Modal Verbs and Mental Stirs

Studying short-term effects of modal verb production on cognition

Astraea F.A. Blonk (s1026793)
Supervisor: Laura J. Speed
Second reader: Saskia van Putten
Bachelor's Thesis in Linguistics
Radboud University, Faculty of Arts
Completion Date: 15-06-2021

Preface

When I found out that my bachelor's thesis in Linguistics was about to be supervised by a researcher who seemed to be teaching primarily at the Communication and Information Sciences program, I wondered if the subject for which I had tried so hard to find a supervisor was actually a topic that fit a thesis in Linguistics, rather than it being more suited for students of the program where my supervisor was teaching. However, after nearly five months of diving into the subject and incorporating many things I learned at courses about psycholinguistics, neurolinguistics and language analysis, among others, I can state with reasonable certainty that my initial doubts were wrong and that the document presented here is worthy of concluding my Linguistics bachelor.

I want to thank my supervisor, Laura Speed, for the time and effort she put in guiding the thesis process from start to finish and for all her comments during various stages of this process, which have been very helpful. I would also like to thank my father, Arie Blonk, thanks to whom I was introduced to Neuro-Linguistic Programming, which inspired the subject of this thesis. I want to thank my sisters Nadia and Paloma Blonk and my mother, Rineke Blonk, for helping me pre-test some anagrams and later the experiment itself. In addition, I want to express my thanks to the following people who have helped to spread the experiment to as many interested people as possible, in alphabetical order: Anita Hunter, Annet van Berkel, Arie Baudoin, Arie Blonk, Aurora Blonk, Carmen Piscador, Corrie Papeleu, Elke Teurlings, Floris Cos, Georgia Vasilaras, Jan Stam, Joke Scheffers, Laura Speed, Laurine Blonk, Marc van der Wijst, Myrthe Spierings, Nadia Blonk, Paloma Blonk, Stef Grondelaers and Tamara Mauro, as well as all the people who reshared a post or shared a link without my knowledge and perhaps one or two people that I did know about but forgot to mention (sorry). Last but not least, my gratitude goes, of course, to all the 357 anonymous people who filled in the experiment, either partly or completely. I very much appreciate that all of you took your time to help me without expecting much in return.

Table of Contents

Abstract
Research into the possible effects of language on thought has focused predominantly on effects of language structure and language perception (i.e. hearing and reading). Few studies have addressed potential cognitive consequences of the words we speak, and those who did, often found effects that could be traced back to mechanisms separate from the language production process. In addition, tests of functional effects of language on the mind have typically employed content words, which are relatively concrete in comparison to function words and thus provide little opportunity to explore the challenges of representing abstract concepts that come with theories of embodied cognition. It would therefore be interesting to explore the effects of language production while using a more structural category of words: modal verbs. Participants in the present experiment generated sentences containing a Dutch modal verb that pertained either to intrinsic motivation ("willen" [*to want to*]), extrinsic motivation ("moeten" [*to have to*]) or neither ("zullen" [*will*]). After the sentence production task, they were asked to solve anagrams and rate their experience of the experiment. Producing different modal verbs did not appear to influence participants' performance or persistence on the anagram task, nor did it result in differences in how they rated their intrinsic motivation and sense of autonomy in relation to the experimental tasks. A number of potential explanations for these findings are discussed, including the possibilities that the particular areas of language that have been investigated cannot affect cognition or that the experiment missed genuine effects due to large amounts of noise. It is suggested that more research be undertaken into the duration of functional language effects, into the extent to which abstract words can influence cognition and into effects of language production in general.

A change in language can transform our appreciation of the cosmos.
(Whorf, 1952, p. 182)

1. Introduction

Can a change in what we say, alter our experience of the world around us? While there are still many questions surrounding the exact relationship between words and the mind, claims about influences of language production on our impression of life are not new. One notable source of such claims is the field of Neuro-Linguistic Programming, or NLP[1] (Liekens, 2005). As remarked by its founders, "[w]e as human beings use our language […] to represent our experience – we call this activity reasoning, thinking, fantasying, rehearsing. When we are using language as a representational system, we are creating a model of our experience." (Bandler & Grinder, 1975, pp. 21-22) According to them, the difference between people who face challenges with lots of energy and creativity and those who undergo them with misery "follows primarily from differences in the richness of their models" (Bandler & Grinder, 1975, p. 14).

However, Bandler and Grinder's claims have been met with some resistance. Levelt (1996) argued that NLP disregards science and that practitioners of NLP fall into the 'black hole' of linguistic relativity (p. 27). While this critical essay was published 25 years ago, Levelt's commentary is not any less relevant today. Although a recent surge in research both within and outside the field of linguistic relativity suggests that language may influence how we think after all, there remains a shortage of studies into the specific linguistic claims from the field of NLP, such as that producing nominalizations could stagnate our progress or that an inconvenient choice of modal verbs would have the power to make tasks harder than they could have been if they had been stated differently (Liekens, 2005). Filling

1 Not to be confused with Natural Language Processing, which shares the same acronym.

this gap and figuring out if the processes involved in language production have some of the proclaimed capabilities will help not only the field of NLP but also the field of linguistic relativity more generally.

The remainder of this chapter is structured as follows: Section 1.1 discusses literature from the broader field that concerns itself with the relationship between language and thought. This is followed by Section 1.2, which discusses language production research more specifically. Section 1.3 explains a possible cognitive mechanism that might enable language production to affect cognition. Finally, in Section 1.4, the current experiment is outlined[2], along with an explanation of how it complements other studies done so far.

1.1 Linguistic Relativity Research

A central field in the study of the relationship between language and the mind is the field of *linguistic relativity,* which investigates Whorf's (1956) proposed principle that "all observers are not led by the same physical evidence of the same picture of the universe, unless their linguistic backgrounds are similar, or can in some way be calibrated" (p. 214). In other words, a person's experience of the world around them would be affected by their 'linguistic background'. What kind of background is this, exactly?

Traditionally, this term 'linguistic background' has been taken to refer to the *structure* of the language spoken by a person (Brown, 1976; Lucy, 1997). This interpretation is based on the idea that different languages employ different grammatical structures (Whorf, 1956) and lexical categorizations (Whorf, 1952) and hypothesizes that native speakers of one language (e.g. English) show cognitive differences when compared to native speakers of another language (e.g. Hopi) (Whorf, 1956; Brown, 1976; Lucy, 1997). Among the researchers who tested this reading of the linguistic relativity hypothesis are Davidoff et al. (1999), who conducted a color memory task on the Berinmo people of Papua New Guinea and compared the results to speakers of English who had done the same task. The vocabulary Berinmo people can employ to describe colors differs from the words available for this in English; for example, Berinmo does not make a distinction between 'blue' and 'green', but this language does distinguish between two other shades, 'wor' and 'nol', which would both be described by 'green' by a native speaker of English (Davidoff et al., 1999). The color memory task involved a researcher showing a color to the participants, followed by a 30 seconds delay, after which participants had to pick the color they had seen from two presented options. The authors found that the Berinmo people distinguished more accurately between the two options if one of them was a 'wor' color and the other one a 'nol' color than if one of them was a shade of green and the other one was blue. For the English participants, the effect was opposite. This was but one of the studies that displayed support for the structural reading of the linguistic relativity hypothesis, with similar research showing that differences in the structure of language systems can affect speakers' speed of color discernment (Winawer et al., 2007), spatial reasoning (Levinson et al., 2002) and discrimination of shapes (Lupyan & Spivey, 2008).

However, the structural reading is not the only possible interpretation of 'linguistic background'. Lucy (1997) notes that there are at least two other levels at which language may affect thought: (1) the semiotic level, which relates to the potential influence of speaking *any language at all* in comparison to species who do not use such a system of arbitrary symbols (Hockett, 1960) to communicate, and (2) the functional level, which involves the potential effects of using a language in a particular way (e.g. using certain English words over synonyms). Considering that the claims from the field of NLP about the effects of choosing certain linguistic elements or structures (e.g. nominalizations) over others fall under the latter category, this thesis will be focusing on the functional reading of the linguistic relativity hypothesis. Whereas Lucy (1997) refers to this type of relativity as

---

2 This experiment has been preregistered at the Open Science Framework under DOI 10.17605/OSF.IO/CTEVW.

discursive relativity and prefers to limit the usage of the term 'linguistic relativity' to those cases in which it applies to the structural level of language, for simplicity's sake, the latter term will be used here to denote influences of language on the mind from any level, including the functional.

Under the functional interpretation, the linguistic relativity hypothesis would predict that using language in a particular way has consequences for non-linguistic cognition (Lucy, 1997). Since language is mainly used for communication purposes (Vygotsky, 1987)[3] and a typical communicative situation involves at least one person who *produces* language and one person who *perceives* language (Jakobson, 1960), a distinction can be made between functional effects of language on cognition during production and during perception. There have been a number of studies that have attempted to find support for influences of language perception on the mind; for instance, Meteyard et al. (2007) had participants listen to verbs referring to either upward (e.g. 'rise') or downard (e.g. 'fall') movement while they performed a visual task in which they had to discern vertical motion from random motion, and found that it was harder to detect the vertical motion when its direction was opposite to the direction referred to by the background verbs than when these directions were congruent (cf. Francken et al., 2015). Other language perception research has shown that hearing the name of an object makes it easier to spot it (Lupyan & Ward, 2013; Ostarek & Huettig, 2017) and that grammatical gender of ingredients in fragrance descriptions can influence memory for odors (Speed & Majid, 2019). However, these are all studies that have considered effects of reading or listening to language. While they provide some support for the functional interpretation of the linguistic relativity hypothesis, they do not rule out the possibility that such effects are limited to language perception. Are there any studies that have looked at cognitive influences of language production specifically?

1.2 Language Production Research

There have been a number of prior lines of research that have considered effects of producing language on various parts of cognition. One of them tested the effects of having sports players engage in motivational (e.g. 'I can do this') or instructional self-talk (e.g. 'Keep your arm low'). For instance, Hatzigeorigiadis et al. (2009) found that tennis players who were instructed to use motivational or instructional cue words ended up performing better, showing more self-confidence and being less anxious than a control group who received short tactical lectures instead. Slimani et al. (2014) found that kickboxers who identified negative self-talk statements and turned them into positive, motivational ones had more self-confidence and positive affect and less negative affect after a period of 12 weeks than those who performed physical training during that period instead. What's more, Tod et al. (2011) conducted a meta-analysis and found that the majority of the studies into instructional or motivational self-talk that they analyzed supported an effect on performance in sports. Nonetheless, although these acts of self-talk involved language production of some sort, there is no certainty that the measured effects were caused by the very act of producing language itself. Effects may also have been caused by participants' above average awareness of what they were saying and how they were saying it, a natural consequence of trying to integrate a new form of self-talk into their system. For example, increased awareness of instructional self-talk statements could have helped participants focus more on their task and on what they had to do, thus enhancing performance through heightened concentration. In the case of motivational self-talk, participants being aware of telling themselves that they 'can do it' could have encouraged them to convince themselves that they 'can do it', with that conviction altering their performance rather than the fact that they produced some corresponding utterances (i.e. a self-fulfilling prophecy, see Madon et al., 1997). A pure effect of language production should be one that is present

3 Language also enables *inner speech* (Vygotsky, 1987), or self-talk.

both when one is conscious of their choice of words, such as in the experiments described in this paragraph, as well as when one is not, such as in more typical everyday situations.

A second line of language production research has looked into the occurrence of a phenomenon which the original discoverers dubbed “Saying Is Believing” (Higgins & Rholes, 1978). In their experiment, they had participants read an essay about a stimulus person and summarize it for an addressee who, so they were told, either liked or disliked the person. This resulted in summaries biased according to the recipient attitude, which in turn would influence their own personal judgment of the stimulus person: after a delay, participants were asked to reproduce the contents of the original essay, as well as to rate the likability of the stimulus person, and it turned out that those who had written a relatively positive summary had a subsequent tendency to give more positive ratings and to reproduce a more positively sounding essay than the ones who had produced relatively negative summaries. Interestingly, this bias only turned up for the participants who actually *wrote* a summary, and not for the participants in the control condition who were merely instructed to write a summary but were told that they had been placed in the wrong experimental condition after they finished reading the essay. However, while the authors of the original paper saw their findings as a good reason to coin the term “Saying Is Believing” (SIB), later research showed that there might have been factors other than “saying” which could have caused “believing” in the original experiment. Echterhoff et al. (2005) tested the influence of audience feedback on SIB and discovered that participants who heard that the addressee failed to identify the stimulus person based on their summary did not exhibit SIB. Echterhoff et al. (2008) found that the type of communication goal also affected the occurrence of this effect, with participants who tuned their message to their audience with a goal such as being polite or obtaining a monetary reward showing no subsequent cognitive influences. Both articles propose that the effect discovered by Higgins and Rholes might not be caused by language production directly, but by the creation of a ‘shared reality’: the positive feedback given by the audience (or assumed positive feedback when no explicit feedback has been given) encourages belief in the contents of the produced message, and this increased belief might have been what actually led to an influence on memory and likability ratings.

One final study that has looked at the effects of language production on thinking dealt with emotion-related words. Oosterwijk et al. (2009) had participants come up with words associated either with ‘pride’ or with ‘disappointment’ while their posture was being measured. It turned out that participants decreased their posture height more during the production of disappointment words than during the production of words related to ‘pride’. Moreover, covariance analysis showed that it weren’t increased feelings of disappointment that mediated this effect. Then, what kind of mechanism could have caused these changes in posture? And is this mechanism separable from language, like the increases in concentration and the shared reality effects, or is it inherent to the language production process, i.e. an automatic, context-independent consequence of generating meaningful language?

1.3 Embodied Cognition

When a person produces language (e.g. “chair”), there is typically an underlying situation or concept (e.g. a mental image of a chair) that they have encoded linguistically in order to send it to another person who can decode the utterance and retrieve the original meaning (Schramm, 1954). There is thus a critical connection between language and meaning, or between language and knowledge about concepts in the world. Hence, in order to gain a full understanding of the processes involved in language production and perception, it is also necessary to gain an understanding of the way concepts are represented in the brain. Fodor (1975) argues for the existence of a ‘language of thought’ that stores this kind of information. He contends that any being that is able to make decisions in a particular situation must have a dedicated representational system in order to reason about the options available to them, and he believes that such a system would be like a ‘mental language’, since it shares some characteristic features with natural languages, such as productivity (Hockett, 1960). Fodor’s theory and similar proposals (see e.g. Pylyshyn, 1984) have been called *amodal* theories of knowledge since they suppose a representational system that is separate from the modality-specific areas of the brain, i.e. the areas that process input from the body’s senses (Barsalou, 1999, 2008; Niedenthal et al., 2005).

