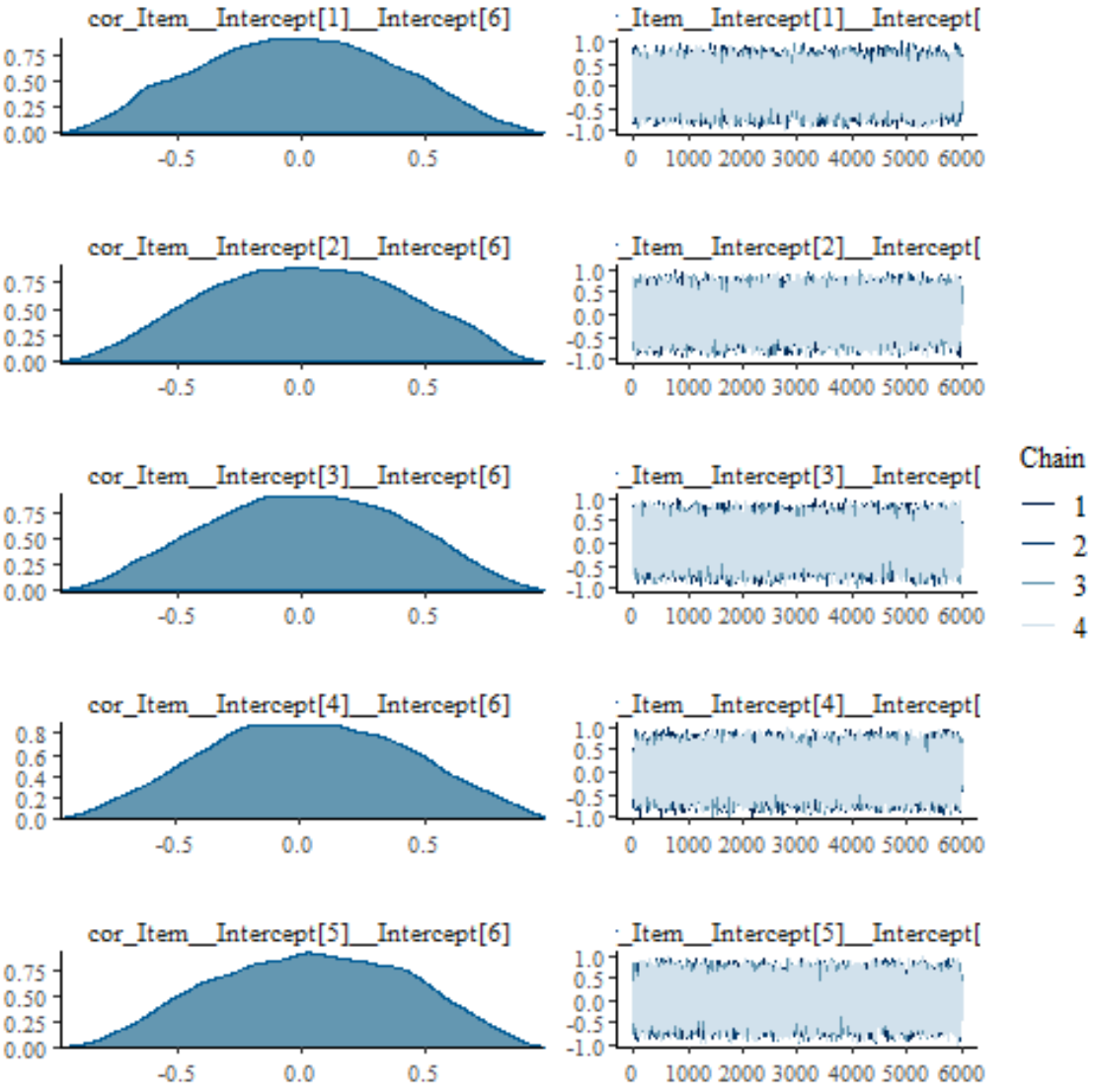


Figure C.44: Density & trace plots for the unequal variance between items model.

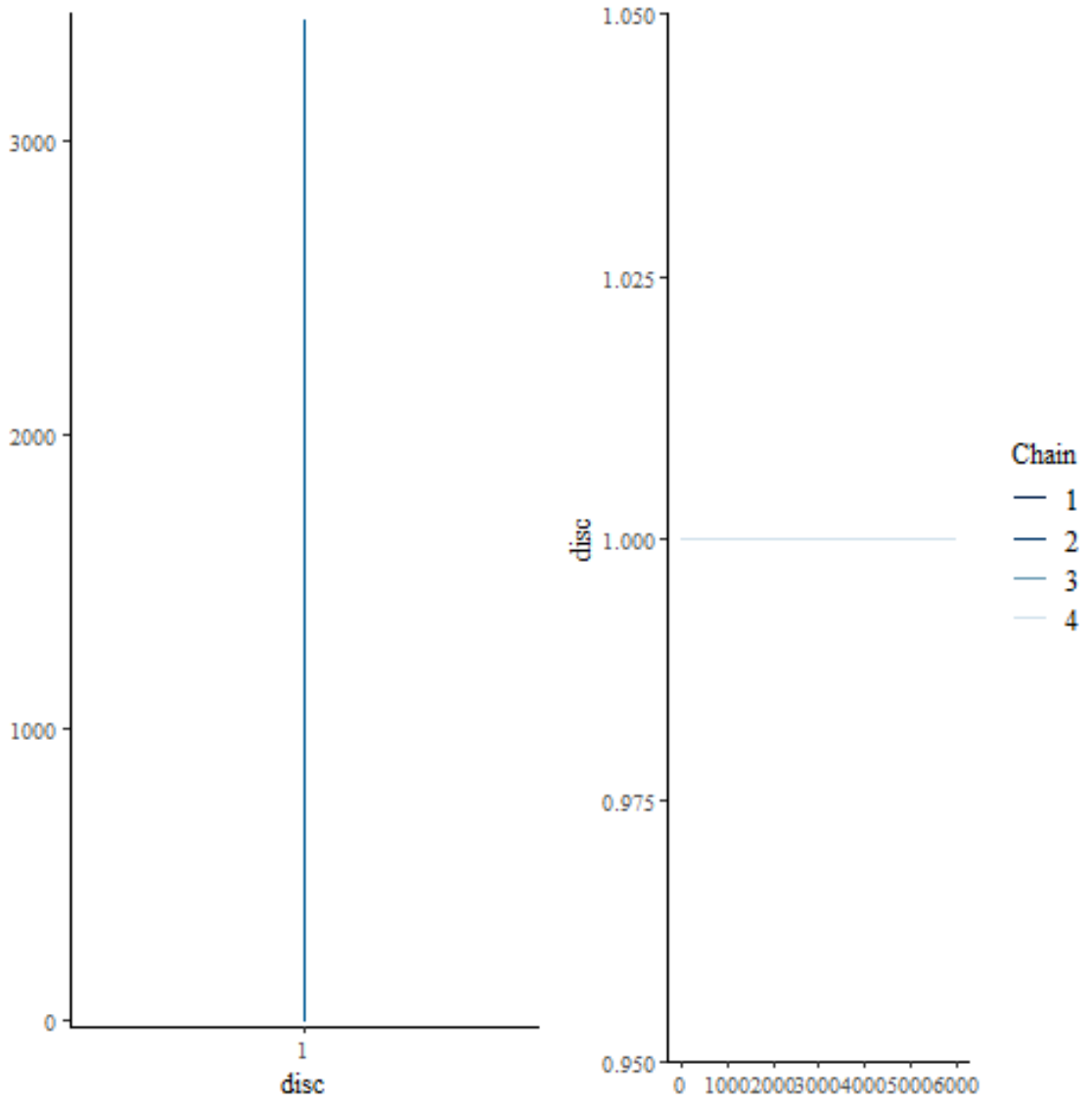


Figure C.45: Density & trace plots for the unequal variance between items model.

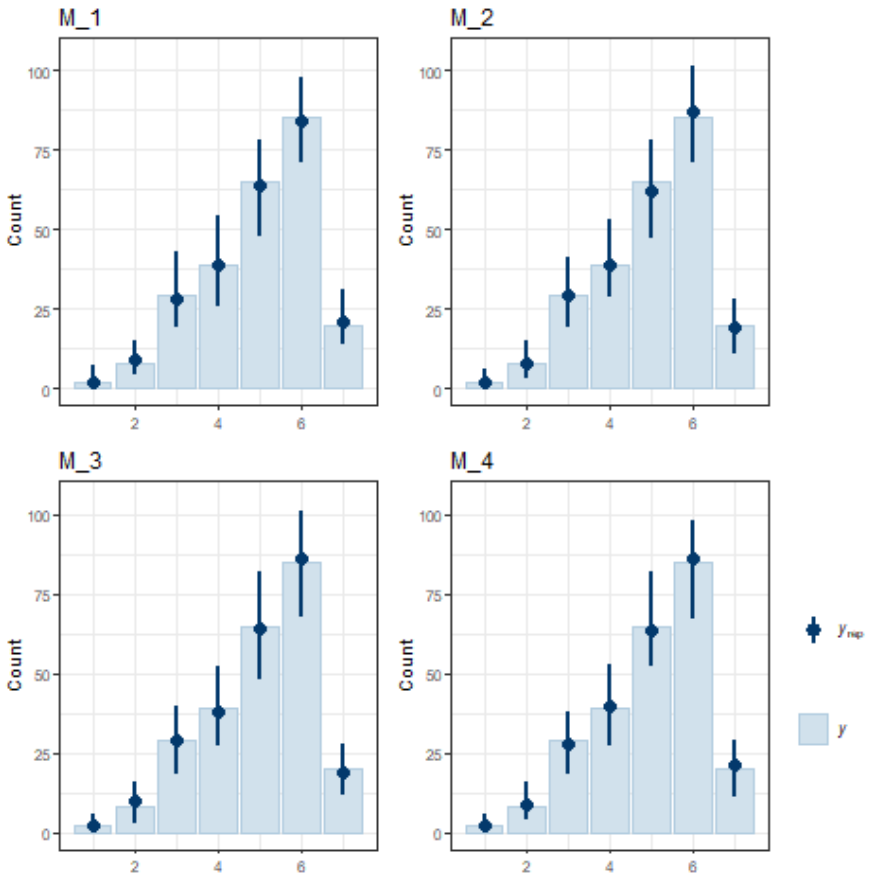


Figure C.46: Predictive posterior checks of the SoA OLR models.

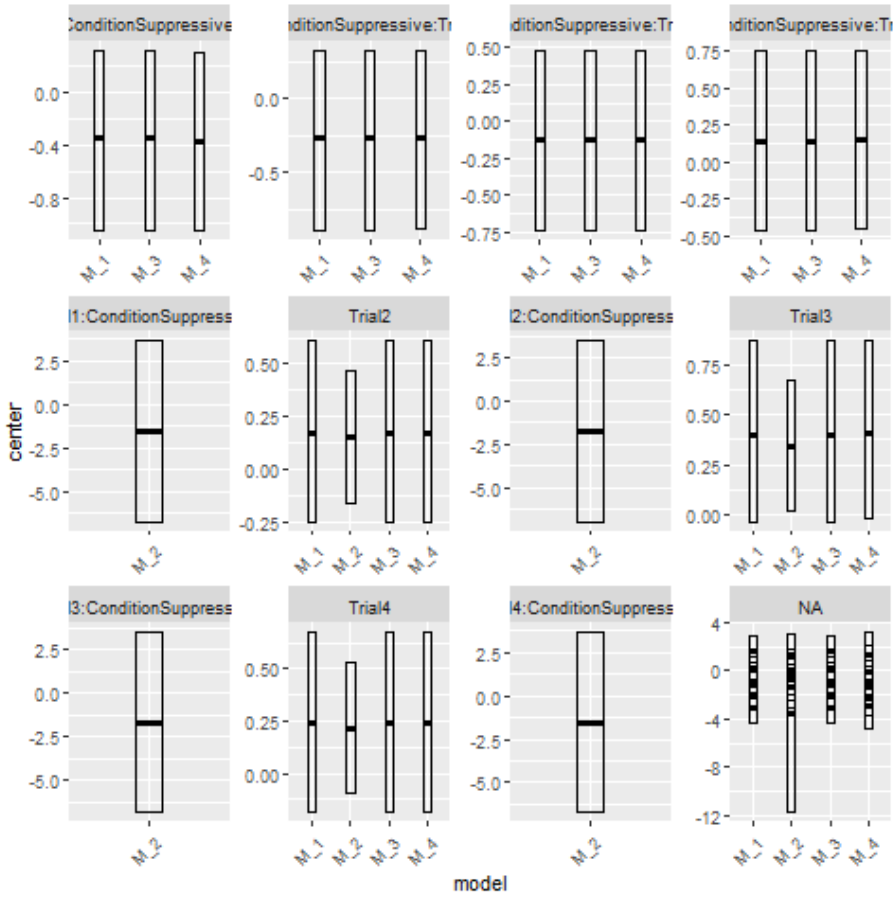


Figure C.47: Center & 95 % HDI bar plots of the SoA OLR models.

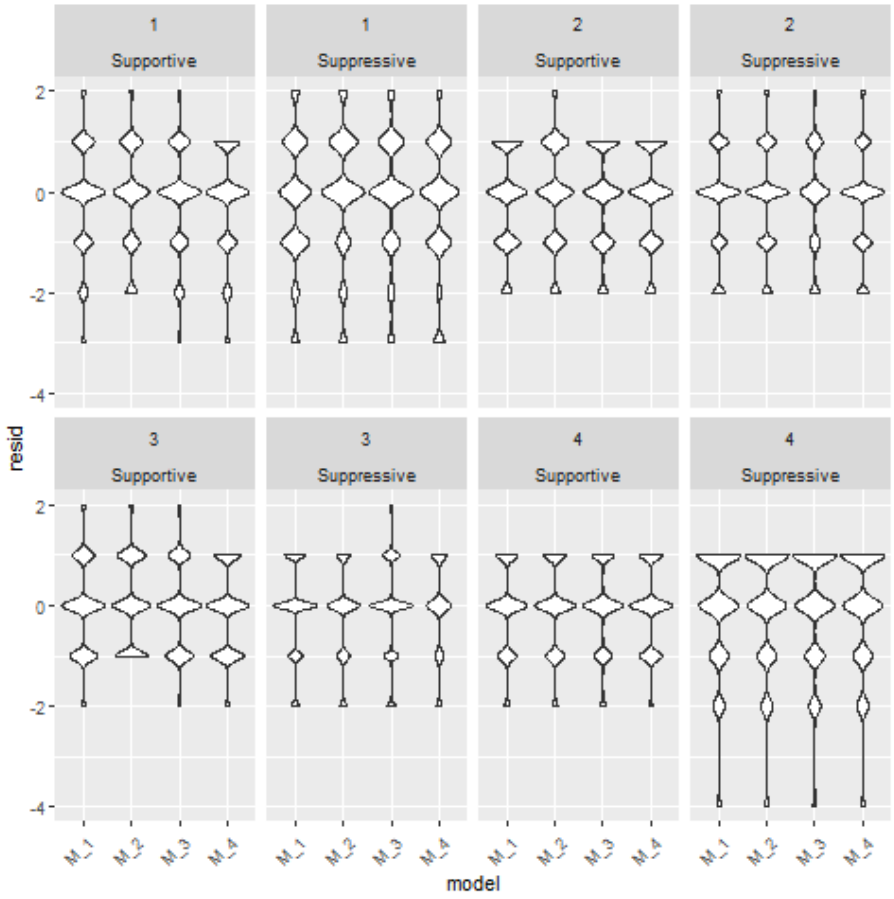


Figure C.48: Violin plots of the residuals of the SoA OLR models.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept[1]	1.0	2027	-3.3	2.9	-10.8	-2.7	0.4
b_Intercept[2]	1.0	1534	-1.9	0.7	-3.5	-1.9	-0.6
b_Intercept[3]	1.0	905	-0.5	0.5	-1.8	-0.5	0.7
b_Intercept[4]	1.0	925	-0.1	0.5	-1.3	-0.1	1.0
b_Intercept[5]	1.0	792	-0.0	0.5	-1.3	-0.0	1.1
b_Intercept[6]	1.0	902	1.0	0.5	-0.3	1.0	2.1
b_Trial2	1.0	6316	0.1	0.1	-0.2	0.1	0.4
b_Trial3	1.0	5752	0.3	0.2	0.0	0.3	0.6
b_Trial4	1.0	6684	0.6	0.2	0.2	0.6	0.9
b_Trial1:ConditionSuppressive	1.0	860	-1.8	2.4	-6.7	-1.8	2.8
b_Trial2:ConditionSuppressive	1.0	847	-1.7	2.4	-6.6	-1.7	2.9
b_Trial3:ConditionSuppressive	1.0	843	-1.7	2.4	-6.7	-1.7	2.9
b_Trial4:ConditionSuppressive	1.0	857	-2.0	2.4	-6.9	-1.9	2.7
bcs_ConditionSuppressive[1]	1.0	1352	0.2	3.6	-7.8	0.6	6.1
bcs_ConditionSuppressive[2]	1.0	889	0.3	2.5	-4.4	0.3	5.3
bcs_ConditionSuppressive[3]	1.0	861	0.9	2.4	-3.8	0.9	5.8
bcs_ConditionSuppressive[4]	1.0	856	1.3	2.4	-3.4	1.2	6.2
bcs_ConditionSuppressive[5]	1.0	867	1.2	2.4	-3.5	1.1	6.1
bcs_ConditionSuppressive[6]	1.0	867	0.8	2.5	-3.9	0.8	5.8
sd_Item__Intercept	1.0	617	0.6	0.8	0.0	0.3	3.0
sd_Participant__Intercept	1.0	3818	0.4	0.1	0.3	0.4	0.6
disc		14000	1.0	0.0	1.0	1.0	1.0
r_Item[Q1,Intercept]	1.0	698	-0.1	0.5	-1.3	-0.0	1.1
r_Item[Q2,Intercept]	1.0	711	-0.0	0.5	-1.3	0.0	1.1

Table C.12: Parameter table for the category-adjacent model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept[1]	1.0	483	-3.4	0.5	-4.4	-3.4	-2.4
b_Intercept[2]	1.0	291	-2.0	0.4	-2.7	-2.0	-1.0
b_Intercept[3]	1.0	311	-1.0	0.4	-1.7	-1.1	-0.2
b_Intercept[4]	1.0	310	-0.7	0.4	-1.4	-0.7	0.2
b_Intercept[5]	1.0	315	-0.1	0.4	-0.8	-0.2	0.7
b_Intercept[6]	1.0	392	1.3	0.4	0.6	1.3	2.2
b_ConditionSuppressive	1.0	2762	-1.2	0.3	-1.9	-1.2	-0.6
b_Trial2	1.0	2028	0.1	0.2	-0.3	0.1	0.5
b_Trial3	1.0	2001	0.4	0.2	-0.1	0.4	0.8
b_Trial4	1.0	2535	0.7	0.2	0.3	0.7	1.1
b_ConditionSuppressive:Trial2	1.0	2507	0.2	0.3	-0.4	0.1	0.8
b_ConditionSuppressive:Trial3	1.0	1970	0.3	0.3	-0.3	0.3	0.9
b_ConditionSuppressive:Trial4	1.0	1833	-0.0	0.3	-0.7	-0.0	0.6
sd_Item__Intercept	1.0	240	0.5	0.6	0.0	0.3	2.3
sd_Participant__Intercept	1.0	1294	0.7	0.1	0.5	0.7	0.9
disc		24000	1.0	0.0	1.0	1.0	1.0
r_Item[Q1,Intercept]	1.0	280	0.0	0.3	-0.5	-0.0	0.7
r_Item[Q2,Intercept]	1.0	281	0.1	0.3	-0.4	0.1	0.8

Table C.13: Parameter table for the unequal variances model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept[1]	1.0	1027	-3.1	1.6	-5.9	-3.3	0.7
b_Intercept[2]	1.0	164	-1.9	1.1	-3.7	-2.0	0.8
b_Intercept[3]	1.0	1901	-1.0	0.7	-2.4	-1.0	0.7
b_Intercept[4]	1.0	1447	-0.7	0.7	-2.1	-0.7	0.9
b_Intercept[5]	1.0	1911	-0.2	0.7	-1.9	-0.2	1.4
b_Intercept[6]	1.0	2455	1.1	1.2	-1.5	1.2	3.4
b_ConditionSuppressive	1.0	2032	-1.2	0.3	-1.8	-1.2	-0.5
b_Trial2	1.0	2857	0.1	0.2	-0.3	0.1	0.5
b_Trial3	1.0	783	0.4	0.2	0.0	0.4	0.9
b_Trial4	1.0	3028	0.7	0.2	0.3	0.7	1.2
b_ConditionSuppressive:Trial2	1.0	2380	0.1	0.3	-0.5	0.1	0.8
b_ConditionSuppressive:Trial3	1.0	3200	0.2	0.3	-0.4	0.2	0.8
b_ConditionSuppressive:Trial4	1.0	2699	-0.1	0.3	-0.7	-0.1	0.5
sd_Item__Intercept[1]	1.0	1296	2.1	2.0	0.1	1.6	6.9
sd_Item__Intercept[2]	1.0	505	1.2	1.3	0.0	0.8	4.7
sd_Item__Intercept[3]	1.0	2601	0.9	1.0	0.0	0.5	3.7
sd_Item__Intercept[4]	1.0	1321	0.9	1.0	0.0	0.5	3.7
sd_Item__Intercept[5]	1.0	399	1.0	1.3	0.0	0.6	4.2
sd_Item__Intercept[6]	1.0	1933	1.9	1.3	0.4	1.5	5.4
sd_Participant__Intercept	1.0	2436	0.7	0.1	0.5	0.7	1.0
cor_Item__Intercept[1]__Intercept[2]	1.0	1759	-0.0	0.4	-0.7	-0.1	0.7
cor_Item__Intercept[1]__Intercept[3]	1.0	3509	-0.0	0.4	-0.7	-0.0	0.7
cor_Item__Intercept[2]__Intercept[3]	1.0	3452	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[1]__Intercept[4]	1.0	3127	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[2]__Intercept[4]	1.0	4065	-0.0	0.4	-0.7	-0.0	0.7
cor_Item__Intercept[3]__Intercept[4]	1.0	781	0.0	0.4	-0.7	-0.0	0.8
cor_Item__Intercept[1]__Intercept[5]	1.0	933	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[2]__Intercept[5]	1.0	1052	-0.0	0.4	-0.7	-0.0	0.7
cor_Item__Intercept[3]__Intercept[5]	1.0	2351	-0.0	0.4	-0.7	-0.0	0.7
cor_Item__Intercept[4]__Intercept[5]	1.0	1851	-0.0	0.4	-0.7	-0.0	0.7
cor_Item__Intercept[1]__Intercept[6]	1.0	2761	0.1	0.4	-0.7	0.1	0.7
cor_Item__Intercept[2]__Intercept[6]	1.0	1046	-0.0	0.4	-0.7	-0.1	0.7
cor_Item__Intercept[3]__Intercept[6]	1.0	1686	-0.0	0.4	-0.7	-0.0	0.7
cor_Item__Intercept[4]__Intercept[6]	1.0	2589	0.0	0.4	-0.7	0.0	0.7
cor_Item__Intercept[5]__Intercept[6]	1.0	4685	0.0	0.4	-0.7	0.0	0.7
disc		24000	1.0	0.0	1.0	1.0	1.0
r_Item[Q1,Intercept[1]]	1.0	1114	0.2	1.5	-2.6	-0.0	3.8
r_Item[Q2,Intercept[1]]	1.0	1138	1.2	1.7	-1.2	0.8	5.5
r_Item[Q1,Intercept[2]]	1.0	155	0.3	1.1	-1.4	0.1	3.0
r_Item[Q2,Intercept[2]]	1.0	155	0.1	1.1	-1.8	-0.0	2.6
r_Item[Q1,Intercept[3]]	1.0	1911	0.1	0.7	-1.2	0.1	1.9
r_Item[Q2,Intercept[3]]	1.0	1921	-0.0	0.7	-1.4	-0.0	1.7
r_Item[Q1,Intercept[4]]	1.0	1593	-0.0	0.7	-1.5	-0.0	1.5
r_Item[Q2,Intercept[4]]	1.0	1515	0.1	0.7	-1.3	0.0	1.7
r_Item[Q1,Intercept[5]]	1.0	1868	-0.1	0.7	-1.8	-0.0	1.4
r_Item[Q2,Intercept[5]]	1.0	1922	0.1	0.7	-1.5	0.1	1.7
r_Item[Q1,Intercept[6]]	1.0	2570	-0.9	1.2	-3.6	-0.8	1.4
r_Item[Q2,Intercept[6]]	1.0	2531	0.2	1.2	-2.4	0.3	2.6

Table C.14: Parameter table for the unequal item variances model.

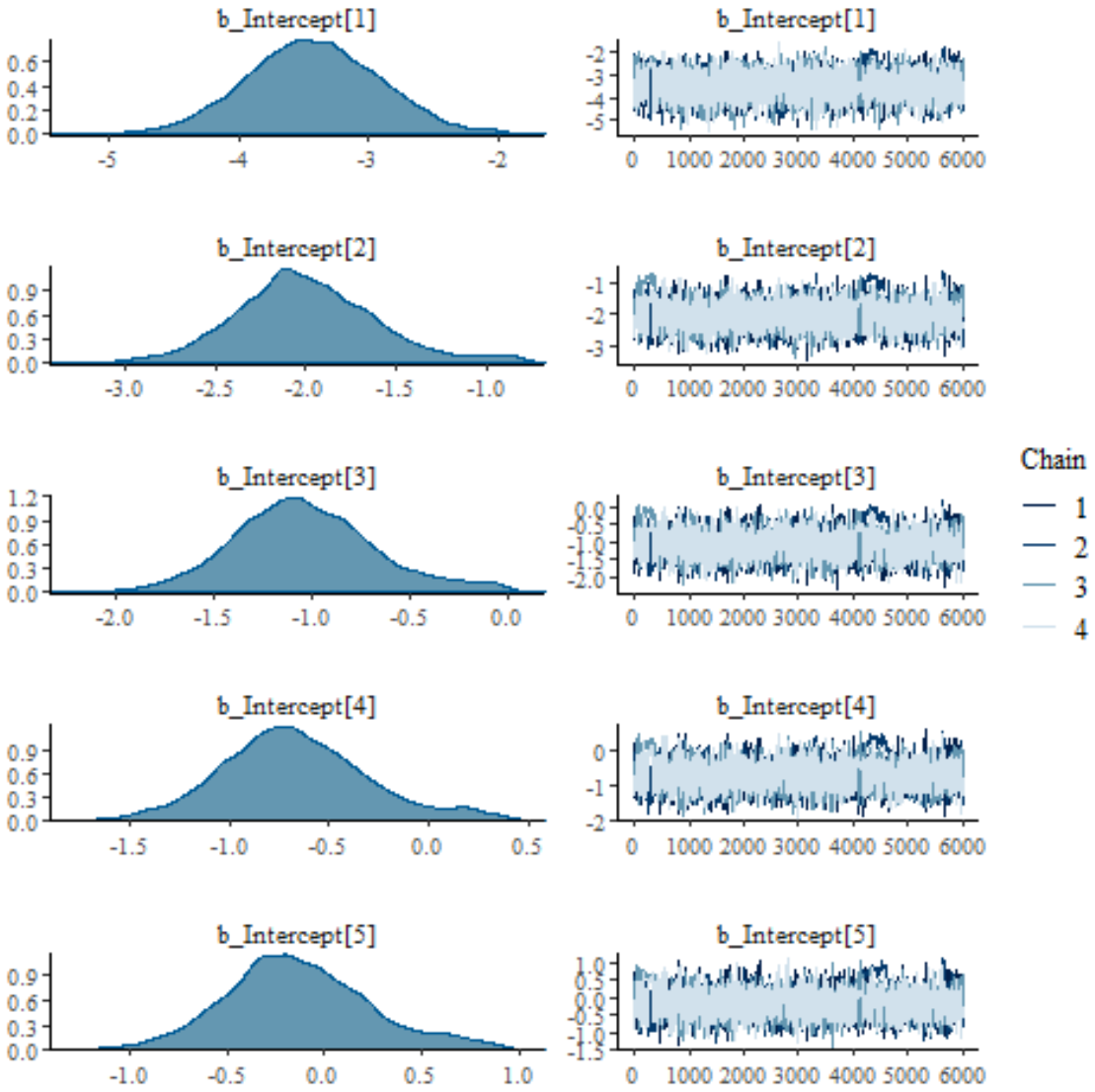


Figure C.49: Density & trace plots for the cumulative model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	5198	0.6	0.1	0.4	0.6	0.8
b_ConditionSuppressive	1.0	4852	0.0	0.2	-0.3	-0.0	0.3
b_TrialTrial_2	1.0	5763	-0.2	0.2	-0.5	-0.2	0.1
b_ConditionSuppressive:TrialTrial_2	1.0	4992	0.4	0.2	-0.1	0.3	0.8
sd_Participant__Intercept	1.0	2082	0.2	0.1	0.0	0.2	0.4
sigma	1.0	8000	1.6	0.1	1.5	1.6	1.8

Table C.15: Parameter table of the Gaussian PPD model.

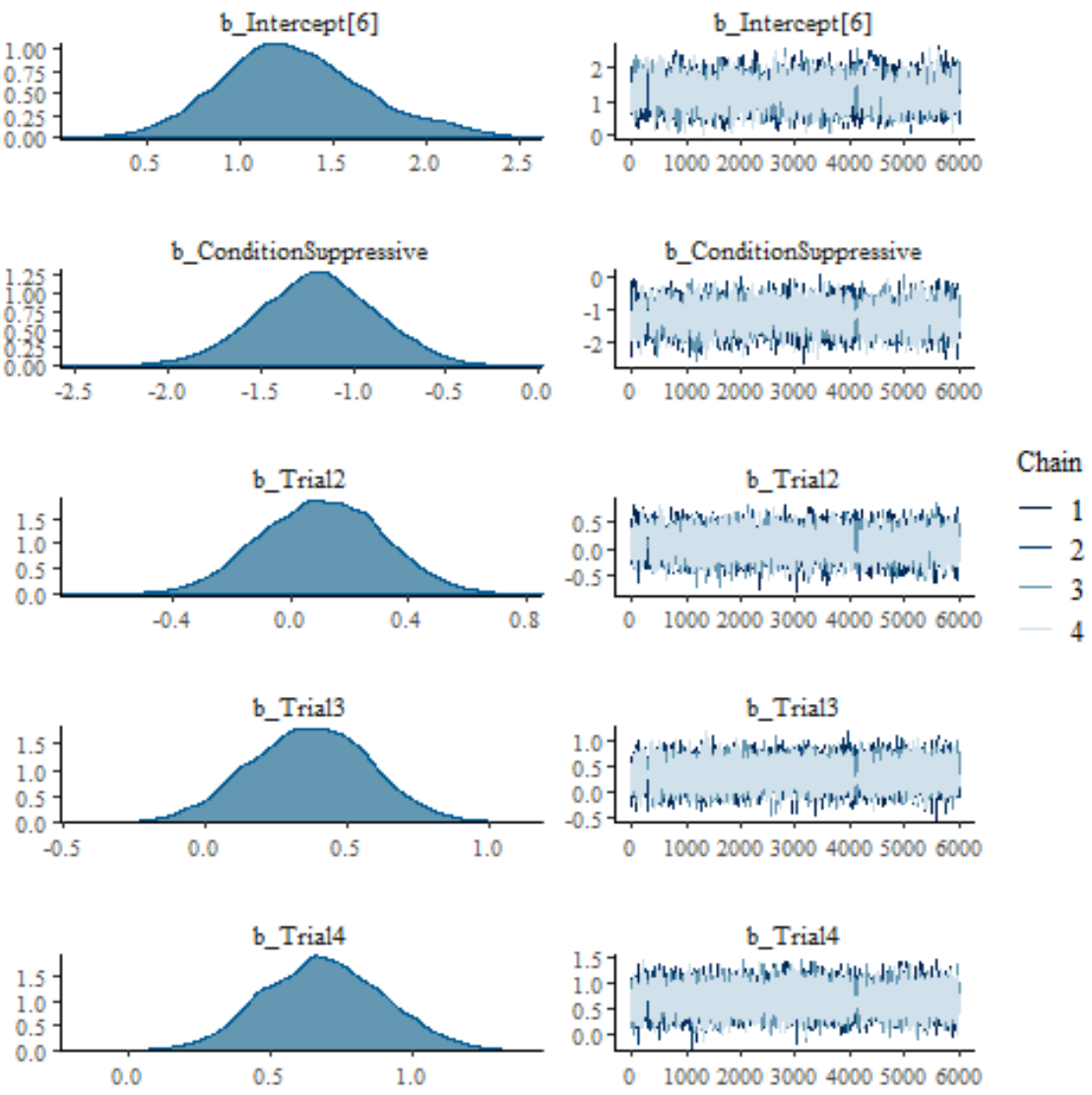


Figure C.50: Density & trace plots for the cumulative model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	6193	0.6	0.1	0.4	0.6	0.8
b_ConditionSuppressive	1.0	5486	0.0	0.2	-0.3	0.0	0.3
b_TrialTrial_2	1.0	6793	-0.2	0.1	-0.5	-0.2	0.1
b_ConditionSuppressive:TrialTrial_2	1.0	5818	0.3	0.2	-0.1	0.3	0.7
sd_Participant__Intercept	1.0	2338	0.2	0.1	0.0	0.2	0.3
shape	1.0	8000	1.4	0.1	1.2	1.4	1.6

Table C.16: Parameter table of the Gamma PPD model.

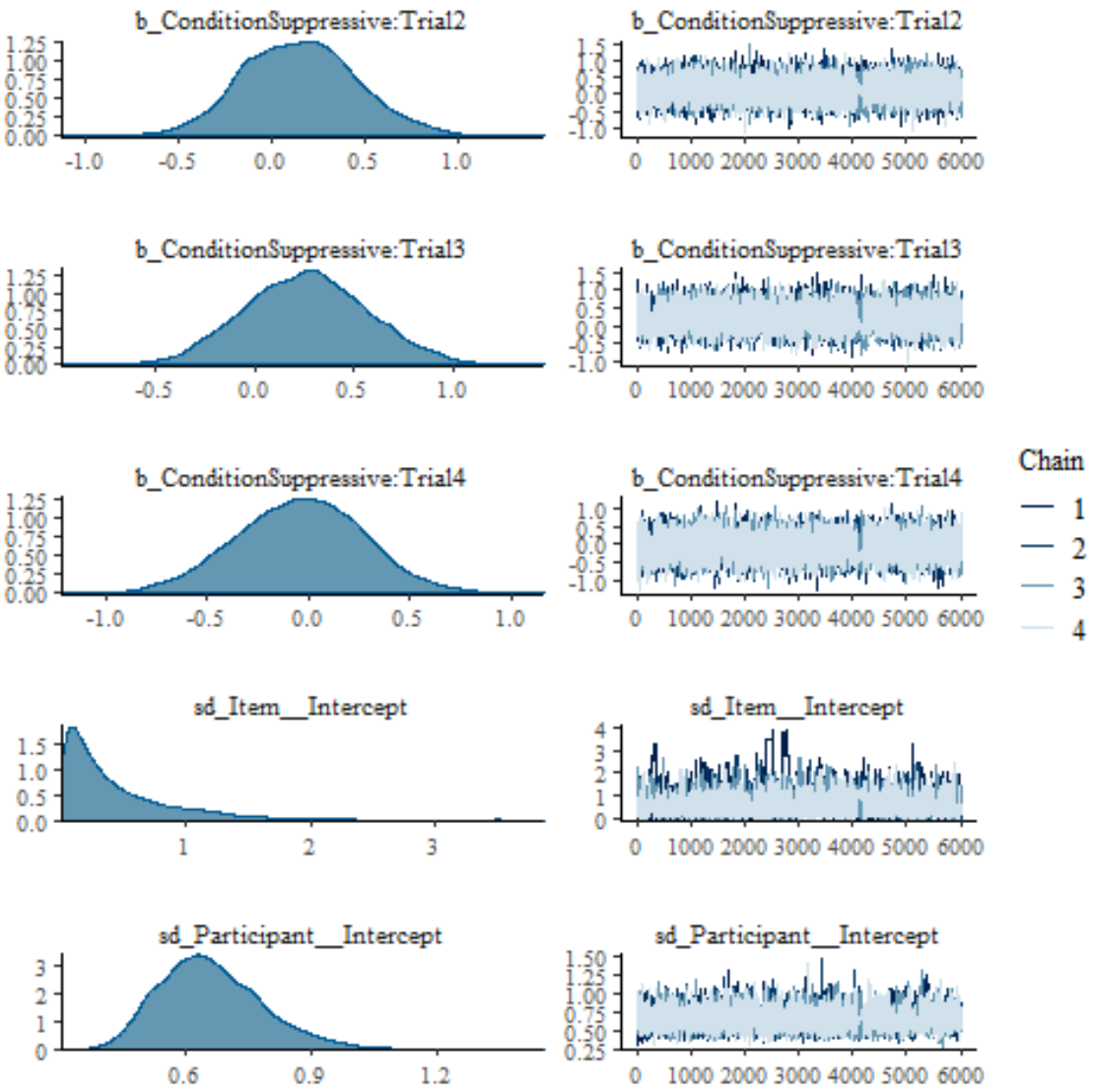


Figure C.51: Density & trace plots for the cumulative model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	4377	0.6	0.1	0.5	0.6	0.8
b_ConditionSuppressive	1.0	4109	0.0	0.1	-0.1	0.0	0.1
b_TrialTrial_2	1.0	3646	0.0	0.1	-0.1	0.0	0.1
b_ConditionSuppressive:TrialTrial_2	1.0	3887	-0.1	0.1	-0.2	-0.1	0.1
sd_Participant__Intercept	1.0	2087	0.0	0.0	0.0	0.0	0.1
sigma	1.0	2012	0.2	0.1	0.0	0.2	0.3
beta	1.0	4311	1.7	0.1	1.5	1.7	1.9

Table C.17: Parameter table of the exponentially modified Gaussian PPD model.