However, this has not been the only view of concept representation in the brain. If the process of encoding meaning into language or decoding it from language involved only the interaction with the part of the brain where this “metalanguage” (Fodor, 1975, p. 65) would be located, it would be difficult to explain the effects of several language perception studies discussed in Section 1.1, such as the study by Meteyard et al. (2007), who showed that motion detection could be impaired by hearing motion verbs that were directionally incongruent. Indeed, neuroimaging studies show that language processing activates brain areas specific to the meaning of the words being processed, such as auditory areas in the case of words with strong acoustic features (Binder & Desai, 2011; Kiefer et al., 2008) or motor areas for words referring to bodily actions (Binder & Desai, 2011; Hauk et al., 2004; Pulvermüller, 2005). Considering that such effects generally happened within 200 milliseconds after a word was presented to participants (Kiefer et al., 2008; Pulvermüller et al., 2005), it is unlikely that these activations are byproducts of language processing. Rather, they might be essential steps in the translation of words to concepts and vice versa, as proposed by theories of *embodied cognition* (Barsalou, 1999, 2008; Lakoff & Johnson, 1999; Niedenthal et al., 2005). Such theories assert that concepts are not stored as abstract symbols in a dedicated part of the mind, but rather in modality-specific systems, represented by a selection of the neural activations that took place during earlier encounters with each concept (Barsalou, 1999). For example, when a person encounters a ‘chair’, they may record visual information about what it looks like in the visual areas of the brain, acoustic information about the sound it makes when it moves across the floor in the auditory areas, tactile information about what it feels like to sit on a chair in the haptic areas, etc. Whereas amodal theories seem to suggest that these streams of information are then transduced into and stored as an abstract symbolic representation of ‘chair’ (Pylyshyn, 1984), embodied cognition theories propose that these streams are stored *as is* in the perceptual, motor and introspective areas of the brain where they were activated (Barsalou, 2008; Niedenthal et al., 2005). Then later, when a person wants to think about the ‘chair’ in its absence, they would be able to create an internal representation based on their prior experiences with the concept by reactivating the activation patterns that have been stored, a process called ‘simulation’ (Barsalou, 1999, 2008; Niedenthal et al., 2005). It has been theorized that this process of simulation is also used to understand linguistic utterances; when a person hears a sentence about an event that is not taking place around them at that moment, they can still comprehend what is being said by creating a mental simulation of this event, guided by the concepts and predicates contained in the sentence (Barsalou, 1999, 2008). Thus, if cognition were embodied, it would be expected that language processing required the activation of neural areas related to the meaning of the words involved, which would retrieve sensory, motor and introspective information associated with each concept. For instance, if a person who knew English heard words like ‘rise’ or ‘fall’, these words would activate ideas about upward or downward movement in the brain’s motor ares, which would inhibit the processing of concurrent motion in the opposite direction, thereby explaining the effect found by Meteyard et al. (2007). The results from Oosterwijk et al. (2009) can be explained in a similar way: by producing words related to ‘pride’ and ‘disappointment’, the body postures that co-define these concepts would be reactivated in the parts of the brain that handle proprioceptive information, which, in turn, could lead to unconscious physical re-enactment of those postures.

As mentioned before, there is an essential connection between language and conceptual knowledge, such that any successful decoding or any meaningful production of language must be accompanied by activating information about concepts. Hence, although a concept is in itself not a linguistic thing, any effect on cognition that appears to be caused by the activation of embodied information about concepts during language production or perception will be considered an effect of language on cognition for the purposes of this thesis, since activations of conceptual knowledge are inherent to the language production and perception processes. This contrasts with the mechanisms proposed as alternative explanations for some of the language production studies in Section 1.2, which are not inherent to processes related to language use (as shown, for example, by Echterhoff et al., 2005, 2008 for the shared reality effects). If a supposed effect of language production occurs only under specific circumstances, it is not an effect of language use, but rather an effect of that which distinguishes between the contexts in which the effect does and does not occur.

1.4 Current Experiment

It appears that embodiment provides a vehicle through which language production and perception might possibly affect cognition. However, other than the study performed by Oosterwijk et al. (2009), evidence that such effects occur in language production seems to be limited. This contrasts with the extensive support for such effects in language perception (e.g. by Francken et al., 2015; Lupyan & Ward, 2013; Meteyard et al., 2007; Ostarek & Huettig, 2017; Speed & Majid, 2019). Without more language production research, it is difficult to estimate whether embodiment effects are generally confined to the language perception process or whether such effects appear systematically in language production as well.

In order to address this issue and extend the body of available studies into the potential effects of language production on cognition, the experiment presented here has investigated a specific claim from the field of NLP: the effect of producing particular modal verbs on one's state of mind. Liekens (2005) argues that the use of the Dutch modal verb 'moeten' [*to have to*], which is associated with obligation and orders from others (Den Boon & Geeraerts, 2005), would activate a state of revulsion in the speaker, which would close them off from their enthusiasm and creativity. He suggests to replace usages of this verb with a different modal verb, 'willen' [*to want to*][4], associated with desires and personally decided goals (Den Boon & Geeraerts, 2005); this, he believes, would push the unconscious mind of the speaker to reach their goals as fast and as enjoyably as possible. To put this claim to the test, an experiment was set up that assessed short term cognitive effects after participants had produced a number of sentences that contained either 'moeten' (henceforth, the Extrinsic verb), 'willen' (the Intrinsic verb) or a Control verb with a meaning unrelated to a speaker's type of motivation. The modal verb used in the Control condition was 'zullen' [*will*], which is used to express that a speaker thinks it is highly probable that the event expressed in the sentence is true, as in "Marie zal een afspraak met hem maken" [*"Mary will make an appointment with him"*] (see Verkuyl & Broekhuis, 2013 for a more thorough explanation)[5].

Testing embodiment effects of language production in the domain of modal verbs has two main advantages for the broader field of functional linguistic relativity research. The first one has already been mentioned; many studies that have considered effects of language production cognition so far have been shown to be explainable using factors separate from the language production process (see Section 1.2), therefore, additional language production research may help answer the question of whether embodiment effects can occur both during the production as well as during the perception of language, or whether their appearance is generally limited to the latter process (with the exception of

4 Note that the English translations of these verbs, 'to want to' and 'to have to', are generally not considered to be modal verbs (Aarts, 2011), while their Dutch counterparts are (Haeseryn et al., 2019).

5 'Zullen' has also been described as the 'auxiliary of the future tense' (e.g. by Den Boon & Geeraerts, 2005). However, see Verkuyl and Broekhuis (2013) for arguments against this interpretation.

Oosterwijk et al., 2009). Although it has been argued in Section 1.3 that the activation of conceptual knowledge is inherent to both of these processes, there is a crucial directional difference between the two that could make for more evident cognitive influences during language perception than during its production: whereas the perception process activates concepts based on a linguistic message (Hemforth & Konieczny, 2006), which makes the question of which concepts are activated highly dependent upon the linguistic content, the production process is said to create a message based on an already activated idea (Levelt, 2001), which might lead to the expectation that the specific words chosen to represent this idea (e.g the Intrinsic vs. the Extrinsic verb) cannot cause separate embodiment effects if the underlying idea is the same. If, however, differences in word choice are reflected by differences in which concepts are activated mentally, language production should be able to affect cognition in a manner similar to what seems to happen during language perception.

The second way in which the present experiment is potentially relevant for answering broader research questions has to do with the type of words used to test for functional linguistic relativity effects. Contrary to many earlier studies that have investigated the cognitive influence of content words, such as nouns (e.g. Lupyan & Ward, 2013; Ostarek & Huettig, 2017; Speed & Majid, 2019) or non-auxiliary verbs (e.g. Francken et al., 2015; Meteyard et al., 2007), the current experiment elicited modal verbs, which have been classed as function words (Fries, 1952). Considering that function words generally have a less clearly defined lexical meaning than content words (Fries, 1952), there exists a possibility that they are also less strongly embodied than the latter. Finding out whether modal verbs can show embodiment effects could therefore aid in understanding whether such relatively abstract words (Brysbaert et al., 2014) are represented in the brain similarly to words from more concrete lexical classes.

If the production of the Intrinsic or the Extrinsic verb shows effects of embodiment and results in the mental states Liekens (2005) claims they activate, there are four potential ways in which such effects could be made visible. Firstly, since using the Intrinsic verb expresses an internal desire to do something whereas the Extrinsic verb typically conveys that something has been ordered by someone else (Den Boon & Geeraerts, 2005), it could be expected that those who produce the former verb experience increased intrinsic motivation, while users of the latter show decreases in their intrinsic motivation compared to the Control group (see hypotheses 1a and 1b below). Secondly, considering that an internal desire usually points at something that someone would do out of free will, while the Extrinsic verb describes things that a speaker feels obligated to do (Den Boon & Geeraerts, 2005), those who use the Intrinsic verb may be expected to feel more autonomous than the Control group whereas for the Extrinsic verb users, the opposite may hold (hypotheses 2a and 2b). Thirdly, it has been demonstrated that intrinsically motivated people persevere longer at optional tasks than extrinsically motivated people do (Deci et al., 1999); therefore, a possible consequence of the mental states activated by the Intrinsic and the Extrinsic verb could be that they increase or, respectively, decrease persistence at an experimental task in comparison to the Control group (hypotheses 3a and 3b). Lastly, high levels of intrinsic motivation have been shown to correlate with better creative problem solving (Glucksberg, 1962); hence, it might be expected that the use of the Intrinsic verb makes it easier to perform well at tasks that require creativity, whereas producing the Extrinsic verb could make it harder to come up with creative solutions compared to the Control group (hypotheses 4a and 4b). To summarize: if producing particular modal verbs can have differential effects on cognition, the following hypotheses would be expected to be true:

H1a: The difference between pre-test and post-test Intrinsic Motivation will be most positive (i.e. highest increase or lowest decrease) in the Intrinsic group, compared to the two other experimental groups (Extrinsic and Control).

H1b: Out of the two remaining groups, the difference between pre-test and post-test Intrinsic Motivation will be more positive in the Control group than in the Extrinsic group.

H2a: There will be a more positive difference between pre-test and post-test Sense of Autonomy in the Intrinsic group than in the other two groups.
H2b: Participants from the Control group will experience a more positive difference between pre-test and post-test Sense of Autonomy than those from the Extrinsic group.

H3a: The Intrinsic group will display more Persistence than the two other groups.
H3b: The Control group will display more Persistence than the Extrinsic group.

H4a: The Intrinsic group will show the highest Creative Performance out of all three experimental groups.
H4b: The Control group will show a higher Creative Performance than the Extrinsic group.

To test the above hypotheses, a digital experiment was set up that included two tasks and a questionnaire that was issued once before and once after the two tasks. During the first task, participants had to create and type in sentences by combining sets of words and pictures, with some sets requiring that participants used one of the three modal verbs, depending on the condition they had been placed in. This was followed by a linguistic puzzle solving task, which was used to test hypotheses 3 and 4. Finally, during the pre-test/post-test questionnaire, participants rated statements about their Intrinsic Motivation and Sense of Autonomy in relation to the experiment (e.g. “I would describe these exercises as very interesting”) on a scale of 1 to 7; this would serve as a basis for testing hypotheses 1 and 2. The experiment was designed in such a way as to exclude alternative explanations as much as possible: there should have been no effects of participants becoming more concentrated on the experiment or convincing themselves that they are or are not intrinsically motivated as a result of producing sentences about the experiment itself, since the words and pictures that had to be combined pushed participants to create sentences about entirely unrelated topics (e.g. swimming). Moreover, participants had to type out their sentences to a digital screen without a clear audience to address, which should have prevented any effects of shared realities. This way, the experiment was maximally focused on testing if cognition could be affected by the language production process itself.

2. Methodology

2.1 Research design

The aim of this study was to find out whether variation in the production of modal verbs could cause effects on cognition. A between subjects design was used, with participants being part of either (1) the Intrinsic group, who produced a Dutch modal verb associated with free will and an internal source of motivation (“willen”), (2) the Extrinsic group, who produced a verb linked to obligation and externally founded motivation (“moeten”) or (3) the Control group, who produced a modal verb denoting high probability (“zullen”). In order to measure different components of cognition expected to be influenced by the incorporated modal verbs, four different outcome variables were used: (1) Intrinsic Motivation, (2) Sense of Autonomy, (3) Persistence and (4) Creative Performance. The former two variables are self-reported measures, based on the difference between the pre-test and post-test Likert ratings of statements related to each scale. The latter two were constructed using results of a linguistic puzzle solving task, with Creative Performance relating to the number of solutions participants could come up with and Persistence to the amount of time spent thinking after participants had filled in their last answer. In addition, initial self-reported Competence was included as a control variable.

2.2 Participants

Participants for the online experiment were recruited using a combination of convenience sampling and snowball sampling. No compensation was given out to those who participated. People within the network of the author who didn't yet know about the research purpose were given an anonymous link, which allowed them to partake in the experiment using their phone, computer, laptop or tablet. Upon finishing the experiment, participants would get a message asking them to spread the link to more people they knew. In addition, the link was also spread through a post on the author's LinkedIn, via a tweet on the supervisor's Twitter, and by distributing flyers in the author's hometown.

During the four-week period in which the experiment was open, a total of 191 participants finished it completely. However, the data of three participants had to be removed: one had produced less than 8 sentences in which the verb corresponding to her experimental condition was used in its most prototypical word sense, one displayed excessive awareness of the research goal and one reached the 240 seconds time limit for the linguistic puzzles multiple times in a row without performing any clicks, indicating that she had been distracted.

The final sample thus consisted of 62 people in the Intrinsic group, 64 in the Extrinsic group and 62 in the Control group, making for a total of 188 individuals who were included in the analysis. Participants were between 16 and 75 years old (M = 38.5, SD = 16.1), encompassing 47 males, 137 females, 1 person who identified themselves in a different manner and 3 people who did not disclose any information about their gender. None of the participants reported having been diagnosed with color blindness, and all were native speakers of Dutch, although 11 respondents reported having more than one native language.

2.3 Materials

The experiment that respondents took part in consisted of three main tasks: (1) a sentence generation task, (2) a linguistic puzzle solving task and (3) two questionnaires that were issued before and after these two tasks. Each of these components required separate materials, which will be discussed in order below.