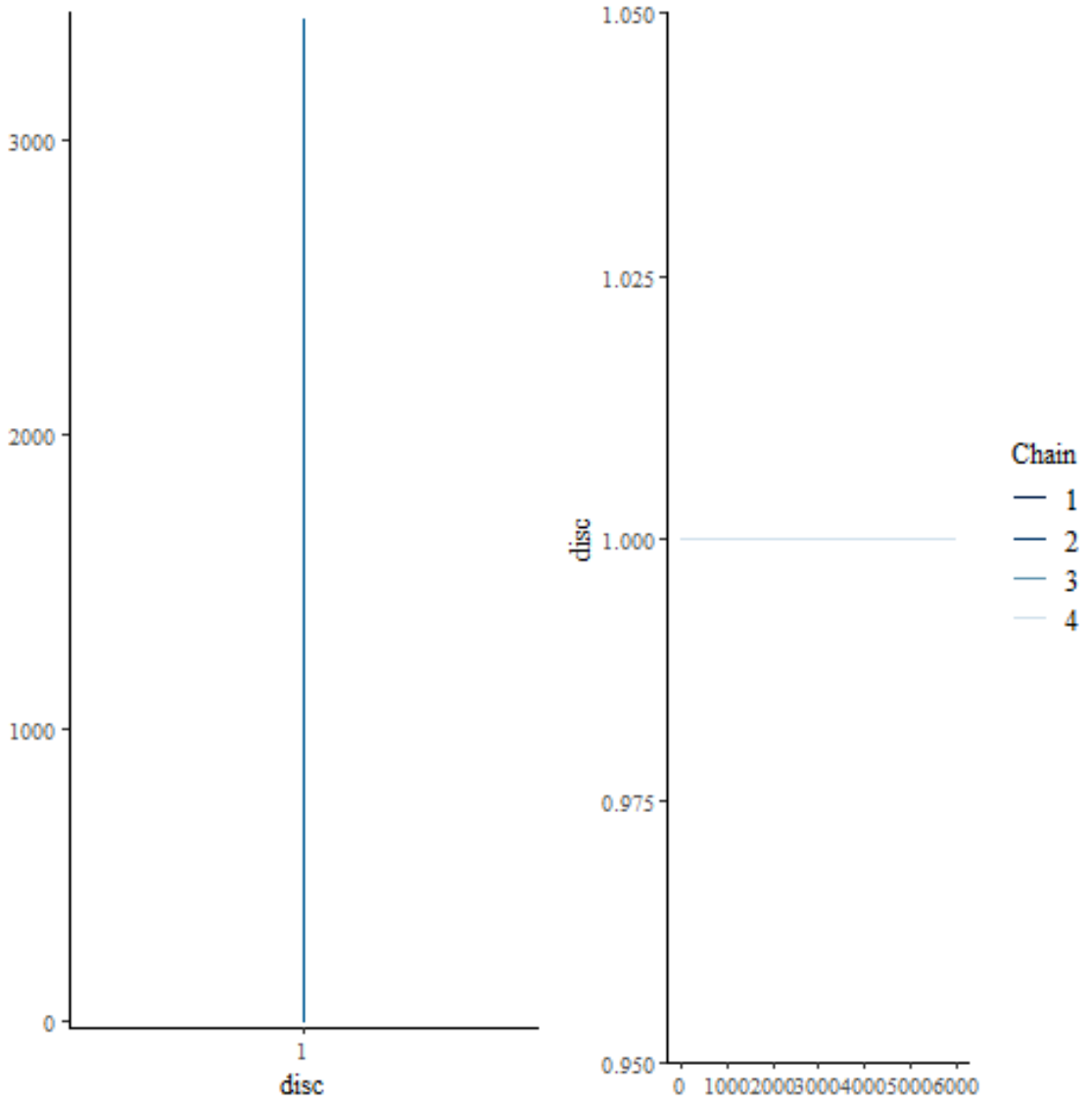


Figure C.52: Density & trace plots for the cumulative model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	6456	0.6	0.1	0.4	0.6	0.8
b_ConditionSuppressive	1.0	5759	-0.2	0.2	-0.5	-0.2	0.1
b_TrialTrial_2	1.0	8000	-0.2	0.1	-0.5	-0.2	0.0
b_ConditionSuppressive:TrialTrial_2	1.0	6426	0.3	0.2	-0.1	0.3	0.7
sd_Participant__Intercept	1.0	2546	0.2	0.1	0.1	0.2	0.4
sigma	1.0	8000	1.4	0.1	1.2	1.4	1.5

Table C.18: Parameter table of the Gaussian PPD model.

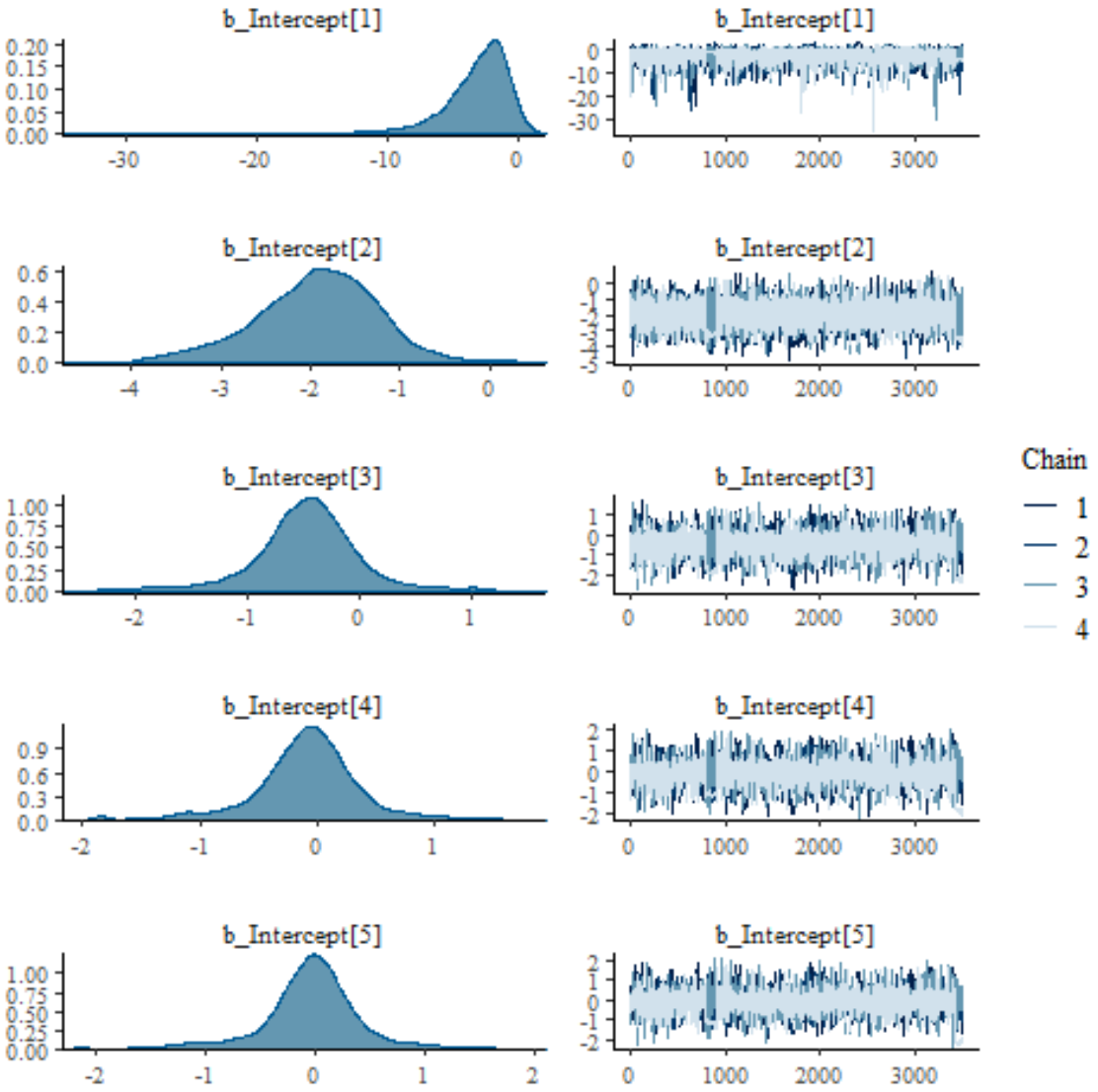


Figure C.53: Density & trace plots for the categorical model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	7413	0.6	0.1	0.4	0.6	0.8
b_ConditionSuppressive	1.0	7282	-0.2	0.2	-0.5	-0.2	0.1
b_TrialTrial_2	1.0	7192	-0.2	0.1	-0.5	-0.2	0.1
b_ConditionSuppressive:TrialTrial_2	1.0	6976	0.3	0.2	-0.1	0.3	0.7
sd_Participant__Intercept	1.0	2775	0.1	0.1	0.0	0.1	0.3
shape	1.0	8000	1.3	0.1	1.1	1.3	1.5

Table C.19: Parameter table of the Gamma PPD model.

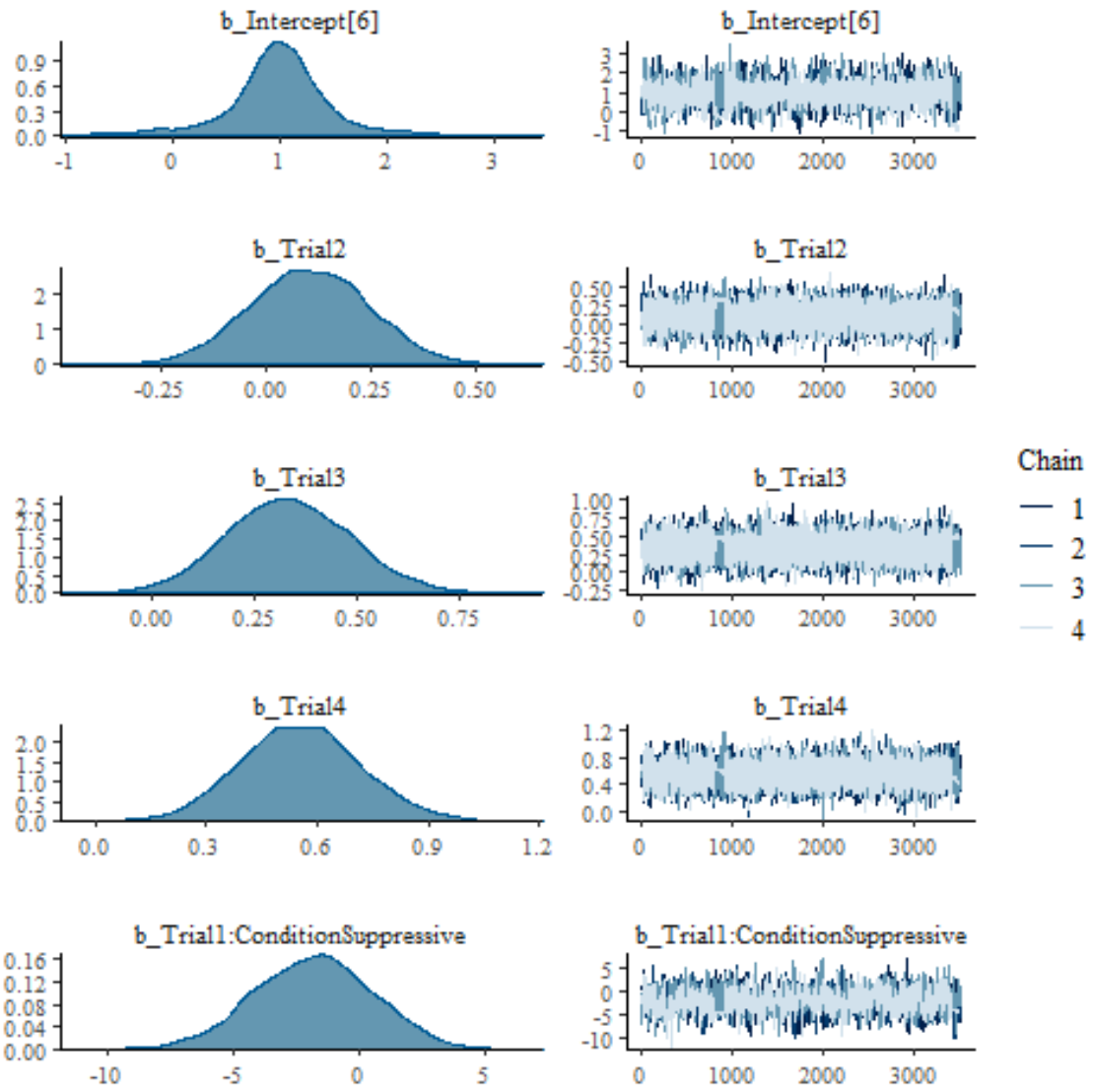


Figure C.54: Density & trace plots for the categorical model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	2764	0.5	0.1	0.4	0.5	0.7
b_ConditionSuppressive	1.0	1887	-0.1	0.1	-0.2	-0.0	0.1
b_TrialTrial_2	1.0	1649	-0.1	0.1	-0.2	-0.1	0.0
b_ConditionSuppressive:TrialTrial_2	1.0	1223	0.1	0.1	-0.0	0.1	0.3
sd_Participant__Intercept	1.0	867	0.1	0.0	0.0	0.1	0.2
sigma	1.0	1087	0.1	0.1	0.0	0.1	0.2
beta	1.0	3073	1.5	0.1	1.3	1.5	1.7

Table C.20: Parameter table of the exponentially modified Gaussian PPD model.

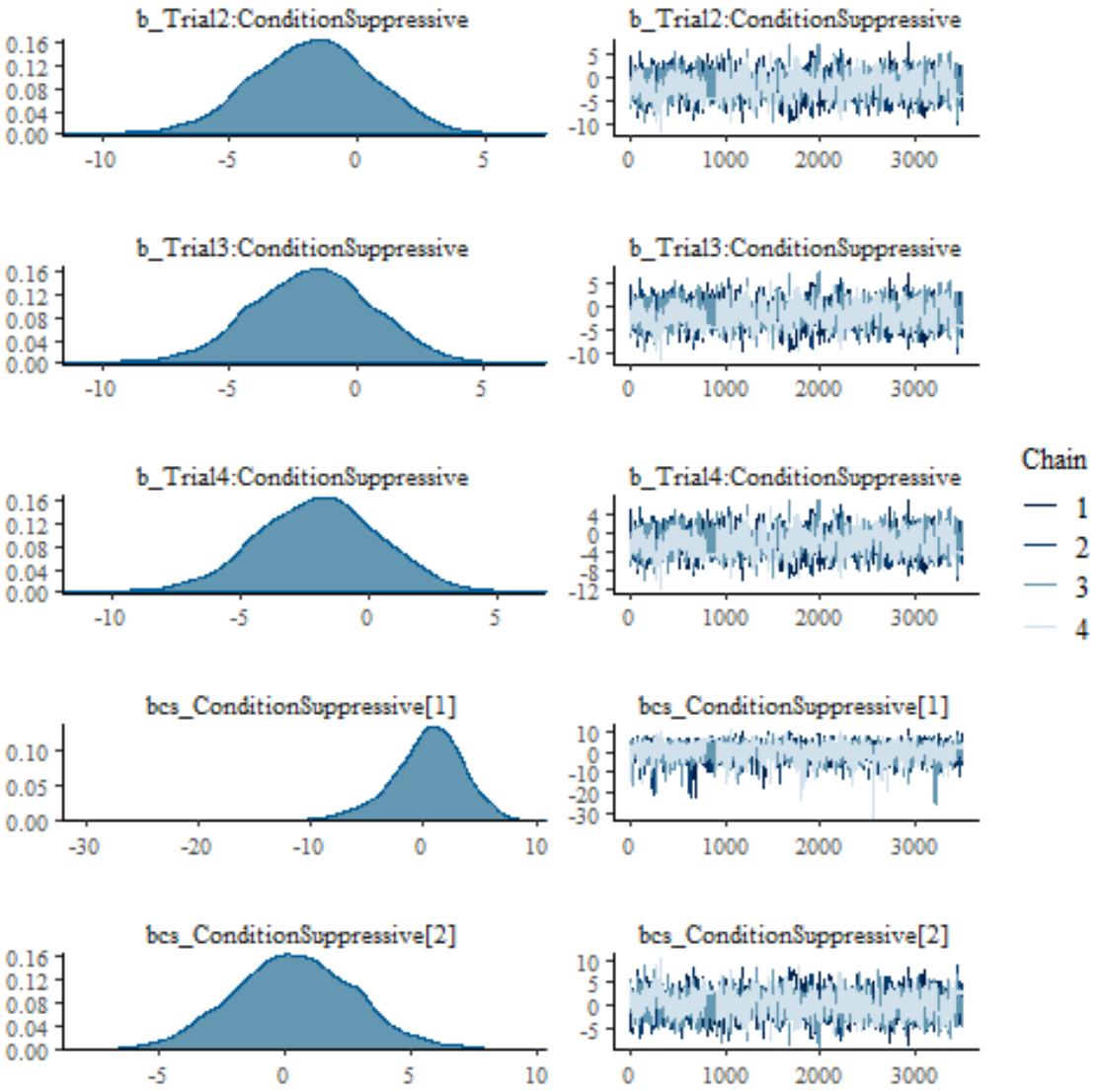


Figure C.55: Density & trace plots for the categorical model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	4440	0.8	0.1	0.5	0.8	1.1
b_ConditionSuppressive	1.0	4411	0.2	0.2	-0.1	0.2	0.5
b_TrialTrial_2	1.0	4980	0.1	0.2	-0.2	0.1	0.4
b_ConditionSuppressive:TrialTrial_2	1.0	4774	0.1	0.2	-0.3	0.1	0.5
sd_Participant__Intercept	1.0	1788	0.2	0.1	0.0	0.2	0.4
sigma	1.0	8000	2.4	0.1	2.2	2.4	2.6

Table C.21: Parameter table of the Gaussian PPD model.

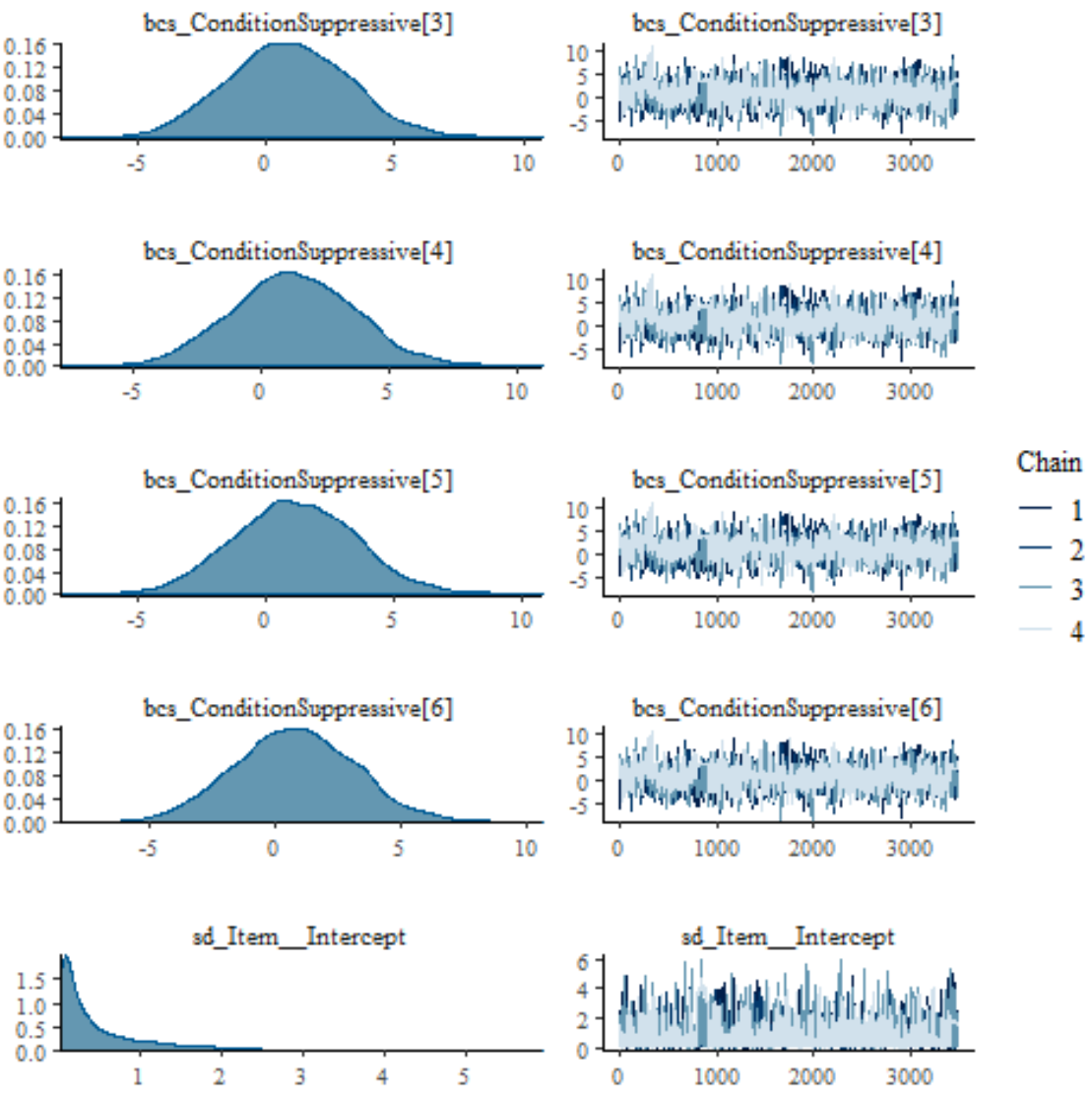


Figure C.56: Density & trace plots for the categorical model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	6345	0.9	0.1	0.7	0.9	1.1
b_ConditionSuppressive	1.0	6056	0.2	0.2	-0.1	0.2	0.5
b_TrialTrial_2	1.0	6615	0.1	0.1	-0.2	0.1	0.4
b_ConditionSuppressive:TrialTrial_2	1.0	5640	0.0	0.2	-0.4	0.0	0.4
sd_Participant__Intercept	1.0	2423	0.1	0.1	0.0	0.1	0.3
shape	1.0	8000	1.3	0.1	1.1	1.3	1.5

Table C.22: Parameter table of the Gamma PPD model.

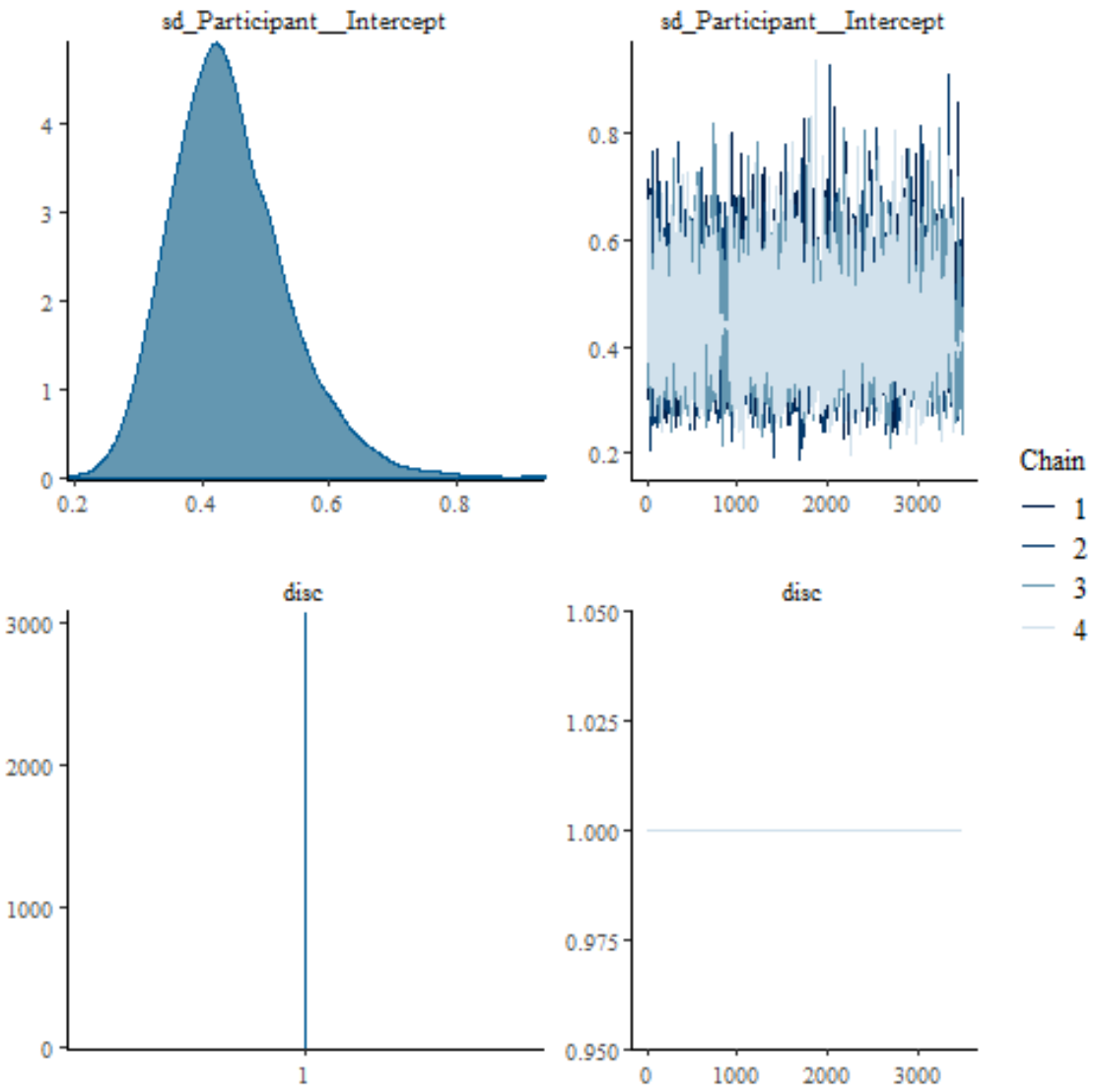


Figure C.57: Density & trace plots for the categorical model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	4134	1.0	0.1	0.9	1.0	1.2
b_ConditionSuppressive	1.0	3441	-0.0	0.0	-0.1	-0.0	0.1
b_TrialTrial_2	1.0	4102	-0.1	0.0	-0.1	-0.0	0.0
b_ConditionSuppressive:TrialTrial_2	1.0	3966	0.0	0.1	-0.1	0.0	0.2
sd_Participant__Intercept	1.0	1412	0.0	0.0	0.0	0.0	0.1
sigma	1.0	1223	0.1	0.1	0.0	0.1	0.3
beta	1.0	3994	2.6	0.2	2.3	2.6	2.9

Table C.23: Parameter table of the exponentially modified Gaussian PPD model.

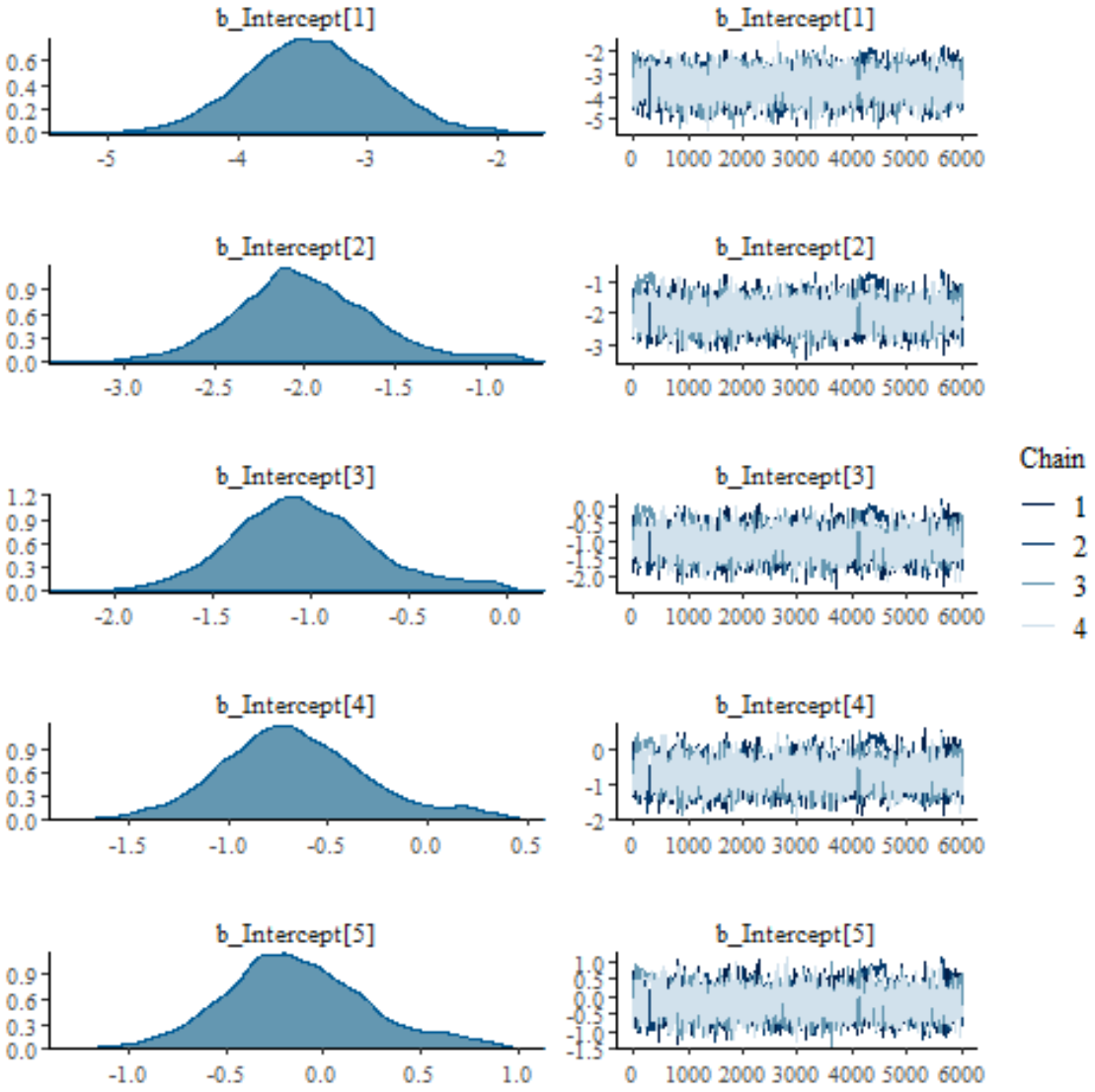


Figure C.58: Density & trace plots for the unequal variance between conditions model.

Parameter	Prior
Intercept	student-t(3, 0, 2.5)
sd	student-t(3, 0, 2.5)
sigma	student-t(3, 0, 2.5)
phi(beta-binomial)	gamma(0.01, 0.01)

Table C.24: Default priors for Poisson, binomial and beta-binomial models.

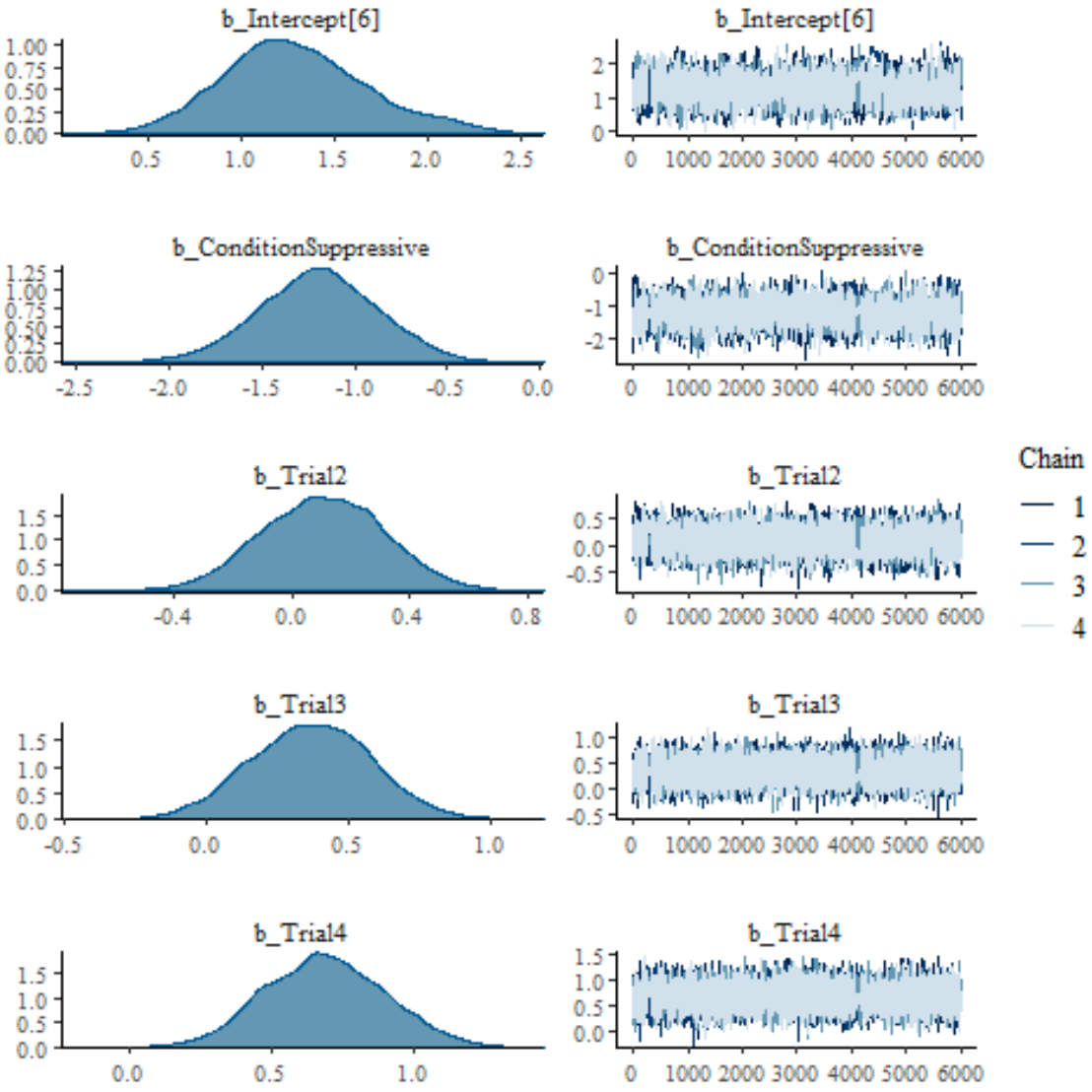


Figure C.59: Density & trace plots for the unequal variance between conditions model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	12000	1.2	0.1	1.1	1.2	1.4
b_ConditionSuppressive	1.0	12000	-0.1	0.1	-0.4	-0.1	0.1
b_Trial1	1.0	12000	-0.1	0.1	-0.3	-0.1	0.1
b_Trial2	1.0	12000	0.0	0.1	-0.2	0.0	0.3
b_Trial3	1.0	12000	-0.0	0.1	-0.2	-0.0	0.2
b_ConditionSuppressive:Trial1	1.0	12000	0.1	0.2	-0.3	0.1	0.5
b_ConditionSuppressive:Trial2	1.0	12000	-0.1	0.2	-0.5	-0.1	0.2
b_ConditionSuppressive:Trial3	1.0	12000	-0.0	0.2	-0.4	-0.0	0.3

Table C.25: Parameter table for the Poisson model.

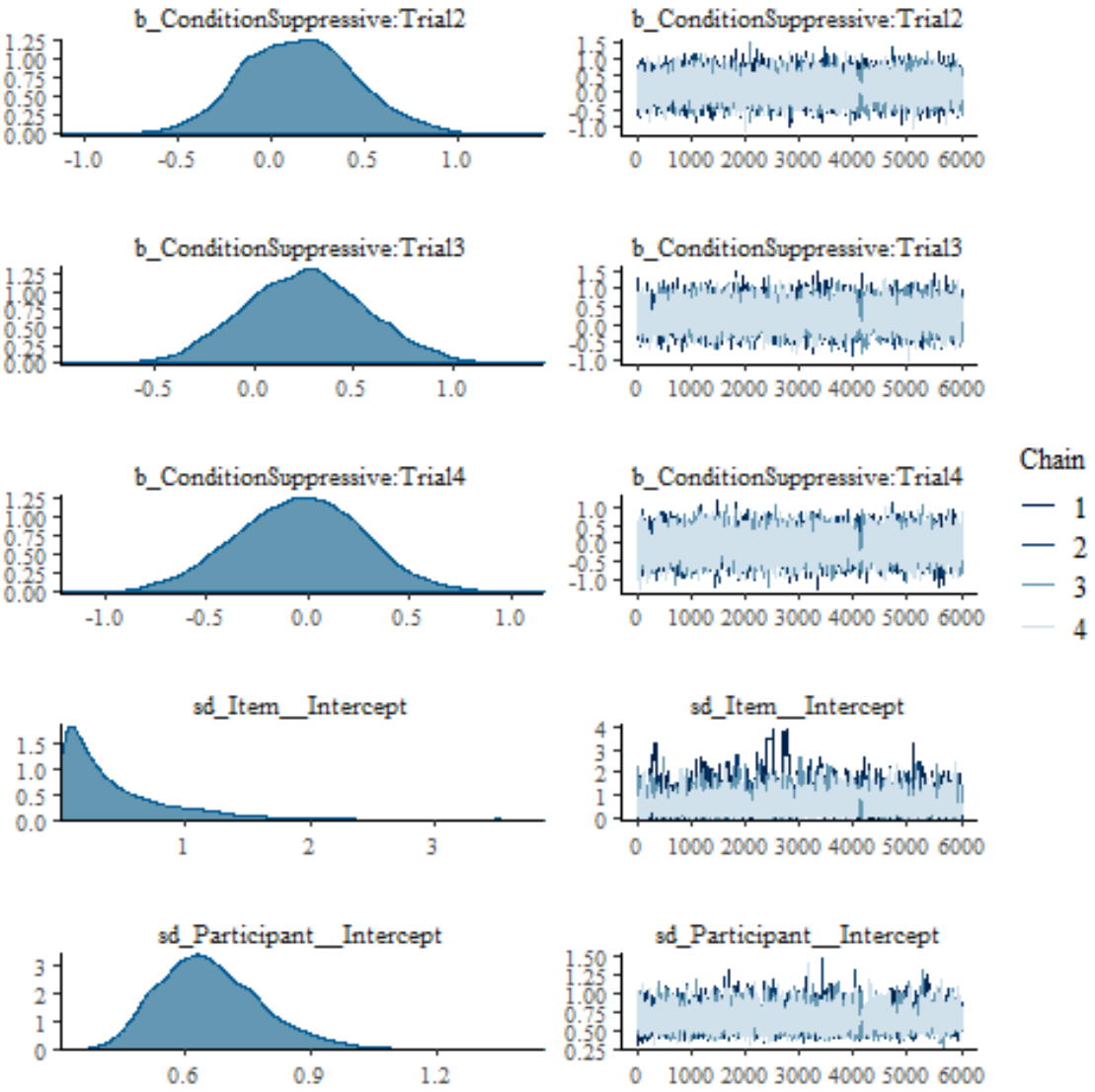


Figure C.60: Density & trace plots for the unequal variance between conditions model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	1661	1.5	0.4	0.7	1.5	2.4
b_ConditionSuppressive	1.0	1582	-0.8	0.6	-1.9	-0.8	0.3
b_TrialTrial_2	1.0	2612	0.2	0.5	-0.7	0.2	1.2
b_TrialTrial_3	1.0	2357	0.2	0.5	-0.7	0.2	1.2
b_TrialTrial_4	1.0	1903	0.6	0.5	-0.3	0.6	1.7
b_ConditionSuppressive:TrialTrial_2	1.0	2718	-0.3	0.6	-1.5	-0.3	0.9
b_ConditionSuppressive:TrialTrial_3	1.0	2680	-0.1	0.6	-1.4	-0.1	1.1
b_ConditionSuppressive:TrialTrial_4	1.0	2584	-0.2	0.7	-1.6	-0.2	1.1

Table C.26: Parameter table for the Binomial model.

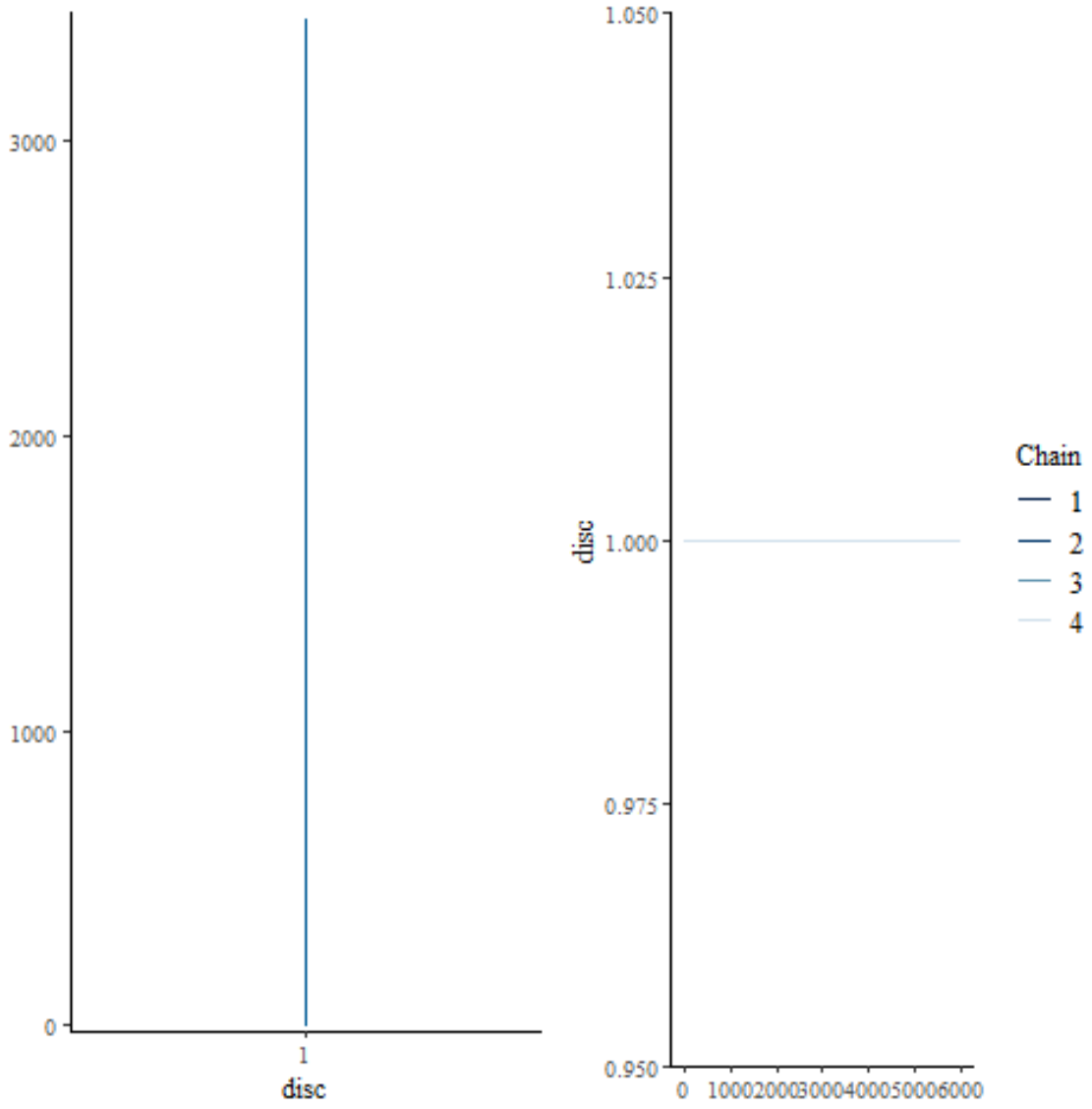


Figure C.61: Density & trace plots for the unequal variance between conditions model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	8526	1.9	0.2	1.4	1.9	2.4
b_ConditionSuppressive	1.0	9800	-0.7	0.3	-1.4	-0.7	-0.1
b_Trial1	1.0	8515	-0.5	0.3	-1.1	-0.5	0.1
b_Trial2	1.0	7759	0.0	0.4	-0.6	0.0	0.8
b_Trial3	1.0	7991	-0.1	0.3	-0.8	-0.1	0.6
b_ConditionSuppressive:Trial1	1.0	8868	0.5	0.4	-0.4	0.5	1.3
b_ConditionSuppressive:Trial2	1.0	8307	-0.5	0.5	-1.4	-0.5	0.3
b_ConditionSuppressive:Trial3	1.0	8209	-0.1	0.4	-0.9	-0.1	0.8
phi	1.0	12000	86.7	72.7	15.2	64.5	286.0

Table C.27: Parameter table for the Beta-Binomial model.

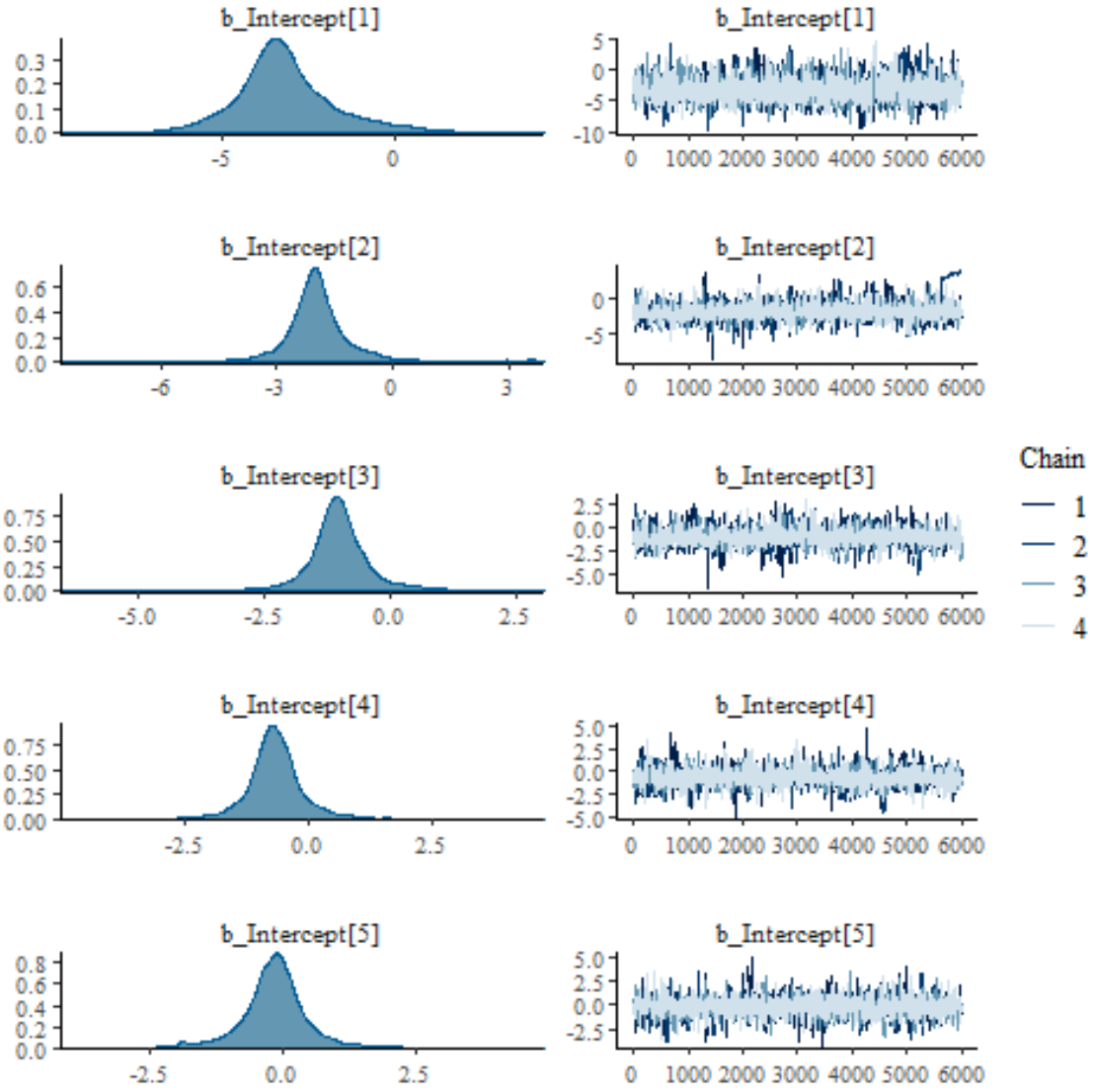


Figure C.62: Density & trace plots for the unequal variance between items model.

Parameter	**Prior**
Intercept	student-t(3, 4.4, 2.5)
sd	student-t(3, 0, 2.5)
sigma	student-t(3, 0, 44.7)
beta (Exp. Gaussian)	gamma(1, 0.1)
shape (Gamma)	gamma(0.01, 0.01)

Table C.28: Weakly-informative (default) priors for Gaussian, exp. Gaussian, and Gamma models.

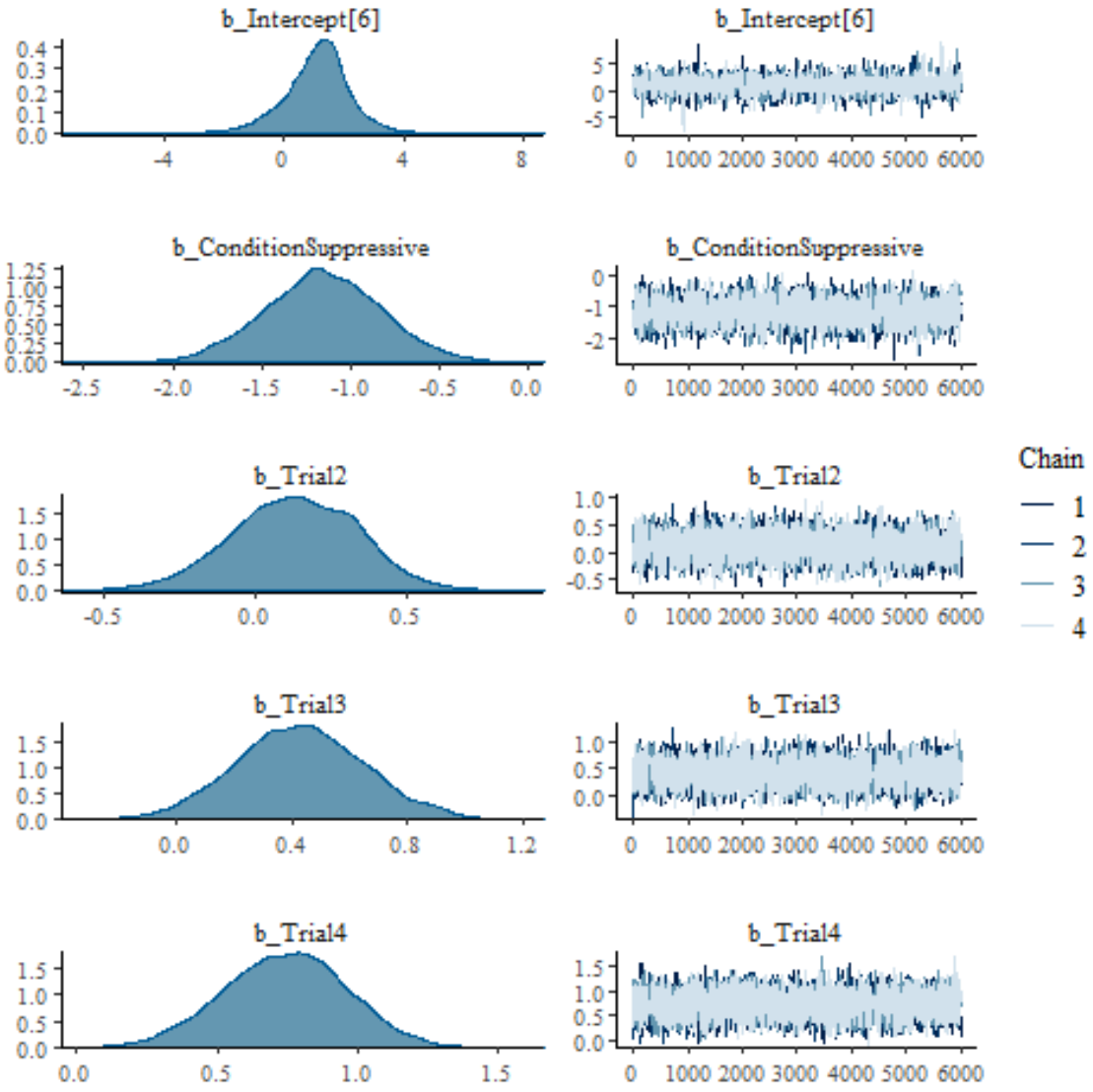


Figure C.63: Density & trace plots for the unequal variance between items model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	2790	4.5	0.1	4.3	4.5	4.6
b_ConditionSuppressive	1.0	2579	0.2	0.1	-0.1	0.2	0.4
b_TrialTrial_2	1.0	3393	0.1	0.1	-0.2	0.1	0.3
b_TrialTrial_3	1.0	3399	-0.1	0.1	-0.3	-0.1	0.1
b_TrialTrial_4	1.0	3347	-0.0	0.1	-0.2	-0.0	0.2
b_ConditionSuppressive:TrialTrial_2	1.0	3863	-0.1	0.2	-0.4	-0.1	0.2
b_ConditionSuppressive:TrialTrial_3	1.0	3028	-0.0	0.2	-0.4	-0.0	0.3
b_ConditionSuppressive:TrialTrial_4	1.0	3123	-0.1	0.2	-0.4	-0.1	0.2
sigma	1.0	8000	48.9	2.0	45.1	48.8	52.9

Table C.29: Parameter table for the Gaussian model.

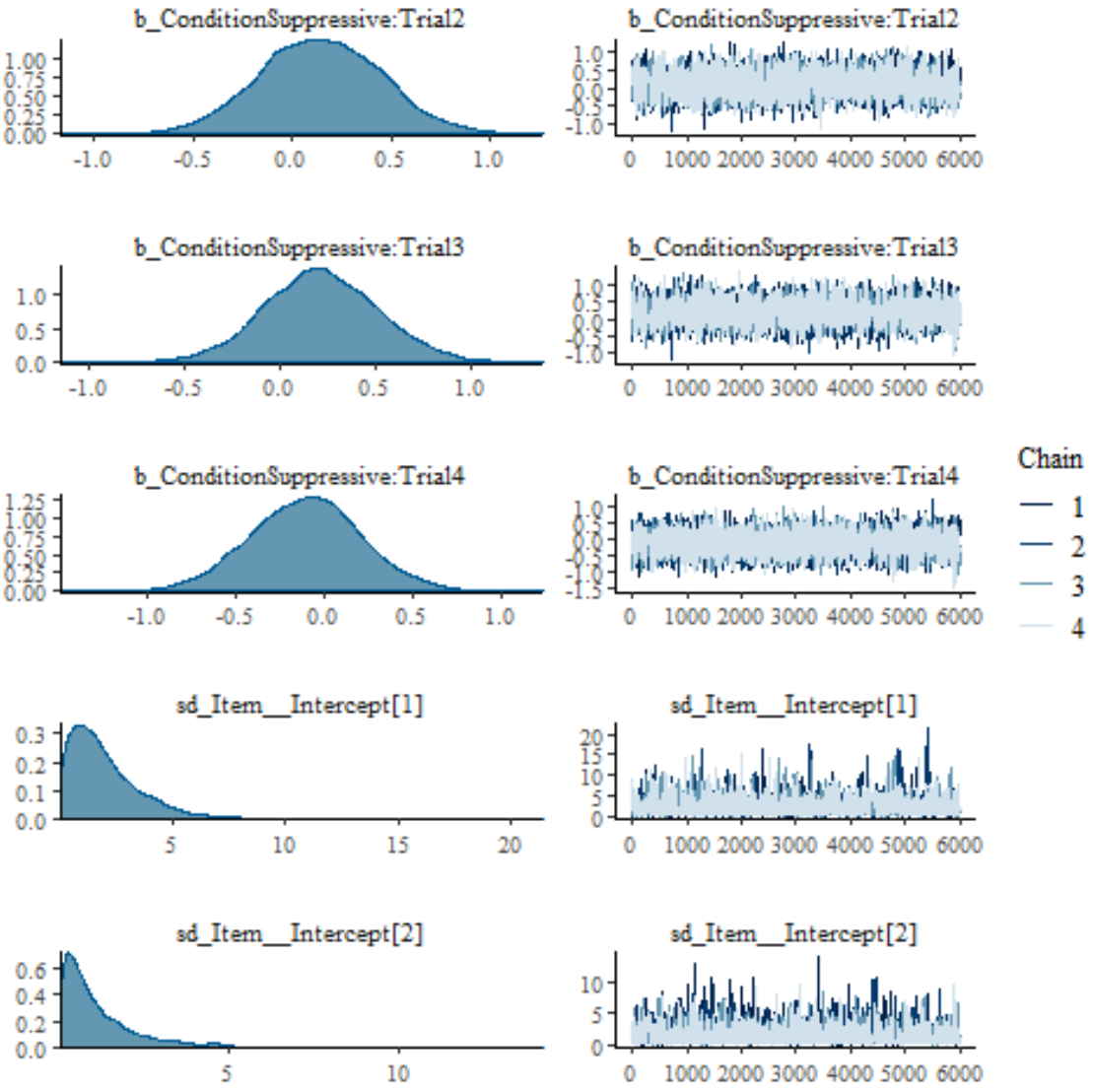


Figure C.64: Density & trace plots for the unequal variance between items model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	3883	4.5	0.1	4.3	4.5	4.6
b_ConditionSuppressive	1.0	3541	0.2	0.1	-0.1	0.2	0.4
b_TrialTrial_2	1.0	5123	0.1	0.1	-0.1	0.1	0.3
b_TrialTrial_3	1.0	5257	-0.1	0.1	-0.3	-0.1	0.1
b_TrialTrial_4	1.0	4799	0.0	0.1	-0.2	0.0	0.2
b_ConditionSuppressive:TrialTrial_2	1.0	4514	-0.1	0.2	-0.4	-0.1	0.2
b_ConditionSuppressive:TrialTrial_3	1.0	4342	-0.0	0.2	-0.4	-0.0	0.3
b_ConditionSuppressive:TrialTrial_4	1.0	4051	-0.1	0.2	-0.4	-0.1	0.2
shape	1.0	8000	3.8	0.3	3.2	3.8	4.4

Table C.30: Parameter table for the Gamma model.

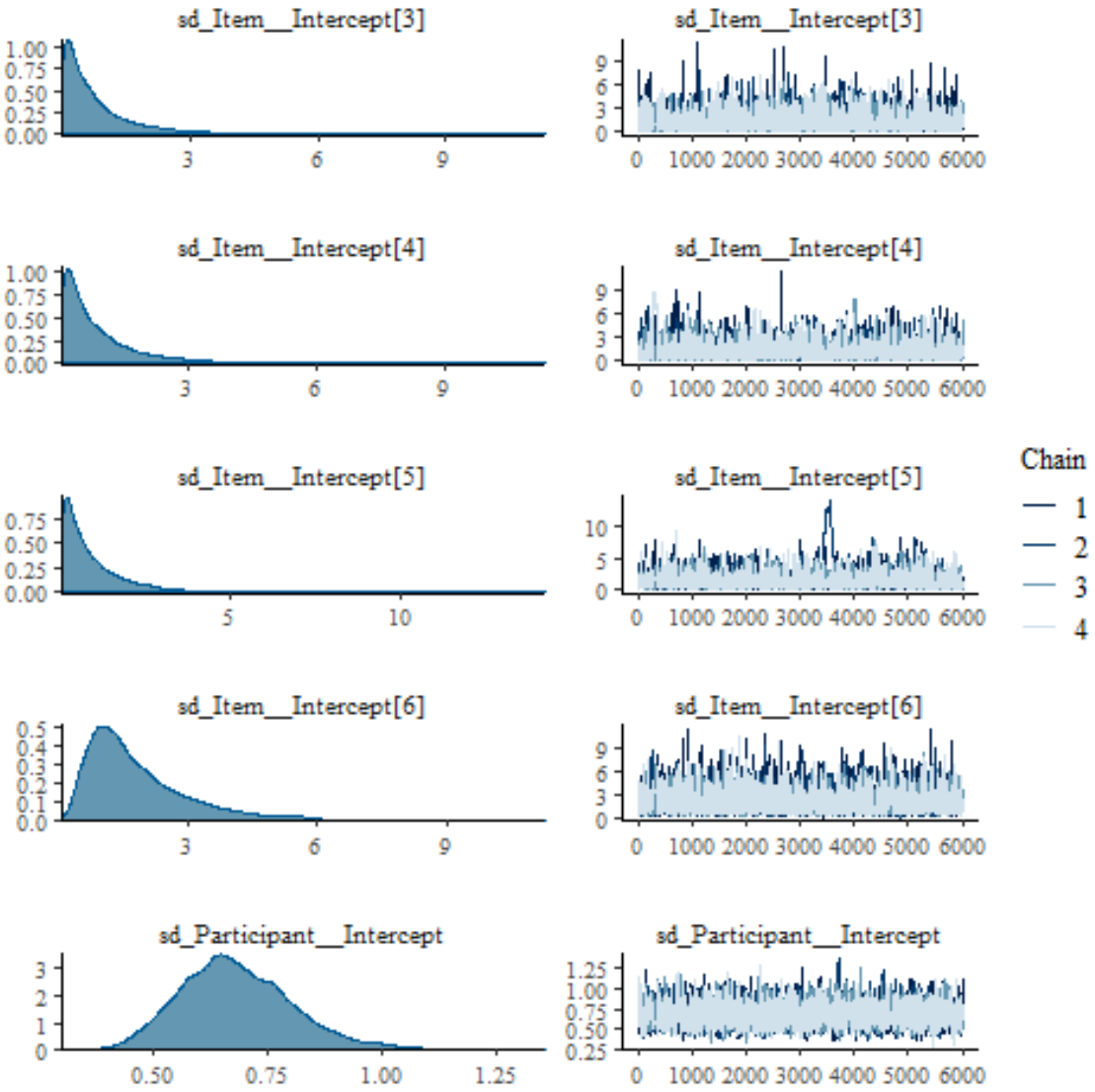


Figure C.65: Density & trace plots for the unequal variance between items model.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
b_Intercept	1.0	2513	4.4	0.1	4.2	4.4	4.5
b_ConditionSuppressive	1.0	2326	0.2	0.1	-0.0	0.2	0.3
b_TrialTrial_2	1.0	2610	0.2	0.1	0.1	0.2	0.4
b_TrialTrial_3	1.0	2903	0.1	0.1	-0.0	0.1	0.3
b_TrialTrial_4	1.0	2926	0.1	0.1	-0.1	0.1	0.2
b_ConditionSuppressive:TrialTrial_2	1.0	2635	-0.1	0.1	-0.3	-0.1	0.1
b_ConditionSuppressive:TrialTrial_3	1.0	2752	-0.1	0.1	-0.4	-0.1	0.1
b_ConditionSuppressive:TrialTrial_4	1.0	2704	-0.1	0.1	-0.3	-0.1	0.1
sigma	1.0	5095	14.2	2.2	10.3	14.1	18.7
beta	1.0	5327	49.0	3.3	42.7	48.9	55.7

Table C.31: Parameter table for the exponentially modified Gaussian model.

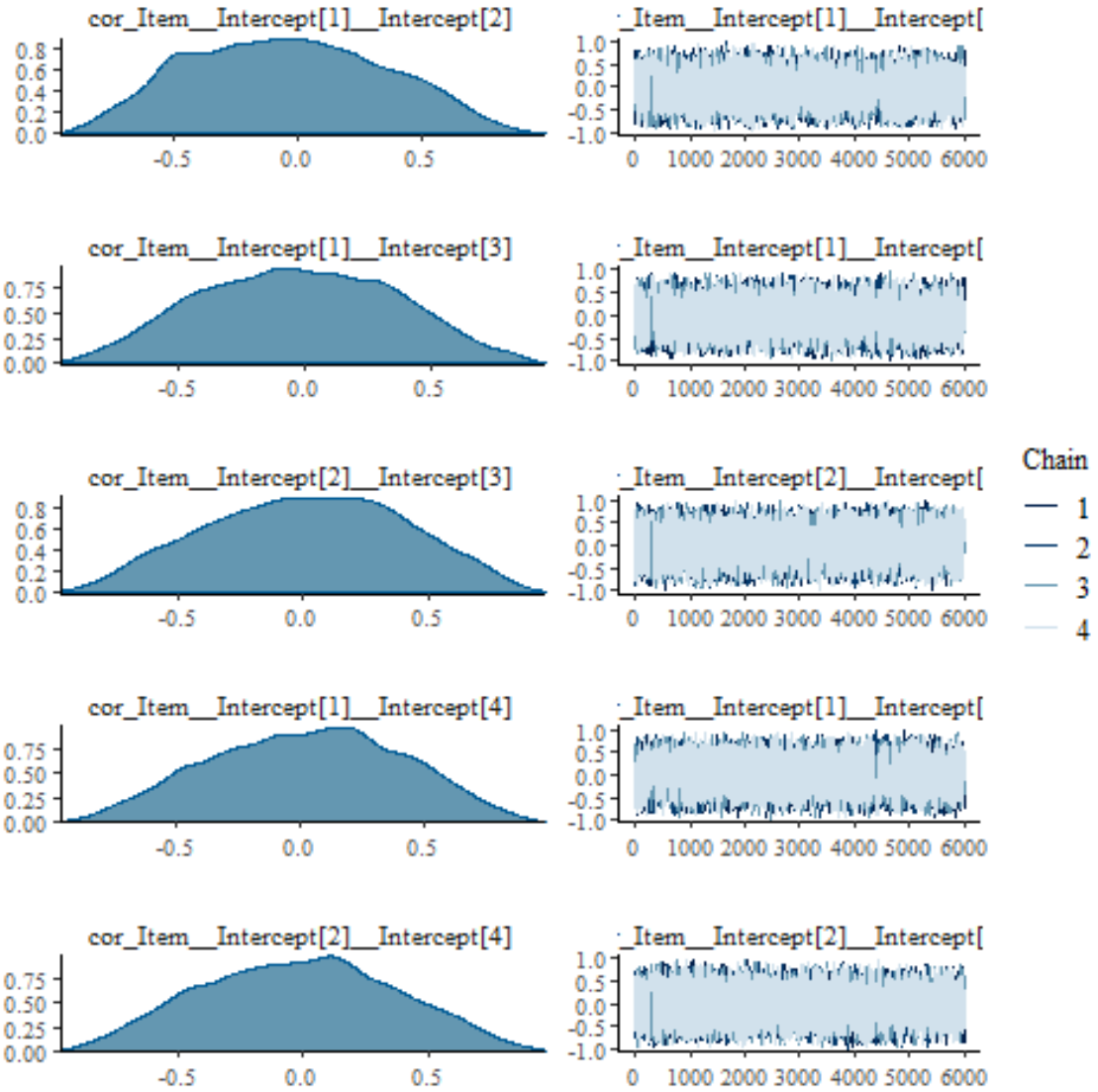


Figure C.66: Density & trace plots for the unequal variance between items model.

Model	2.5 % HDI < 0
Gaussian	0.751
Exp. Gaussian	0.000
Gamma	0.000

Table C.32: Proportions of the predictive posterior samples with a negative 2.5% credibility limit.

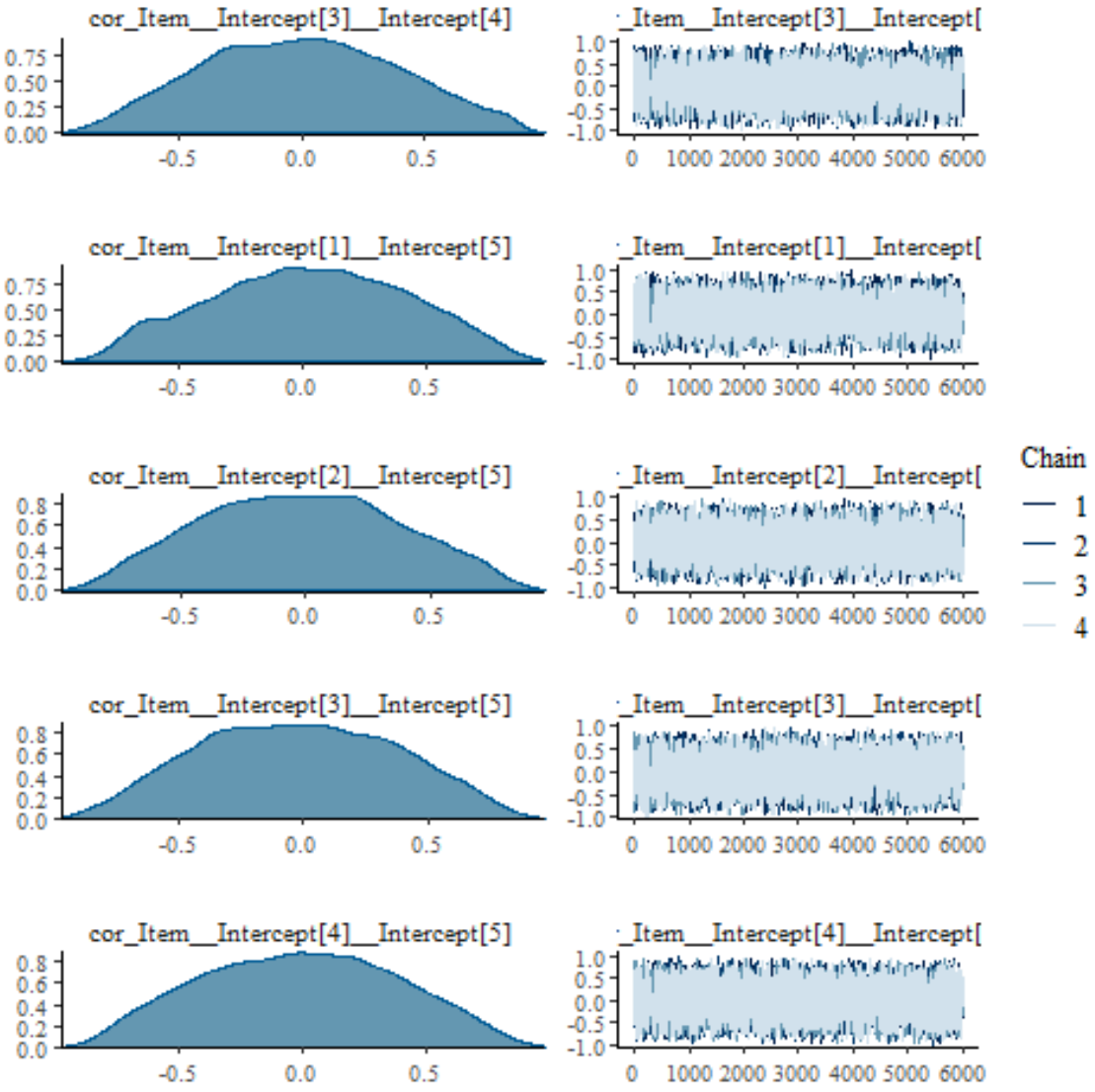


Figure C.67: Density & trace plots for the unequal variance between items model.

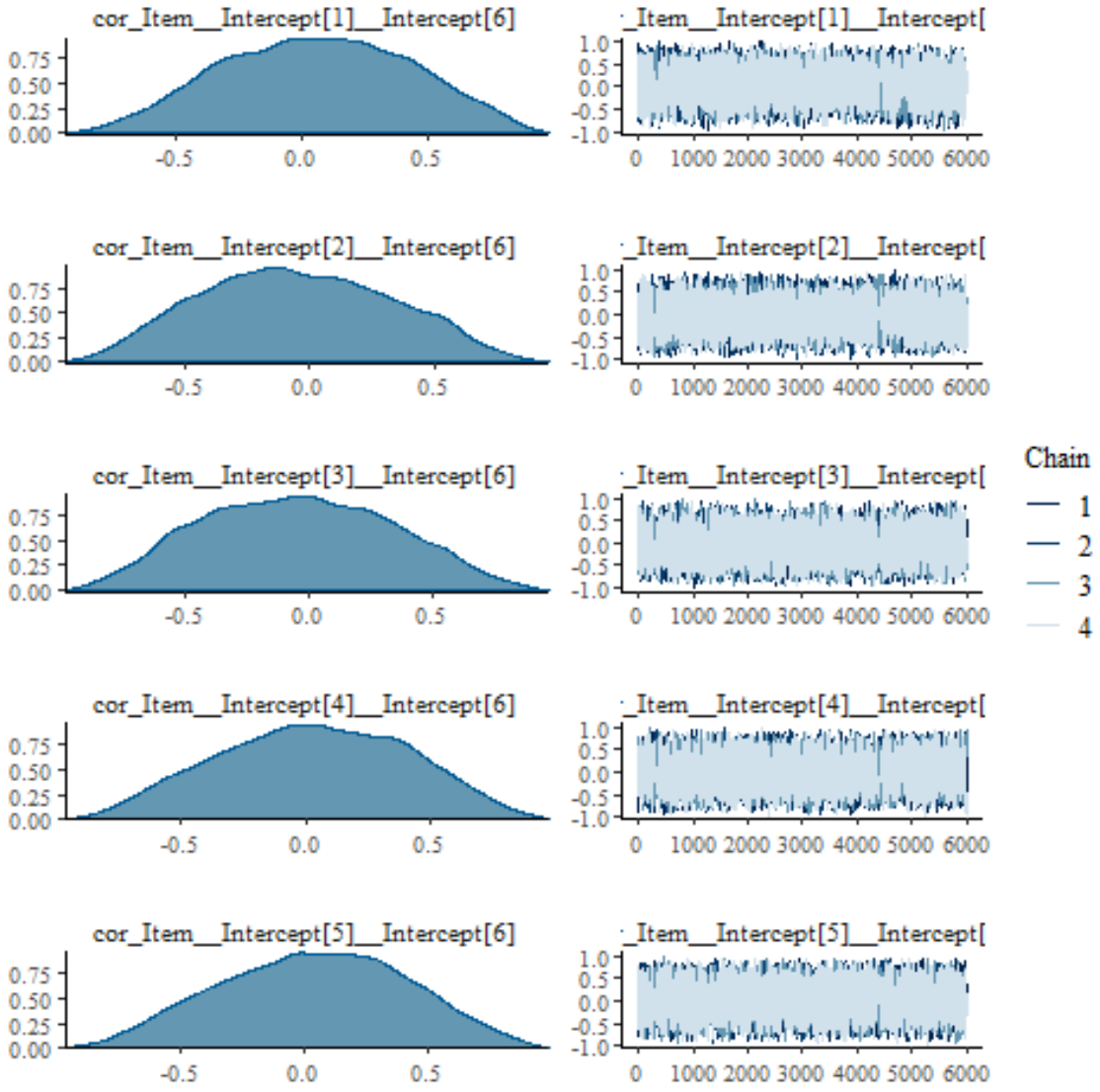


Figure C.68: Density & trace plots for the unequal variance between items model.