The first component was the sentence generation task, which was used to elicit up to 10 uses of the modal verb corresponding to participants' experimental condition ("willen" [*to want to*] in the Intrinsic group, "moeten" [*to have to*] in the Extrinsic group and "zullen" [*will*] in the Control group). This was done by asking participants to create sentences using sets of stimuli that were different for each item but always included either the condition-dependent verb, or a filler verb ("blijven" [*to keep*] or "gaan" [*to go*]). With 10 experimental items and 10 filler items, each participant would produce 20 different sentences in total. Besides a verb, each set of stimuli also contained an image of a color, an image depicting an action, a subject pronoun which the verb from the set was inflected to agree with, and a word limit for the sentence to be produced, which could be either a maximum limit (e.g. 10 or less) or a minimum limit (e.g. 16 or more). To illustrate this with two examples: one of the filler items asked participants to combine an image of the color "purple", an image of someone "stirring" the pot, the pronoun "zij" [*she*] and the verb "blijft" [*keeps*] while using at least 12 words; one of the experimental items asked participants to create a sentence that integrated an image of the color "grey", an image of someone "swimming", the pronoun "jij" [*you, singular*] and one of the verbs "wilt" [*want to*] / "moet" [*have to*] or "zult" [*will*], while using at most 9 words. To have a look at the sets of stimuli for the other 18 items, please see Appendix I.

The sentence generation task was set up in such a way as to strike a balance between, on the one hand, approaching the process of language production and "thinking of what to say" as closely as possible by allowing creative freedom, and on the other hand, reducing inter-participant noise and ensuring that the condition-dependent verb was elicited by limiting creative freedom. In addition, it was vital that participants did not figure out what the real purpose of the task was. This balancing act determined the choice for this specific task and the specific stimuli that have been shown to the respondents. Where possible, images have been used rather than words in order to mimic real life situations in which the words needed for a sentence are not readily available either, but have to be obtained by "translating" a mental concept into its linguistic form (Levelt, 2001). The experiment included 10 different color images for creativity purposes, and these images were balanced in such a way that each of them was coupled with exactly one filler item and exactly one experimental item. The color images were created in a simple graphics editor program using shades judged by the author to be prototypical of the 10 basic colors "black", "brown", "red", "orange", "yellow", "green", "blue", "purple", "pink" and "grey". Action images were included because the verbs to be elicited (both experimental and filler) were all auxiliaries, which need a second non-auxiliary verb in order to form a complete sentence, and specifically incorporating action images that would be translated into verbs was expected to keep participants from using the auxiliary verbs in a non-auxiliary way. The 20 different action images were taken from the Object and Action Naming Battery (Druks & Masterson, 2000). Subject pronouns were added to the sets of stimuli to keep the verbs from standing out too much (and thus revealing the manipulation), as otherwise they would have been the only stimuli that were shown using a word rather than an image. Apart from the third person singular neutral pronoun, "het" [*it*], which was excluded to limit epistemic uses of some of the condition-dependent verbs, all other nominative forms of Dutch personal pronouns could show up in the subject pronoun slot, which resulted in the following 7 pronouns appearing two to three times each over the course of the task: "ik" [*I*], "jij" [*you, singular*], "hij" [*he*], "zij" [*she*], "wij" [*we*], "jullie" [*you, plural*] and "zij" [*they*][6]. Filler verbs were included to draw away attention from the experimental manipulation. The two non-modal auxiliary verbs, "blijven" [*to keep*] and "gaan" [*to go*] were linked to 5 items each and were chosen over other non-modal auxiliary verbs because of the similarities between the sentence structure associated with these two verbs and the structure of a typical Dutch modal verb sentence: contrary to many other auxiliaries and similar to most modal verbs, they do not require the particle "te" [*to*][7], and, in contrast to verbs describing sensory experiences, they do not require the introduction of a second agent (e.g. "her" in "I see her walking"). Finally, word limits were added to push participants to use different sentence structures instead of applying the same matrix sentence to each of the 20 items. Possible maximum word limits included 7, 8, 9, 10 or 11 words, possible minimum word limits were 12, 14, 16, 18 or 20 words, and these 10 limits were balanced in such a way that each of them occurred once during an experimental item and once during a filler item.

The second component that made up the experiment, was a linguistic puzzle solving task. This task, which followed right after the sentence generation task, was intended to measure the effect of differential modal verb production on two outcome measures: Creative Performance and Persistence. To this end, participants were presented with strings of letters (e.g. "areht") that had to be reordered to produce existing words (e.g. "heart", "earth", "hater"). These types of puzzles are called *anagrams*. Some studies that tested Persistence have used anagrams that had only one or zero solutions (e.g.

6 It may be noticed that "zij" is listed twice with different translations: [*she*] and [*they*]. While it's true that an isolated instance of this pronoun is ambiguous between these two meanings, different associated inflections on the finite verb typically solve this ambiguity. The fact that the stimulus verbs in the current experiment were inflected in accordance with the stimulus subject pronouns, was therefore considered to warrant treating them as two separate pronouns.

7 There is one modal verb that does require the particle "te": "hoeven" [*don't have to*]. However, this is an exception.

Aspinwall & Richter, 1999; Gignac & Wong, 2020; Van Sintemaartensdijk & Righetti, 2019). However, because the current task was meant to assess not just Persistence but also Creative Performance, a different set-up was chosen in which participants weren't limited to finding "the" one solution, but were instead asked to type in as many solutions as they could think of (cf. Faddegon et al., 2009). To reflect this approach, the material for this task was designed to include anagrams that had between 2 and 5 solutions (although additional solutions found by participants later raised the upper limit to 7). In order to design this set of anagrams, 40 strings of 3 to 5 letters that had the required number of solutions were collected and subjected to a small pre-test to determine their relative difficulty. The collection of these initial anagrams and their solutions was done using a Python script that generated anagrams based on words from the *Corpus of Spoken Dutch*, or CGN (Oostdijk, 2000); this script can be found under Appendix II. Word frequency information of each solution provided by the same corpus was used as one of the two measures of anagram difficulty. The other measure of difficulty was obtained by asking three of the author's family members to solve the anagrams. Any anagram that had at least one solution that none of the family members had thought of and at least one solution (not necessarily the same one) with a CGN word frequency below 10, was considered to be "hard". Anagrams to which neither of these conditions applied, were considered "easy". If an anagram met one of the conditions but not the other, it was considered to be of "intermediate" difficulty. A final set of 10 anagrams was then extracted from these 40 initial strings of letters that was balanced for difficulty by incorporating 3 easy anagrams, 4 intermediate anagrams 3 hard ones. In addition, it was also balanced for word length by including 5 anagrams with a length of four letters and 5 anagrams with a a length of five letters; it was decided not to use anagrams consisting of three letters as it was deemed too easy for participants to see that such an anagram had no possible solutions left and this was expected to inhibit their Persistence. The set was also intended to be balanced for number of solutions, with 2 anagrams having two solutions, 3 anagrams having three solutions, 3 anagrams having four solutions and 2 anagrams having five solutions. However, this last balance did not hold as some participants with abundant vocabularies came up with additional solutions that had not appeared in the CGN. To view the final set of anagrams, including their initial solutions and the additional ones found by participants, see Appendix III.

The final component that was present in the experiment was a questionnaire that was issued twice to each participant; once before starting the two tasks described above and once after finishing them. This questionnaire contained 6 statements related to Intrinsic Motivation (e.g. "I quite enjoyed these exercises") and 6 statements related to Sense of Autonomy (e.g. "I did these exercises because I had no choice") that participants had to rate on a scale of 1 to 7 for how much they applied to them. By probing ratings both at the start and at the end of the experiment rather than performing only a single post-test measurement, the experiment intended to control for the influence of participants' initial Intrinsic Motivation and Sense of Autonomy on their final scores for these constructs. To obtain the 12 statements, items from the Intrinsic Motivation Inventory (IMI; Ryan, 1982) related to the subscales of Interest/Enjoyment and Perceived Choice were translated and adapted to measure respectively participants' Intrinsic Motivation and Sense of Autonomy in relation to the current experiment. For each statement, two versions were written that differed slightly in their wording and their use of tense, with one version being created for use in the pre-test and one for use during the post-test (e.g. "I think I will be quite enjoying these exercises" vs. "I quite enjoyed these exercises"). In addition, 5 further statements (e.g. "I think I am very good at these kinds of exercises") were added only to the pre-test which were based on a third subscale of the IMI, Perceived Competence. The ratings of these statements would be used to yield a score for the control variable, Competence, and were taken at the start of the experiment to prevent influences of the experimental manipulation. The 17 statements that thus constituted the pre-test and the 12 statements that formed the post-test can be found under Appendix IV.

2.4 Procedure

The experiment was conducted using the online survey platform Qualtrics (http://www.qualtrics.com). Participants received an anonymous link through which they could take part in the experiment using their phone, computer, laptop or tablet, without a researcher overseeing the process. If anything came up while participants were taking the experiment, they could pause it and return to where they left off at a later moment, provided the time between closing the experiment and revisiting it did not exceed the limit of one week.

Upon opening the link, participants were met by some general information about the contents of the experiment as well as essential ethical information such as their right to stop the experiment at any time. They were told that they would be partaking in a study ostensibly about ‘linguistic creativity’ and were given a short description of the two tasks that they would be doing. After reading this information, participants were asked for their consent.

If respondents agreed to take part in the experiment, some introductory questions followed that asked them about their age, their gender, whether or not Dutch was (one of) their native language(s) and whether they had ever been diagnosed with color blindness. Those who weren’t native speakers of Dutch or mentioned they had been diagnosed with color blindness were notified that they failed to meet the requirements for participation and were redirected to the end of the survey. All other participants could go to the next part of the experiment.

After the introductory questions, respondents received the pre-test questionnaire. 6 statements relating to Intrinsic Motivation (e.g. “I think I will be quite enjoying these exercises”), 6 statements relating to Sense of Autonomy (e.g. “I do these exercises because I have no choice”) and 5 statements relating to Competence (e.g. “I think I am very good at these kinds of exercises”) were shown in a pre-randomized fixed order, along with sliders that respondents could move to indicate how well the statements applied to them on a scale of 1 to 7. Participants were told that 7 indicated full agreement (“helemaal mee eens”), 4 indicated partial agreement (“een beetje mee eens”) and 1 indicated full disagreement (“helemaal mee oneens”).

When the first questionnaire was finished, participants were automatically assigned to a random experimental group (Intrinsic, Extrinsic or Control) and received the corresponding version of the sentence generation task. Each version consisted of 20 items, shown one by one in a random order, in which respondents were told to create and type in sentences that incorporated an item-specific (1) color, (2) action, (3) personal pronoun and (4) modal/non-modal auxiliary verb, while also keeping with a specific (5) minimum or maximum word limit for the sentence. Color and action requirements were shown using images, which participants had to convert into words in order to put them into their sentence; all other requirements were communicated verbally, with the personal pronoun and the auxiliary verb being presented as a single combined unit (e.g. a sentence that had to contain the pronoun “he” and the verb “go” was presented as having to contain the phrase “he goes”). An example: one of the items showed an image of a black color, an image of someone drinking, the phrase “Ik blijf” [*I keep*] and a minimum word limit of 14 words; a possible sentence that participants could come up with was “I keep drinking coffee from my black mug because the hot weather makes me thirsty” (15 words). In addition to images and text, slides from the sentence generation contained a dynamic word counter that updated as participants typed their sentence.

When a respondent finished typing in a sentence and submitted their answer, their sentence was subjected to a quick automatic validation process. If this validation failed, respondents would be sent back to alter their answer until it complied with the validation requirements. This validation process checked (1) if participants had not submitted empty answers and (2) in about the first half of the participants, it also checked if they had conformed to an item's individual word limit. The automatic validation did not check whether the colors, actions, pronouns or verbs had been included; this was done to prevent participants from receiving erroneous warnings if they had used different words to describe the color/action image or if they had made a spelling mistake. Thus, in theory, respondents could get away with writing sentences that did not contain the required stimuli[8]. The second half of the participants could also get away with answers that did not conform to the word limit. This was done because the regular expression[9] used to check for word limit violations had some issues dealing with answers containing a break (<enter>), much to the frustration of a small number of participants in the first half who kept getting sent back to alter their genuinely correct sentences, which caused the decision to remove the word limit validation altogether to prevent further irritation.

Other than the slight mid-way change to the experiment described above and the fact that the order in which the items appeared was randomized, all people who participated in the experiment received a mostly similar version of the sentence generation task. There was, however, one difference, depending on the experimental group a participant was assigned to: those from the Intrinsic group would encounter 10 items that asked them to use a form of the verb “willen” [*to want to*], whereas participants in the Extrinsic and Control groups were asked to incorporate “moeten” [*to have to*] or “zullen” [*will*], respectively, for those same items instead. This was the experimental manipulation. There were no other differences between participants of different experimental groups.

Upon finishing the sentence generation task, participants were taken to a linguistic puzzle solving task in which they were asked to write down as many words as possible that could be made by rearranging the letters of an anagram. 10 such anagrams were included in the task and they appeared in a fixed order to ensure that easy, intermediate and hard anagrams were distributed similarly for all participants, as differences in distributions were expected to affect Persistence. The order in which the anagrams appeared was: Intermediate, Intermediate, Easy, Intermediate, Hard, Intermediate, Hard, Easy, Hard, Easy (see also Appendix III). For each anagram, participants were provided with 6 answer fields to type in any solutions they could come up with. When building the experiment, the maximum number of solutions to any item from the set of 10 anagrams was expected to be 5 (although this would later turn out to be 7), and so the sixth field was added to give participants always a reason to persist until they couldn't come up with any answers anymore. To prevent particularly eager participants from spending excessive amounts of time on this task, each item also contained a hidden timer that would automatically move respondents to the next item if they had spent 240 seconds on the same anagram.

When participants finished the last anagram, a post-test questionnaire followed in which participants had to rate 6 statements related to their Intrinsic Motivation and 6 statements related to their Sense of Autonomy on a scale of 1 to 7, similar to the questionnaire at the start of the experiment. The statements were shown in a fixed order that had been randomized before the experiment was opened. This order differed from the order in which the pre-test statements had appeared.

After completing the questionnaire, participants received one last question that asked them what they thought the purpose of the research was. This was followed by a final statement that debriefed

8 In practice, this rarely happened. After the experiment, participants' responses to the 10 experimental items were read by the author to count the number of items in which the condition-dependent verb was used in its prototypical word sense. All of these sentences appeared to be genuine attempts at complying with the requirements. This also follows from the fact that only one participant had to be removed for producing less than 8 prototypical instances of her condition-dependent verb.

9 A regular expression is a pattern of characters that can be compared against a piece of text to see if there's a “match”, i.e. the piece of text contains the regular expression. For example, the regular expression “ook” matches with “book” and “looks” but not with “okay” or “kangaroo”.

them about the true goal of the experiment, asked them if they were willing to spread the experiment link to other people and provided the e-mail address of the researcher in case they had any questions or were curious about the results. This statement was the only piece of text in the experiment that contained modal verbs, apart from the phrases in the sentence generation task and two statements related to Sense of Freedom in the questionnaires ("I do these exercises because I **have to**" and "I do these exercises because I **want to**"). All other text had been carefully written not to include modal verbs in order to maximize the effect of modal verb production in the sentence generation task. To see how the instructions for the different parts of the experiment were formulated without the availability of modal verbs, please consult Appendix V.