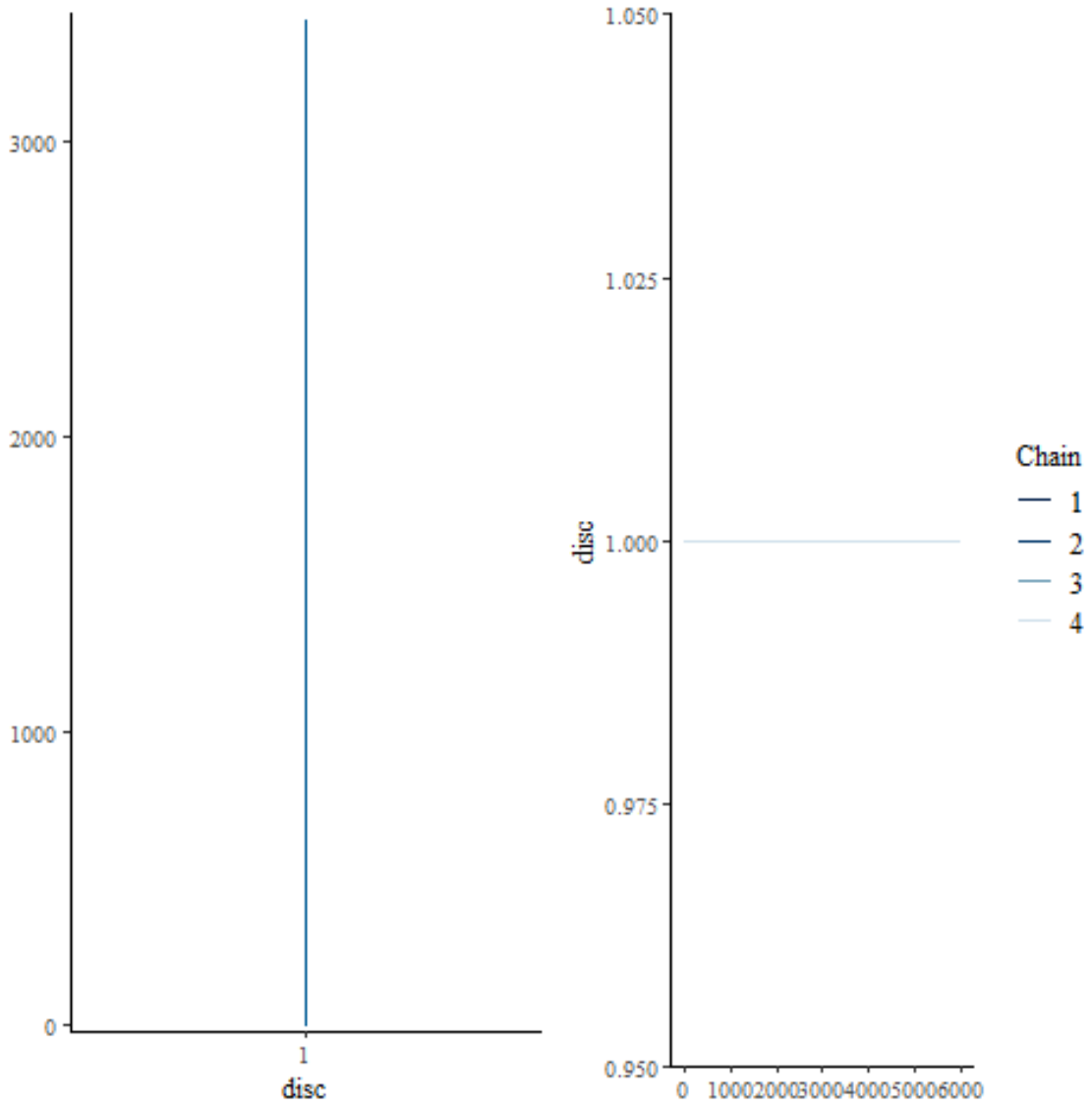


Figure C.69: Density & trace plots for the unequal variance between items model.

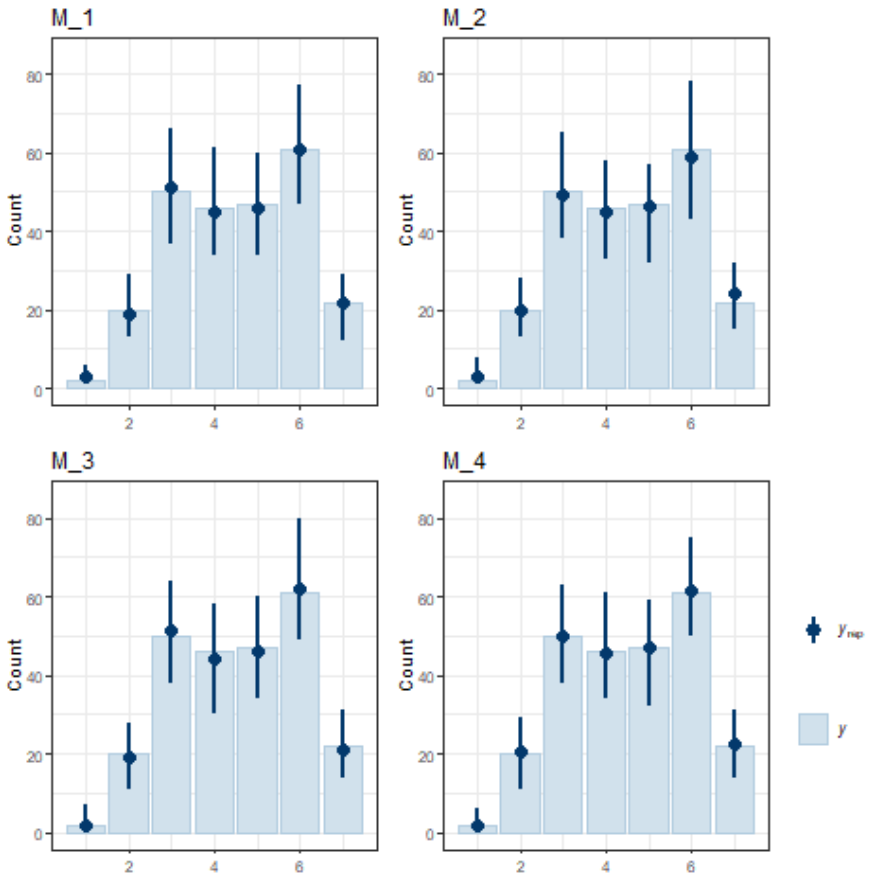


Figure C.70: Predictive posterior checks of the SoS OLR models.

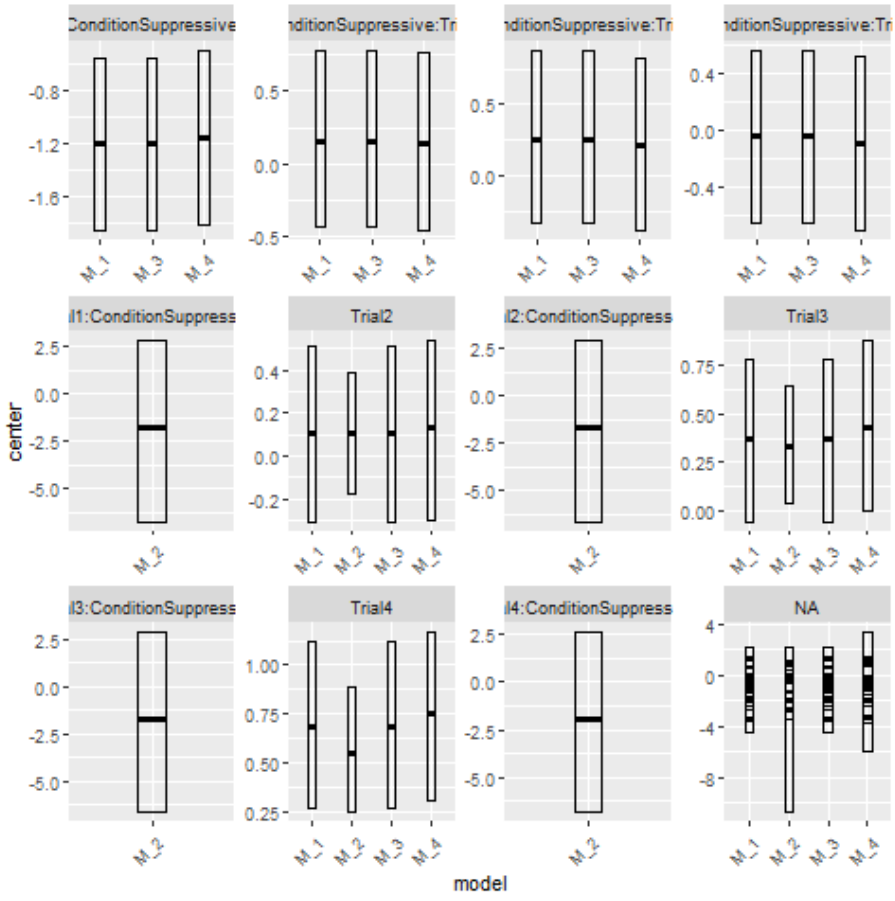


Figure C.71: Center & 95 % HDI bar plots of the SoS OLR models.

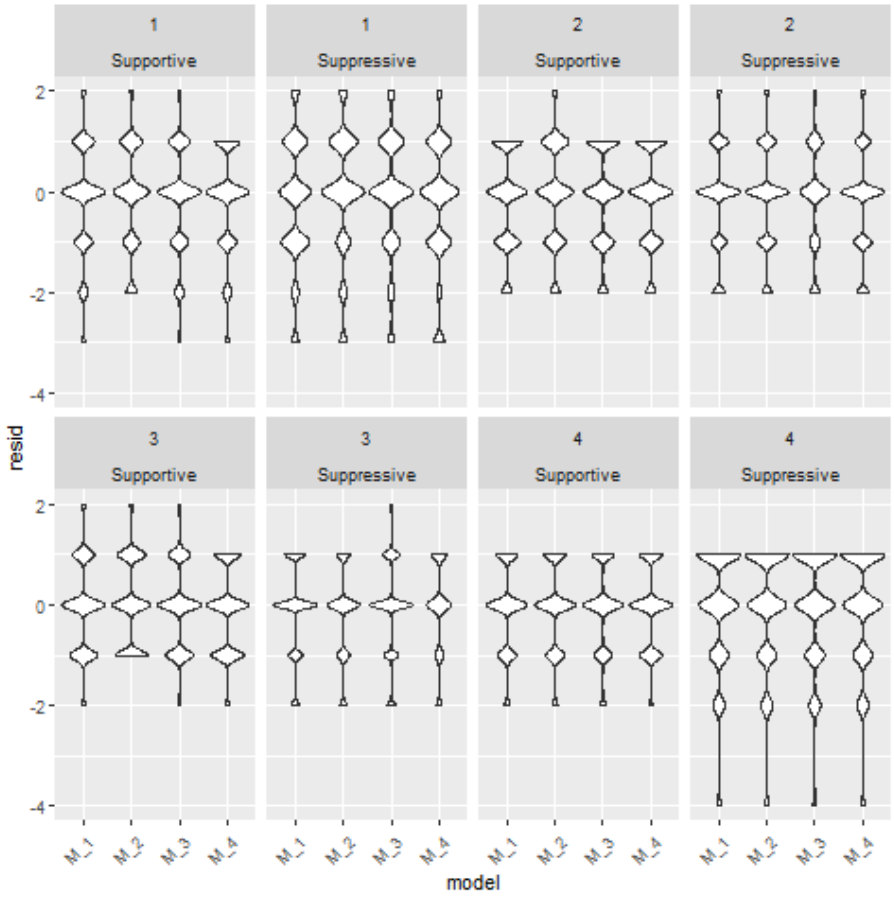


Figure C.72: Violin plots of the residuals of the SoO OLR models.

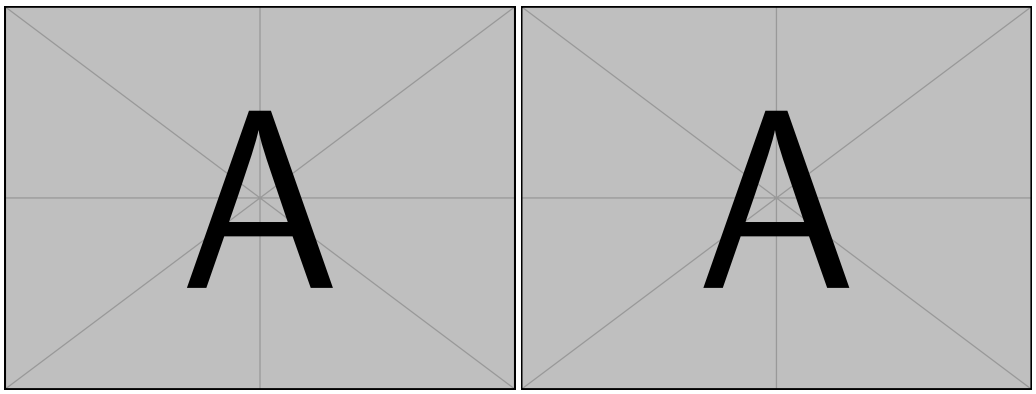


Figure C.73: Priors for learning rate on the normal scale (left) and inverse-logit scale (right).

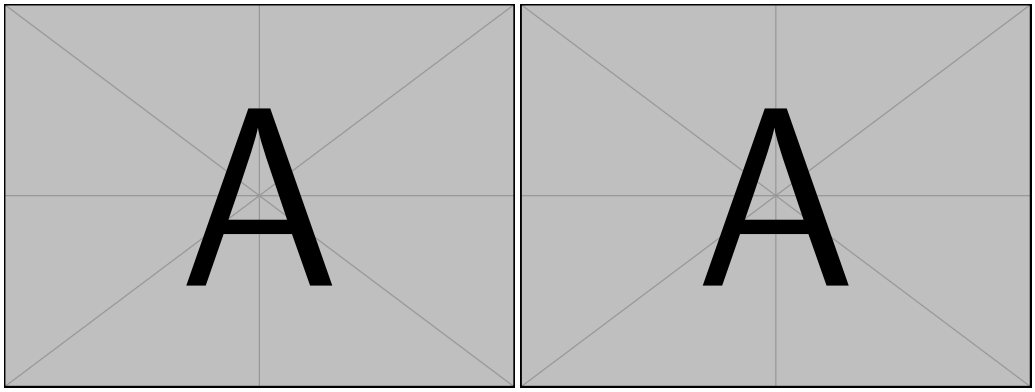


Figure C.74: Priors for the asymptote on the normal scale (left) and inverse-logit scale (right) for the variable number of inserted pegs.

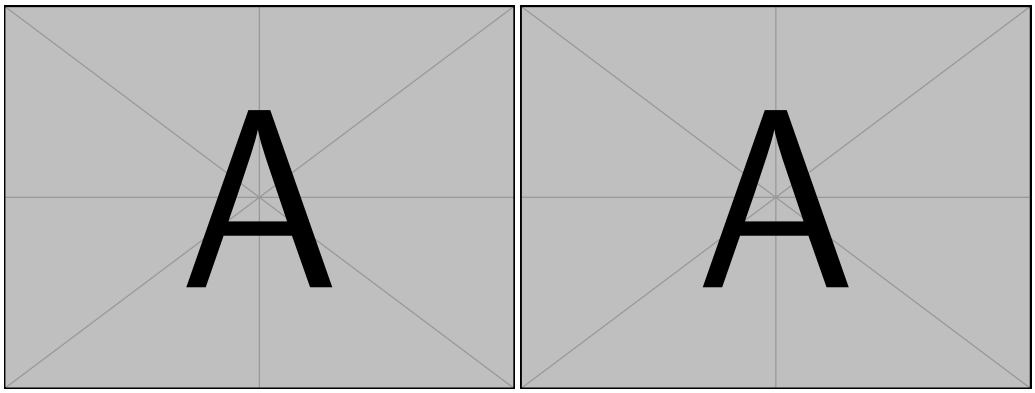


Figure C.75: Priors for the amplitude on the normal scale (left) and inverse-logit scale (right) for the variable number of inserted pegs.

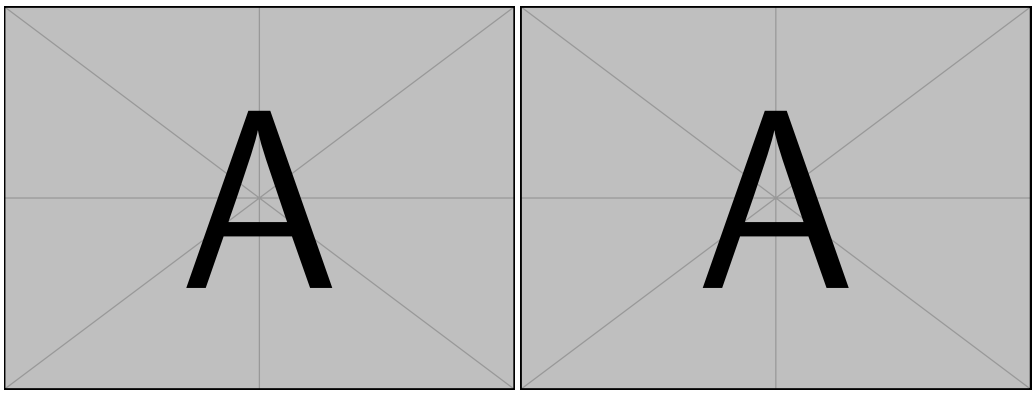


Figure C.76: Priors for the asymptote on the normal scale (left) and inverse-logit scale (right) for the variable time per peg.

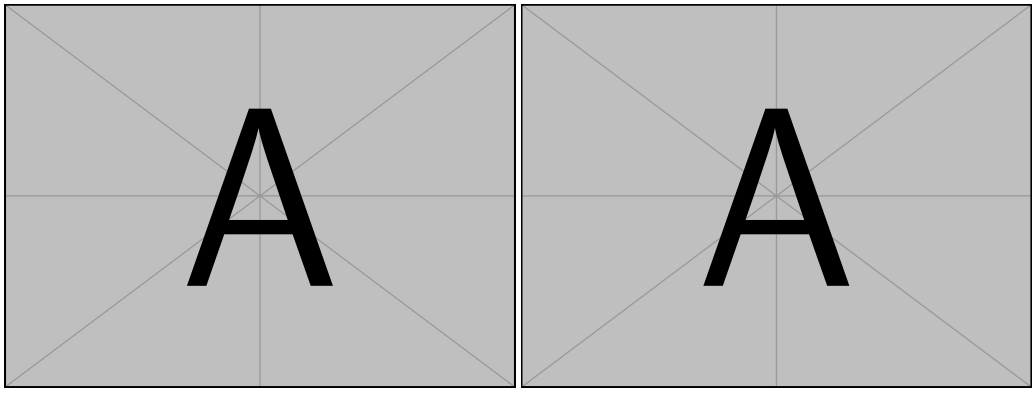


Figure C.77: Priors for the amplitude on the normal scale (left) and inverse-logit scale (right) for the variable time per peg.

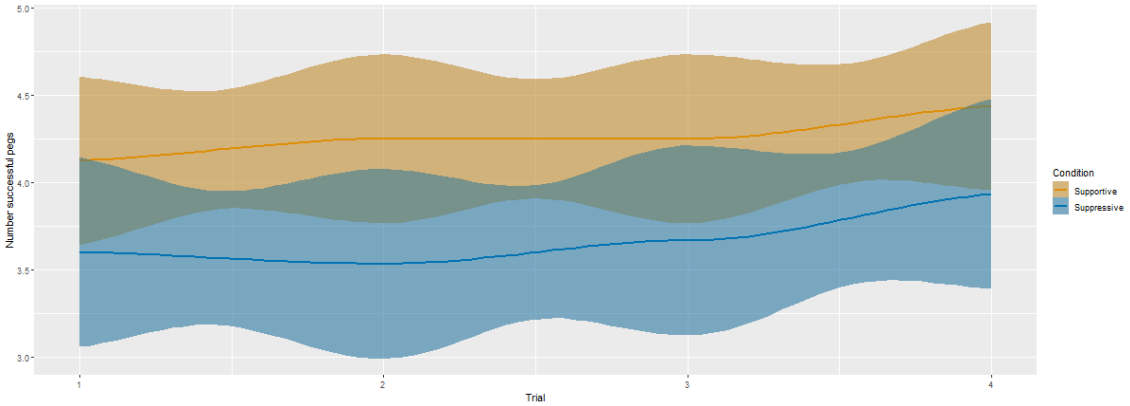


Figure C.78: Visualization of learning behavior on the population level for the count variable.

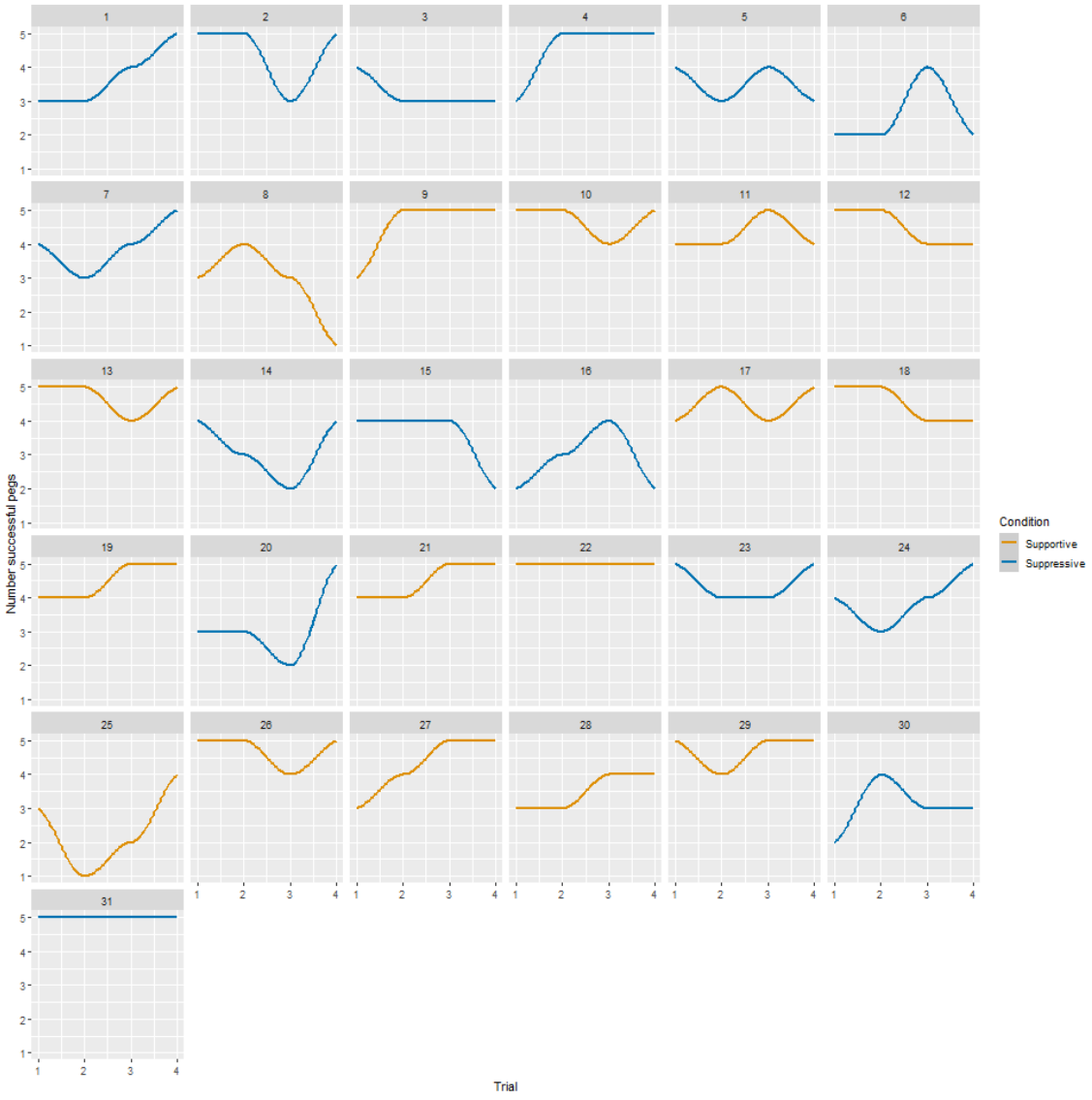


Figure C.79: Visualization of learning behavior on the participant level for the count variable.

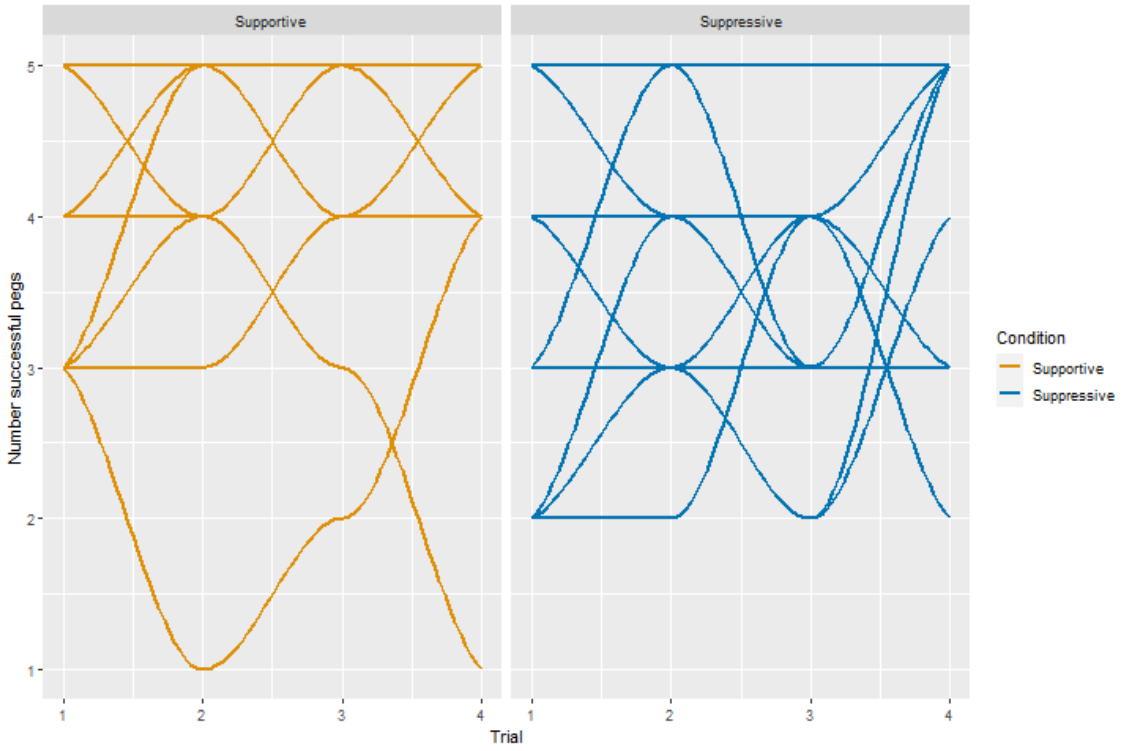


Figure C.80: Visualization of learning behavior on the participant level, grouped, for the count variable.

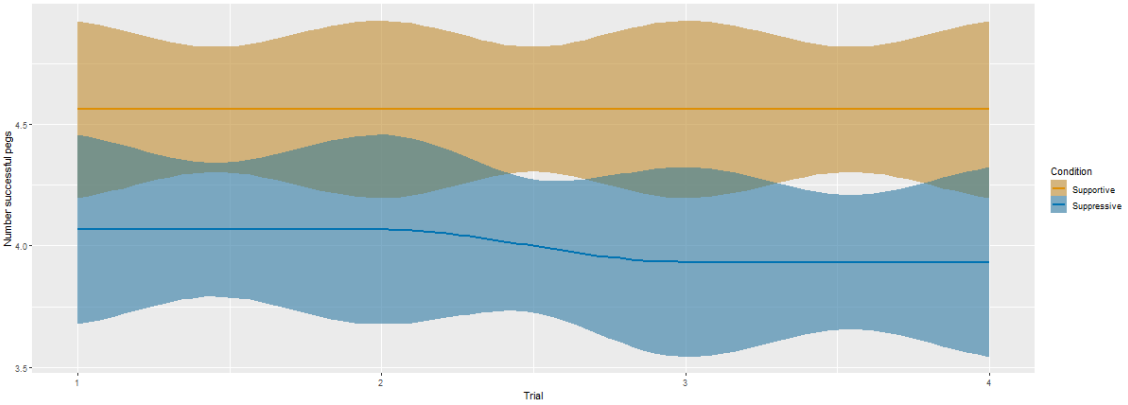


Figure C.81: Visualization of the predicted learning behavior on the population level for the count variable.

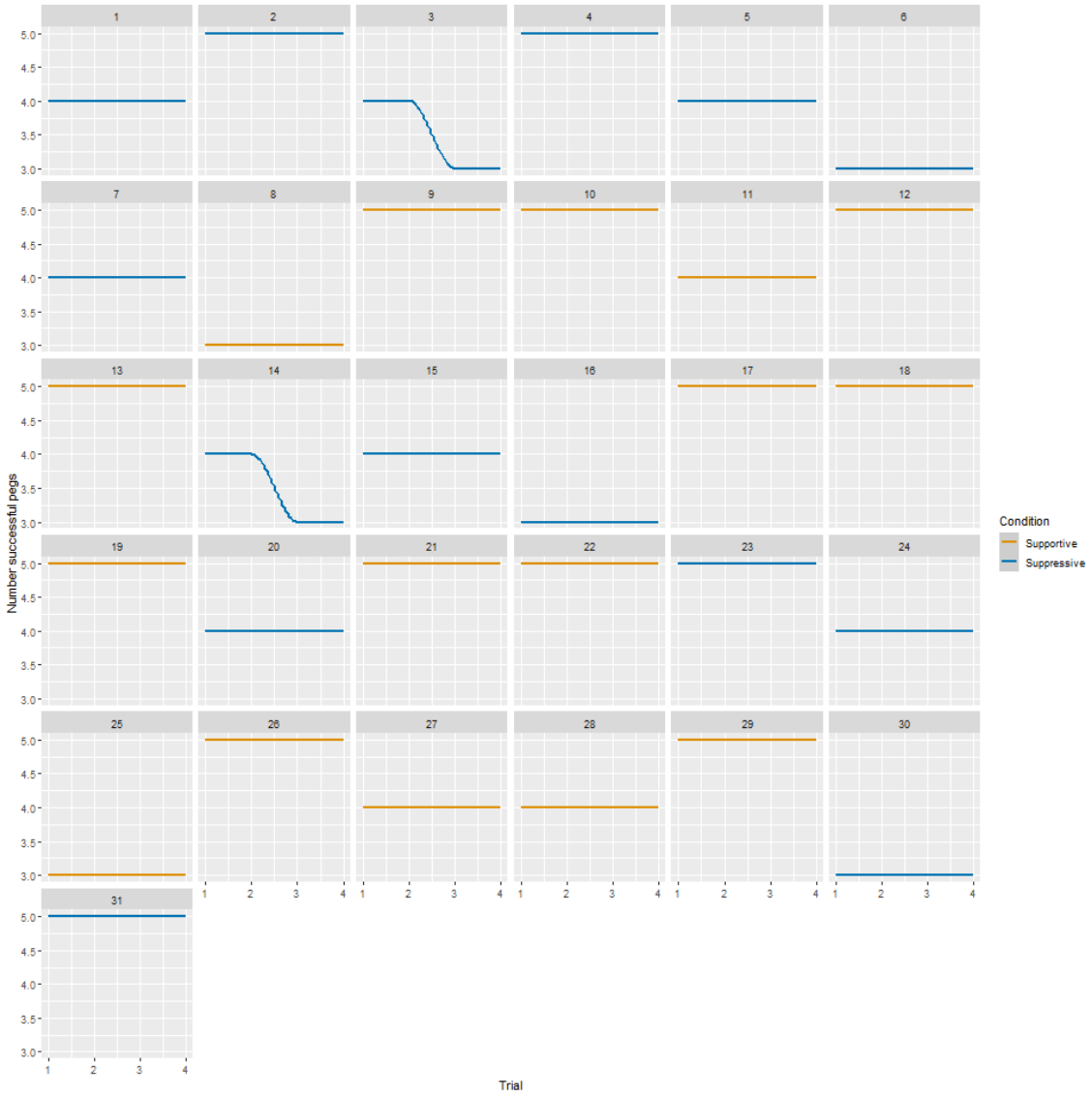


Figure C.82: Visualization of the predicted learning behavior on the participant level for the count variable.

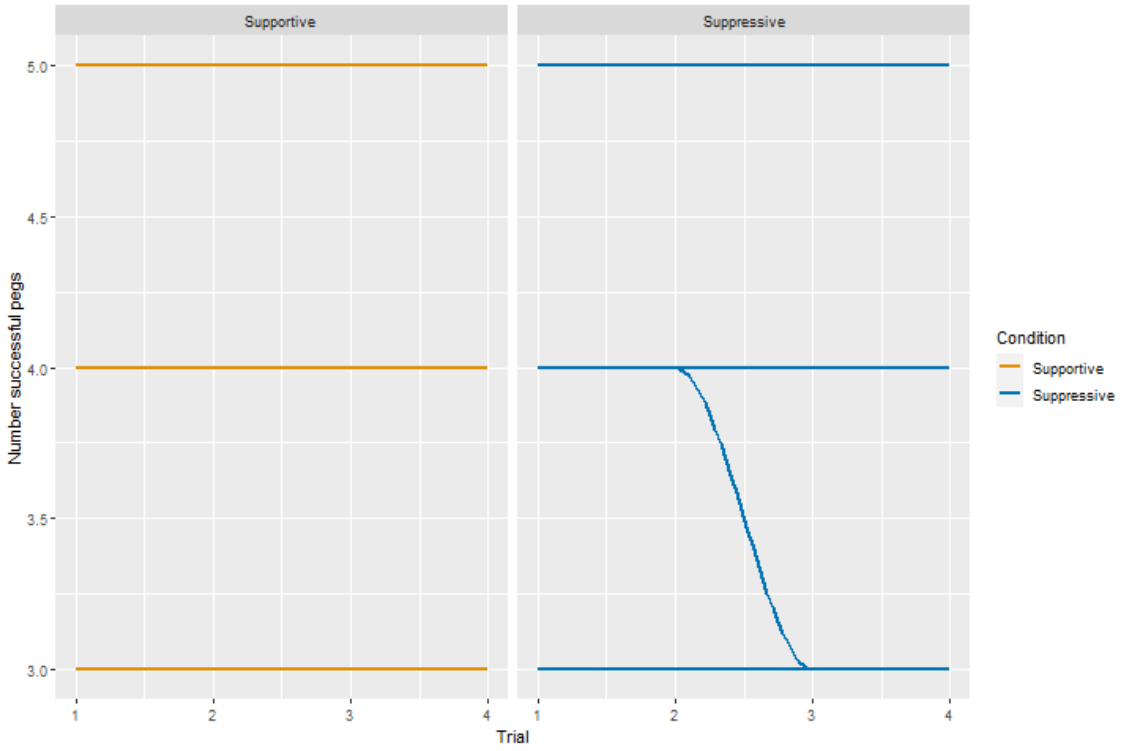


Figure C.83: Visualization of the predicted learning behavior on the participant level, grouped, for the count variable.

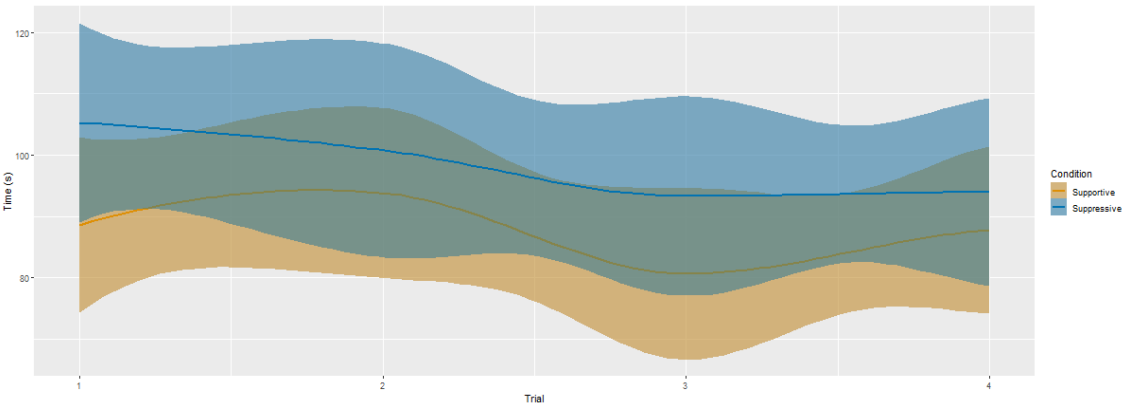


Figure C.84: Visualization of learning behavior on the population level for the ToT variable.

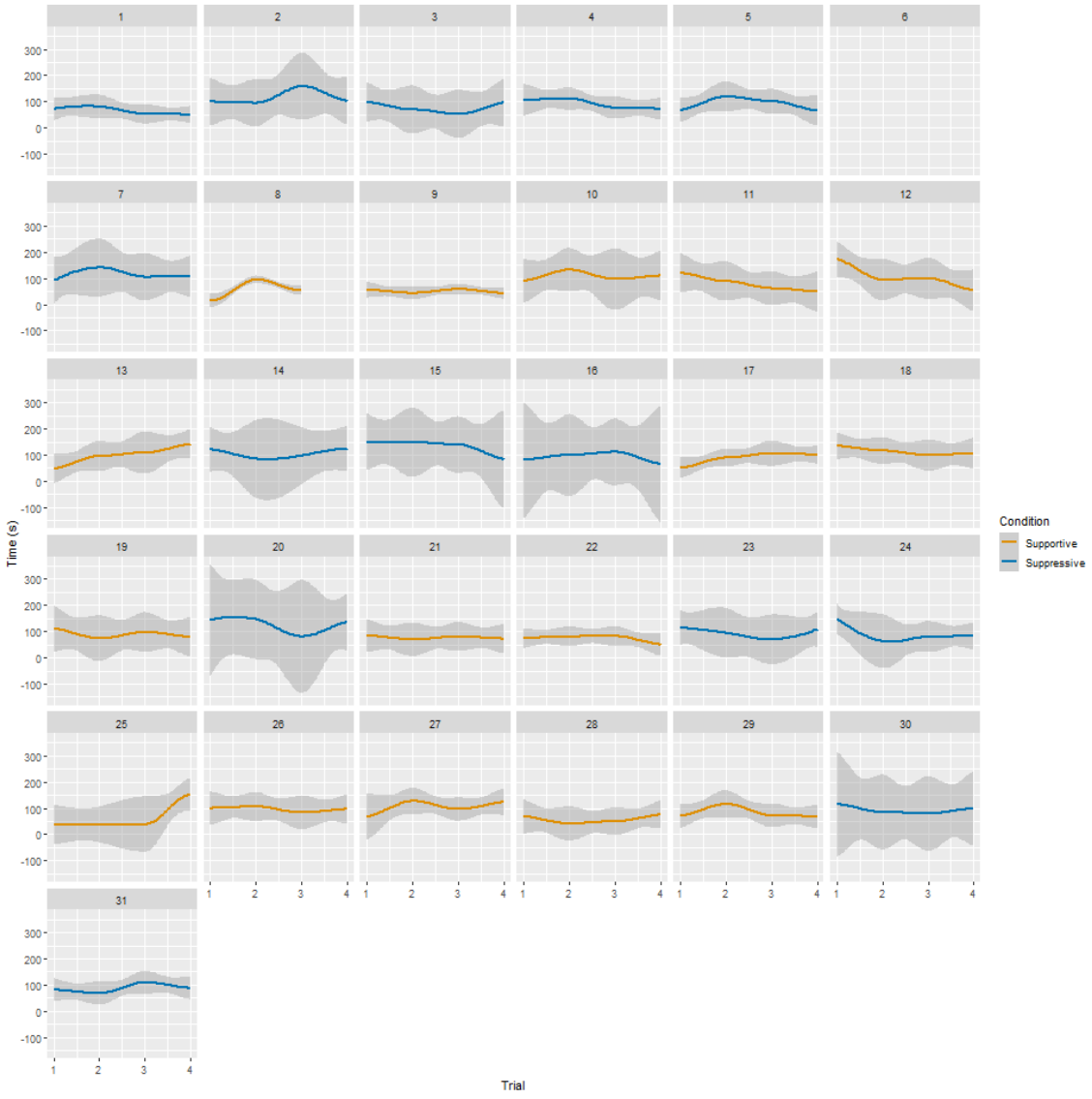


Figure C.85: Visualization of learning behavior on the participant level for the ToT variable.

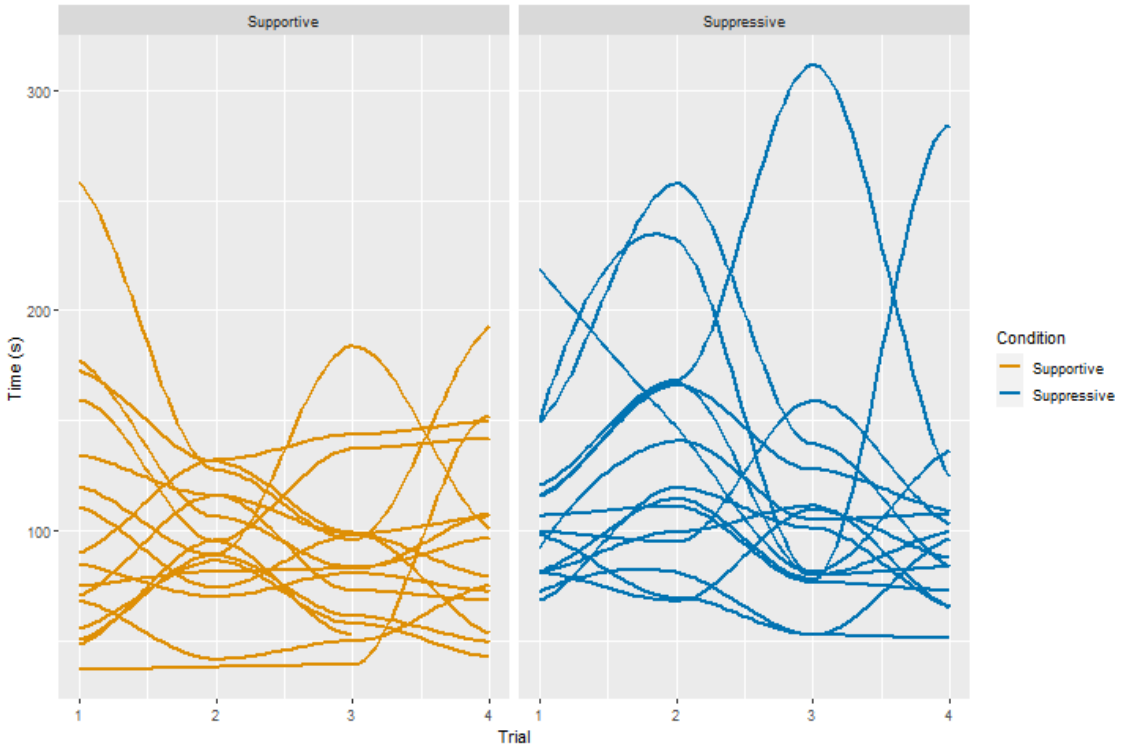


Figure C.86: Visualization of learning behavior on the participant level, grouped, for the ToT variable.

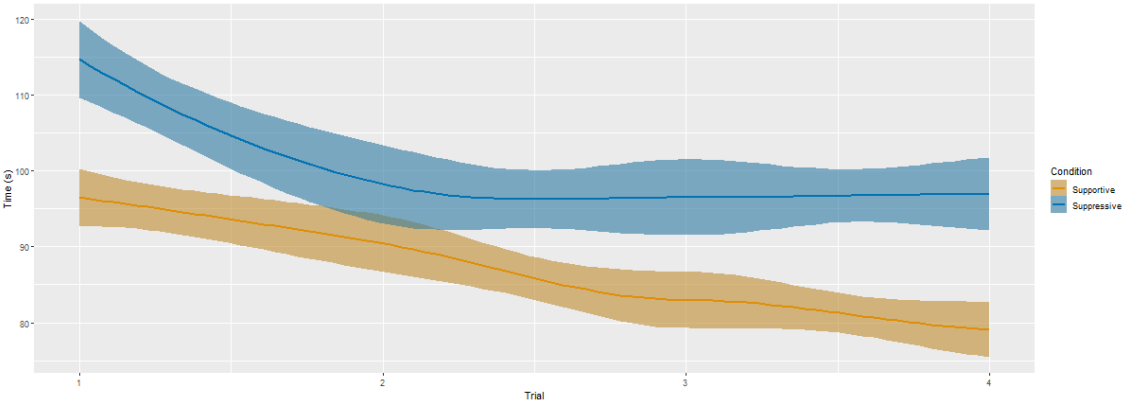


Figure C.87: Visualization of the predicted learning behavior on the population level for the ToT variable.

Figure C.88: Visualization of the predicted learning behavior on the participant level for the ToT variable.

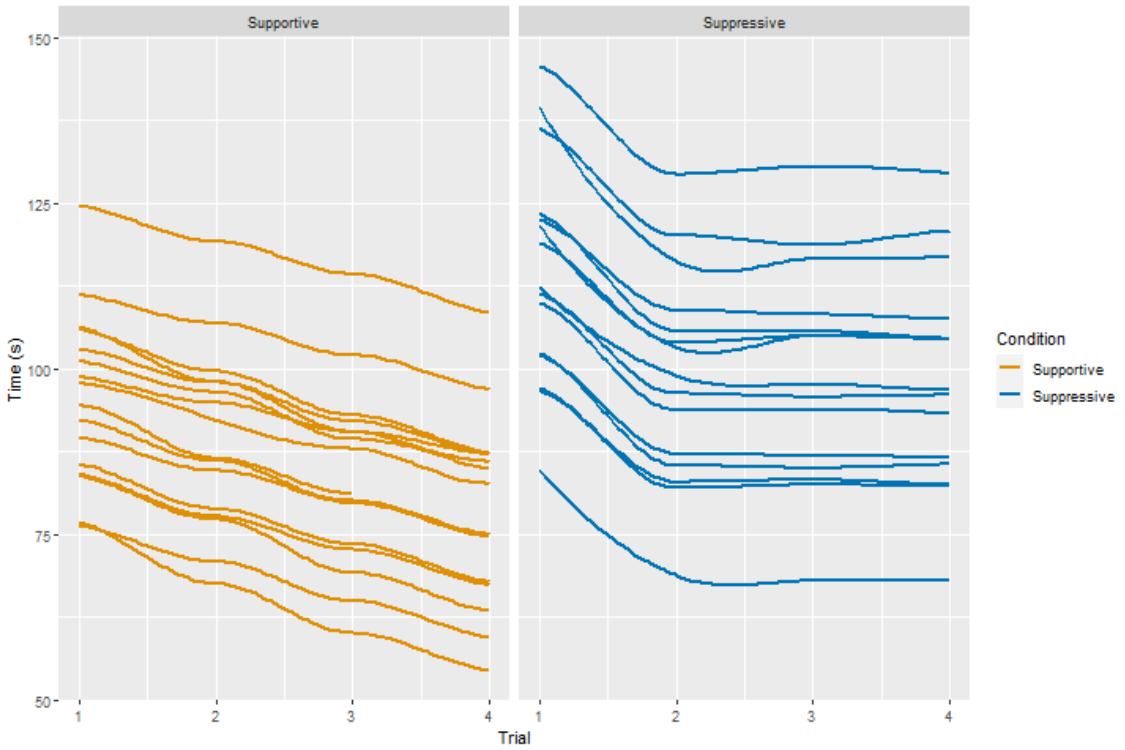


Figure C.89: Visualization of the predicted learning behavior on the participant level, grouped, for the ToT variable.

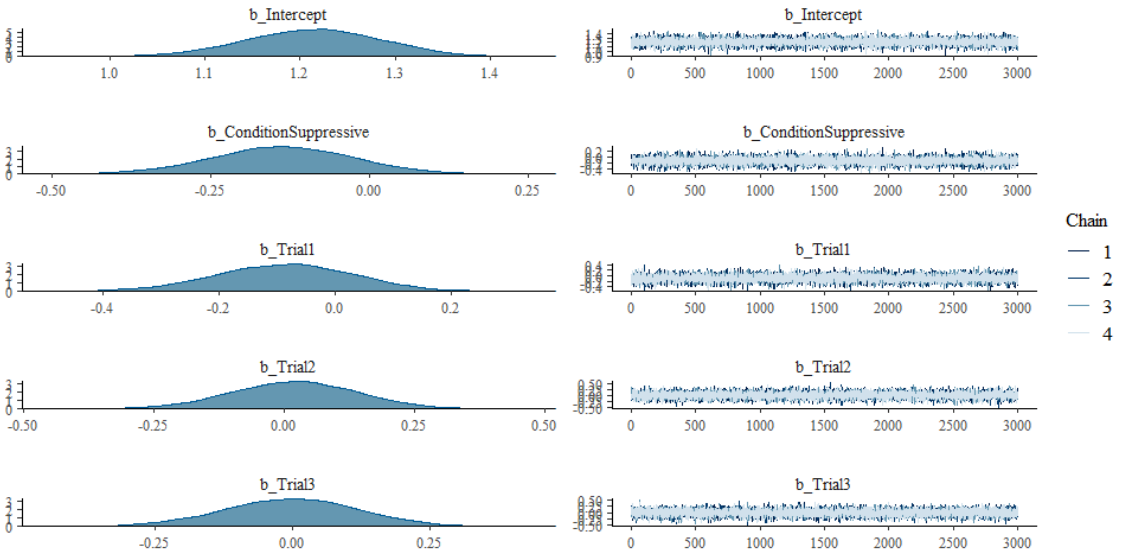


Figure C.90: Density & trace plots for the Poisson model .

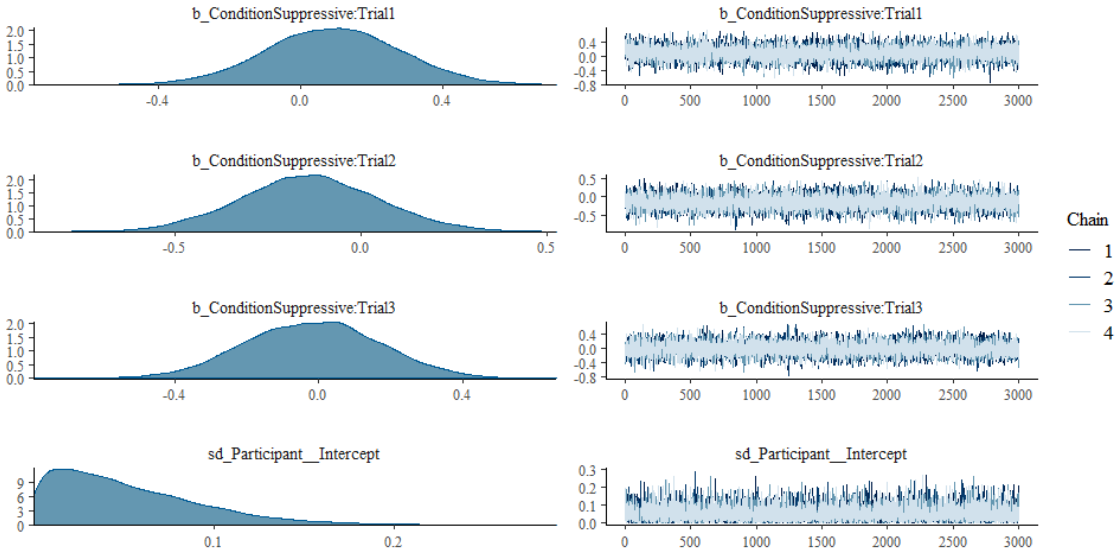


Figure C.91: Density & trace plots for the Poisson model .

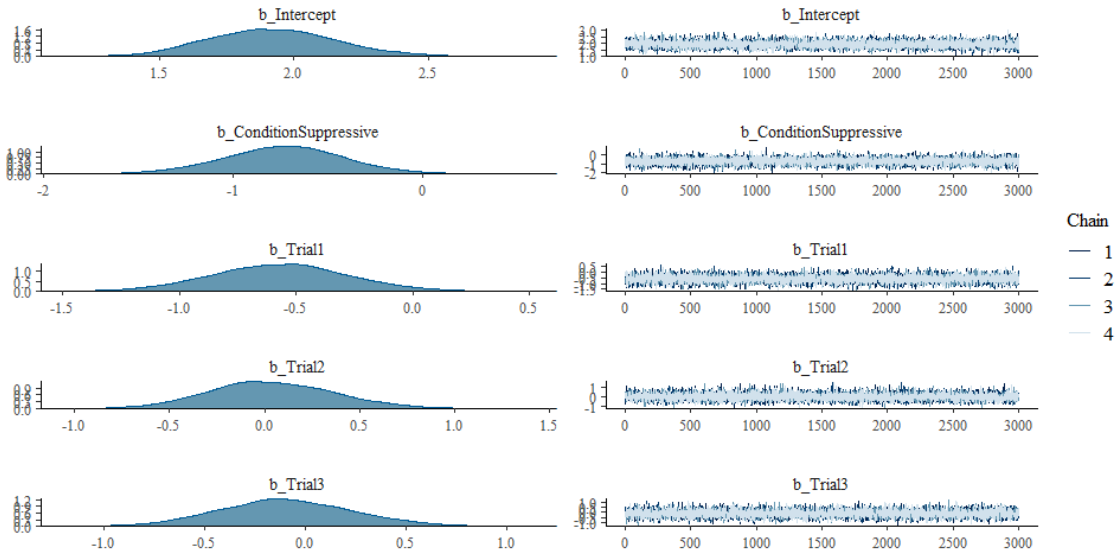


Figure C.92: Density & trace plots for the Binomial model .

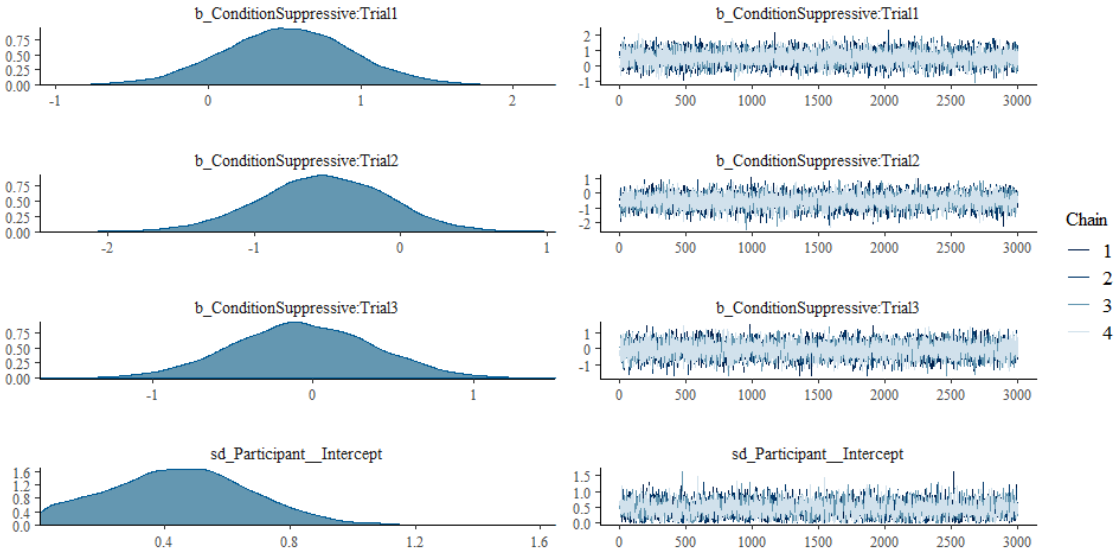


Figure C.93: Density & trace plots for the Binomial model .

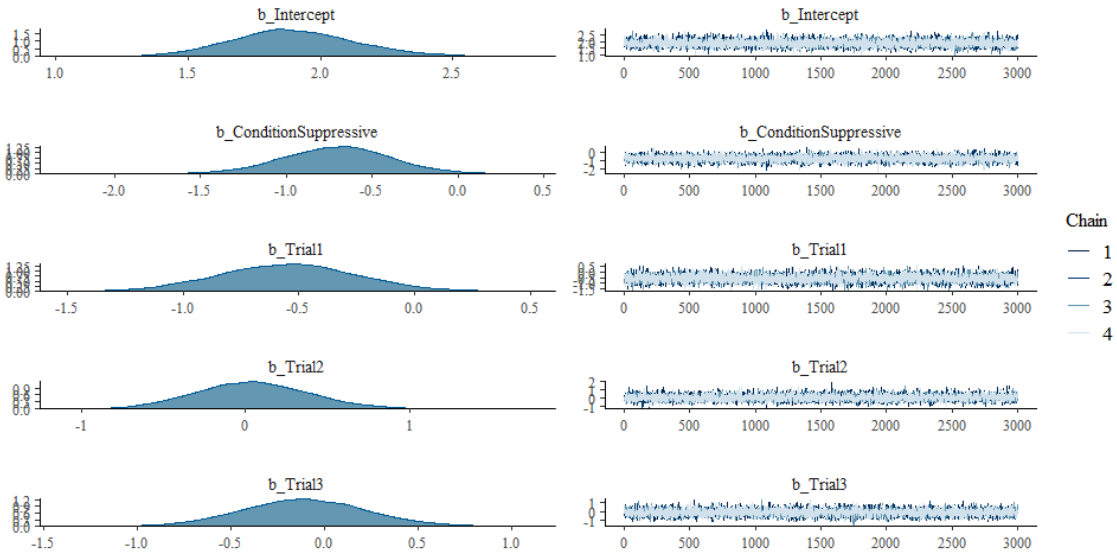


Figure C.94: Density & trace plots for the Beta-Binomial model .

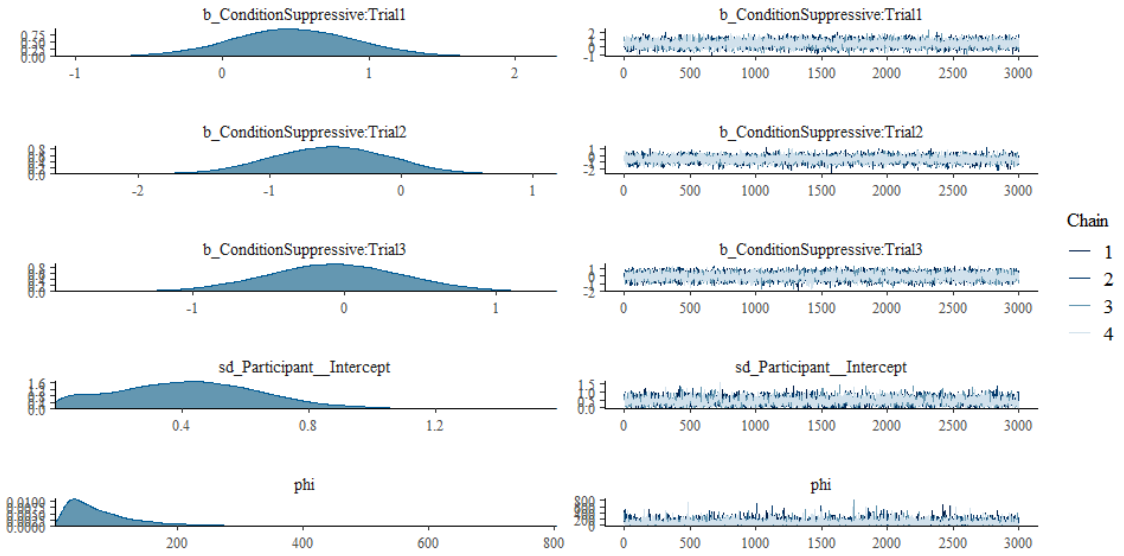


Figure C.95: Density & trace plots for the Beta-Binomial model .

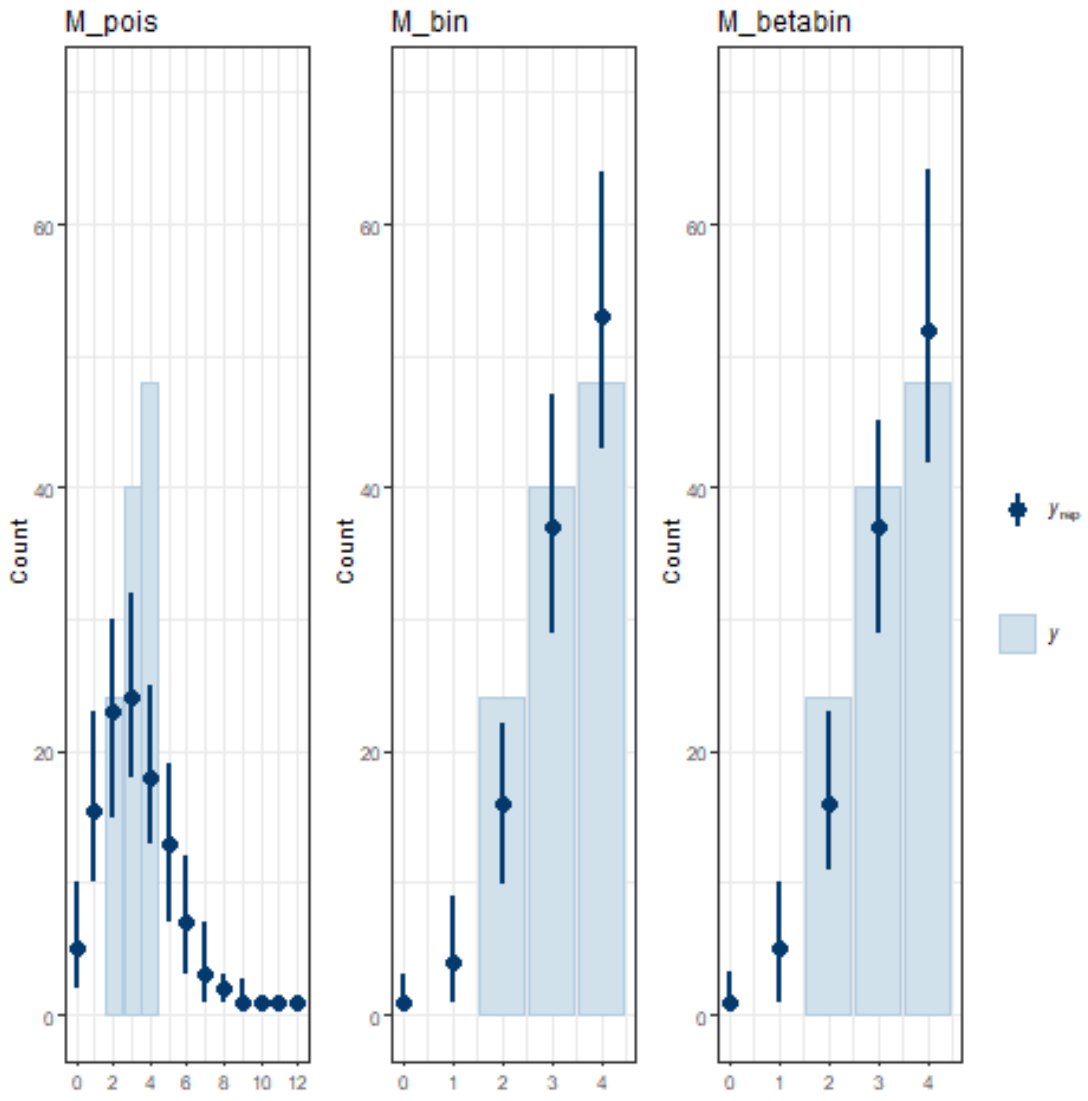


Figure C.96: Predictive posterior checks for Poisson, binomial and beta-binomial models.

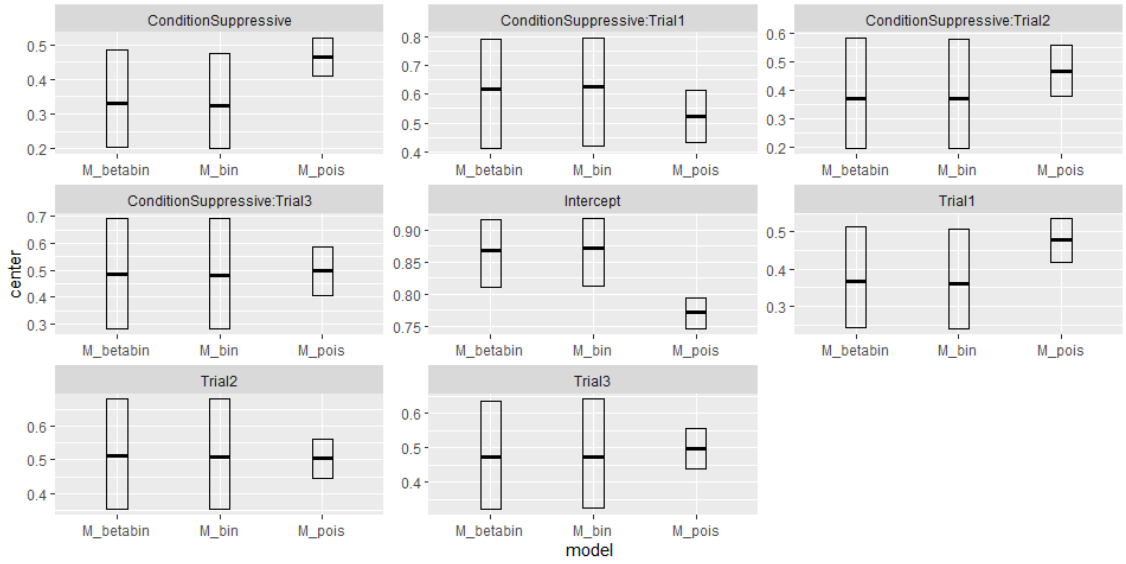


Figure C.97: Bar plots for Poisson, binomial and beta-binomial models.

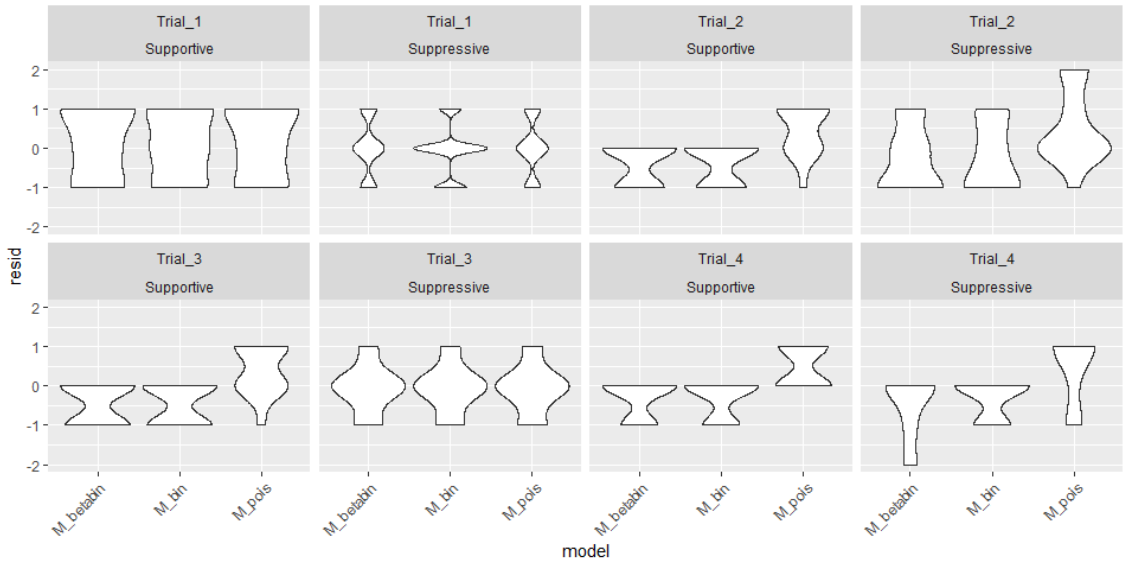


Figure C.98: Violin plots of the residuals for Poisson, binomial and beta-binomial models.

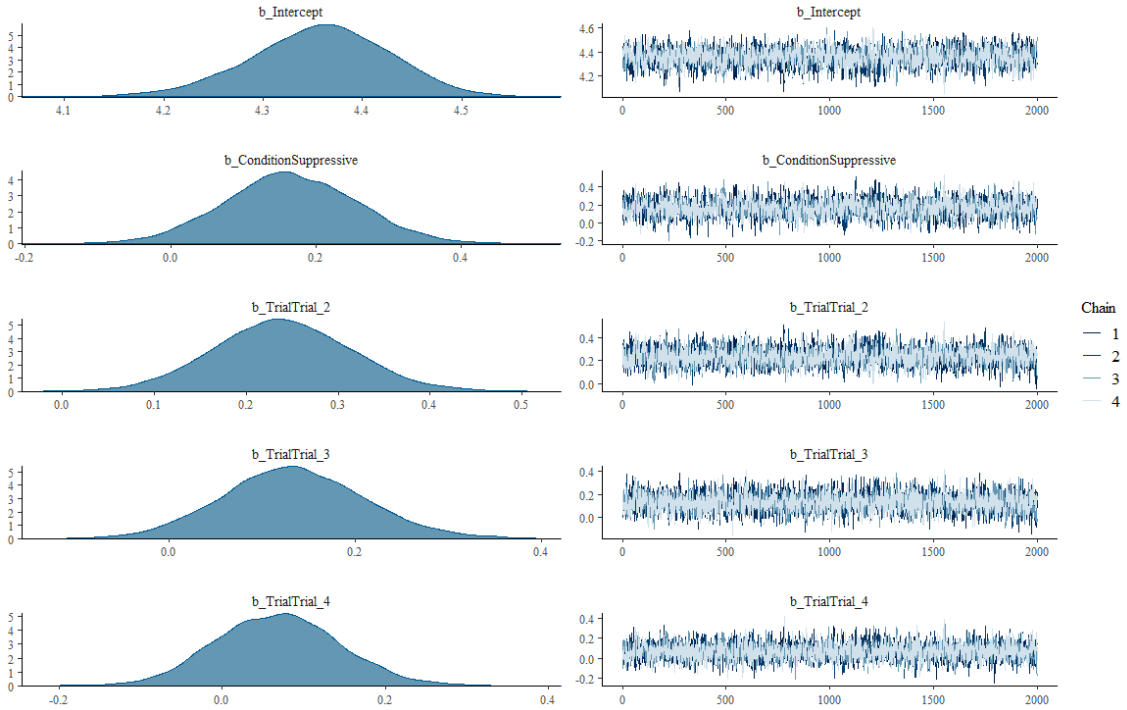


Figure C.99: Density & trace plots for the Gaussian model .

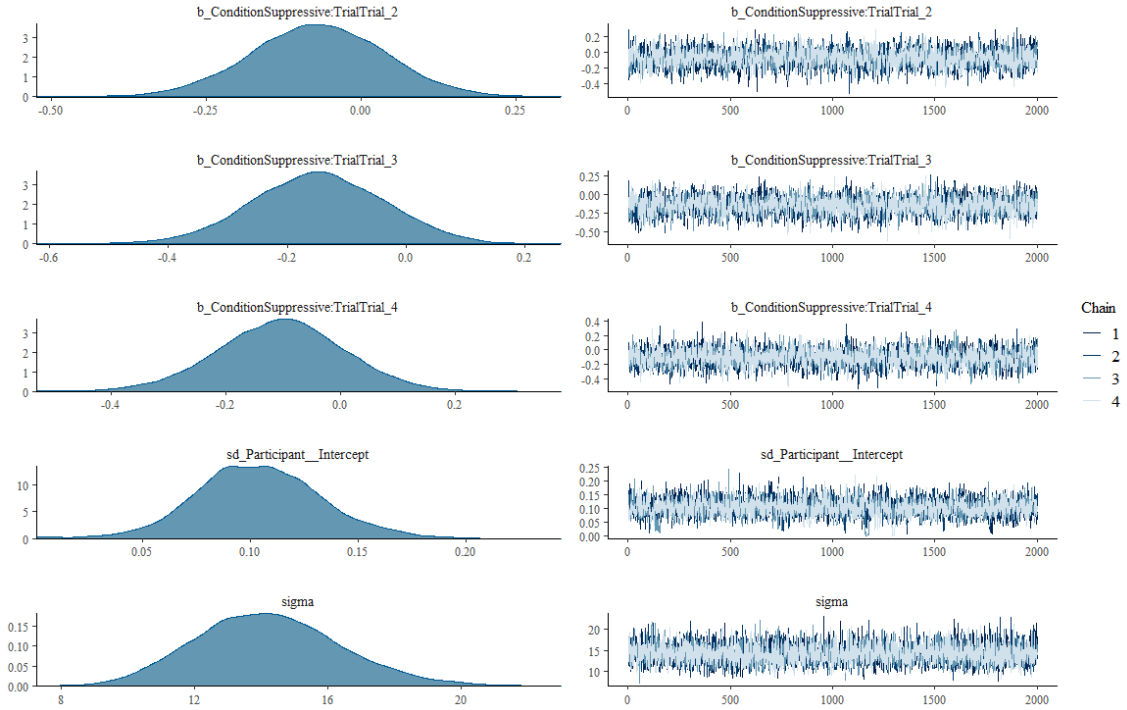


Figure C.100: Density & trace plots for the Gaussian model .