2.5 Data Analysis

The data gathered in this experiment was analyzed using R (Version 3.6.3; R Core Team, 2020). All code can be found in Appendix VI. The statistical analysis essentially consisted of three parts: (1) Factor Analysis, (2) Main AN(C)OVAs and (3) Alternative ANOVA.

The first step involved conducting a Factor Analysis on the items of the pre-test questionnaire to ensure that the three factors that would be drawn from it (Intrinsic Motivation, Sense of Autonomy and Competence) had sufficient inter-item reliability. This analysis used the Principal Component Analysis method with oblimin rotation and was conducted using the R package "psych" (Version 2.1.3; Revelle, 2020). While the original IMI has been subject to tests of validation and reliability in the past (e.g. by McAuley et al., 1987), the version employed in this study used translated items and included only three out of its six subscales. On top of that, items had been reformulated to create pre-test counterparts, while the original statements had been intended to be used as post-test measures of various constructs related to Intrinsic Motivation. All this warranted additional reliability measures.

After performing a Factor Analysis, the four outcome variables and the control variable were constructed and four Analyses of (Co)variance were performed at an α-level of .0125 (Bonferroni correction). Construction of the analysis variables went as follows: scores for Intrinsic Motivation and for Sense of Autonomy were created by taking the average difference between pre-test and post-test Likert ratings of statements related to each of the concepts; a Competence score was derived by averaging Likert ratings of statements related to competence, which were only present in the pre-test; Persistence was calculated by averaging the amount of time spent between a participant's penultimate click on one of the anagram slides (which was considered to be them clicking the answer box to fill in the last solution they could come up with) and their final click on the button to go to the next page; and Creative Performance was equated to the total number of *correct* solutions participants came up with for all the anagrams combined, with a solution being correct if it or an inflected form appeared in the Van Dale Great Dictionary of the Dutch Language (Den Boon & Geeraerts, 2005). These variables were then subjected to tests of assumptions related to the four ANCOVAs that were planned to analyze the effect of Verb group on each of the four outcome measures whilst controlling for participant's Competence. These tests showed that two assumptions had been violated by some of the intended analyses: firstly, there was no significant linear correlation between the covariate and the different levels of the outcome measures Intrinsic Motivation, Persistence and Performance; secondly, the residuals for the Intrinsic Motivation, Sense of Autonomy and Persistence scores were significantly non-normally distributed. A lack of linear influence from the covariate was solved by removing Competence as a control variable and thereby turning the affected analyses into regular ANOVAs, while normality assumptions were solved by creating a new bootstrapped distribution consisting of 1000 subsamples from the original outcome measures; see Table 1 for an overview of what the final analyses looked like. AN(C)OVAs were conducted using the packages "stats" (Version 3.6.3; R Core

Team, 2020), and “car” version (Third edition; Fox & Weisberg, 2019), while bootstrapping was made possible by the “boot” package (Version 1.3-24; Canty & Ripley, 2019).

Table 1: Overview of the four statistical tests that made up the main analysis.

Outcome Variable	Type of Analysis	Was bootstrap used?	α-level	Description
Intrinsic Motivation	ANOVA	yes (1000 subsamples)	0.0125	Testing the effect of Verb group on a bootstrapped sample of Intrinsic Motivation scores.
Sense of Autonomy	ANCOVA	yes (1000 subsamples)	0.0125	Testing the effect of Verb group on a bootstrapped sample of Sense of Autonomy scores, whilst controlling for participant’s Competence.
Persistence	ANOVA	yes (1000 subsamples)	0.0125	Testing the effect of Verb group on a bootstrapped sample of Persistence scores.
Creative Performance	ANOVA	no	0.0125	Testing the effect of Verb group on Performance.

In addition to the four Analyses of (Co)variance described above, an alternative exploratory ANOVA was performed to test whether the delay between the intervention (first task) and the measurement of the outcome variables (second task and post-test questionnaire) had had any influence on the results. This analysis involved testing if there were any differences in the time needed by participants to come up with sentences containing the Intrinsic, Extrinsic or Control verb during the first task. Because this new outcome measure violated the same two assumptions as the ones mentioned in the description of the main analysis, the assertion was tested using a regular ANOVA on a set of 1000 bootstrapped subsamples of the sentence generation time variable.

3. Results

3.1 Factor Analysis

A Factor Analysis was performed to test if the 17 statements from the pre-test questionnaire could be reliably explained by three psychological constructs. The initial loadings are shown in Table 2 below. To see the statements corresponding to each statement number, see Appendix IV. It was hypothesized (based on the original IMI) that 6 specific statements would correlate strongly to the Intrinsic Motivation factor, 6 other statements to the Sense of Autonomy factor and yet 5 other statements to the Competence factor. Indeed, the results show that the statements unanimously loaded highest on their hypothesized constructs. However, these loadings were less than .6 for a number of statements, which have been given a red color in Table 2. Some of them also showed very high communality values. The red statements have been removed from further analysis; the scores for Intrinsic Motivation and Sense of Autonomy would thus both be based on 3 rather than 6 statements, while Competence kept its 5 statements. The loadings after removing these statements are shown in Table 3. Deleting these items resulted in increased reliability for Intrinsic Motivation (from α = .80 to α = .82) and Sense of Autonomy (from α = .79 to α = .86), while Competence was already quite reliable (α = .85).

Table 2: Loadings for 17 statements from the pre-test questionnaire on three hypothesized factors. Statements shown in red were removed after this factor analysis.

Statement no.	**Hypothesized Construct**	**Factor: Intrinsic Motivation**	**Factor: Sense of Autonomy**	**Factor: Competence**	**Communality**
1	Intrinsic Motivation	**0.45**	0.4	0.03	2
2	Intrinsic Motivation	**0.56**	0.23	0.14	1.5
3	Intrinsic Motivation	**0.82**	-0.12	0.13	1.1
4	Intrinsic Motivation	**0.86**	0.01	-0.08	1
5	Intrinsic Motivation	**0.38**	0.08	0	1.1
6	Intrinsic Motivation	**0.78**	0.03	0.06	1
7	Sense of Autonomy	0.13	**0.56**	-0.05	1.1
8	Sense of Autonomy	-0.07	**0.86**	0.01	1
9	Sense of Autonomy	0.16	**0.31**	0.18	2.2
10	Sense of Autonomy	0.44	**0.48**	-0.04	2
11	Sense of Autonomy	0.03	**0.8**	0.03	1
12	Sense of Autonomy	0	**0.87**	0	1
13	Competence	0.13	0.07	**0.66**	1.1
14	Competence	-0.17	0.27	**0.71**	1.4
15	Competence	-0.05	-0.03	**0.87**	1
16	Competence	0.08	0.02	**0.73**	1
17	Competence	0.1	-0.13	**0.87**	1.1

Table 3: Loadings for 11 statements from the pre-test questionnaire on three hypothesized factors, after removing statements that failed to load > .6 on their hypothesized factors.

Statement no.	**Hypothesized Construct**	**Factor: Intrinsic Motivation**	**Factor: Sense of Autonomy**	**Factor: Competence**	**Communality**
3	Intrinsic Motivation	**0.81**	-0.05	0.13	1.1
4	Intrinsic Motivation	**0.86**	0.08	-0.07	1
6	Intrinsic Motivation	**0.81**	0.1	0.06	1
8	Sense of Autonomy	-0.07	**0.85**	0.03	1
11	Sense of Autonomy	0.08	**0.85**	0.01	1
12	Sense of Autonomy	0.07	**0.92**	-0.03	1
13	Competence	0.13	0.04	**0.67**	1.1
14	Competence	-0.25	0.24	**0.74**	1.4
15	Competence	0	-0.02	**0.84**	1
16	Competence	0.06	-0.02	**0.76**	1
17	Competence	0.15	-0.11	**0.85**	1.1

3.2 Main Analysis

The main analysis consisted of four Analyses of (Co)variance, see Table 1. These analyses showed no evidence of an effect of verb group on either Intrinsic Motivation ($F(2,185) = .01$, $p = .987$, $\eta^2 < .01$), Sense of Autonomy ($F(2,185) = .31$, $p = .731$, $\eta^2 < .01$), Persistence ($F(2,185) = .92$, $p = .400$, $\eta^2 = .01$), or Performance ($F(2,185) = 1.93$, $p = .148$, $\eta^2 < .01$); see also Figure 1.

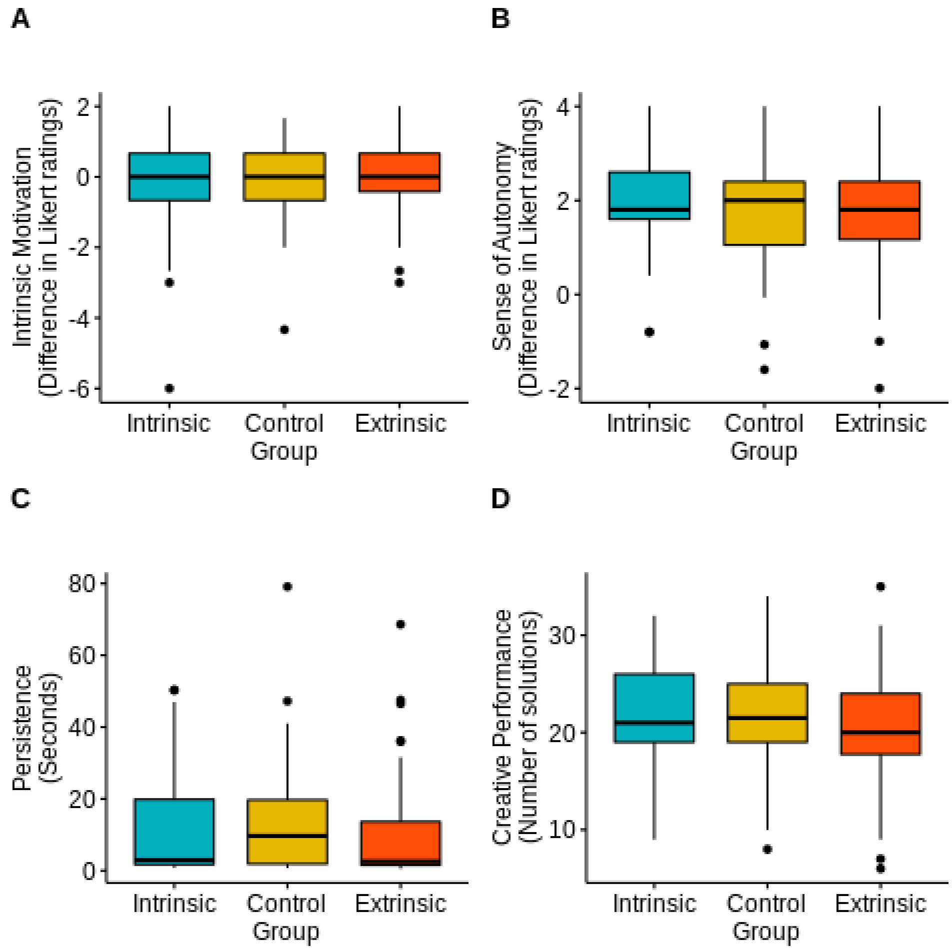

Figure 1: Scores for (A) Intrinsic Motivation, (B) Sense of Autonomy, (C) Persistence and (D) Creative Performance for participants who had produced an Intrinsic, Extrinsic or a Control verb.

3.3 Alternative Analysis

A separate analysis showed that there was no evidence ($F(2,185) = .50$, $p = .608$, $\eta^2 < .01$) that the time participants needed to come up with sentences containing their experimental verb differed as a result of producing either Intrinsic (M = 72.8, SD = 38.1), Extrinsic (M = 70.9, SD = 46.6) or Control verbs (M = 78.5, SD = 47.6).

4. Discussion

Whether or not language can affect our thinking and under which circumstances such effects could or could not occur, are two major questions that linguistic relativity researchers have attempted to answer over the years (Lucy, 2016). The experiment described in this thesis was administered in order to assess whether such effects might appear in the language production process and whether function words might bear the same cognitive power that has been attributed predominantly to content words so far (such as by Francken et al., 2015; Lupyan & Ward, 2013; Meteyard et al., 2007; Oosterwijk et al., 2009; Ostarek & Huettig, 2017; Speed & Majid, 2019). In doing so, it tested a specific claim from the field of NLP that modal verb production could influence a person's state of mind (Liekens, 2005). The three Dutch modal verbs used in this experiment were each associated with either intrinsic motivation/free choice ("willen" [*to want to*]), extrinsic motivation/will of others ("moeten" [*to have* *to*]) or something unrelated to a speaker's attitude ("zullen" [*will*])[10] (Den Boon & Geeraerts, 2005). If it's true that producing modal verbs can influence a speaker's state of mind, it was hypothesized that this would be detectable in the form of (1a) more positive differences between pre-test and post-test Intrinsic Motivation, (2a) more positive differences in Sense of Autonomy, (3a) higher levels of Persistence and (4a) more Creative Performance in the Intrinsic group than in the other two groups. Under the same condition, it was expected that a verb denoting obligation would yield opposite effects, such that the Extrinsic group would show (1b) more negative differences in Intrinsic Motivation, (2b) more negative differences in Sense of Autonomy, (3b) lower levels of Persistence and (4b) less Creative Performance than the Control group would.

Strictly speaking, the results show no support for any of the above hypotheses. The analyses of (co)variance for Intrinsic Motivation, Sense of Autonomy, Persistence as well as Creative Performance each failed to reach significance, advocating the absence of a relationship between the production of modal verbs and cognition. This finding is at odds with what was expected based on earlier research into similar topics (e.g. Francken et al., 2015; Lupyan & Ward, 2013; Meteyard et al., 2007; Oosterwijk et al., 2009; Ostarek & Huettig, 2017; Speed & Majid, 2019). Below, four theoretical explanations for this discrepancy will be reviewed: (1) language cannot affect cognition in general, (2) modal verb production in particular cannot affect cognition, (3) effects of modal verb production were too short-lived to be captured by the outcome variables or (4) the experiment lacked enough power to detect genuine effects.