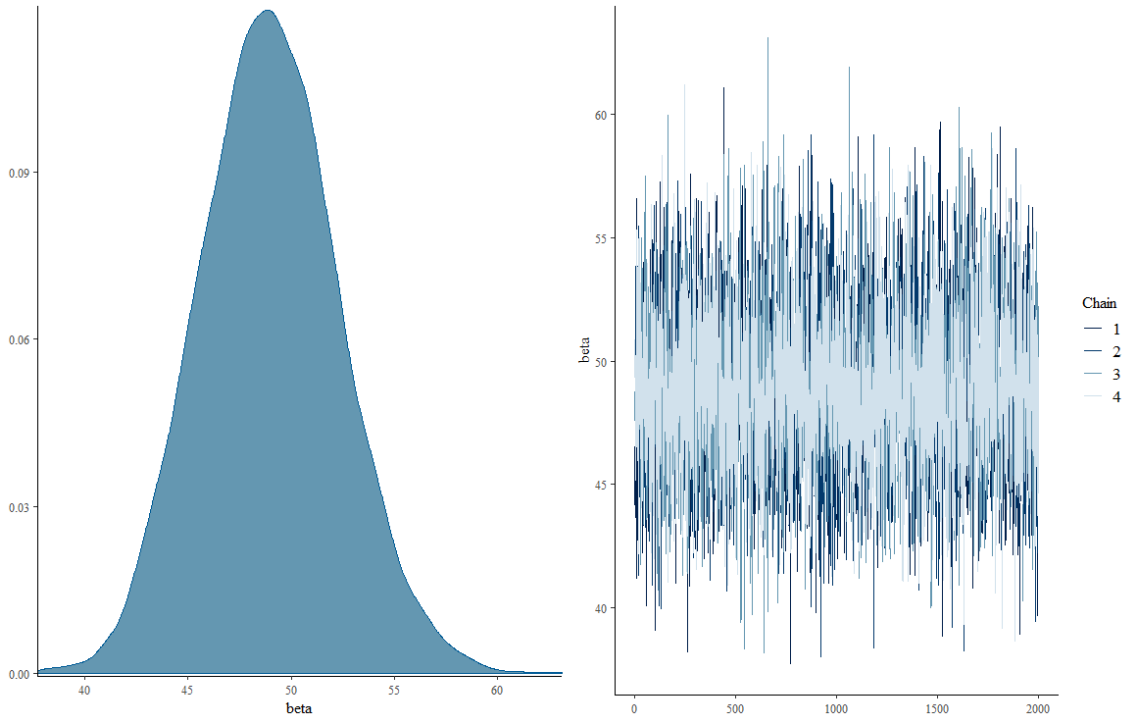


Figure C.101: Density & trace plots for the Gaussian model .

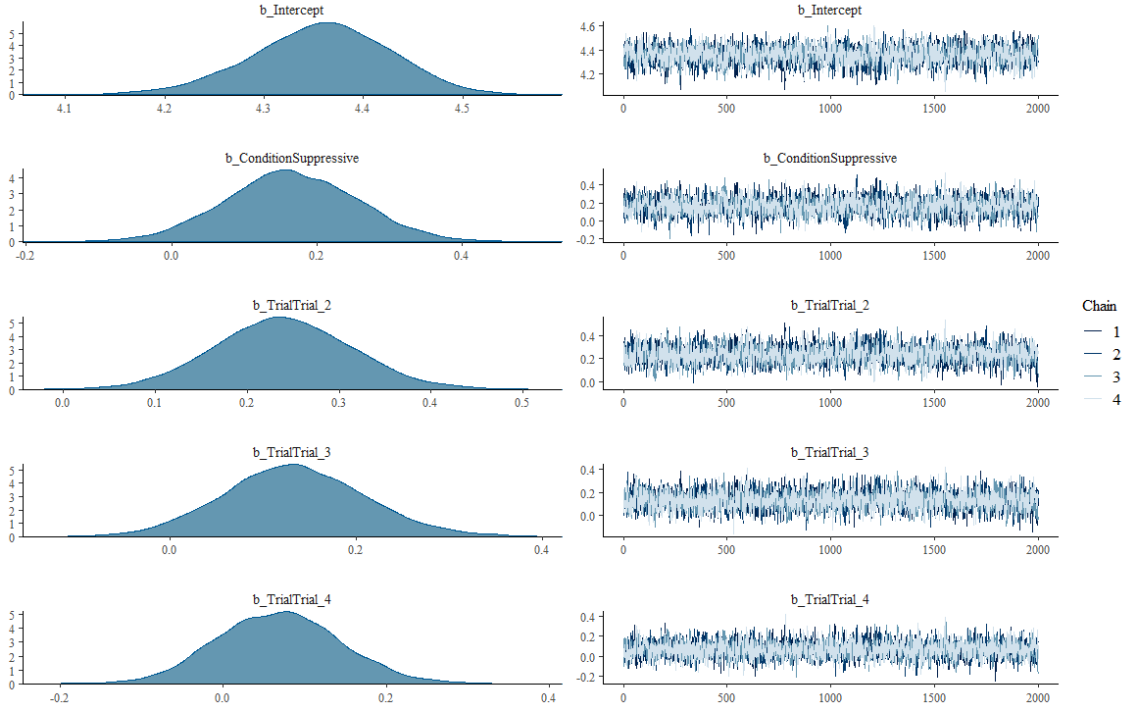


Figure C.102: Density & trace plots for the Exponentially modified Gaussian model .

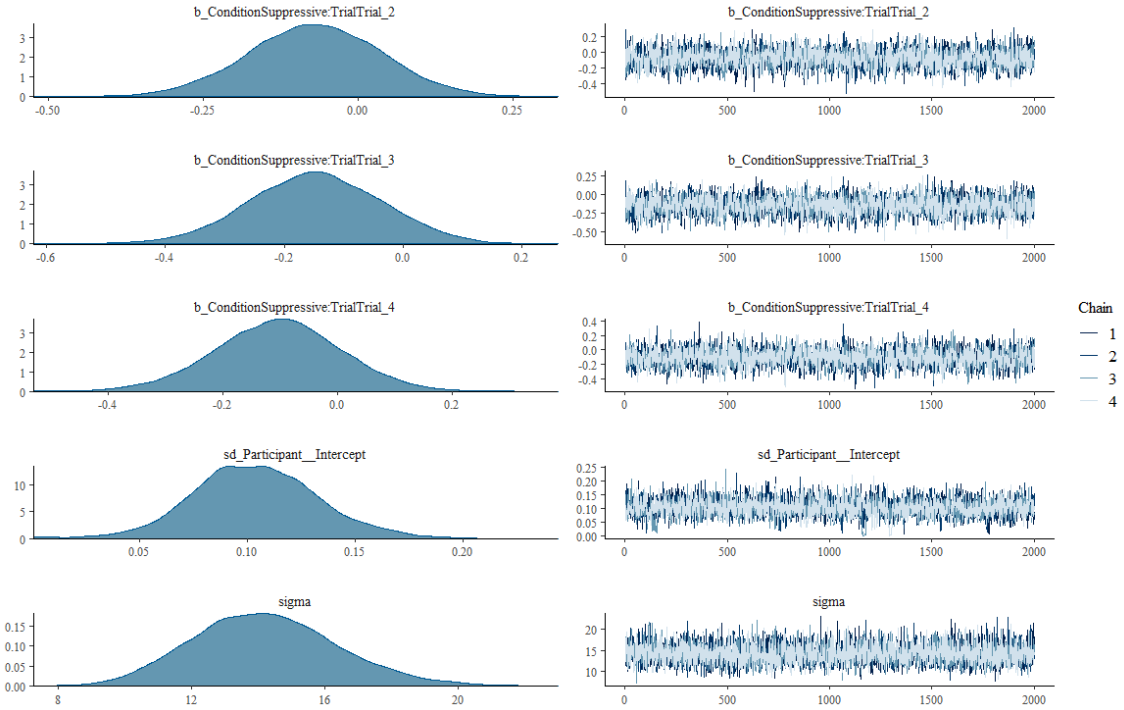


Figure C.103: Density & trace plots for the Exponentially modified Gaussian model .

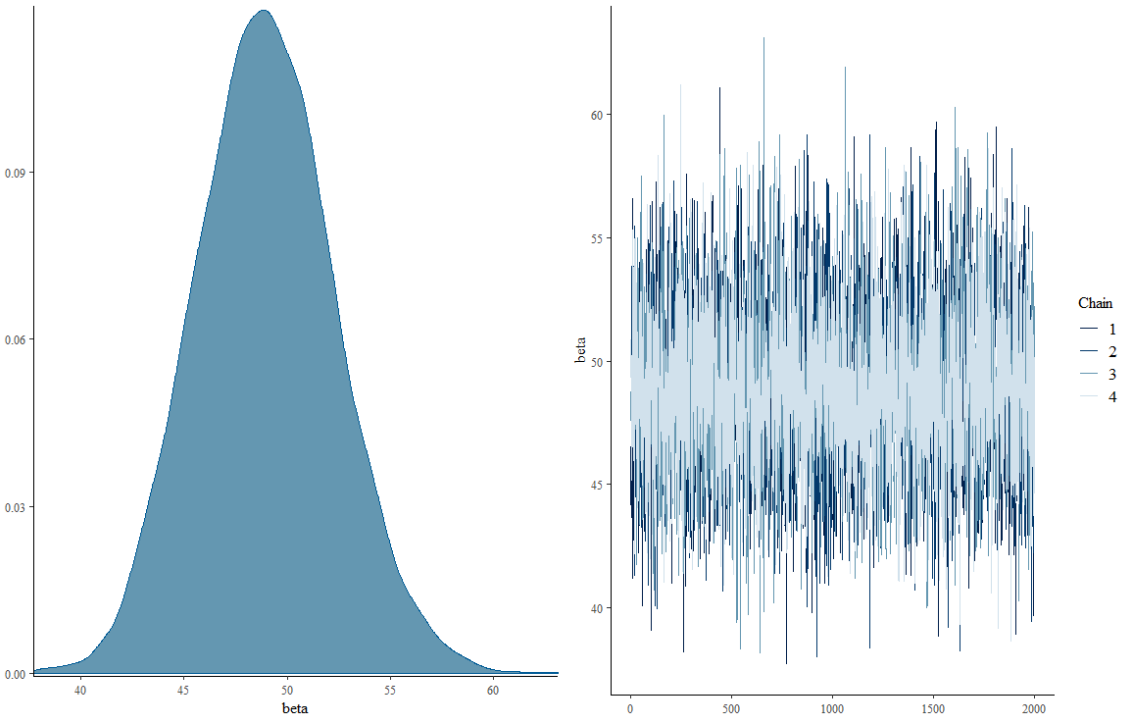


Figure C.104: Density & trace plots for the Exponentially modified Gaussian model .

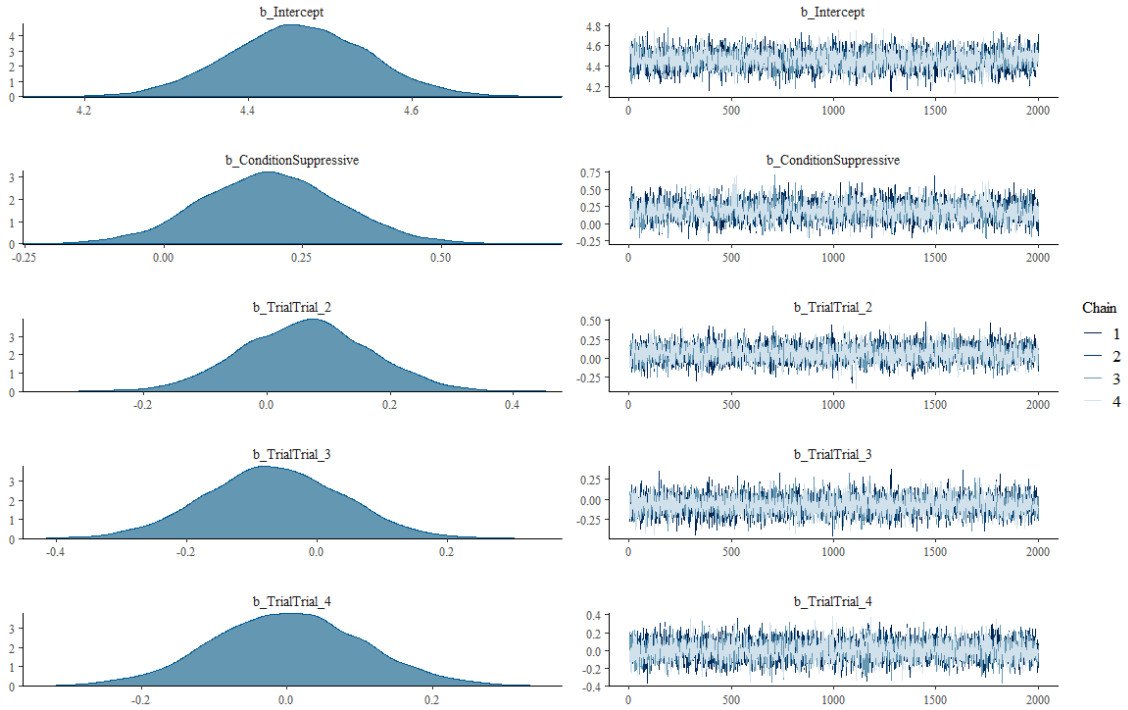


Figure C.105: Density & trace plots for the Gamma model .

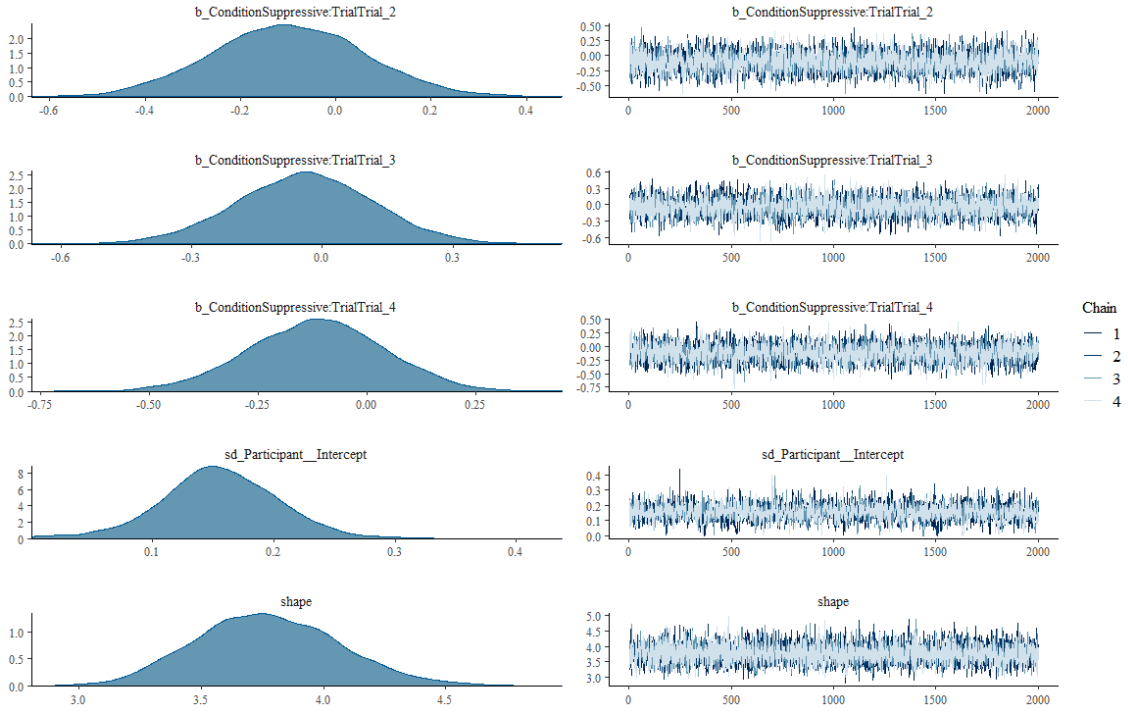


Figure C.106: Density & trace plots for the Gamma model .

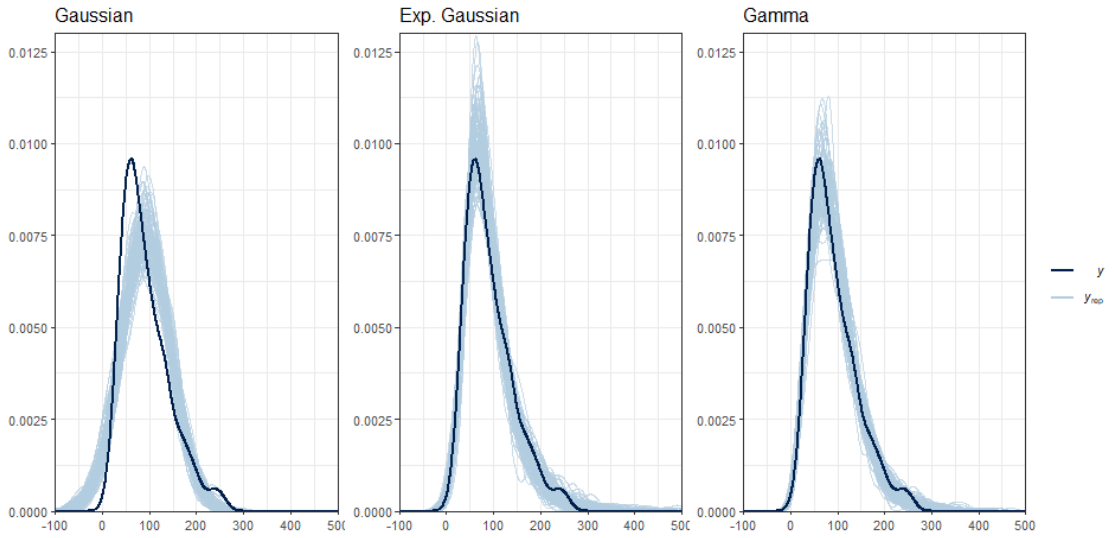


Figure C.107: Predictive posterior checks for Gaussian, exp. Gaussian, and Gamma models (Time on the x axis).

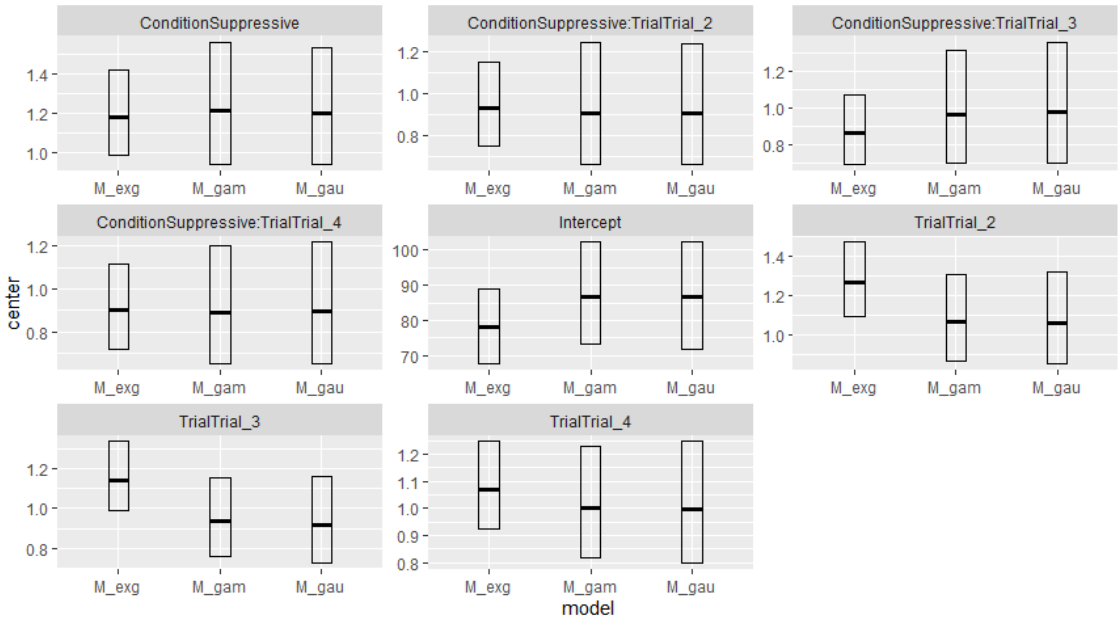


Figure C.108: Center & 95 % HDI bar plots for Gaussian, exp. Gaussian, and Gamma models.

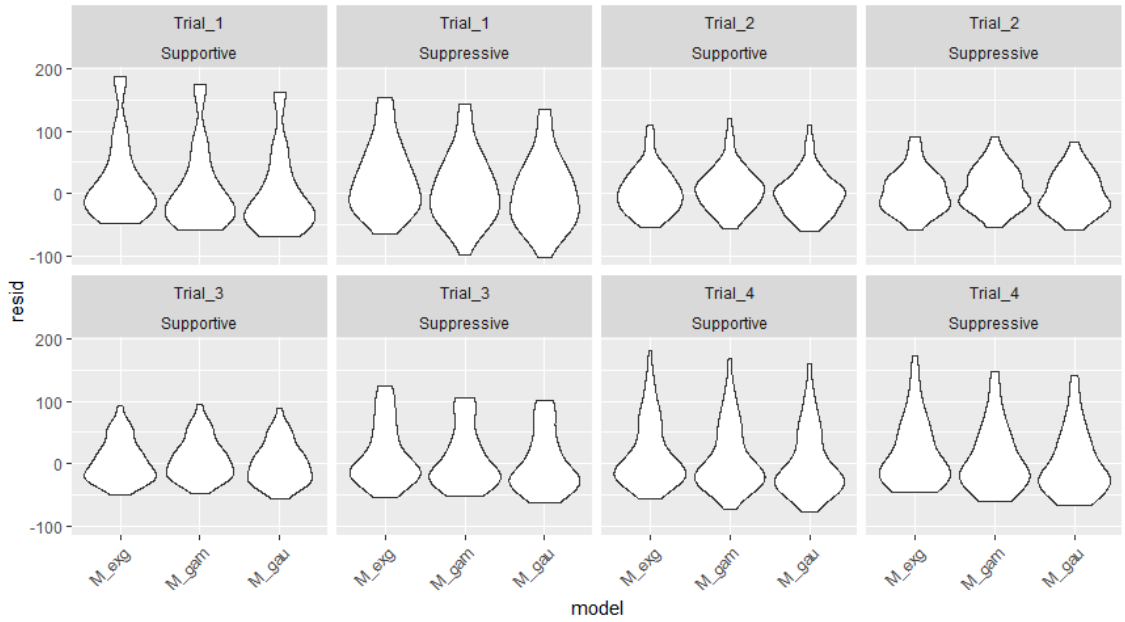


Figure C.109: Violin plots of the residuals for Gaussian, exp. Gaussian, and Gamma models.

D Appendix: Physiological Data

This appendix includes time series plots of raw and standardized physiological data, namely skin conductance, heart rate and pupil dilation. Raw data is meant to compare groups directly against each other, while standardized data (here Z scores) are meant to show changes over time, i.e. over the various peg and hole trials. Pupil dilation has been recorded via the HTC VIVE Pro Eye whereas heart rate and skin conductance have been recorded via the E4 wristband. The features of the E4 wrist band have been generated using the `FLIRT` package [264].

Initially the goal was to use this data as a measure of cognitive workload as well as as a measure of embodiment. The data has not been included in the main analysis of the thesis as 1) no baseline recordings have been made prior to the experiments, 2) no subjective measures of workload have been recorded and 3) no threat had been inserted into the experiment to record visible spikes in physiological data that would allow an assessment of embodiment. Overall it can be observed that are no significant differences over time except in the skin conductance data (mean tonic). Furthermore, an offset of approximately 10 bpm can be observed in the raw heart rate data between the suppressive and supportive group, however as this data appears to be relatively constant over time, this offset is likely due to individual differences of participants. In the standardized data, no significant changes within the trials can be observed.

D.1 Skin Conductance

It can be observed that first the supportive group has a higher SCR level, however the embodiment groups essentially switch places of the course of the 4 trials.

D.2 Heart Rate

It appears that heart rate of suppressive stays relative constant while it slightly decreases for supportive towards the last trial.

D.3 Pupil Dilation

Pupil dilation has been recorded via the HMD and the SRanpipal SDK. The calibration procedure in [265] has been remade in Unity to compensate for any incoming light affecting pupil dilation, however the resulting calibration did not lead to a linear relationship between pupil size and measured light, therefore the raw pupil dilation values have been plotted below. Blinks have been removed and linearly interpolated.

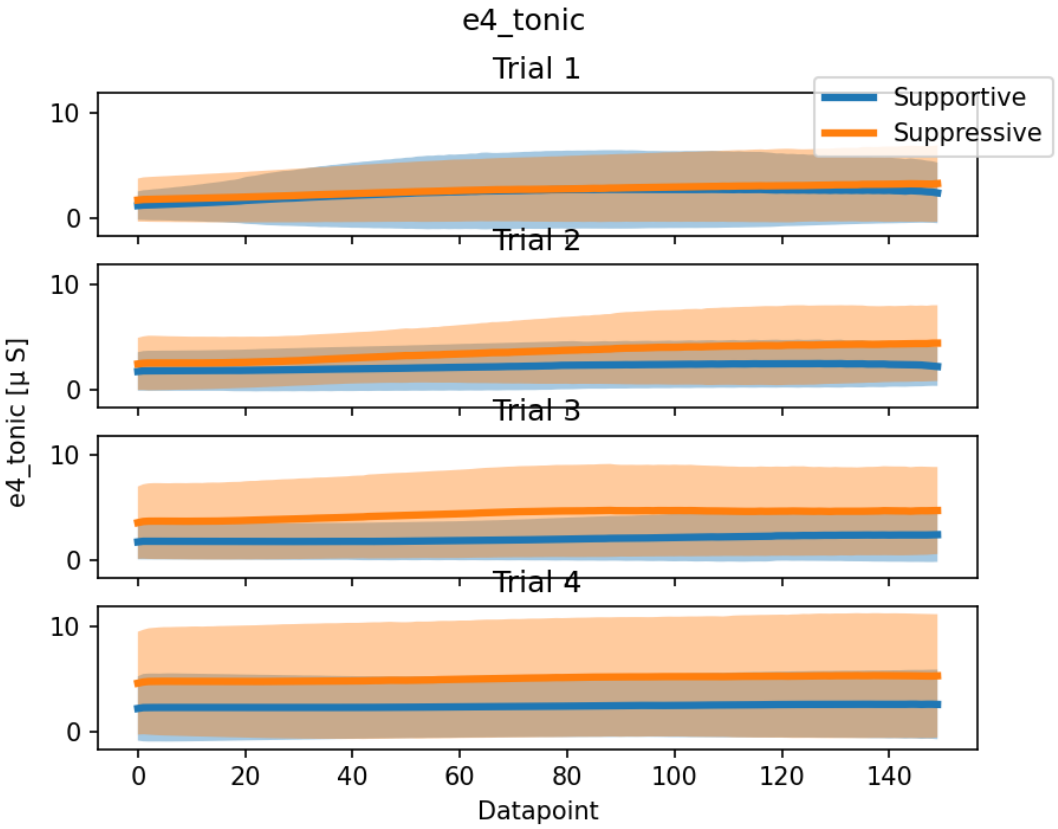


Figure D.1: Raw tonic data.

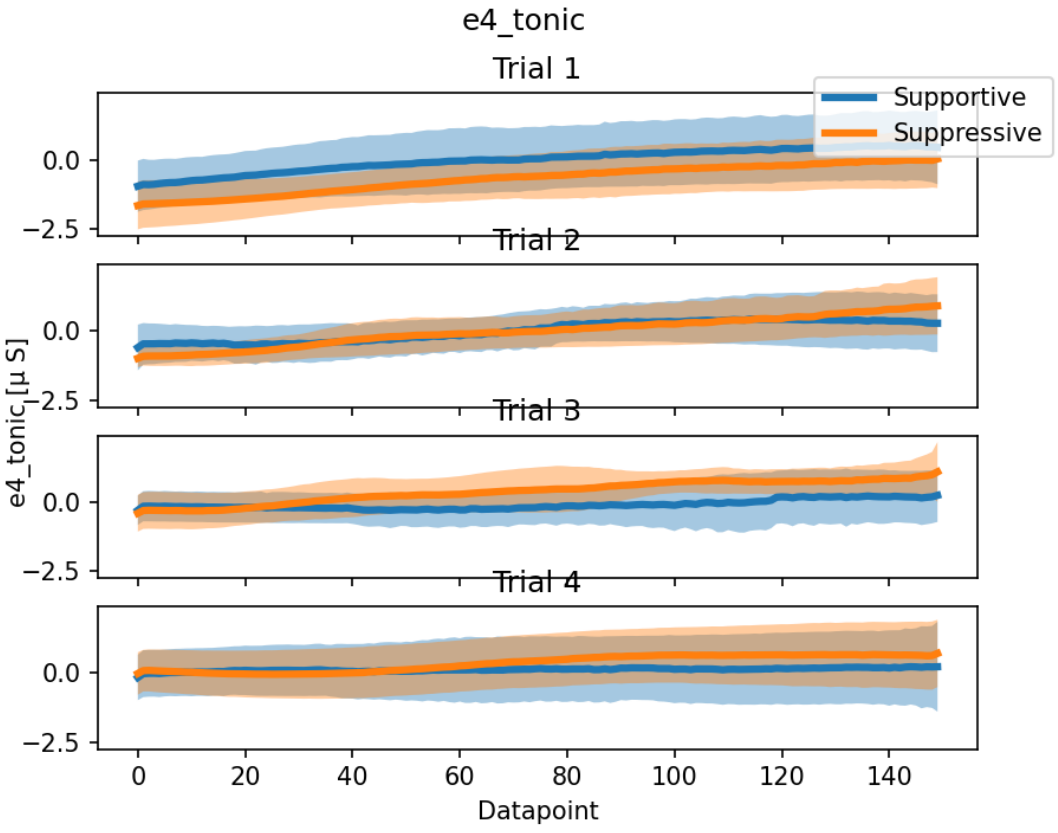


Figure D.2: Z-scores of tonic data.

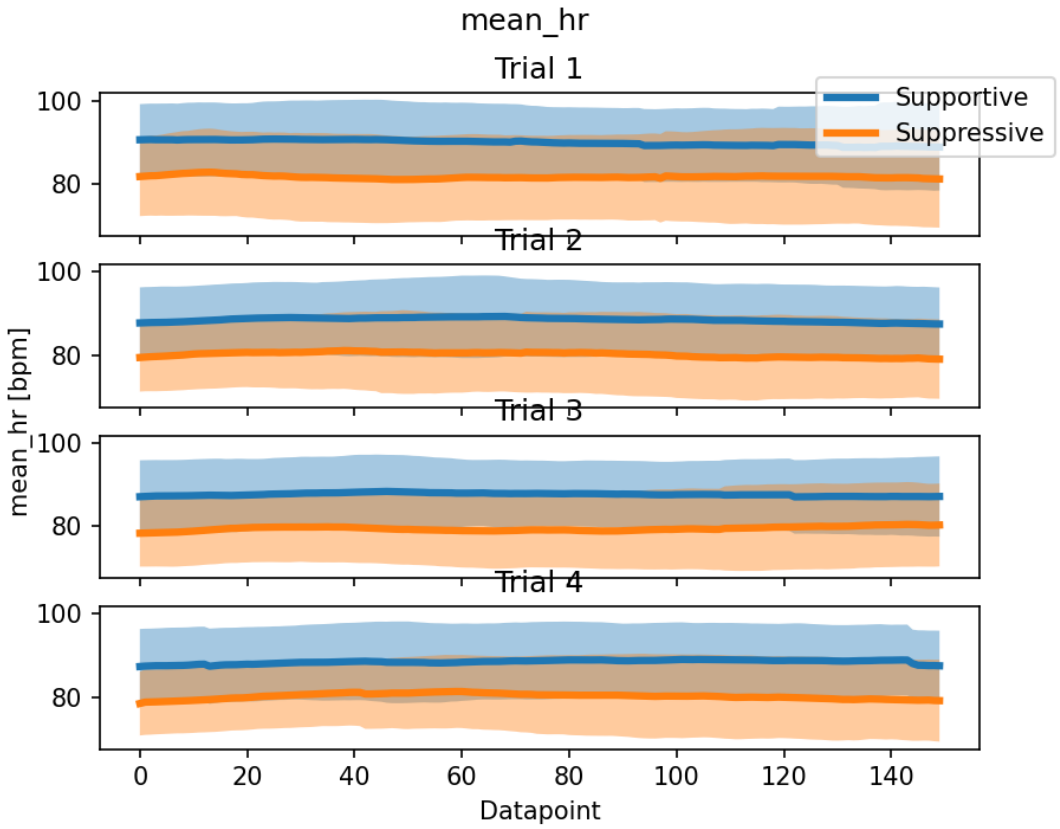


Figure D.3: Raw heart rate data.

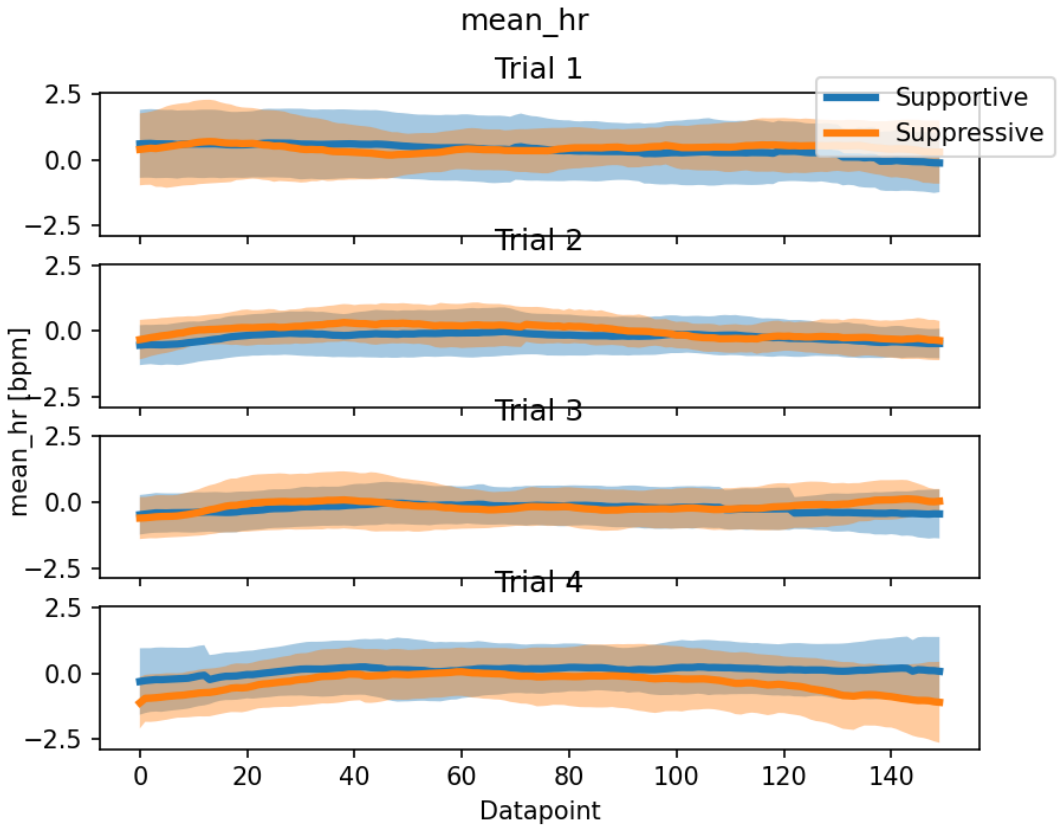


Figure D.4: Z-scores of heart rate data.

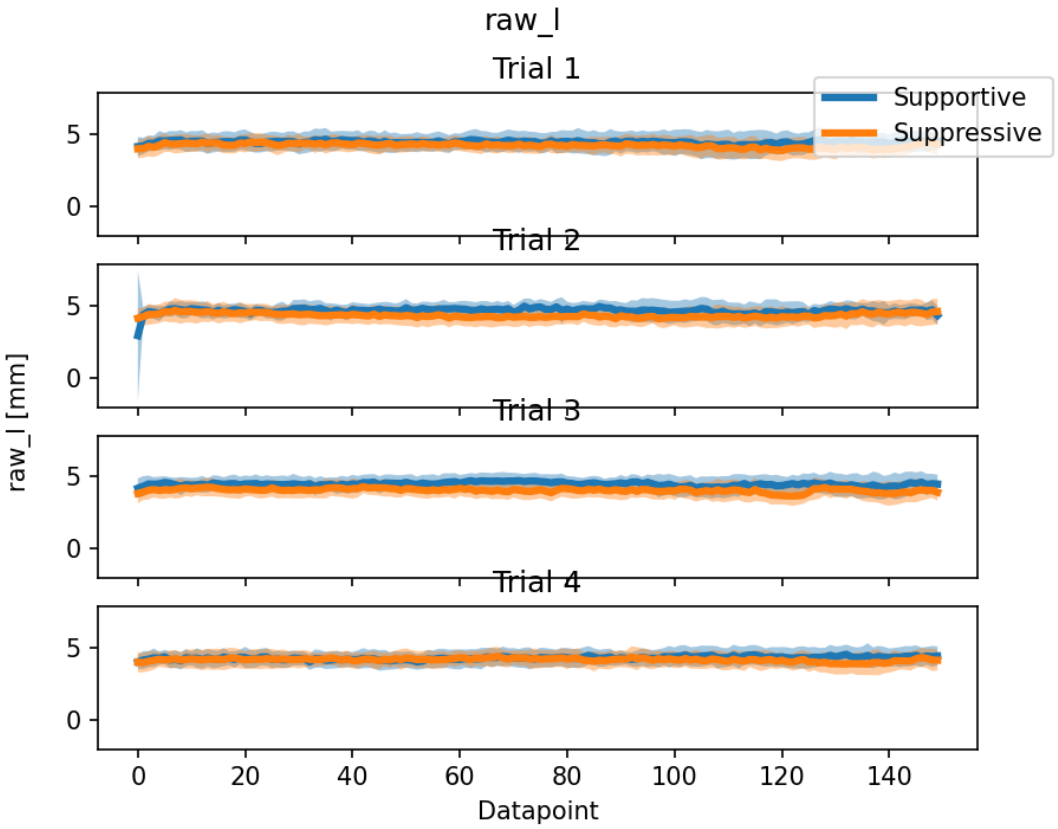


Figure D.5: Raw pupil dilation data (left eye).

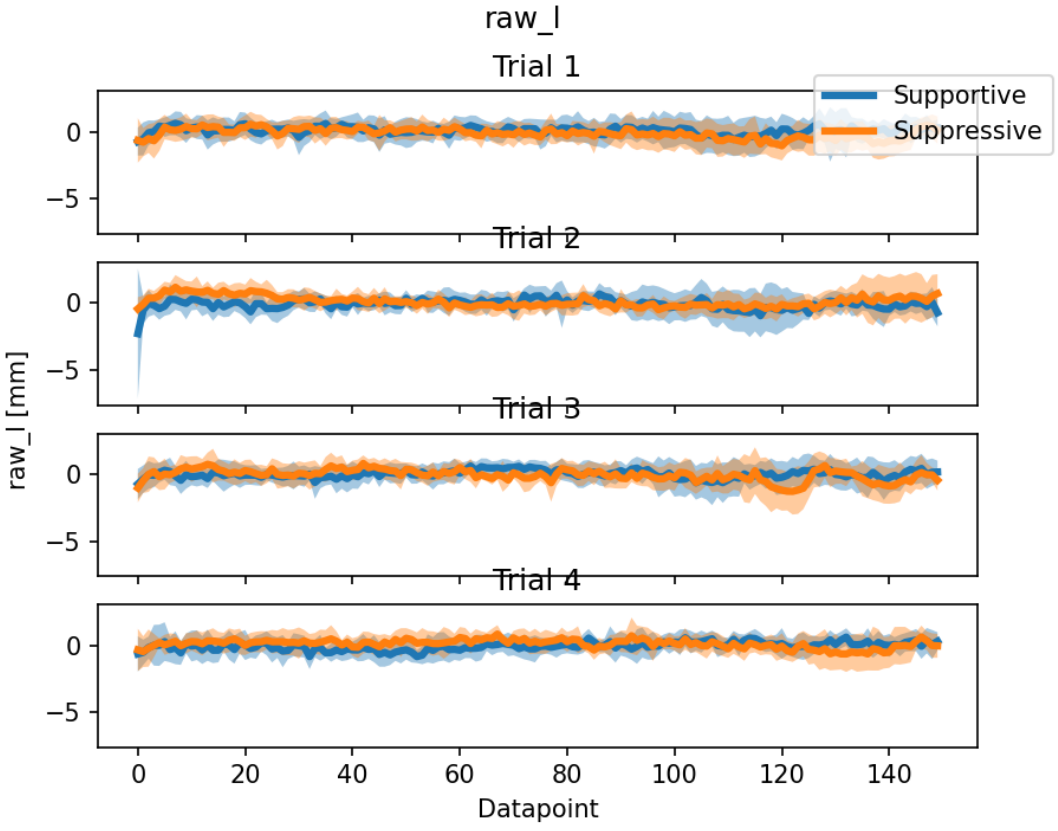


Figure D.6: Z-scores of pupil dilation data (left eye).

E Appendix: Correlations

This section examines the correlations between various SoE components, as well as the correlation between SoE components and proprioceptive drift and task performance measures. To visualize the correlations, a scatter plot combined with a regression line has been created for each variable pair. Additionally, the distributions of the embodiment groups are also added to the plot.

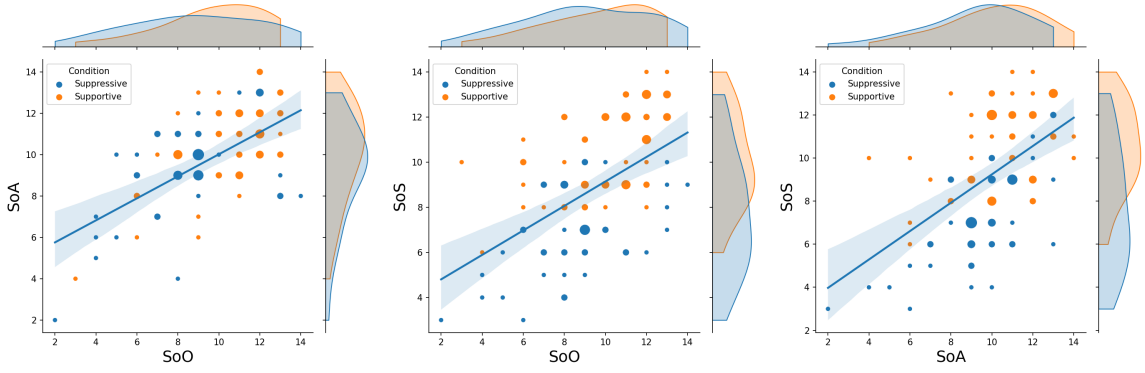
To calculate the correlations, the scores of the items of each SoE scale have been summed (where the control item has been inverted). Furthermore, the data points of proprioceptive drift and time per peg have been averaged to obtain a single number respectively, so it can be correlated to any other variable of interest. Note that for the averaged data, potential outliers have been removed based on the interquartile range (IQR). The correlation used in this analysis is the Shepherd's Pi correlation [209], which is a more robust version of the Spearman's Rho correlation, that removes outliers by bootstrapping the Mahalanobis distance of each observation from the bivariate mean and excluding points that are greater than a certain threshold. To asses the values of the resulting correlation r, [266] has been used as guideline with the following naming conventions as shown in table E.1. If the the correlation value was between two categories, e.g. 0.34, then it has been rounded up or down to the first decimal, depending on to which category it was closer.

r	Interpretation
1	Perfect
0.7 - 0.9	Strong
0.6 - 0.4	Moderate
0.1 - 0.3	Weak
0	None

Table E.1: Correlation values and interpretation. Note that these hold for negative and positive correlation.

E.1 Correlations between SoE Components

To analyze the the correlations between the individual SoE components correlation plots have been generated in figures E.1.



(a) SoA & SoO **(b)** SoS & SoO **(c)** SoS & SoA

Figure E.1: Correlations plots between SoE components.

For every combination of the SoE components, a positive correlation can be observed with a similar slope. Below in table E.2, the correlation r between the SoE components can be found,

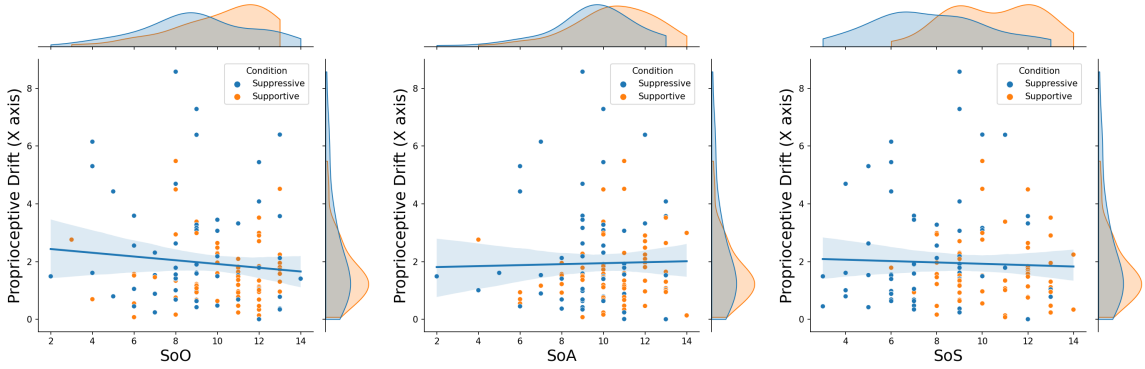
as well as the number of removed outliers n, the p-value p and corresponding power. As can be seen, for all SoE components a significant, moderate correlation can be found.

Variable 1	Variable 2	n	r	95 % CI	p	Power
SoO	SoA	8	0.576	[0.44 0.69]	0.0	1.0
SoO	SoS	3	0.609	[0.48 0.71]	0.0	1.0
SoA	SoS	7	0.624	[0.5 0.72]	0.0	1.0

Table E.2: Correlation between SoE components.

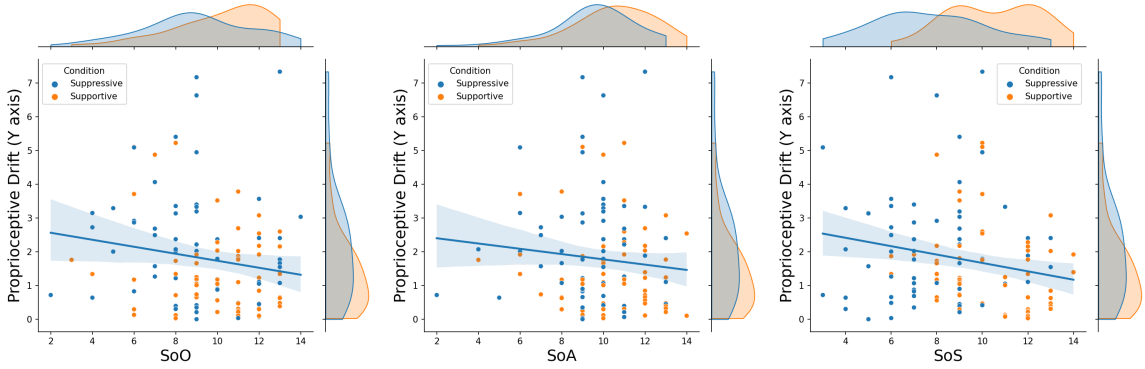
E.2 Correlations between SoE Components and Proprioceptive Drift

To analyze the the correlations between the individual SoE components and proprioceptive drift (PD), correlation plots have been generated in figures E.2, E.3 and E.4.



(a) PD (X) & SoO **(b)** PD (X) & SoA **(c)** PD (X) & SoA

Figure E.2: Correlations plots between SoE components and Proprioceptive drift (X axis).

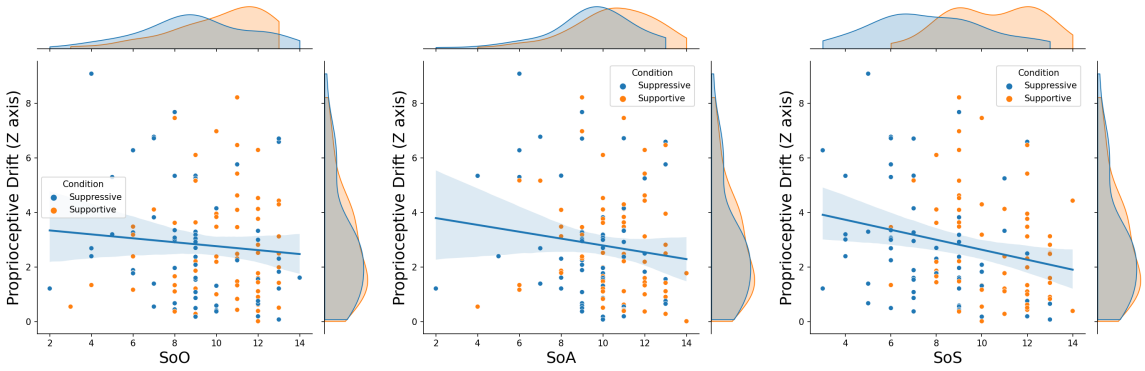


(a) PD (Y) & SoO **(b)** PD (Y) & SoA **(c)** PD (Y) & SoA

Figure E.3: Correlations plots between SoE components and Proprioceptive drift (Y axis).

As can be seen, barely any correlations can be found on the X axis, some correlations on the Y axis, and "some more" on the Z axis. Below in table E.2, the correlation values between the SoE components and proprioceptive drift axes can be found.

The only significant correlations are on the Y axis and the Z axis. On the Y axis, there was a weak, negative correlation can be found with SoO and SoS, whereas on the Z axis, weak, negative correlations can be found for SoA and SoS. Note however, that the CI's include values very close to 0. Moreover, the power for all of these correlations is smaller than 0.8, with exception of the correlation between SoS and PD (Z).



(a) PD (Z) & SoO **(b)** PD (Z) & SoA **(c)** PD (Z) & SoS

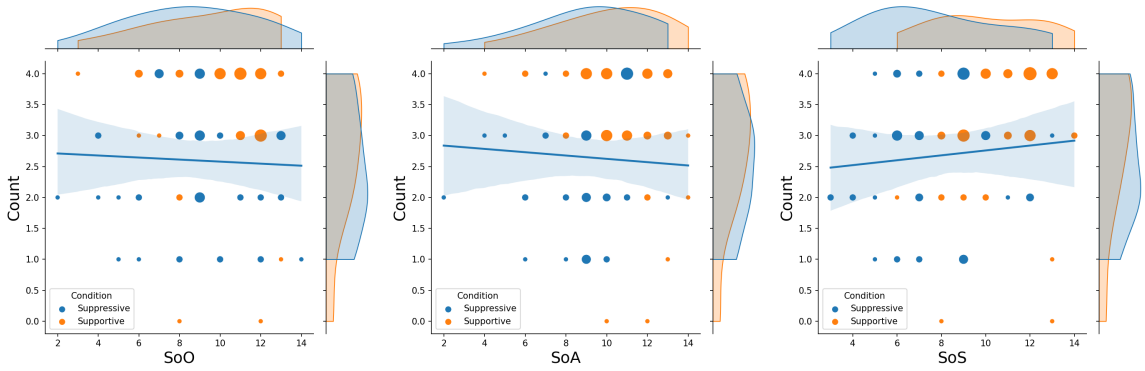
Figure E.4: Correlations plots between SoE components and Proprioceptive drift (Z axis).

Variable 1	Variable 2	n	r	95 % CI	p	Power
PD (X)	SoO	9	-0.006	[-0.19 0.18]	0.951	0.05
PD (X)	SoA	10	0.142	[-0.04 0.32]	0.133	0.325
PD (X)	SoS	9	0.057	[-0.13 0.24]	0.543	0.093
PD (Y)	SoO	7	-0.223	[-0.39 -0.04]	0.016	0.68
PD (Y)	SoA	8	-0.139	[-0.31 0.04]	0.136	0.321
PD (Y)	SoS	6	-0.22	[-0.39 -0.04]	0.017	0.674
PD (Z)	SoO	7	-0.152	[-0.32 0.03]	0.101	0.376
PD (Z)	SoA	9	-0.164	[-0.34 0.02]	0.079	0.421
PD (Z)	SoS	7	-0.258	[-0.42 -0.08]	0.005	0.808

Table E.3: Correlation between SoE components and proprioceptive drift.

E.3 Correlations between SoE Components and Task Performance measures

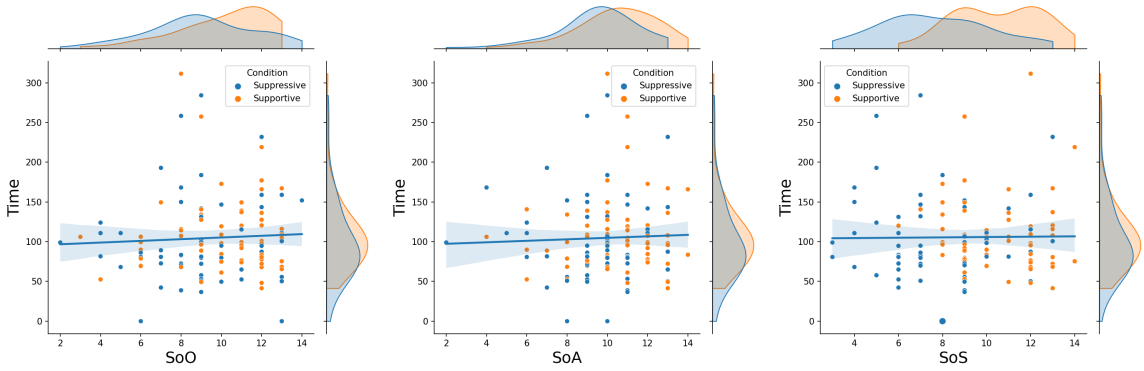
To analyze the the correlations between the individual SoE components and task performance measures, correlation plots have been generated in figures E.5 and E.6.



(a) Count & SoO **(b)** Count & SoA **(c)** Count & SoS

Figure E.5: Correlations plots between SoE components and task performance (count).

For the count variable, no correlation can be observed with SoO, a slight correlation with SoA and "some" correlation with SoS. For the time variable, a negative correlation can be observed with SoO, no correlation with SoA and some negative correlation with SoS. Below in table E.2, the correlation values between the SoE components and task performance measures can be found.



(a) Time & SoO **(b)** Time & SoA **(c)** Time & SoS

Figure E.6: Correlations plots between SoE components and task performance (count).

Variable 1	Variable 2	n	r	95 % CI	p	Power
Count	SoO	6	-0.014	[-0.19 0.17]	0.878	0.053
Count	SoA	8	0.161	[-0.02 0.33]	0.084	0.41
Count	SoS	3	0.324	[0.15 0.48]	0.0	0.956
Time	SoO	0	0.103	[-0.08 0.28]	0.277	0.193
Time	SoA	8	0.087	[-0.1 0.27]	0.351	0.154
Time	SoS	6	-0.0	[-0.18 0.18]	0.999	0.05

Table E.4: Correlation between SoE components and task performance measures.

As can be seen, only for the count variable and SoS a significant correlation can be found, which is in this case a weak correlation.

Bibliography

[1] T. B. Sheridan, "MIT research in telerobotics," in *Proc. Workshop on Space Telerobotics*, vol. 2, pp. 403–412, 1987.

[2] S. Falcone, "Design challenges to achieve the sense of embodiment while developing a teleoperation system," in *First international workshop on Designerly HRI Knowledge*, 2020.

[3] M. Sanchez-Vives and M. Slater, "From presence to consciousness through virtual reality," *Nature reviews. Neuroscience*, vol. 6, pp. 332–9, 05 2005.

[4] S. Falcone, A.-M. Brouwer, I. Cocu, K. Gijsbertse, D. Heylen, and J. Erp, "The relative contribution of five key perceptual cues and their interaction to the sense of embodiment.," *Technology, Mind, and Behavior*, vol. 3, 03 2022.

[5] N. Rublein, "Robust and transparent interaction in tele-robotics in the presence of time delays." Student Report, University of Twente, 2020.

[6] A. Toet, I. A. Kuling, B. N. Krom, and J. B. F. van Erp, "Toward enhanced teleoperation through embodiment," *Frontiers in Robotics and AI*, vol. 7, p. 14, 2020.

[7] B. Hannaford and J. Ryu, "Time-domain passivity control of haptic interfaces," *IEEE Transactions on Robotics and Automation*, vol. 18, no. 1, pp. 1–10, 2002.

[8] M. Franken, S. Stramigioli, S. Misra, C. Secchi, and A. Macchelli, "Bilateral telemanipulation with time delays: A two-layer approach combining passivity and transparency," *IEEE Transactions on Robotics*, vol. 27, no. 4, pp. 741–756, 2011.

[9] X. Xu, B. Cizmeci, A. Al-Nuaimi, and E. Steinbach, "Point cloud-based model-mediated teleoperation with dynamic and perception-based model updating," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 11, pp. 2558–2569, 2014.

[10] K. Kilteni, R. Groten, and M. Slater, "The sense of embodiment in virtual reality," vol. 21, p. 373–387, Dec. 2012.

[11] H. H. Ehrsson, C. Spence, and R. E. Passingham, "That's my hand! activity in premotor cortex reflects feeling of ownership of a limb," *Science*, vol. 305, no. 5685, pp. 875–877, 2004.

[12] H. H. Ehrsson, N. P. Holmes, and R. E. Passingham, "Touching a rubber hand: Feeling of body ownership is associated with activity in multisensory brain areas," *The Journal of Neuroscience*, vol. 25, pp. 10564 – 10573, 2005.

[13] M. Tsakiris, M. D. Hesse, C. Boy, P. Haggard, and G. R. Fink, "Neural Signatures of Body Ownership: A Sensory Network for Bodily Self-Consciousness," *Cerebral Cortex*, vol. 17, pp. 2235–2244, 11 2006.

[14] M. Schiefer, D. Tan, S. M. Sidek, and D. J. Tyler, "Sensory feedback by peripheral nerve stimulation improves task performance in individuals with upper limb loss using a myoelectric prosthesis," *Journal of Neural Engineering*, vol. 13, p. 016001, dec 2015.

[15] G. Rognini, F. M. Petrini, S. Raspopovic, G. Valle, G. Granata, I. Strauss, M. Solcà, J. Bello-Ruiz, B. Herbelin, R. Mange, E. D'Anna, R. Di Iorio, G. Di Pino, D. Andreu, D. Guiraud, T. Stieglitz, P. M. Rossini, A. Serino, S. Micera, and O. Blanke, "Multisensory bionic limb

to achieve prosthesis embodiment and reduce distorted phantom limb perceptions," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 90, no. 7, pp. 833–836, 2019.

[16] D. W. Tan, M. A. Schiefer, M. W. Keith, J. R. Anderson, J. Tyler, and D. J. Tyler, "A neural interface provides long-term stable natural touch perception," *Science Translational Medicine*, vol. 6, no. 257, pp. 257ra138–257ra138, 2014.

[17] G. Valle, A. Mazzoni, F. Iberite, E. D'Anna, I. Strauss, G. Granata, M. Controzzi, F. Clemente, G. Rognini, C. Cipriani, T. Stieglitz, F. Petrini, P. Rossini, and S. Micera, "Biomimetic intraneural sensory feedback enhances sensation naturalness, tactile sensitivity, and manual dexterity in a bidirectional prosthesis," *Neuron*, vol. 100, 09 2018.

[18] P. D. Marasco, J. S. Hebert, J. W. Sensinger, C. E. Shell, J. S. Schofield, Z. C. Thumser, R. Nataraj, D. T. Beckler, M. R. Dawson, D. H. Blustein, S. Gill, B. D. Mensh, R. Granja-Vazquez, M. D. Newcomb, J. P. Carey, and B. M. Orzell, "Illusory movement perception improves motor control for prosthetic hands," *Science Translational Medicine*, vol. 10, no. 432, p. eaao6990, 2018.

[19] K. Grechuta, J. Guga, G. Maffei, B. Ballester, and P. Verschure, "Visuotactile integration modulates motor performance in a perceptual decision-making task," *Scientific Reports*, vol. 7, 06 2017.

[20] G. Gorisse, O. Christmann, E. A. Amato, and S. Richir, "First- and third-person perspectives in immersive virtual environments: Presence and performance analysis of embodied users," *Frontiers in Robotics and AI*, vol. 4, 2017.

[21] K. Friston and S. Kiebel, "Predictive coding under the free-energy principle," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1521, pp. 1211–1221, 2009.

[22] K. Friston, "Prediction, perception and agency," *International Journal of Psychophysiology*, vol. 83, no. 2, pp. 248–252, 2012. Predictive information processing in the brain: Principles, neural mechanisms and models.

[23] J. Hohwy, *The Predictive Mind*. United Kingdom: Oxford University Press, 2013.

[24] N. Rublein, "Research topics: Embodiment and task performance in telerobotics," 2022.

[25] M. Gonzalez-Franco and T. C. Peck, "Avatar embodiment. towards a standardized questionnaire," *Frontiers in Robotics and AI*, vol. 5, p. 74, 2018.

[26] T. C. Peck and M. Gonzalez-Franco, "Avatar embodiment. a standardized questionnaire," *Frontiers in Virtual Reality*, vol. 1, 2021.

[27] S. Falcone, G. Englebienne, J. van Erp, and D. Heylen, "Toward standard guidelines to design the sense of embodiment in teleoperation applications: A review and toolbox," *Human-computer interaction*, Mar. 2022.

[28] K. Lee, "Presence, explicated," *Communication Theory*, vol. 14, pp. 27 – 50, 02 2004.

[29] N. Nostadt, D. Abbink, O. Christ, and P. Beckerle, "Embodiment, presence, and their intersections: Teleoperation and beyond," *ACM Transactions on Human-Robot Interaction*, vol. 9, 05 2020.

[30] L. Almeida, P. Menezes, and J. Dias, "Interface transparency issues in teleoperation," *Applied Sciences*, vol. 10, 09 2020.

[31] S. C. Pritchard, R. Zopf, V. Polito, D. M. Kaplan, and M. A. Williams, "Non-hierarchical influence of visual form, touch, and position cues on embodiment, agency, and presence in virtual reality," *Frontiers in Psychology*, vol. 7, 2016.

[32] M. Riemer, J. Trojan, M. Beauchamp, and X. Fuchs, "The rubber hand universe: On the impact of methodological differences in the rubber hand illusion," *Neuroscience Biobehavioral Reviews*, vol. 104, 07 2019.

[33] M. Botvinick and J. Cohen, "Rubber hands feel touch that eyes see," *Nature*, vol. 391, pp. 756–756, 1998.

[34] A. Kalckert and H. Ehrsson, "Moving a rubber hand that feels like your own: A dissociation of ownership and agency," *Frontiers in Human Neuroscience*, vol. 6, 2012.

[35] B. N. Krom, M. Catoire, A. Toet, R. J. E. van Dijk, and J. B. van Erp, "Effects of likeness and synchronicity on the ownership illusion over a moving virtual robotic arm and hand," in *2019 IEEE World Haptics Conference (WHC)*, pp. 49–54, 2019.

[36] D. Romano, E. Caffa, A. Hernandez-Arieta, P. Brugger, and A. Maravita, "The robot hand illusion: Inducing proprioceptive drift through visuo-motor congruency," *Neuropsychologia*, vol. 70, pp. 414–420, 2015.

[37] M. Riemer, D. Kleinboehl, R. Hoelzl, and J. Trojan, "Action and perception in the rubber hand illusion," *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, vol. 229, 01 2013.

[38] "What is embodiment? a psychometric approach," *Cognition*, vol. 107, no. 3, pp. 978–998, 2008.

[39] J. Zbinden, E. Lendaro, and M. Ortiz-Catalan, "Prosthetic embodiment: systematic review on definitions, measures, and experimental paradigms," *Journal of NeuroEngineering and Rehabilitation*, vol. 19, no. 1, pp. 1–16, 2022.

[40] N. David, B. Bewernick, M. Cohen, A. Newen, S. Lux, G. Fink, N. Shah, and K. Vogeley, "Neural representations of self versus other: Visual-spatial perspective taking and agency in a virtual ball-tossing game," *Journal of cognitive neuroscience*, vol. 18, pp. 898–910, 07 2006.

[41] H. Debarba, S. Bovet, R. Salomon, O. Blanke, B. Herbelin, and R. Boulic, "Characterizing first and third person viewpoints and their alternation for embodied interaction in virtual reality," *PLOS ONE*, vol. 12, p. e0190109, 12 2017.

[42] M. Matamala-Gomez, E. Brivio, A. Chirico, C. Malighetti, O. Realdon, S. Serino, A. Dakanalis, G. Corno, N. Polli, C. Cacciatore, G. Riva, and F. Mantovani, "Assessing user experience of a virtual reality training in patients with anorexia nervosa: insigths from a pilot study," *Annual Review of CyberTherapy and Telemedicine*, 05 2021.

[43] A. Krekhov, S. Cmentowski, and J. Krüger, "Vr animals: Surreal body ownership in virtual reality games," pp. 503–511, 10 2018.

[44] H. H. Ehrsson, "The experimental induction of out-of-body experiences," *Science*, vol. 317, pp. 1048 – 1048, 2007.

[45] K. Baxter, C. Courage, and K. Caine, *Understanding Your Users: A Practical Guide to User Research Methods*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2 ed., 2015.

[46] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[47] H. Ziadeh, D. Gulyas, L. D. Nielsen, S. Lehmann, T. B. Nielsen, T. K. K. Kjeldsen, B. I. Hougaard, M. Jochumsen, and H. Knoche, ""mine works better": Examining the influence of embodiment in virtual reality on the sense of agency during a binary motor imagery task with a brain-computer interface," *Frontiers in Psychology*, vol. 12, 2021.

[48] J. L. Smith, "Semi-structured interviewing and qualitative analysis," 1995.

[49] E. Lewis and D. Lloyd, "Embodied experience: A first-person investigation of the rubber hand illusion," *Phenomenology and the Cognitive Sciences*, vol. 9, pp. 317–339, 09 2010.

[50] S. Gallagher, "How the body shapes the mind," *How the Body Shapes the Mind*, 01 2005.

[51] P. Haggard and D. M. Wolpert, "Disorders of body scheme," in *In Freund, H.J., Jeannerod, M., Hallett, M., Leiguarda R., (Eds.), Higher-Order Motor Disorders*, University Press, 2005.

[52] M. Tsakiris and P. Haggard, "The rubber hand illusion revisited: visuotactile integration and self-attribution.," *Journal of experimental psychology. Human perception and performance*, vol. 31 1, pp. 80–91, 2005.

[53] M. P. Kammers, F. de Vignemont, L. Verhagen, and H. C. Dijkerman, "The rubber hand illusion in action," *Neuropsychologia*, vol. 47, no. 1, pp. 204–211, 2009.

[54] A. Kalckert and H. H. Ehrsson, "The spatial distance rule in the moving and classical rubber hand illusions," *Consciousness and Cognition*, vol. 30, pp. 118–132, 2014.

[55] R. Zopf, S. Truong, M. Finkbeiner, J. Friedman, and M. A. Williams, "Viewing and feeling touch modulates hand position for reaching," *Neuropsychologia*, vol. 49, no. 5, pp. 1287–1293, 2011.

[56] N. P. Holmes, H. J. Snijders, and C. Spence, "Reaching with alien limbs: Visual exposure to prosthetic hands in a mirror biases proprioception without accompanying illusions of ownership," *Perception & psychophysics*, vol. 68, no. 4, pp. 685–701, 2006.

[57] T. Dummer, A. Picot-Annand, T. Neal, and C. Moore, "Movement and the rubber hand illusion," *Perception*, vol. 38, no. 2, pp. 271–280, 2009. PMID: 19400435.

[58] G. Tieri, E. Tidoni, E. Pavone, and S. Aglioti, "Body visual discontinuity affects feeling of ownership and skin conductance responses," *Scientific Reports*, vol. 5, p. 17139, 11 2015.

[59] G. Tieri, A. Gioia, M. Scandola, E. Pavone, and S. Aglioti, "Visual appearance of a virtual upper limb modulates the temperature of the real hand: a thermal imaging study in immersive virtual reality," *European Journal of Neuroscience*, vol. 45, 2017.

[60] M. Slater, B. Spanlang, M. Sanchez-Vives, and O. Blanke, "First person experience of body transfer in virtual reality," *PloS one*, vol. 5, p. e10564, 05 2010.

[61] S. Falcone, A.-M. Brouwer, D. Heylen, J. Van Erp, L. Zhang, S. S. Pradhan, I. V. Stuldreher, I. Cocu, M. Heuvel, P. S. de Vries, *et al.*, "Pupil diameter as implicit measure to estimate sense of embodiment," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 44, 2022.

[62] H. Ehrsson, K. Wiech, N. Weiskopf, R. Dolan, and R. Passingham, "Threatening a rubber hand that you feel is yours elicits a cortical anxiety response," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 9828–33, 07 2007.

[63] V. Petkova and H. Ehrsson, "If i were you: Perceptual illusion of body swapping," *PloS one*, vol. 3, p. e3832, 02 2008.

[64] A. Haans, F. G. Kaiser, D. G. Bouwhuis, and W. A. IJsselsteijn, "Individual differences in the rubber-hand illusion: Predicting self-reports of people's personal experiences," *Acta Psychologica*, vol. 141, no. 2, pp. 169–177, 2012.

[65] E. Capelari, C. Uribe, and J. Brasil-Neto, "Feeling pain in the rubber hand: Integration of visual, proprioceptive, and painful stimuli," *Perception*, vol. 38, pp. 92–9, 02 2009.

[66] G. K. Essick, F. McGlone, C. Dancer, D. Fabricant, Y. Ragin, N. Phillips, T. Jones, and S. Guest, "Quantitative assessment of pleasant touch," *Neuroscience Biobehavioral Reviews*, vol. 34, no. 2, pp. 192–203, 2010. Touch, Temperature, Pain/Itch and Pleasure.

[67] M. L. Filippetti, L. P. Kirsch, L. Crucianelli, and A. Fotopoulou, "Affective certainty and congruency of touch modulate the experience of the rubber hand illusion," *Scientific Reports*, vol. 9, 2019.

[68] D. Lloyd, V. Gillis, E. Lewis, M. Farrell, and I. Morrison, "Pleasant touch moderates the subjective but not objective aspects of body perception," *Frontiers in behavioral neuroscience*, vol. 7, p. 207, 12 2013.

[69] H. E. van Stralen, M. J. van Zandvoort, S. S. Hoppenbrouwers, L. M. Vissers, L. J. Kappelle, and H. C. Dijkerman, "Affective touch modulates the rubber hand illusion," *Cognition*, vol. 131, no. 1, pp. 147–158, 2014.

[70] J. Ward, A. Mensah, and K. Jünemann, "The rubber hand illusion depends on the tactile congruency of the observed and felt touch," *Journal of experimental psychology. Human perception and performance*, vol. 41, 07 2015.

[71] R. C. White, A. M. A. Davies, T. J. Halleen, and M. Davies, "Tactile expectations and the perception of self-touch: An investigation using the rubber hand paradigm," *Consciousness and Cognition*, vol. 19, no. 2, pp. 505–519, 2010.

[72] J. R. de Jong, A. Keizer, M. M. Engel, and H. C. Dijkerman, "Does affective touch influence the virtual reality full body illusion?," *Experimental Brain Research*, vol. 235, pp. 1781 – 1791, 2017.

[73] L. Aymerich-Franch, D. Petit, G. Ganesh, and A. Kheddar, "Non-human looking robot arms induce illusion of embodiment," *International Journal of Social Robotics*, vol. 9, 09 2017.

[74] T. V. Huynh, R. Bekrater-Bodmann, J. Fröhner, J. Vogt, and P. Beckerle, "Robotic hand illusion with tactile feedback: Unravelling the relative contribution of visuotactile and visuomotor input to the representation of body parts in space," *PLOS ONE*, vol. 14, pp. 1–20, 01 2019.

[75] C. Armel and V. Ramachandran, "Projecting sensations to external objects: Evidence from skin conductance response," *Proceedings. Biological sciences / The Royal Society*, vol. 270, pp. 1499–506, 08 2003.

[76] M. Slater, D. Pérez Marcos, H. Ehrsson, and M. Sanchez-Vives, "Towards a digital body: the virtual arm illusion," *Frontiers in Human Neuroscience*, vol. 2, 2008.

[77] A. Maselli and M. Slater, "The building blocks of the full body ownership illusion," *Frontiers in Human Neuroscience*, vol. 7, p. 83, 2013.

[78] A. Keizer, A. Elburg, R. Helms, and C. Dijkerman, "A virtual reality full body illusion improves body image disturbance in anorexia nervosa," *PLOS ONE*, vol. 11, p. e0163921, 10 2016.

[79] M. Alimardani, S. Nishio, and H. Ishiguro, "Humanlike robot hands controlled by brain activity arouse illusion of ownership in operators," *Scientific reports*, vol. 3, p. 2396, 08 2013.

[80] M. Alimardani, S. Nishio, and H. Ishiguro, "BCI-teleoperated androids; a study of embodiment and its effect on motor imagery learning," in *2015 IEEE 19th International Conference on Intelligent Engineering Systems (INES)*, pp. 347–352, 2015.