The first theoretical possibility that could explain the lack of evidence for linguistic relativity effects found in the current experiment, would be that it is not possible for language to influence thought in general, or at least not in the functional way. Theories of embodied cognition that are supposed to underlie such effects (Barsalou, 1999, 2008) are still a topic of debate (Mahon, 2014). Additionally, a number of studies that have differentiated language production and measured effects on cognition have been shown to be explainable using language-external mechanisms (e.g. Hatzigeorgiadis et al., 2009; Higgins & Rholes, 1978; Slimani et al., 2014; Tod et al., 2001). However, a serious defense of this possibility would exceed the scope of this thesis. It has been mentioned here for theoretical purposes, but see Firestone and Scholl (2016) for arguments against language effects on perception, Mahon (2014) for arguments against embodied cognition and Chatterjee (2010) for alternative ways of interpreting evidence from language production and perception experiments.

A second possible explanation for the absence of evidence of linguistic relativity effects in the reported experiment would be that language *can* influence thought, just not in the particular way that the current research design has tried to test. It has been mentioned before that many studies into effects of language on the way we experience the world have focused on the structural reading (Lucy, 1997) of the linguistic relativity hypothesis (e.g. Davidoff et al., 1999; Levinson et al., 2003; Lupyan & Spivey, 2008; Winawer et al., 2007). Those that investigated functional linguistic relativity mainly seem to have focused on effects of language perception on cognition (e.g. Francken et al., 2015; Lupyan & Ward, 2013; Meteyard et al., 2007; Ostarek & Huettig, 2017; Speed & Majid, 2019). Studies investigating the influence of language production have been relatively scarce, and those who have been taken to do so did not always mention an intent to study functional linguistic relativity or embodied cognition effects specifically (e.g. Hatzigeorgiadis et al., 2009; Slimani et al., 2014; Tod et al., 2001). From this, one might attempt to conclude that cognitive influences of language are limited to language structures and the processes that occur during the perception and comprehension of language. Indeed, as has been argued in Section 1.4, language perception and language production are two distinct processes that might interact with conceptual knowledge in different ways, which has potential consequences for their

10 Although, as will be discussed later on, the meaning of the Control verb was not *entirely* unrelated.

ability to influence cognition; whereas the former translates linguistic utterances into concepts (Hemforth & Konieczny, 2006), the latter does the opposite and translates a target concept into a word (Levelt, 2001). Although some models of language production contend that it is a bidirectional process in which 'later' levels (e.g. word) can affect earlier ones (e.g. mental concept), this assertion has not been shared by all models (Warren, 2013). If it is considered to be a unidirectional process, there are indeed reasons to believe that the particular words used to formulate an idea may not result in differential effects on cognition during language production, while they may do so during language perception, when the question of which concepts are activated can be expected to depend more strongly on the language used. This may appear to provide an explanation for the discrepancy between the current study and the studies that showed support for effects of language perception on cognition (i.e. Francken et al., 2015; Lupyan & Ward, 2013; Meteyard et al., 2007; Ostarek & Huettig, 2017; Speed & Majid, 2019). However, some arguments against this point need to be discussed. Participants in the current study did not just produce sentences containing one of the three modal verbs, they also read the word corresponding to their experimental condition, once for every sentence they were asked to produce, as part of the set of stimuli that went along with each sentence[11]. If embodiment effects are more apparent in language perception than in language production, then how come the results of this experiment did not reflect an effect of reading the stimulus verbs? Furthermore, although studies that show support for functional linguistic relativity in language production appear not to be manifold, they do exist, such as the one by Oosterwijk et al. (2009) who found effects of producing words related to "pride" and "disappointment" on participants' mood and posture. The notion of linguistic influences on cognition being limited to effects of structure and effects during perception, thus poses some complications. Future language production research will hopefully shed some light on this.

Alternatively, it could be that it's not the focus on production of language that inhibited linguistic-relativistic effects, but the type of words used for the intervention. Part of the goal of this study was to see if function words could deliver effects similar to those of more commonly studied content words. The fact that the former typically alter other words in a sentence instead of bearing clear lexical content of their own (Fries, 1952) might make them harder to simulate, diminishing potential effects on cognition. An argument against this would be that Speed and Majid (2019) found that not just a word's lexical content, but also its more grammatical properties such as gender, can have an impact on cognition; in their experiments, they paired masculine and feminine fragrances with names of ingredients that were either masculine or feminine in their grammatical gender, and found that subsequent memory was affected by whether there was a match or a mismatch between an odor's marketed gender and the gender of the words in the description. It appears that clear lexical content may not be required for a word to affect cognition. At the same time, explaining how abstract concepts are represented in the brain has been known to be a challenge for theories of embodied cognition (Chatterjee, 2010). The study of abstract words thus remains an open field in which many questions are yet to be resolved. Perhaps more research into the effects of producing or perceiving abstract words and words with mostly structural meanings will help to find out how such words may be represented mentally and under which circumstances – if any – they may be expected to influence thinking.

A third possible way of explaining the findings of the current experiment is not by saying that either language production or function words may be unable to affect cognition, but rather that their effects might last for only a short time, and that the delay between the intervention (first task) and the test of effects (second task and post-test questionnaire) was too long to detect the effects of the mental states the Intrinsic and the Extrinsic verb have been claimed to induce (Liekens, 2005). Indeed, many of the functional linguistic relativity studies that have been reviewed here combine an intervention with

11 Which is, of course, a design flaw, although it was difficult to prevent.

an almost simultaneous measurement of its effects (e.g. Francken et al., 2015; Lupyan & Ward, 2013; Meteyard et al., 2007; Oosterwijk et al., 2009; Ostarek & Huettig, 2017). In contrast, not only did the present experiment contain a separation between the task used to elicit the experimental verb and the measures of its influence, it was also conducted via an online questionnaire that respondents could pause at any time, giving them up to one week to return and finish it; this may have resulted in even bigger delays for some participants. Figure 2 below shows how much time the participants that were included in the final analysis needed between first opening the survey and finishing it. Although none of the participants ended up taking an entire week – the biggest amount of time anyone took was 25 hours, 47 minutes and 33 seconds – there were quite a few respondents whose survey time was well above the median of 40 minutes and 6 seconds, suggesting that at least a part of the participants took an extended break somewhere during the experiment. To see if it had made any difference if a simultaneous measurement was used over delayed outcome measures, an exploratory ANOVA was conducted on the average time needed by each group to come up with sentences containing the condition-dependent verb in the sentence generation task, the results of which have been reported in Section 3.3. The fact that this did not result in significant between-group differences may be taken as a reason to believe that delays were not the real reason the experiment failed to find evidence for its hypotheses. However, considering that the sentence time variable was neither part of the original research design nor specifically included in the pre-formulated hypotheses, there may well be other reasons for why there were no significant differences in the time spent thinking of sentences. For instance, a potential increase in Intrinsic Motivation during the creation of a sentence with an Intrinsic verb may be reflected by either a *reduction* in the time needed by participants, who would become more eager to finish the item as soon as possible, or an *increase* in the time they spent, since they might have been extra intent on coming up with the best (e.g. most beautiful) sentence possible; such differential effects could have canceled each other out. Participants may have also been differentially distracted. Regardless, it cannot be fully ruled out that the delayed nature of the outcome measures may have played some role in weakening potential effects of modal verb production on cognition. One way in which a future study could address this issue, is by having participants create sentences with the Intrinsic or the Extrinsic verb and testing if this affects a dedicated simultaneous measurement, such as e.g. levels of the hormone cortisol in participants' blood, which is a biological indicator of stress (Levine et al., 2007). It could also be of interest to try and replicate some of the studies that did find effects of language production or perception on simultaneously taken outcome measures (e.g. Francken et al., 2015; Lupyan & Ward, 2013; Meteyard et al., 2007; Oosterwijk et al., 2009; Ostarek & Huettig, 2017), but while introducing delays of different durations and seeing how long such effects of embodiment remain active.

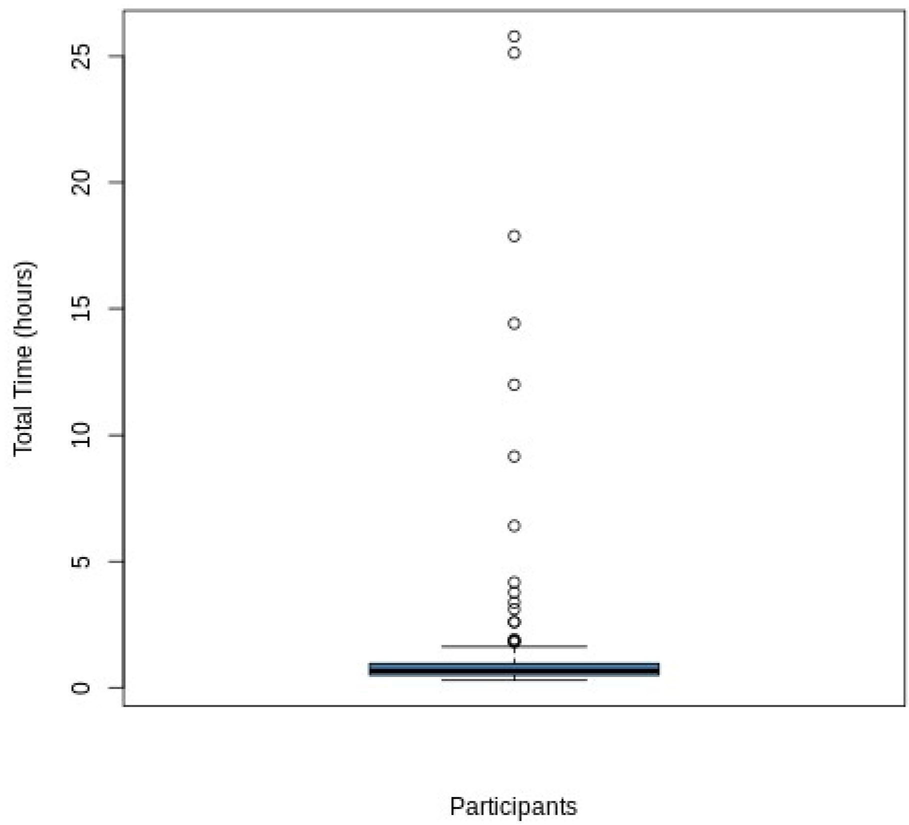

Figure 2: Time spent by participants between opening the survey for the first time and submitting their answers.

One final possibility that might explain why the experiment failed to find evidence for its hypothesized effects could have to do with the various kinds of experimental noise that influenced the outcome measures other than the experimental variable. The most severe one is most likely the fact that the experiment was not conducted in a controlled environment, but from wherever a participant happened to be when they opened the survey; not only does this introduce a great amount of variability in the kind of room a participant was in and the type of device on which the survey was made, there was also not a researcher nearby to ask questions to regarding any unclarities in the instruction texts. It turned out, for instance, that some participants in the anagram task thought that not all letters from the original word string had to be used to create solutions (consider e.g. an English anagram “lalb”, for which a correct solution would be “ball”, but in this case participants also submitted words such as “all” or “lab”). Since there were only 6 answer fields, expending them on answers that would never be deemed correct may have had a diminishing influence on their Creative Performance. If there had been a researcher nearby, misinterpreted instructions such as these could have been pointed out.

There were other sources of noise as well. Elaborating on all of them would make for a very long discussion section, but some prominent ones will be mentioned here shortly, in hopes that they may aid future researchers to create more robust research designs: (1) pre-test and post-test Likert ratings of statements relating to participants’ experience of the experimental tasks were compared, but participants only really got to know the tasks after the pre-test, (2) the experiment had a slight mid-way change concerning the removal of the automatic sentence length validation (see Section 2.4); not being confronted with a red warning if a sentence was incorrect may have influenced participants' Intrinsic Motivation and Sense of Autonomy, (3) Persistence was operationalized as the time between a participant's penultimate click on an anagram slide and their final click on the button to go to the next item, but there have been other, perhaps more effective ways to measure it, such as tracking the time spent on insolvable anagrams (Aspinwall & Richter, 1999; Gignac & Wong, 2020; Van Sintemaartensdijk & Righetti, 2019), (4) Creative Performance depended not only on participants' creativity but also on the size of their vocabulary and their willingness to use external sources (e.g. search engines) to find additional solutions and (5) the Control verb with a supposedly unrelated definition may not have been entirely neutral. The last point requires a bit more explanation: as was mentioned in Section 1.4, the Control verb, "zullen" [*will*], was chosen because it does not typically refer to a person's motivation to perform the action described in a sentence, but rather to express that a speaker thinks it is highly probable that the event expressed in the sentence is true, will be true or has been true (Verkuyl & Broekhuis, 2013). In this light, the following example sentence was mentioned briefly, which was taken from Verkuyl and Broekhuis (2013):

(1)	Marie	zal	een	afspraak	met	hem	maken.
	Mary	will	an	appointment	with	him	make.

Mary will make an appointment with him.

Sentence (1) outlines a prototypical use of "zullen" (inflected here as "zal"), in which the speaker expresses their strong belief that the mentioned event (Mary making an appointment) is going to happen sometime in the future. It does not seem as if a word that appears in sentences like these, should affect a speaker's own motivation. However, consider what happens when the speaker starts talking about themselves, as in (2):

(2)	Ik	zal	een	afspraak	met	hem	maken.
	I	will	an	appointment	with	him	make.

I will make an appointment with him.

Now, the speaker expresses a strong belief that they themselves will perform the action, sometime in the future, that had been ascribed to Mary in sentence (1). Since the production of a sentence like (2) commits the producer to its truth (Grice, 1975), it might be expected that the person who said this would make sure that that appointment will be made, even if they don't necessarily feel intrinsically motivated to do so. Intuitively, this commitment is of a much stronger kind than merely saying "I want to make an appointment" or "I have to make an appointment". This might therefore explain why participants from the Control group turned out to have the highest average Persistence (13.4 seconds vs. 11.2 seconds in the Intrinsic group and 10.0 seconds in the Extrinsic group); if the lack of evidence for linguistic relativity effects was due to excessive noise and genuine effects have been missed, the production of the Control verb may have activated a state of 'strong commitment' in the minds of participants that led them to be extra persistent in the anagram task that followed. In a future experiment, it may then be better to opt for a different verb for the Control condition; perhaps a non-modal auxiliary such as 'blijven' [*to keep*] or 'gaan' [*to go*] will do. In addition, such an experiment could be conducted in a controlled environment and it could seek to find solutions to the other sources of noises that have been mentioned, so that the fourth possible explanation of the results found in this study may also be scrutinized.

5. Conclusion

In this thesis, an experiment has been described that aimed to find out if support for effects of language on cognition could be found in the relatively unexplored region of language production, using a specific category of function words as opposed to the more generally used content words (e.g. by Francken et al., 2015; Lupyan & Ward, 2013; Meteyard et al., 2007; Oosterwijk et al., 2009; Ostarek & Huettig, 2017; Speed & Majid, 2019). In doing so, it tested a specific claim from the field of NLP that the production of modal verbs such as “willen” [*to want to*] or “moeten” [*to have to*] could affect a speaker’s mental state (Liekens, 2005). The lack of confirmation that was found for the postulated hypotheses has been discussed in light of four theoretical explanations: (1) language cannot affect cognition in general, (2) language production or function words specifically cannot affect cognition, (3) effects of modal verb production were too short-lived to be detected and (4) the experiment failed to capture genuine effects due to a plethora of noise. Some concrete suggestions for future linguistic relativity research have been made, such as investigating the duration of functional language effects and conducting additional research in the areas of abstract words and language production in general. However, the results of the experiment also have implications for the field of NLP: on the one hand, it appears that one of their linguistic claims may be hard to support empirically; on the other hand, however, there may be language-external mechanisms that have not been tested here, such as shared realities (Echterhoff et al., 2005, 2008) and self-fulfilling prophecies (Mahon et al., 1997), that may still enable effects of modal verb production under certain circumstances. Could Levelt (1996) have been right when he criticized NLP practitioners for falling into the ‘black hole’ of linguistic relativity (p. 27)? After 25 years, this question remains short of a definitive answer.

References

Aarts, B. (2011). *Oxford modern English grammar*. Oxford University Press.

Aspinwall, L. G. & Richter L. (1999). Optimism and self-mastery predict more rapid disengagement from unsolvable tasks in the presence of alternatives. *Motivation and Emotion, 23*(3). 221–245. https://doi.org/10.1023/A:1021367331817

Bandler, R. & Grinder, J. (1975). *The structure of magic I: A book about language and therapy*. Science and Behavior Books, Inc.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–660. https://doi.org/10.1017/s0140525x99002149

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617– 645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15*, 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Brown, R. (1976). Reference: In memorial tribute to Eric Lenneberg. *Cognition, 4*(2), 125–153. https://doi.org/10.1016/0010-0277(76)90001-9

Brysbaert, M., Warriner, A.B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*, 904-911. https://doi.org/10.3758/s13428-013-0403-5

Canty, A., & Ripley, B. D. (2019). Boot: Bootstrap r (s-plus) functions (R package version 1.3-24) [Computer software]. *The Comprehensive R Archive Network*. Retrieved from https://CRAN.R-project.org/package=boot

Chatterjee, A. (2010). Disembodying cognition. *Language and Cognition, 2*(1), 79–116. https://doi.org/10.1515/LANGCOG.2010.004

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature, 398*(6724), 203–204. https://doi.org/10.1038/18335

Deci, E. L., Koestner, R. & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic reward on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668. https://doi.org/10.1037/0033-2909.125.6.627

Den Boon, C. A. & Geeraerts, D. (2005). *Van Dale Groot woordenboek van de Nederlandse taal.* Van Dale Lexicografie.

Druks, J., & Masterson, J. (2000). *An Object and Action Naming Battery.* Psychology Press.

Echterhoff, G., Higgins, E. T., & Groll, S. (2005). Audience-tuning effects on memory: The role of shared reality. *Journal of Personality and Social Psychology, 89*, 257–276. https://doi.org/10.1037/0022-3514.89.3.257

Echterhoff, G., Higgins, E. T., Kopietz, R., & Groll, S. (2008). How communication goals determine when audience tuning biases memory. *Journal of Experimental Psychology: General, 137*, 3–21. https://doi.org/10.1037/0096-3445.137.1.3

Faddegon, K., Ellemers, N. & Scheepers, D. (2009). Eager to be the best, or vigilant not to be the worst: The emergence of regulatory focus in disjunctive and conjunctive group tasks. *Group Processes & Intergroup Relations, 12*(5), 653–671. https://doi.org/10.1177/1368430209339922

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences, 39*, 1–77. https://doi.org/10.1017/S0140525X15000965

Fodor, J. A. (1975). *The language of thought*. Thomas Y Crowell Company.

Fox, J., & Weisberg, S. (2019). An {R} Companion to Applied Regression (Third edition) [Computer software]. *The Comprehensive R Archive Network*. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Francken, J. C., Kok, P., Hagoort, P., & de Lange, F. P. (2015). The behavioral and neural effects of language on motion perception. *Journal of Cognitive Neuroscience, 27*(1), 175–184. https://doi.org/10.1162/jocn_a_00682

Fries, C. C. (1952). *The structure of English: An introduction to the construction of English sentences.* Harcourt, Brace & World Inc.

Gignac, G. E. & Wong, K. K. (2020). A psychometric examination of the anagram persistence task: More than two unsolvable anagrams may not be better. *Assessment, 27*(6), 1198–1212. https://doi.org/10.1177/1073191118789260

Glucksberg, S. (1962). The influence of strength of drive on functional fixedness and perceptual recognition. *Journal of Experimental Psychology, 63*(1), 36–41. https://doi.org/10.1037/h0044683

Grice, H. P. (1975). Logic and Conversation. In P. Cole, & J. L. Morgan. (Eds.), *Syntax and Semantics 3: Speech Acts*, pp. 41–58. Academic Press.

Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J. & van den Toorn, M.C. (2019, January). 18.5.4.4.i Inleiding. *Algemene Nederlandse Spraakkunst*. Retrieved from https://e-ans.ivdnt.org/topics/pid/ans1805040401lingtopic

Hauk, O., Johnsrude, I. & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*, 301–307. https://doi.org/10.1016/S0896-6273(03)00838-9

Hatzigeorgiadis, A., Zourbanos, N., Mpoumpaki, S., & Theodorakis, Y. (2009). Mechanisms underlying the self-talk-performance relationship: The effects of motivational self-talk on self-confidence and anxiety. *Psychology of Sport and Exercise, 10*, 186–192. https://doi.org/10.1016/j.psychsport.2008.07.009

Hemforth, B. & Konieczny, L. (2006). Language processing: Construction of mental models or more? In C. Held, M. Knauff & G. Vosgerau (Eds.), *Mental models in cognitive psychology, neuroscience and philosophy of mind*, pp. 189–204. Elsevier. https://doi.org/10.1016/S0166-4115(06)80035-X
Higgins, E. T., & Rholes, W. S. (1978). ‘Saying is believing’: Effects of message modification on memory and liking for the person described. *Journal of Experimental Social Psychology, 14*, 363–378. https://doi.org/10.1016/0022-1031(78)90032-X
Hockett, C. F. (1960). The origin of speech. *Scientific American, 203*(3), 88–97. https://doi.org/10.1038/scientificamerican0960-88
Jakobson, R. (1960). Linguistics and poetics. In T. Sebeok (Ed.), *Style in Language* (pp. 350–377). M.I.T. Press.
Kiefer, M., Sim, E., Herrnberger, B., Grothe, J. & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *The Journal of Neuroscience, 28*(47), 12224 – 12230. https://doi.org/10.1523/JNEUROSCI.3579-08.2008
Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. Basic Books.
Levelt, W. J. M. (1996). Hoedt u voor neuro-linguïstisch programmeren! *Skepter, 9*(3), 26–28.
Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. https://doi.org/10.1073/pnas.231459498
Levine, A., Zagoory, S., Feldman, R., Lewis, J. G., Weller, A. (2007). Measuring cortisol in human psychogiological studies. *Psychology & Behaviour, 90*, 43–53. https://doi.org/10.1016/j.physbeh.2006.08.025
Levinson, S. C., Kita, S., Haun, D. B. M. & Rash, B. H. (2002). Returning the tables: language affects spational reasoning. *Cognition, 84*. 155–188. https://doi.org/10.1016/S0010-0277(02)00045-8
Liekens, P. (2005). *Gevangen in de taal: Hoe woorden je kunnen beperken*. Ankh-Hermes.
Lucy, J. A. (1997). Linguistic relativity. *The Annual Review of Anthropology, 26*, 291–312. https://doi.org/10.1146/annurev.anthro.26.1.291
Lucy, J. A. (2016). Recent advances in the study of linguistic relativity in historical context: A critical assessment. *Language Learning, 66*(3), 487–515. https://doi.org/10.1111/lang.12195
Lupyan, G., & Spivey, M. J. (2008). Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Current Biology, 18*(10), R410–R412.
Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences, 110*(35), 14196–14201. https://doi.org/10.1073/pnas.1303312110
Madon, S., Jussim, L., & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology, 72*(4), 791–809. https://doi.org/10.1037/0022-3514.72.4.791
Mahon, B. Z. (2014). What is embodied about cognition? *Language, Cognition and Neuroscience, 30*(4). 420–429. https://doi.org/10.1080/23273798.2014.987791
McAuley, E., Duncan, T., & Tammen, V. V. (1987). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, *60*, 48-58. https://doi.org/10.1080/02701367.1989.10607413
Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs: Language affects low-level visual perception. *Psychological Science*, *18*(11), 1007–1013. https://doi.org/10.1111/j.1467-9280.2007.02016.x

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S. & Ric, F. (2005). Embodiment in attitudes, social perception and emotion. *Personality and Social Psychology Review, 9*(3), 184–211. https://doi.org/10.1207/s15327957pspr0903_1

Oostdijk, N. (2000). The spoken dutch corpus: Overview and first evaluation. *Proceedings of LREC-2000*.

Oosterwijk, S., Rotteveel, M., Fischer, A. H. & Hess, U. (2009). Embodied emotion concepts: How generating words about pride and disappointment influences posture. *European Journal of Social Psychology, 39*, 475-466. https://doi.org/10.1002/ejsp.584

Ostarek, M., & Huettig, F. (2017). Spoken words can make the invisible visible? Testing the involvement of low-level visual representations in spoken word processing. *Journal of Experimental Psychology: Human Perception and Performance, 43*(3), 499–508. https://doi.org/10.1037/xhp0000313

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature reviews Neuroscience, 6*(7), 567–582. https://doi.org/10.1038/nrn1706

Pulvermüller, F., Shtyrov, Y. & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience, 17*(6), 884–892. https://doi.org/10.1162/0898929054021111

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. The MIT Press.

R Core Team (2020). R: A language and environment for statistical computing (R version 3.6.3) [Computer software]. *The Comprehensive R Archive Network.* Retrieved from https://www.R-project.org/

Revelle, W. (2020). Psych: Procedures for psychological, psychometric, and personality research (R package version 2.1.3) [Computer software]. *The Comprehensive R Archive Network*. Retrieved from https://CRAN.R-project.org/package=psych

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43*(3), 450-461. https://doi.org/10.1037/0022-3514.43.3.450

Schramm, W. (1954). *The process and effects of mass communication*. University of Illinois Press.

Slimani M., Hentati, A., Bouazizi, M., Boudhiba, D., Ben Amar, I. & Chéour F. (2014). Effects of self-talk and mental training package on self-confidence and positive and negative affects in male kickboxers. *IOSR Journal of Humanities and Social Sciences*, *19*(5), 31-34. https://doi.org/10.9790/0837-19513134

Speed, L. J. & Majid, A. (2019). Linguistic features of fragrances: The role of grammatical gender and gender associations. *Attention, Perception & Psychophysics*, *81*, 2063-2077. https://doi.org/10.3758/s13414-019-01729-0

Tod D., Hardy J. & Oliver E. (2011). Effects of self-talk: A systematic review. *Journal of Sport and Exercise Psychology*, *33*, 666-687. https://doi.org/10.1123/jsep.33.5.666

Van Sintemaartensdijk, I. & Righetti, F. (2019). Who does most of the work? High self-control individuals compensate for low self-control partners. *Journal of Theoretical Social Psychology, 3*, 209–215. https://doi.org/10.1002/jts5.47

Verkuyl, H. J. & Broekhuis, H. (2013). Temporaliteit en modaliteit. *Nederlandse Taalkunde*, *18*(3), 306–323. https://doi.org/10.5117/NEDTAA2013.3.VERK

Vygotsky, L. S. (1987). Thinking and Speech. In R. W. Rieber, & A. S. Carton (Eds.), *The Collected Works of L. S. Vygotsky 1: Problems of General Psychology* (pp. 39-285). Plenum Press. (Original work published 1934)

Warren, P. (2013). *Introducing psycholinguistics*. Cambridge University Press.

Whorf, B. L. (1952). Language, mind and reality. *ETC: A Review of General Semantics*, *9*(3), 167–188. (Original work published in 1942)

Whorf, B. L. (1956). *Language, thought and reality: Selected writings of Benjamin Lee Whorf*. The Massachusetts Institute of Technology.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104

Appendix I: Items from the sentence generation task

Below is an overview of the 20 items that constituted the sentence generation task. Each sentence had to include a specific pronoun (column 1), a specific verb (column 2) inflected to agree with the pronoun (column 3), a specific action depicted on a picture from the Object and Action Naming Battery (Druks & Masterson, 2000; column 4), a specific color shown on a picture as well (column 5), while keeping with a certain minimum or maximum word limit (column 6). In addition, the task counted 10 filler items with a fixed non-modal auxiliary verb and 10 experimental items where participants were asked to use either "willen" [t*o want to*], "moeten" [*to have to*] or "zullen" [*will*]. This type is shown in column 7.