[81] W. Steptoe, A. Steed, and M. Slater, "Human tails: Ownership and control of extended humanoid avatars," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 4, pp. 583–590, 2013.

[82] G. Rognini, A. Sengül, J. E. Aspell, R. Salomon, H. Bleuler, and O. Blanke, "Visuo-tactile integration and body ownership during self-generated action," *European Journal of Neuroscience*, vol. 37, no. 7, pp. 1120–1129, 2013.

[83] M. V. Sanchez-Vives, B. Spanlang, A. Frisoli, M. Bergamasco, and M. Slater, "Virtual hand illusion induced by visuomotor correlations," *PLOS ONE*, vol. 5, pp. 1–6, 04 2010.

[84] E. Kokkinara and M. Slater, "Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion," *Perception*, vol. 43, no. 1, pp. 43–58, 2014.

[85] K. Kilteni, A. Maselli, K. P. Kording, and M. Slater, "Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception," *Frontiers in Human Neuroscience*, vol. 9, 2015.

[86] M. Tsakiris, G. Prabhu, and P. Haggard, "Having a body versus moving your body: How agency structures body-ownership," *Consciousness and Cognition*, vol. 15, no. 2, pp. 423–432, 2006.

[87] M. Rohde, M. Di Luca, and M. O. Ernst, "The rubber hand illusion: Feeling of ownership and proprioceptive drift do not go hand in hand," *PLOS ONE*, vol. 6, pp. 1–9, 06 2011.

[88] A. Maselli and M. Slater, "Sliding perspectives: dissociating ownership from self-location during full body illusions in virtual reality," *Frontiers in Human Neuroscience*, vol. 8, p. 693, 2014.

[89] V. Petkova, M. Björnsdotter, G. Gentile, T. Jonsson, T.-Q. Li, and H. Ehrsson, "From part- to whole-body ownership in the multisensory brain," *Current Biology*, vol. 21, no. 13, pp. 1118–1122, 2011.

[90] C. Brozzoli, G. Gentile, and H. H. Ehrsson, "That's near my hand! parietal and premotor coding of hand-centered space contributes to localization and self-attribution of the hand," *Journal of Neuroscience*, vol. 32, no. 42, pp. 14573–14582, 2012.

[91] K. J. Blom, J. Arroyo-Palacios, and M. Slater, "The effects of rotating the self out of the body in the full virtual body ownership illusion," *Perception*, vol. 43, no. 4, pp. 275–294, 2014. PMID: 25109018.

[92] D. M. Lloyd, "Spatial limits on referred touch to an alien limb may reflect boundaries of visuo-tactile peripersonal space surrounding the hand," *Brain and Cognition*, vol. 64, no. 1, pp. 104–109, 2007.

[93] I. Bergström, K. Kilteni, and M. Slater, "First-person perspective virtual body posture influences stress: A virtual reality body ownership study," *PLOS ONE*, vol. 11, pp. 1–21, 02 2016.

[94] N. Franck, C. Farrer, N. Georgieff, M. Marie-Cardine, J. Daléry, T. d'Amato, and M. Jeannerod, "Defective recognition of one's own actions in patients with schizophrenia," *The American journal of psychiatry*, vol. 158, pp. 454–9, 04 2001.

[95] G. Richard, T. Pietrzak, F. Argelaguet, A. Lécuyer, and G. Casiez, "Studying the role of haptic feedback on virtual embodiment in a drawing task," *Frontiers in Virtual Reality*, vol. 1, 2021.

[96] P. Beckerle, *Virtual Hand Experience*, pp. 41–53. Cham: Springer International Publishing, 2021.

[97] C. Krogmeier, C. Mousas, and D. Whittinghill, "Human–virtual character interaction: Toward understanding the influence of haptic feedback," *Computer Animation and Virtual Worlds*, vol. 30, no. 3-4, p. e1883, 2019.

[98] J. Fröhner, G. Salvietti, P. Beckerle, and D. Prattichizzo, "Can wearable haptic devices foster the embodiment of virtual limbs?," *IEEE Transactions on Haptics*, vol. 12, no. 3, pp. 339–349, 2019.

[99] J. Kreimeier, S. Hammer, D. Friedmann, P. Karg, C. Bühner, L. Bankel, and T. Götzelmann, "Evaluation of different types of haptic feedback influencing the task-based presence and performance in virtual reality," in *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, PETRA 2019, (New York, NY, USA), p. 289–298, Association for Computing Machinery, 2019.

[100] I. S. MacKenzie and C. Ware, "Lag as a determinant of human performance in interactive systems," in *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pp. 488–493, 1993.

[101] T. B. Sheridan and W. R. Ferrell, "Remote manipulative control with transmission delay," *IEEE Transactions on Human Factors in Electronics*, no. 1, pp. 25–29, 1963.

[102] R. S. Allison, J. E. Zacher, D. Wang, and J. Shu, "Effects of network delay on a collaborative motor task with telehaptic and televisual feedback," in *Proceedings of the 2004 ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry*, VRCAI '04, (New York, NY, USA), p. 375–381, Association for Computing Machinery, 2004.

[103] S. Hirche, A. Bauer, and M. Buss, "Transparency of haptic telepresence systems with constant time delay," in *Proceedings of 2005 IEEE Conference on Control Applications, 2005. CCA 2005.*, pp. 328–333, 2005.

[104] C. Jay, M. Glencross, and R. Hubbold, "Modeling the effects of delayed haptic and visual feedback in a collaborative virtual environment," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 14, no. 2, pp. 8–es, 2007.

[105] A. Sengül, F. Rivest, M. van Elk, O. Blanke, and H. Bleuler, "Visual and force feedback time-delays change telepresence: Quantitative evidence from crossmodal congruecy task," in *2013 World Haptics Conference (WHC)*, pp. 577–582, 2013.

[106] J. Fröhner, P. Beckerle, S. Endo, and S. Hirche, "An embodiment paradigm in evaluation of human-in-the-loop control," *IFAC-PapersOnLine*, vol. 51, no. 34, pp. 104–109, 2019. 2nd IFAC Conference on Cyber-Physical and Human Systems CPHS 2018.

[107] E. A. Caspar, A. Cleeremans, and P. Haggard, "The relationship between human agency and embodiment," *Consciousness and Cognition*, vol. 33, pp. 226–236, 2015.

[108] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[109] S. Endo, J. Fröhner, S. Musić, S. Hirche, and P. Beckerle, "Effect of external force on agency in physical human-machine interaction," *Frontiers in Human Neuroscience*, vol. 14, p. 114, 2020.

[110] B. Lenggenhager, T. Tadi, T. Metzinger, and O. Blanke, "Video ergo sum: Manipulating bodily self-consciousness," *Science (New York, N.Y.)*, vol. 317, pp. 1096–9, 09 2007.

[111] A. Pomes and M. Slater, "Drift and ownership toward a distant virtual body," *Frontiers in Human Neuroscience*, vol. 7, p. 908, 2013.

[112] G. Gorisse, O. Christmann, E. A. Amato, and S. Richir, "First- and third-person perspectives in immersive virtual environments: Presence and performance analysis of embodied users," *Frontiers in Robotics and AI*, vol. 4, p. 33, 2017.

[113] P. Pozeg, G. Galli, and O. Blanke, "Those are your legs: The effect of visuo-spatial viewpoint on visuo-tactile integration and body ownership," *Frontiers in Psychology*, vol. 6, p. 1749, 2015.

[114] I. V. Piryankova, H. Y. Wong, S. A. Linkenauger, C. Stinson, M. R. Longo, H. H. Bülthoff, and B. J. Mohler, "Owning an overweight or underweight body: Distinguishing the physical, experienced and virtual body," *PLOS ONE*, vol. 9, pp. 1–13, 08 2014.

[115] F. Pavani, C. Spence, and J. Driver, "Visual capture of touch: Out-of-the-body experiences with rubber gloves," *Psychological Science*, vol. 11, no. 5, pp. 353–359, 2000. PMID: 11228904.

[116] S. Serino, E. Pedroli, A. Keizer, S. Triberti, A. Dakanalis, F. Pallavicini, A. Chirico, and G. Riva, "Virtual reality body swapping: A tool for modifying the allocentric memory of the body," *Cyberpsychology, Behavior, and Social Networking*, vol. 19, no. 2, pp. 127–133, 2016. PMID: 26506136.

[117] B. Lenggenhager, M. Mouthon, and O. Blanke, "Spatial aspects of bodily self-consciousness," *Consciousness and Cognition*, vol. 18, no. 1, pp. 110–117, 2009.

[118] J. E. Aspell, B. Lenggenhager, and O. Blanke, "Keeping in touch with one's self: Multisensory mechanisms of self-consciousness," *PLOS ONE*, vol. 4, pp. 1–10, 08 2009.

[119] V. Petkova, M. Khoshnevis, and H. Ehrsson, "The perspective matters! multisensory integration in ego-centric reference frames determines full-body ownership," *Frontiers in psychology*, vol. 2, p. 35, 03 2011.

[120] P. Wittkopf, D. Lloyd, and M. Johnson, "Changing the size of a mirror reflected hand moderates the experience of embodiment but not proprioceptive drift: A repeated measures study on healthy human participants.," *Experimental Brain Research*, vol. 235, 06 2017.

[121] W. Ijsselsteijn, H. de Ridder, R. Hamberg, D. Bouwhuis, and J. Freeman, "Perceived depth and the feeling of presence in 3dtv," *Displays*, vol. 18, no. 4, pp. 207–214, 1998.

[122] W. IJsselsteijn, H. d. Ridder, J. Freeman, S. E. Avons, and D. Bouwhuis, "Effects of Stereoscopic Presentation, Image Motion, and Screen Size on Subjective and Objective Corroborative Measures of Presence," *Presence: Teleoperators and Virtual Environments*, vol. 10, pp. 298–311, 06 2001.

[123] W. Luu, B. Zangerl, M. Kalloniatis, and J. Kim, "Effects of stereopsis on vection, presence and cybersickness in head-mounted display (hmd) virtual reality," *Scientific reports*, vol. 11, no. 1, pp. 1–10, 2021.

[124] C. Mroczkowski and E. Niechwiej-Szwedo, "Stereopsis contributes to the predictive control of grip forces during prehension," *Experimental Brain Research*, vol. 239, pp. 1–14, 04 2021.

[125] J. Tresilian, M. Mon-Williams, and B. Kelly, "Increasing confidence in vergence as a cue to distance," *Proceedings. Biological sciences / The Royal Society*, vol. 266, pp. 39–44, 02 1999.

[126] E. Niechwiej-Szwedo, M. Cao, and M. Barnett-Cowan, "Binocular viewing facilitates size constancy for grasping and manual estimation," *Vision*, vol. 6, no. 2, 2022.

[127] D. R. Melmoth, M. Storoni, G. D. Todd, A. L. Finlay, and S. Grant, "Dissociation between vergence and binocular disparity cues in the control of prehension," *Experimental Brain Research*, vol. 183, pp. 283–298, 2007.

[128] D. A. Gonzalez and E. Niechwiej-Szwedo, "The effects of monocular viewing on hand-eye coordination during sequential grasping and placing movements," *Vision Research*, vol. 128, pp. 30–38, 2016.

[129] M. Schmettow, *New Statistics For Design Researchers: A Bayesian Workflow in Tidy R*. Springer Cham, 1 ed., 2021.

[130] P.-C. Bürkner and M. Vuorre, "Ordinal regression models in psychology: A tutorial," *Advances in Methods and Practices in Psychological Science*, vol. 2, no. 1, pp. 77–101, 2019.

[131] S. S. Stevens, "On the theory of scales of measurement," *Science*, vol. 103, no. 2684, pp. 677–680, 1946.

[132] T. M. Liddell and J. K. Kruschke, "Analyzing ordinal data with metric models: What could possibly go wrong?," *Journal of Experimental Social Psychology*, vol. 79, pp. 328–348, 2018.

[133] A. Solomon Kurz, "Notes on the Bayesian cumulative probit," Aug 2022.

[134] E.-J. Wagenmakers, M. Lee, T. Lodewyckx, and G. J. Iverson, *Bayesian Versus Frequentist Inference*. New York, NY: Springer New York, 2008.

[135] J. Kruschke and T. Liddell, "The bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective," *Psychonomic Bulletin Review*, vol. 25, pp. 1–29, 02 2017.

[136] R. Schoot and S. Depaoli, "Bayesian analyses: Where to start and what to report," *European Health Psychologist*, vol. 16, pp. 75–84, 01 2014.

[137] A. Kumar, "Bayes theorem," 2020.

[138] R. D. Morey, J.-W. Romeijn, and J. N. Rouder, "The philosophy of bayes factors and the quantification of statistical evidence," *Journal of Mathematical Psychology*, vol. 72, pp. 6–18, 2016. Bayes Factors for Testing Hypotheses in Psychological Research: Practical Relevance and New Developments.

[139] R. Kelter, "Bayesian alternatives to null hypothesis significance testing in biomedical research: a non-technical introduction to bayesian inference with jasp," *BMC Medical Research Methodology*, vol. 20, 06 2020.

[140] J. K. Kruschke, "Chapter 12 - bayesian approaches to testing a point ("null") hypothesis," in *Doing Bayesian Data Analysis (Second Edition)* (J. K. Kruschke, ed.), pp. 335–358, Boston: Academic Press, second edition ed., 2015.

[141] J. Kruschke, "Bayesian analysis reporting guidelines," *Nature Human Behaviour*, vol. 5, 08 2021.

[142] P.-C. Bürkner, "Advanced bayesian multilevel modeling with the r package brms," *R Journal*, vol. 10, pp. 395–411, 07 2018.

[143] P. Fitts and M. Posner, *Motor Learning*. Brooks/Cole, 1967.

[144] R. A. Schmidt, "A schema theory of discrete motor skill learning.," *Psychological Review*, vol. 82, pp. 225–260, 1975.

[145] J. R. Anderson, "Acquisition of cognitive skill," *Psychological Review*, vol. 89, pp. 369–406, 1982.

[146] A. Heathcote, S. Brown, and D. Mewhort, "The power law repealed: The case for an exponential law of practice," *Psychonomic bulletin review*, vol. 7, pp. 185–207, 07 2000.

[147] L. Posen, "A secondary data analysis using bayesian statistics to explore the influence of gender and initial performance on skill acquisition using a laparoscopy simulator," November 2020.

[148] L. David, "Towards a novel approach to applied research: The role of motor sequence learning in the process of mastering complex motor procedures," April 2018.

[149] V. L. Kaschub, "Learning complex motor procedures : can the ability to learn dexterity games predict a person's ability to learn a complex task?," June 2016.

[150] C. O. H. Weimer, "Towards an effective mis simulator-based training with basic laparoscopic tasks: The impact of time pressure on the learning process," June 2019.

[151] A. Küpper, "Individual learning curves in bronchoscopy : exploring skill acquisition with a low-fidelity prototype.," August 2018.

[152] M. Voskes, "The effectiveness of an online driving simulator: Transfer effects of driving skills examined with the tweak-finder model," 2022.

[153] E. Villalobos Becerril, "Learning effect in driving simulators : online driving simulators and driving simulators learning curves analysis," January 2022.

[154] P.-C. Bürkner, "brms: An r package for bayesian multilevel models using stan," *Journal of Statistical Software*, vol. 80, no. 1, p. 1–28, 2017.

[155] K. Friston, "The free-energy principle: a unified brain theory?," *Nature reviews neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.

[156] L. T. Trujillo, "Mental effort and information-processing costs are inversely related to global brain free energy during visual categorization," *Frontiers in Neuroscience*, vol. 13, 2019.

[157] K. J. Friston, J. Daunizeau, and S. J. Kiebel, "Reinforcement learning or active inference?," *PLOS ONE*, vol. 4, pp. 1–13, 07 2009.

[158] D. C. Knill and A. Pouget, "The bayesian brain: the role of uncertainty in neural coding and computation," *Trends in Neurosciences*, vol. 27, no. 12, pp. 712–719, 2004.

[159] A. Berti, G. Bottini, M. Gandola, L. Pia, N. Smania, A. Stracciari, I. Castiglioni, G. Vallar, and E. Paulesu, "Shared cortical anatomy for motor awareness and motor control," *science*, vol. 309, no. 5733, pp. 488–491, 2005.

[160] K. Kilteni and H. H. Ehrsson, "Body ownership determines the attenuation of self-generated tactile sensations," *Proceedings of the National Academy of Sciences*, vol. 114, no. 31, pp. 8426–8431, 2017.

[161] K. Grechuta, L. Ulysse, B. Rubio Ballester, and P. F. M. J. Verschure, "Self beyond the body: Action-driven and task-relevant purely distal cues modulate performance and body ownership," *Frontiers in Human Neuroscience*, vol. 13, 2019.

[162] B. Laha, J. N. Bailenson, A. S. Won, and J. O. Bailey, "Evaluating control schemes for the third arm of an avatar," *Presence: Teleoperators and Virtual Environments*, vol. 25, no. 2, pp. 129–147, 2016.

[163] M. C. Egeberg, S. L. Lind, S. Serubugo, D. Skantarova, and M. Kraus, "Extending the human body in virtual reality: Effect of sensory feedback on agency and ownership of virtual wings," in *Proceedings of the 2016 Virtual Reality International Conference*, pp. 1–4, 2016.

[164] N. Rosa, R. C. Veltkamp, W. Hürst, A.-M. Brouwer, K. Gijsbertse, I. Cocu, and P. Werkhoven, "Embodiment and performance in the supernumerary hand illusion in augmented reality," *Frontiers in Computer Science*, vol. 3, 2021.

[165] D. Medeiros, R. Anjos, D. Mendes, J. Madeiras Pereira, A. Raposo, and J. Jorge, "Keep my head on my shoulders!: why third-person is bad for navigation in vr," pp. 1–10, 11 2018.

[166] R. Bellman, "On the theory of dynamic programming," *Proceedings of the national Academy of Sciences*, vol. 38, no. 8, pp. 716–719, 1952.

[167] M. Botvinick, S. Ritter, J. Wang, Z. Kurth-Nelson, C. Blundell, and D. Hassabis, "Reinforcement learning, fast and slow," *Trends in Cognitive Sciences*, vol. 23, 04 2019.

[168] N. Sajid, P. J. Ball, T. Parr, and K. J. Friston, "Active inference: Demystified and compared," *Neural Computation*, vol. 33, pp. 674–712, mar 2021.

[169] Haption, "Virtuose 6d." https://www.haption.com/en/products-en/virtuose-6d-en.html. Online, accessed: 21.06.2021.

[170] Haption, "HGlove." https://www.haption.com/en/products-en/hglove-en.html. Online, accessed: 21.06.2021.

[171] Franka Emika, "Panda." http://www.imperial.ac.uk/robot-intelligence/robots/franka-emika-panda/. Online, accessed: 21.06.2021.

[172] QB Robotics, "QB hand." https://qbrobotics.com/products/qb-softhand-research/. Online, accessed: 21.06.2021.

[173] N. Nadgere, "Haptic rendering/control for underactuated qb softhand using haption hglove," 2020.

[174] A. Ajoudani, S. Godfrey, M. Bianchi, M. Catalano, G. Grioli, N. Tsagarakis, and A. Bicchi, "Exploring teleimpedance and tactile feedback for intuitive control of the pisa/iit softhand," *Haptics, IEEE Transactions on*, vol. 7, pp. 203–215, 04 2014.

[175] A. Brygo, I. Sarakoglou, G. Grioli, and N. Tsagarakis, "Synergy-based bilateral port: A universal control module for tele-manipulation frameworks using asymmetric master–slave systems," *Frontiers in Bioengineering and Biotechnology*, vol. 5, 04 2017.

[176] M. Santello, M. Flanders, and J. Soechting, "Postural hand synergies for tool use," *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 18, pp. 10105–15, 12 1998.

[177] R. Lieftink, "Bi-directional impedance reflection for cartesian space telemanipulation control to reduce the effect of time delay," 2020.

[178] G. Niemeyer, C. Preusche, S. Stramigioli, and D. Lee, *Telerobotics*, pp. 1085–1108. Cham: Springer International Publishing, 2016.

[179] "HTC VIVE Pro Eye." https://summit.vive.com/eu/product/vive-pro-eye/overview/.

[180] "Zed mini." https://www.stereolabs.com/zed-mini/. Online, accessed: 08.10.2021.

[181] W. Choi, L. Li, S. Satoh, and K. Hachimura, "Multisensory integration in the virtual hand illusion with active movement," *BioMed Research International*, vol. 2016, pp. 1–9, 01 2016.

[182] M. P. M. Kammers, M. R. Longo, M. Tsakiris, H. C. Dijkerman, and P. Haggard, "Specificity and coherence of body representations," *Perception*, vol. 38, no. 12, pp. 1804–1820, 2009.

[183] M. Catoire, B. Krom, and J. Erp, *Towards a Test Battery to Benchmark Dexterous Performance in Teleoperated Systems*, pp. 440–451. 06 2018.

[184] C. Preston, "The role of distance from the body and distance from the real hand in ownership and disownership during the rubber hand illusion," *Acta Psychologica*, vol. 142, no. 2, pp. 177–183, 2013.

[185] L. Aymerich-Franch, R. F. Kizilcec, and J. N. Bailenson, "The relationship between virtual self similarity and social anxiety," *Frontiers in Human Neuroscience*, vol. 8, p. 944, 2014.

[186] D. L. Sackett, "Bias in analytic research," *Journal of Chronic Diseases*, vol. 32, no. 1, pp. 51–63, 1979.

[187] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, "Stan: A probabilistic programming language," *Journal of Statistical Software*, vol. 76, no. 1, p. 1–32, 2017.

[188] J. C. Nunnally, *An Overview of Psychological Measurement*, pp. 97–146. Boston, MA: Springer US, 1978.

[189] J. Hair, W. Black, B. Babin, and R. Anderson, *Multivariate Data Analysis: A Global Perspective*. 03 2010.

[190] M. Allen, D. Poggiali, K. Whitaker, T. Marshall, J. van Langen, and R. Kievit, "Raincloud plots: a multi-platform tool for robust data visualization (version 2; peer review: 2 approved)," *Wellcome Open Research*, vol. 4, no. 63, 2021.

[191] P. Bürkner, "set_prior: Prior definitions for brms models." https://www.rdocumentation.org/packages/brms/versions/2.17.0/topics/set_prior. Online, accessed: 02.06.2022.

[192] A. Chakraborty, D. J. Nott, and M. Evans, "Weakly informative priors and prior-data conflict checking for likelihood-free inference," 2022.

[193] A. Gelman and D. B. Rubin, "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, vol. 7, no. 4, pp. 457 – 472, 1992.

[194] A. Vehtari, A. Gelman, and J. Gabry, "Practical bayesian model evaluation using leave-one-out cross-validation and waic," *Statistics and Computing*, vol. 27, pp. 1413–1432, 2017.

[195] H. Akaike, *Information Theory and an Extension of the Maximum Likelihood Principle*, pp. 199–213. New York, NY: Springer New York, 1998.

[196] S. Watanabe, "Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory," *J. Mach. Learn. Res.*, vol. 11, p. 3571–3594, dec 2010.

[197] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, pp. 297–334, 1951.

[198] J. Cohen, *Statistical power analysis for the behavioral sciences*. Routledge, 2013.

[199] S. Gallagher, "Philosophical conceptions of the self: implications for cognitive science," *Trends in Cognitive Sciences*, vol. 4, no. 1, pp. 14–21, 2000.

[200] M. Pyasik, D. Burin, and L. Pia, "On the relation between body ownership and sense of agency: A link at the level of sensory-related signals," *Acta Psychologica*, vol. 185, pp. 219–228, 2018.

[201] Y. Stern, D. Koren, R. Moebus, G. Panishev, and R. Salomon, "Assessing the relationship between sense of agency, the bodily-self and stress: Four virtual-reality experiments in healthy individuals," *Journal of Clinical Medicine*, vol. 9, 2020.

[202] Z. Abdulkarim and H. Ehrsson, "No causal link between changes in hand position sense and feeling of limb ownership in the rubber hand illusion," *Attention, perception psychophysics*, vol. 78, 11 2015.

[203] S. Shibuya, S. Unenaka, and Y. Ohki, "Body ownership and agency: task-dependent effects of the virtual hand illusion on proprioceptive drift," *Experimental Brain Research*, vol. 235, 01 2017.

[204] M. Tsakiris, "My body in the brain: A neurocognitive model of body-ownership," *Neuropsychologia*, vol. 48, no. 3, pp. 703–712, 2010. The Sense of Body.

[205] N. Braun, S. Debener, N. Spychala, E. Bongartz, P. Sörös, H. H. O. Müller, and A. Philipsen, "The senses of agency and ownership: A review," *Frontiers in Psychology*, vol. 9, p. 535, 2018.

[206] N. Braun, J. D. Thorne, H. Hildebrandt, and S. Debener, "Interplay of agency and ownership: The intentional binding and rubber hand illusion paradigm combined," *PLOS ONE*, vol. 9, pp. 1–10, 11 2014.

[207] N. Braun, R. Emkes, J. D. Thorne, and S. Debener, "Embodied neurofeedback with an anthropomorphic robotic hand," *Scientific Reports*, vol. 6, 2016.

[208] M. Tsakiris, "The multisensory basis of the self: From body to identity to others," *Quarterly Journal of Experimental Psychology*, vol. 70, no. 4, pp. 597–609, 2017. PMID: 27100132.

[209] D. S. Schwarzkopf, B. De Haas, and G. Rees, "Better ways to improve standards in brain-behavior correlation analysis," *Frontiers in human neuroscience*, vol. 6, p. 200, 2012.

[210] S. Hamasaki, A. Yamashita, and H. Asama, "A three-dimensional evaluation of body representation change of human upper limb focused on sense of ownership and sense of agency," in *2018 International Symposium on Micro-NanoMechatronics and Human Science (MHS)*, pp. 1–5, 2018.

[211] J. M. Loomis and J. W. Philbeck, "Is the anisotropy of perceived 3-d shape invariant across scale? perception psychophysics," 1999.

[212] J. F. Norman, J. T. Todd, V. J. Perotti, and J. S. Tittle, "The visual perception of three-dimensional length.," *Journal of experimental psychology: Human Perception and Performance*, vol. 22, no. 1, p. 173, 1996.

[213] S. Sharples, S. Cobb, A. Moody, and J. R. Wilson, "Virtual reality induced symptoms and effects (vrise): Comparison of head mounted display (hmd), desktop and projection display systems," *Displays*, vol. 29, no. 2, pp. 58–69, 2008.

[214] P. Kourtesis, S. Collina, L. A. A. Doumas, and S. E. MacPherson, "Validation of the virtual reality neuroscience questionnaire: Maximum duration of immersive virtual reality sessions without the presence of pertinent adverse symptomatology," *Frontiers in Human Neuroscience*, vol. 13, 2019.

[215] K. Grechuta, J. De La Torre Costa, B. R. Ballester, and P. Verschure, "Challenging the boundaries of the physical self: Distal cues impact body ownership," *Frontiers in Human Neuroscience*, vol. 15, 2021.

[216] R. Dickstein and J. E. Deutsch, "Motor imagery in physical therapist practice," *Physical therapy*, vol. 87, no. 7, pp. 942–953, 2007.

[217] F. Di Rienzo, U. Debarnot, S. Daligault, E. Saruco, C. Delpuech, J. Doyon, C. Collet, and A. Guillot, "Online and offline performance gains following motor imagery practice: a comprehensive review of behavioral and neuroimaging studies," *Frontiers in human neuroscience*, vol. 10, p. 315, 2016.

[218] D. Wolpert and M. Kawato, "Multiple paired forward and inverse models for motor control," *Neural Networks*, vol. 11, no. 7, pp. 1317–1329, 1998.

[219] R. C. Miall and D. M. Wolpert, "Forward models for physiological motor control," *Neural networks*, vol. 9, no. 8, pp. 1265–1279, 1996.

[220] K. Kilteni, B. Andersson, C. Houborg, and H. Ehrsson, "Motor imagery involves predicting the sensory consequences of the imagined movement," *Nature Communications*, vol. 9, 04 2018.

[221] P. Morasso, M. Casadio, V. Mohan, F. Rea, and J. Zenzeri, "Revisiting the body-schema concept in the context of whole-body postural-focal dynamics," *Frontiers in Human Neuroscience*, vol. 9, 2015.

[222] D. M. Wolpert and J. R. Flanagan, "Motor prediction," *Current biology*, vol. 11, no. 18, pp. 729–732, 2001.

[223] S. Tanaka, "Body schema and body image in motor learning: refining Merleau-Ponty's notion of body schema," in *Body Schema and Body Image: New Directions*, Oxford University Press, 07 2021.

[224] G. Ruggiero, T. Iachini, F. Ruotolo, and V. P. Senese, *Spatial memory: The role of egocentric and allocentric frames of reference*. 01 2009.

[225] D. Colombo, S. Serino, C. Tuena, E. Pedroli, A. Dakanalis, P. Cipresso, and G. Riva, "Egocentric and allocentric spatial reference frames in aging: A systematic review," *Neuroscience Biobehavioral Reviews*, vol. 80, pp. 605–621, 2017.

[226] A. Maselli, "Allocentric and egocentric manipulations of the sense of self-location in full-body illusions and their relation with the sense of body ownership," *Cognitive Processing*, vol. 16, no. 1, pp. 309–312, 2015.

[227] Y. Ataria and S. Tanaka, "When body image takes over the body schema: The case of frantz fanon," *Human Studies*, vol. 43, 12 2020.

[228] S. Giurgola, C. Crico, A. Farnè, and N. Bolognini, "The sense of body ownership shapes the visual representation of body size.," *Journal of Experimental Psychology: General*, vol. 151, no. 4, p. 872, 2022.

[229] S. Ionta, D. Perruchoud, B. Draganski, and O. Blanke, "Body context and posture affect mental imagery of hands," *PLOS ONE*, vol. 7, pp. 1–7, 03 2012.

[230] N. Evans and O. Blanke, "Shared electrophysiology mechanisms of body ownership and motor imagery," *NeuroImage*, vol. 64, pp. 216–228, 2013.

[231] B. Nanay, "Unconscious mental imagery," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 376, no. 1817, p. 20190689, 2021.

[232] D. M. Wolpert and Z. Ghahramani, "Computational principles of movement neuroscience," *Nature neuroscience*, vol. 3, no. 11, pp. 1212–1217, 2000.

[233] C. D. Frith, S.-J. Blakemore, and D. M. Wolpert, "Explaining the symptoms of schizophrenia: abnormalities in the awareness of action," *Brain Research Reviews*, vol. 31, no. 2-3, pp. 357–363, 2000.

[234] A. Maselli, K. Kilteni, J. Lopez-Moliner, and M. Slater, "The sense of body ownership relaxes temporal constraints for multisensory integration," *Scientific Reports*, vol. 6, 08 2016.

[235] A. Shintemirov, T. Taunyazov, B. Omarali, A. Nurbayeva, A. Kim, A. Bukeyev, and M. Rubagotti, "An open-source 7-dof wireless human arm motion-tracking system for use in robotics research," *Sensors*, vol. 20, no. 11, 2020.

[236] K. Kilteni, J.-M. Normand, M. V. Sanchez-Vives, and M. Slater, "Extending body space in immersive virtual reality: A very long arm illusion," *PLoS ONE*, vol. 7, 2012.

[237] A. Guterstam, V. I. Petkova, and H. H. Ehrsson, "The illusion of owning a third arm," *PLOS ONE*, vol. 6, pp. 1–11, 02 2011.

[238] R. Kondo, M. Sugimoto, K. Minamizawa, T. Hoshi, M. Inami, and M. Kitazaki, "Illusory body ownership of an invisible body interpolated between virtual hands and feet via visual-motor synchronicity," *Scientific Reports*, vol. 8, 05 2018.

[239] E. Caspar, A. De Beir, P. Magalhães De Saldanha da Gama, F. Yernaux, A. Cleeremans, and B. Vanderborght, "New frontiers in the rubber hand experiment: when a robotic hand becomes one's own," *Behavior Research Methods*, vol. Behav Res, 06 2014.

[240] A. Guterstam, G. Gentile, and H. H. Ehrsson, "The invisible hand illusion: Multisensory integration leads to the embodiment of a discrete volume of empty space," *Journal of Cognitive Neuroscience*, vol. 25, pp. 1078–1099, 07 2013.

[241] S. Warren and P. Artemiadis, "On the control of human-robot bi-manual manipulation," *Journal of Intelligent Robotic Systems*, vol. 78, 04 2014.

[242] J. Qu, F. Zhang, Y. Wang, and Y. Fu, "Human-like coordination motion learning for a redundant dual-arm robot," *Robotics and Computer-Integrated Manufacturing*, vol. 57, pp. 379–390, 06 2019.

[243] K. Cieslik and M. Lopatka, "Research on speed and acceleration of hand movements as command signals for anthropomorphic manipulators as a master-slave system," *Applied Sciences*, vol. 12, p. 3863, 04 2022.

[244] C. Lenz and S. Behnke, "Bimanual telemanipulation with force and haptic feedback and predictive limit avoidance," *CoRR*, vol. abs/2109.13382, 2021.

[245] M. Schwarz, C. Lenz, A. Rochow, M. Schreiber, and S. Behnke, "Nimbro avatar: Interactive immersive telepresence with force-feedback telemanipulation," 2021.

[246] Xprize, "ANA Avatar Xprize competition." https://www.xprize.org/prizes/avatar. Online, accessed: 24.09.2022.

[247] T. Porssut, O. Blanke, B. Herbelin, and R. Boulic, "Reaching articular limits can negatively impact embodiment in virtual reality," *PLoS One*, vol. 2, no. 17, 2022.

[248] J. Ventre-Dominey, G. Gibert, M. Bosse-Platiere, A. Farne, P. F. Dominey, and F. Pavani, "Embodiment into a robot increases its acceptability," *Scientific reports*, vol. 9, no. 1, pp. 1–10, 2019.

[249] D. Farizon, P. F. Dominey, and J. Ventre-Dominey, "Social attitude towards a robot is promoted by motor-induced embodiment independently of spatial perspective," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9036–9042, 2022.

[250] R. Reilink, L. C. Visser, J. Bennik, R. Carloni, D. M. Brouwer, and S. Stramigioli, "The twente humanoid head," in *2009 IEEE International Conference on Robotics and Automation*, pp. 1593–1594, 2009.

[251] R. van der Zee, "Remote control of a robotic humanoid headby using an oculus vr headset," 2019.

[252] D. Lee, C. Minji, and J. Lee, "Prediction of head movement in 360-degree videos using attention model," *Sensors*, vol. 21, p. 3678, 05 2021.

[253] D. Radziun and H. Ehrsson, "Auditory cues influence the rubber-hand illusion," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 44, pp. 1012–1021, 07 2018.

[254] G. Darnai, T. Szolcsányi, G. Hegedüs, P. Kincses, J. Kállai, M. Kovács, E. Simon, Z. Nagy, and J. Janszky, "Hearing visuo-tactile synchrony – sound-induced proprioceptive drift in the invisible hand illusion," *British Journal of Psychology*, vol. 108, no. 1, pp. 91–106, 2017.

[255] T. Waltemate, I. Senna, F. Hülsmann, M. Rohde, S. Kopp, M. Ernst, and M. Botsch, "The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality," in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, VRST '16, (New York, NY, USA), p. 27–35, Association for Computing Machinery, 2016.

[256] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Human Mental Workload* (P. A. Hancock and N. Meshkati, eds.), vol. 52 of *Advances in Psychology*, pp. 139–183, North-Holland, 1988.

[257] A.-M. Brouwer, M. Hogervorst, M. Holewijn, and J. Erp, "Evidence for effects of task difficulty but not learning on neurophysiological variables associated with effort," *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 05 2014.

[258] SenseGlove, "Nova," August 2022.

[259] HaptX, "Haptx dk2," August 2022.

[260] N. F. Lepora, C. Ford, A. Stinchcombe, A. Brown, J. Lloyd, M. G. Catalano, M. Bianchi, and B. Ward-Cherrier, "Towards integrated tactile sensorimotor control in anthropomorphic soft robotic hands," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1622–1628, 2021.

[261] Psyonic, "Psyonic hand," August 2022.

[262] W. Wen, A. Yamashita, and H. Asama, "The sense of agency during continuous action: Performance is more important than action-feedback association," *PLOS ONE*, vol. 10, pp. 1–16, 04 2015.

[263] A. Weiss, R. Bernhaupt, M. Lankes, and M. Tscheligi, "The usus evaluation framework for human-robot interaction," *Proc. of AISB 09*, vol. 4, pp. 11–26, 01 2009.

[264] S. Föll, M. Maritsch, F. Spinola, V. Mishra, F. Barata, T. Kowatsch, E. Fleisch, and F. Wortmann, "FLIRT: A Feature Generation Toolkit for Wearable Data," *Computer Methods and Programs in Biomedicine*, 2021.

[265] H. Chen, A. Dey, M. Billinghurst, and R. W. Lindeman, "Exploring Pupil Dilation in Emotional Virtual Reality Environments," in *ICAT-EGVE 2017 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments* (R. W. Lindeman, G. Bruder, and D. Iwai, eds.), The Eurographics Association, 2017.

[266] C. P. Dancey and J. Reidy, *Statistics without maths for psychology*. Pearson education, 2007.