Pronoun	**Verb lemma**	**Verb inflected**	**Action (Items from Druks & Masterson, 2000)**	**Color**	**Word Limit**	**Type**
ik (*I*)	blijven (*keep*)	blijf (*keep*)	Item 24 ("drinken" / *to drink*)	Black	Min. 14	Filler
jullie (*you, plural*)	blijven (*keep*)	blijven (*keep*)	Item 40 ("kloppen" / *to knock*)	Pink	Max. 7	Filler
zij (*they*)	blijven (*keep*)	blijven (*keep*)	Item 74 ("schaatsen" / *to ice skate*)	Green	Max. 8	Filler
zij (*she*)	blijven (*keep*)	blijft (*keeps*)	Item 83 ("roeren" / *to stir*)	Purple	Min. 12	Filler
wij (*we*)	blijven (*keep*)	blijven (*keep*)	Item 91 ("typen" / *to type*)	Grey	Min. 16	Filler
hij (*he*)	gaan (*go*)	gaat (*goes*)	Item 15 ("oversteken" / *to cross*)	Yellow	Max. 9	Filler
wij (*we*)	gaan (*go*)	gaan (*go*)	Item 20 ("duiken" / *to dive*)	Brown	Max. 10	Filler
jullie (*you, plural*)	gaan (*go*)	gaan (*go*)	Item 28 ("eten" / *eat*)	Red	Min. 18	Filler
jij (*you, singular*)	gaan (*go*)	gaat (*go*)	Item 55 ("inschenken" / *to pour*)	Orange	Max. 11	Filler
zij (*she*)	gaan (*go*)	gaat (*goes*)	Item 90	Blue	Min. 20	Filler

			("strikken" / *to tie*)			
zij (*they*)	willen (*want to*) moeten (*have to*) zullen (*will*)	willen (*want to*) moeten (*have to*) zullen (*will*)	Item 8 ("bouwen" / *to build*)	Black	Min. 14	Experimental
ik (*I*)	willen (*want to*) moeten (*have to*) zullen (*will*)	wil (*want to*) moet(*have to*) zal (*will*)	Item 9 ("dragen"/"tillen" / *to carry*)	Brown	Min. 16	Experimental
jij (*you, singular*)	willen (*want to*) moeten (*have to*) zullen (*will*)	wilt (*want to*) moet (*have to*) zult (*will*)	Item 14 ("kruipen" / *to crawl*)	Red	Max. 7	Experimental
hij (*he*)	willen (*want to*) moeten (*have to*) zullen (*will*)	wil (*wants to*) moet (*has to*) zal (*will*)	Item 23 ("boren" / *to drill*)	Pink	Max. 8	Experimental
zij (*she*)	willen (*want to*) moeten (*have to*) zullen (*will*)	wil (*wants to*) moet (*has to*) zal (*will*)	Item 32 ("vouwen" / *to fold*)	Orange	Min. 12	Experimental
wij (*we*)	willen (*want to*) moeten (*have to*) zullen (*will*)	willen (*want to*) moeten (*have to*) zullen (*will*)	Item 61 ("lezen" / *to read*)	Yellow	Min. 18	Experimental
jullie (*you, plural*)	willen (*want to*) moeten (*have to*) zullen (*will*)	willen (*want to*) moeten (*have to*) zullen (*will*)	Item 63 ("rinkelen" / *to ring*)	Green	Min. 20	Experimental
zij (*they*)	willen (*want to*) moeten (*have to*) zullen (*will*)	willen (*want to*) moeten (*have to*) zullen (*will*)	Item 71 ("zingen" / *to sing*)	Blue	Max. 11	Experimental

ik (*I*)	willen (*want to*) moeten (*have to*) zullen (*will*)	wil (*want to*) moet (*have to*) zal (*will*)	Item 77 (“slapen” / *to sleep*)	Purple	Max. 10	Experimental
jij (*you, singular*)	willen (*want to*) moeten (*have to*) zullen (*will*)	wilt (*want to*) moet (*have to*) zult (*will*)	Item 86 (“zwemmen” / *to swim*)	Grey	Max. 9	Experimental

Appendix II: Python code used to generate anagrams

To generate an initial list of 40 anagrams that complied with certain criteria (3 to 5 letters, 2 to 5 solutions), a script was written in Python that provided lists of anagrams that had a given number of letters (anagram_length) and given number of solutions (anagram_options). To obtain these solutions, a text file listing all tokens in the *Corpus of Spoken Dutch* (Oostdijk, 2000) was used. See the code below. Some comments have been added, which have been colored blue.

```
#import libraries
from collections import Counter

#Choose what kind of anagram you want by changing these variables
#Anagram_length defines the number of characters that the anagram should have
#Anagram_options defines the number of possible solutions the anagram should have
anagram_length = 5
anagram_options = 5

#Read a file that includes words from the Corpus of Spoken Dutch (Oostdijk, 2000) and their corresponding frequencies
freqlist = {}
with open("./totalph.frq", encoding="ISO-8859-1") as input_file:
    for line in input_file:
        try:
            _, tot, token = line.strip().split(maxsplit=2)
            #Only words that have a number of characters equal to anagram_length are relevant
            #Words that start with a capital should also be left out
            if len(token) == anagram_length and token[0].islower():
                freqlist[token] = tot
        except ValueError:
            pass

#Find out which words consist of the same letters by sorting the letters of all words in freqlist
#If words consist of the same letters (e.g. 'heart'  and 'hater'), they should turn into the same word when all of its letters are sorted
alphabetically (e.g. "aehrt")
sorted_chars_list = {}
for word in freqlist.keys():
    sorted_chars_list[word] = ''.join(sorted(word))
counts = Counter(sorted_chars_list.values())
#Potential anagrams are those that have a number of solutions equal to anagram_options
potential_anagrams = [i[0] for i in counts.items() if i[1] == anagram_options]

#Retrieve the solutions associated with the potential anagrams
associated_words = {}
for an in potential_anagrams:
    associated_words[an] = [i[0] for i in sorted_chars_list.items() if i[1] == an]
keys = list(associated_words.keys())

#Print anagrams and solutions
print("Found %d potential anagrams!" % (len(potential_anagrams)))
for index in range(len(keys)):
    aw_string = ""
    for word in associated_words[keys[index]]:
        aw_string += "%s (%s), " % (word, freqlist[word])
    aw_string = aw_string[:-2]

    print("\nPotential anagram %d: %s" % (index + 1, keys[index]))
    print("Associated solutions: %s" % aw_string)
```

Appendix III: Items from the anagram task

Below is an overview of the 10 anagrams used in the experiment and their characteristics. Anagrams are shown in the order in which they appeared to the participants. The first column (Anagram) displays the string of letters that participants had to rearrange. The second column (Solutions (Word Frequency)) lists the official Dutch words (according to Den Boon & Geeraerts, 2005) that could be formed by rearranging the letters from the first column, along with the frequency of those words in the *Corpus of Spoken Dutch* (CGN; Oostdijk, 2000) between parentheses and an added exclamation mark if none of the three participants from the pre-test came up with that particular solution. Initial solutions have been separated from the solutions that were found by creative participants and that had not been planned for. These solutions are marked with a frequency of 0, because they did not appear in the CGN. For convenience sake, the total number of solutions before and after the experiment is described in columns 3 (Initial number of solutions) and 4 (Final number of solutions) and the number of characters in each anagram in column 5 (Number of characters). The remaining three columns describe the results of the difficulty assessment of each anagram, with column 6 (Initial Difficulty (pre-test)) describing the difficulty of the anagram according to a small pre-test ('Hard' if there was at least one solution that none of the three participants came up with, 'Easy' if otherwise), column 7 (Initial Difficulty (word frequency)) describing the expected difficulty based on word frequency of the solutions ('Hard' if there was at least one solution that occurred less than 10 times in the CGN, 'Easy' if otherwise) and column 8 (Initial Difficulty (combination)) describing the combined difficulty (either 'Hard', 'Easy', or 'Intermediate' if the results of column 6 and 7 were conflicting).

Anagram	**Solutions (word frequency)**[12]	**Initial number of solutions**	**Final number of solutions**	**Number of characters**	**Initial Difficulty (pre-test)**	**Initial Difficulty (word frequency)**	**Initial Difficulty (combination)**
enoz	onze (3757) (!) zoen (36) zone (26) nezo (0)	3	4	4	Hard	Easy	Intermediate
vnele	elven (21) (!) leven (3102) nevel (21) velen (125)	4	4	5	Hard	Easy	Intermediate
adar	aard (166) daar (46721) raad (228)	3	3	4	Easy	Easy	Easy
doer	orde (1060) (!) rode (353)	2	2	4	Hard	Easy	Intermediate
geanl	algen (16) (!) angel (3) lagen (324) lange (1120)	5	6	5	Hard	Hard	Hard

12 An exclamation mark means that none of the three participants in the pre-test came up with that particular solution.

	nagel (16) galen (0)						
sjilt	lijst (379) slijt (12) (!) stijl (211) ijlst (0)	3	4	5	Hard	Easy	Intermediate
nerda	ander (3009) arend (4) (!) nader (135) raden (111) denar (0) narde (0) rande (0)	4	7	5	Hard	Hard	Hard
ools	loos (13) solo (28)	2	2	4	Easy	Easy	Easy
reetd	deert (4) reedt (3) (!) teder (10) trede (8) treed (11) (!) dreet (0)	5	6	5	Hard	Hard	Hard
eelk	elke (2106) keel (187) leek (731) ekel (0)	3	4	4	Easy	Easy	Easy

Appendix IV: Items from the questionnaires

The experiment described in this thesis included two questionnaires, one of which was conducted pre-test and the other post-test. The first 12 statements appeared in both questionnaires and the final 5 statements appeared only in the pre-test. Statements that have been removed after the conduction of a factor analysis are marked in red. To map the statements shown here to the results of the factor analysis shown in Table 2, the statement number displayed in the first column can be used. The texts from the pre-test version and the post-test version of each statement are shown in the second and fourth column, respectively, with English translations provided in the third and fifth columns. Column 6 describes the construct that was hypothesized to be underlying each statement. Finally, column 7 describes whether the rating of a particular statement was reversed during analysis (a rating of 7 would be turned into 1, a rating of 6 would be turned into 2, etc.).

Statement no.	**Dutch text (pre-test)**	**English Translation (pre-test)**	**Dutch text (post-test)**	**English Translation (post-test)**	**Construct**	**Reverse?**
1	Ik verwacht dat ik deze oefeningen saai ga vinden.	I expect that I will think these exercises are boring.	Ik vond deze oefeningen saai.	I thought these exercises were boring.	Intrinsic Motivation	yes
2	Deze oefeningen lijken me leuk om te doen.	These exercises appear to me as fun to do.	Deze oefeningen waren leuk om te doen.	These exercises were fun to do.	Intrinsic Motivation	
3	Ik denk dat ik erg ga genieten van deze oefeningen.	I think I will be quite enjoying these exercises.	Ik heb erg van deze oefeningen genoten.	I quite enjoyed these exercises.	Intrinsic Motivation	
4	De manier waarop deze oefeningen klinken, zou ik omschrijven als erg interessant.	I would describe the way these exercises sound as very interesting.	Deze oefeningen zou ik omschrijven als erg interessant.	I would describe these exercises as very interesting.	Intrinsic Motivation	
5	Ik denk niet dat de oefeningen mijn aandacht helemaal vast	I don't think these exercises will be able to fully hold my attention.	De oefeningen hielden mijn aandacht niet helemaal vast.	These exercises didn't fully hold my attention.	Intrinsic Motivation	yes

	gaan houden.					
6	Ik ga deze oefeningen met plezier maken.	I'm going to enjoy doing these exercises.	Ik heb deze oefeningen met plezier gemaakt.	I've enjoyed doing these exercises.	Intrinsic Motivation	
7	Ik heb het gevoel dat ik er niet zelf voor kies om deze oefeningen te doen.	I feel like I'm not choosing to do these exercises myself.	Ik had het gevoel dat ik er niet zelf voor koos om deze oefeningen te doen.	I felt like I wasn't choosing to do these exercises myself.	Autonomy	yes
8	Ik doe deze oefeningen omdat het moet.	I do these exercises because I have to.	Ik deed deze activiteit omdat het moest.	I did this activity[13] because I had to.	Autonomy	yes
9	Ik geloof dat ik de vrijheid heb om deze oefeningen te maken of niet.	I believe I have the freedom to choose whether or not to do these exercises.	Ik geloof dat ik de vrijheid had om deze oefeningen te maken of niet.	I believe I had the freedom to choose whether or not to do these exercises.	Autonomy	
10	Ik doe deze oefeningen omdat ik dat wil.	I do these exercises because I want to.	Ik deed deze oefeningen omdat ik dat wilde.	I did these exercises because I wanted to.	Autonomy	
11	Ik doe deze oefeningen omdat ik geen keus heb.	I do these exercises because I have no choice.	Ik deed deze oefeningen omdat ik geen keus had.	I did these exercises because I had no choice.	Autonomy	yes
12	Ik heb het idee dat dit verplicht is.	I feel like this is mandatory.	Ik had het idee dat dit verplicht was.	I felt like this was mandatory.	Autonomy	yes
13	Ik verwacht tevreden te zijn met mijn prestaties	I expect to be satisfied with my forthcoming			Competence	

13 This should have been "exercise", but became "activity" due to an unnoticed inconsistency during translation of the statements of the original Intrinsic Motivation Inventory (Ryan, 1982) to Dutch.

	zometeen.	accomplishments.				
14	Dit zijn oefeningen die ik niet zo geweldig ga kunnen maken.	These are exercises I won’t be able to do very amazingly.			Competence	yes
15	Ik ben erg handig met dit soort oefeningen.	I am quite skillful at these kinds of exercises.			Competence	
16	Volgens mij ga ik deze oefeningen best goed doen, vergeleken met leeftijdsgenoten.	I think I’m going to do these exercises pretty well, compared to peers.			Competence	
17	Volgens mij ben ik erg goed in dit soort oefeningen.	I think I am very good at these kinds of exercises.			Competence	

Appendix V: Instruction texts
Instructions used in the experiment had been written in such a way that they did not contain any modal verbs (excluding the final statement). These instructions are shown here in the order in which they appeared in the experiment.

Informed Consent

<u>INFORMATIE EN TOESTEMMING</u>
Welkom bij dit onderzoek naar taalkundige creativiteit. Dit experiment is opgesteld door Astraea Blonk in het kader van een bachelorscriptie Taalwetenschap aan de Radboud Universiteit.

Wat wordt er van je verwacht?
Dit experiment bevat twee oefeningen die je taalkundige creativiteit op de proef stellen. Bij de eerste oefening is het de bedoeling om zinnen te bedenken door plaatjes, kleuren en stukken tekst te combineren. Tijdens de tweede oefening krijg je een aantal lettercombinaties voorgeschoteld, waarbij het aan jou is om woorden te verzinnen met de gegeven letters. Aan het begin en aan het eind van het experiment krijg je daarnaast een aantal vragen te zien. Voorafgaand aan ieder onderdeel is er een uitgebreide uitleg. In totaal duurt het meedoen aan dit onderzoek ongeveer 20 tot 30 minuten. Het is aan te raden om het experiment in één keer te maken.

Vrijwilligheid
Deelname aan dit experiment is geheel vrijwillig. Daarom heb je op elk moment tijdens dit onderzoek de mogelijkheid om ermee te stoppen en je toestemmming in te trekken, ongeacht je reden. Tot twee weken na deelname is het mogelijk om je onderzoeksgegevens te laten verwijderen door een mail te sturen naar <e-mail address>.

Wat gebeurt er met mijn gegevens?
De gegevens uit dit onderzoek worden door wetenschappers gebruikt voor datasets, artikelen en presentaties. De anoniem gemaakte onderzoeksgegevens zijn tenminste 10 jaar beschikbaar voor andere wetenschappers. Als gegevens met andere onderzoekers gedeeld worden, zijn deze dus nooit tot jou persoonlijk te herleiden.
Alle onderzoeksgegevens worden op beveiligde wijze bewaard volgens de richtlijnen van de Radboud Universiteit.

Heb je vragen/opmerkingen over het onderzoek?
Voor meer informatie over dit onderzoek of de mogelijkheid om een klacht in te dienen, stuur je een e-mail naar Astraea Blonk (e-mailadres: <e-mail address>). De Radboud Universiteit heeft ook een Ethische Toetsingscommisie Geesteswetenschappen. Om daar een klacht in te dienen, mail je naar <e-mail address>. Voor vragen over de verwerking van gegevens in dit onderzoek is er de mogelijkheid om te mailen naar <e-mail address>.

<u>TOESTEMMING</u>: Geef hieronder je keuze aan.
Door te klikken op de knop ‘Ik ga akkoord’ geef je aan dat je:
- bovenstaande informatie hebt gelezen
- vrijwillig meedoet aan het onderzoek
- 16 jaar of ouder bent

Introductory Questions

Bedankt dat je meedoet. Om te beginnen, staan hieronder eerst een paar algemene vragen over jezelf. Kies steeds één van de opties door het juiste balkje te markeren of door een antwoord te typen in het invulveld.

Pre-test Questionnaire

Bij de volgende vragen krijg je een aantal stellingen te zien over je houding ten opzichte van de oefeningen in dit onderzoek. Het is steeds de bedoeling dat je op een schaal van 1 tot 7 aangeeft in hoeverre jij vindt dat de stelling op jou van toepassing is. Er zijn geen goede of foute antwoorden.

Beweeg onderstaande sliders naar het cijfer toe dat aangeeft in hoeverre je het met de stelling eens bent. Als de slider al op het juiste cijfer staat, klik er dan een keer op zodat de score geregistreerd wordt.

Ter herinnering: bij de eerste oefening ga je zinnen bedenken door plaatjes, kleuren en stukken tekst te combineren. Bij de tweede oefening krijg je lettercombinaties te zien en ga je op zoek naar de woorden die je ermee kunt bedenken. Het is oké als je beeld van de oefeningen nog wat vaag is.

Sentence Generation Task

Dan gaan we nu door naar de eerste oefening. Zometeen krijg je 3 plaatjes te zien, waarop een **kleur**, een **actie** en een stukje **tekst** zijn afgebeeld. Het is steeds de bedoeling om één Nederlandse zin te bedenken waarin alles wat er op de drie plaatjes staat, gecombineerd wordt. Het maakt niet uit of de zin waar is of niet, zolang de inhoud van de drie afbeeldingen er maar in voorkomt. Om het niet te makkelijk te maken, heeft iedere zin een minimum of maximum woordaantal om aan te voldoen. Hierdoor is er soms best wat creativiteit nodig om een goede zin te maken.
Voordat de oefening begint, komen er eerst twee voorbeelden.

Op het eerste plaatje staat *"hij probeert"*. Op het tweede plaatje is de actie *"water geven"* weergegeven. Het derde plaatje laat de kleur *"rood"* zien. Mogelijk denk je hierbij aan een zin als *"Hij probeert de rode planten water te geven"*, maar die bevat maar 8 woorden, en dat zijn er te weinig. Om de zin goed te maken, verzinnen we er wat dingen bij. *"Hij probeert de planten water te geven die rood kleuren dankzij het licht van de zon"* is hier wel een goede zin. Hij bevat 16 woorden, en het minimum aantal woorden bij deze vraag was 14.

De drie plaatjes laten weer een actie (*"lopen"*), een kleur (*"grijs"*) en een stukje tekst (*"jullie komen"*) zien. Dit keer is het echter de bedoeling om een zo **kort** mogelijke zin te maken, in plaats van een zo lang mogelijke. Om binnen de 8 woorden te blijven, is een zin als *"Jullie komen lopen over het grijze pad"* geschikt. Het bevat de inhoud van alle drie de afbeeldingen en telt precies 7 woorden.

Dan volgen nu de vragen van de oefening, waarbij je zelf aan de slag gaat met het bedenken van zinnen. Hou in gedachte dat het niet uitmaakt of de zinnen waar of niet waar zijn, en het altijd mogelijk is om extra dingen erbij te verzinnen die niet op de plaatjes staan. Veel succes!

Dit is vraag XXX van de 20.

Hieronder zie je drie plaatjes. Bedenk een zin waarin de inhoud van alledrie de plaatjes voorkomt, die bestaat uit **XXX of meer/minder** woorden.

Vul hieronder de zin in die je bedacht hebt van **XXX of meer/minder** woorden.
Aantal woorden: 0

Anagram Task

Tot zover de eerste oefening. Voor het tweede deel van dit experiment ga je aan de slag met *anagrammen*. Een anagram is een reeks letters waaruit woorden te maken zijn wanneer de letters op een bepaalde manier gehusseld worden. Een voorbeeld van een anagram is "ats". Deze drie letters betekenen op deze manier niets, maar wanneer ze in een andere volgorde gezet worden verschijnen ineens de woorden *"tas"* en *"sta"*.

Zometeen krijg je een aantal van dit soort anagrammen te zien. Het is aan jou om bij iedere set letters zoveel mogelijk Nederlandse woorden te verzinnen die je kunt maken door de gegeven letters te husselen, en deze woorden in te vullen in de velden onderaan elke vraag. Sommige anagrammen hebben veel potentiële oplossingen, andere weinig. Daarom staan er soms meer velden dan dat er antwoordmogelijkheden zijn. Doe je best om zoveel mogelijk oplossingen te bedenken, maar aarzel niet om naar de volgende vraag te gaan wanneer er voor een bepaald anagram geen andere oplossingen meer lijken te zijn. Oningevulde velden laat je dan gewoon leeg.

Dit is anagram XXX van de 10.

Welke woorden vallen te maken door de letters uit onderstaande reeks te husselen?

Post-test Questionnaire

Het experiment zit er bijna op! Als laatste krijg je nog een paar stellingen te zien. Deze lijken een beetje op de vragen die je in het begin ook hebt gehad. Geef alsjeblieft weer op een schaal van 1 tot 7 aan in hoeverre je het met de stellingen eens bent.

Beweeg onderstaande sliders weer naar het cijfer toe dat aangeeft in hoeverre je het met iedere stelling eens bent.

Final Statement

Heel erg bedankt voor je deelname! Met dit onderzoek wil ik kijken of het gebruik van de werkwoorden "willen", "moeten" of "zullen" invloed heeft op iemands motivatie en creativiteit. Jij zat in de groep "XXX". Dat betekent dat je bij de eerste oefening zinnen hebt gemaakt met de werkwoorden "blijven", "gaan" en "XXX". Andere mensen hebben in plaats van "XXX" één van de twee andere werkwoorden gekregen om zinnen mee te maken.

Nogmaals bedankt dat je mee wilde doen! Ik hoop dat je zelf ook genoten hebt van de vragen. Mocht je het erg leuk hebben gevonden, zou je dan de link naar het experiment (<link>) verder willen verspreiden (zonder precies uit te leggen wat het doel van het experiment is)? Daar help je me erg mee!

Voor verdere vragen, of als je nieuwsgierig bent naar de uitkomsten van dit onderzoek, kun je mailen naar <e-mail address>.

Appendix VI: R code used for data analysis

The R code below has been used to analyze the results of the experiment. It took a raw results file that could be downloaded from the Qualtrics platform as input, made it ready for statistical analysis, conducted those analyses, and created some tables and graphs that have been put in the thesis. Some comments have been added, which have been colored blue.

```
#Main analysis function
#To execute the entire script at once, Source the script and run this function
analyse <- function()
{
 #Set settings and import libraries
 options(stringsAsFactors = FALSE)
 library(dplyr)
 library(car)
 library(ggpubr)
 library(broom)
 library(rstatix)
 library(ggplot2)
 library(reshape2)
 library(psych)
 library(ltm)
 library(lavaan)
 library(boot)
 library(pwr)
 library(matrixStats)
 library(tidyr)

 #Read datafile
 results_complete <<- read_files("responses_final_21_apr.csv")

 #Preparations
 create_research_purpose_check_table(results_complete) #Check for excessive research purpose awareness
 create_manipulation_check_table(results_complete) #Check if participants have produced at least 8 sentences with the right modal verb in its most prototypical definition
 create_unfinished_responses_table(results_complete) #Check if participants that did not finish the experiment completely can still be included in some analyses
 create_automatically_excluded_participants_table(results_complete) #Some responses lead to automatic exclusion (such as being diagnosed with color blindness)
 participants_to_exclude_fully <- c(1, 2, 9, 15, 16, 19, 26, 62, 71, 82, 85, 86, 87, 88, 89, 90, 92, 93, 94, 95, 97, 99, 100, 101, 102, 103, 105, 110, 112, 113, 114, 116, 120, 121, 123, 124, 125, 126, 127, 128, 135, 142, 148, 149, 154, 158, 160, 161, 163, 164, 166, 167, 169, 170, 171, 173, 174, 175, 176, 179, 180, 182, 185, 187, 188, 189, 190, 191, 192, 193, 194, 196, 197, 199, 200, 201, 203, 204, 205, 206, 207, 208, 209, 210, 212, 213, 214, 215, 216, 217, 218, 219, 221, 222, 224, 231, 233, 234, 237, 242, 247, 265, 266, 267, 269, 270, 271, 275, 278, 283, 288, 289, 290, 292, 293, 294, 295, 296, 297, 298, 299, 301, 302, 303, 305, 306, 307, 308, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 323, 324, 325, 326, 327, 328, 329, 330, 331, 334, 335, 336, 338, 340, 341, 342, 343, 344, 345, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357)
 results_complete <<- filter_excluded_participants(results_complete,participants_to_exclude_fully) #Remove excluded participants
 list_anagram_solutions(results_complete) #List all anagram solutions participants came up with (to check if additional solutions have to be counted as correct)
 correct_solutions <- list(
  c("nezo","onze","zoen","zone"),
  c("elven","leven","nevel","velen"),
  c("aard","daar","raad"),
  c("orde","rode"),
  c("algen","angel","galen","lagen","lange","nagel"),
  c("ijlst","lijst","slijt","stijl"),
  c("ander","arend","denar","nader","narde","raden","rande"),
  c("loos","solo"),
  c("deert","dreet","reedt","teder","trede","treed"),
  c("ekel","elke","keel","leek")
 )

 #Do factor analysis and calculate Cronbach's alpha; after excluding items with low loadings, do factor analysis again
 corr_matrix <- create_correlation_matrix(results_complete)
 do_PCA(corr_matrix)
 calculate_Cronbach_alpha(results_complete)
 items_to_exclude <- c("Q8A_1","Q8A_6","Q8C_2","Q8A_2","Q8B_3","Q8B_6","Q19A_2","Q19A_3","Q19A_4","Q19A_6","Q19B_1","Q19B_4")
 corr_matrix <- create_correlation_matrix(results_complete,items_to_exclude)
 do_PCA(corr_matrix)
```

```
 calculate_Cronbach_alpha(results_complete,items_to_exclude)

 #Create tables for main analysis
 analysis_tbl1 <<- create_indices(results_complete,items_to_exclude) #This table is for the Intrinsic Motivation and Sense of Autonomy analyses
 analysis_tbl2 <<- calculate_persistence(results_complete,analysis_tbl1)
 analysis_tbl2 <<- calculate_correct_solutions(results_complete,analysis_tbl2,correct_solutions) #This table is for the Persistence and Creative
Performance analyses

 #Summaries and plots
 print(summary(analysis_tbl1))
 print(summary(analysis_tbl2))
 show_descriptives(analysis_tbl1,analysis_tbl2)
 show_boxplots(analysis_tbl1,analysis_tbl2)

 #Check assumptions for ANCOVA
 check_covariate_independent_of_predictor(analysis_tbl1)
 check_linear_covariate_relationship_per_group(analysis_tbl1,analysis_tbl2)
 check_normality_of_residuals(analysis_tbl1,analysis_tbl2)
 check_homogeneity_regression_slopes(analysis_tbl1,analysis_tbl2)
 check_levenes(analysis_tbl1,analysis_tbl2)
 check_outliers(analysis_tbl1,analysis_tbl2)

 #Main analysis
 do_main_analysis(analysis_tbl1,analysis_tbl2)

 #Visualize
 create_boxplots_for_paper(analysis_tbl1,analysis_tbl2)

 #Check some alternative explanations
 do_alternative_analysis(results_complete,analysis_tbl2)
 investigate_survey_time(results_complete)
}

#Function to read the datafile
read_files <- function(filename)
{
 results <- read.csv(filename)
 results <- results[3:nrow(results),]
 new_filename <- sub(".csv","_1Header.csv",filename)
 write.csv(results,new_filename,row.names=FALSE)
 results <- read.csv(new_filename)
 ParticipantId <- rownames(results)
 return(as.data.frame(cbind(ParticipantId,results)))
}

#Create a document that focuses solely on participants' answers to the research purpose question
create_research_purpose_check_table <- function(results)
{
 write.table(results[,c("ParticipantId","Q20")],file="Research_Purpose_Check.txt",row.names=FALSE)
}

#Create a document to check if participants produced at least 8 sentences containing the modal verb corresponding to their condition in its most
prototypical word sense
create_manipulation_check_table <- function(results)
{
 X21_Q14 <- vector(mode="character",length=nrow(results))
 X22_Q14 <- X21_Q14
 X23_Q14 <- X21_Q14
 X24_Q14 <- X21_Q14
 X25_Q14 <- X21_Q14
 X26_Q14 <- X21_Q14
 X27_Q14 <- X21_Q14
 X28_Q14 <- X21_Q14
 X29_Q14 <- X21_Q14
 X30_Q14 <- X21_Q14
 Verb_Used <- X21_Q14
 manipulation_table <-
cbind(results[,c("ParticipantId","Group")],X21_Q14,X22_Q14,X23_Q14,X24_Q14,X25_Q14,X26_Q14,X27_Q14,X28_Q14,X29_Q14,X30_Q14,Verb_U
sed)
 for (row in 1:nrow(manipulation_table))
```

```
 {
  if (manipulation_table[row,"Group"] == "willen")
  {
   manipulation_table[row,"X21_Q14"] <- results[row,"X21_Q14"]
   manipulation_table[row,"X22_Q14"] <- results[row,"X22_Q14"]
   manipulation_table[row,"X23_Q14"] <- results[row,"X23_Q14"]
   manipulation_table[row,"X24_Q14"] <- results[row,"X24_Q14"]
   manipulation_table[row,"X25_Q14"] <- results[row,"X25_Q14"]
   manipulation_table[row,"X26_Q14"] <- results[row,"X26_Q14"]
   manipulation_table[row,"X27_Q14"] <- results[row,"X27_Q14"]
   manipulation_table[row,"X28_Q14"] <- results[row,"X28_Q14"]
   manipulation_table[row,"X29_Q14"] <- results[row,"X29_Q14"]
   manipulation_table[row,"X30_Q14"] <- results[row,"X30_Q14"]

   times_verb_used <- 0
   for(loop_num in 21:30)
   {
    #Did the sentence contain “wil”, “wilt” or “willen”? Was this word not surrounded by other letters (or else we’re dealing with a different word such as “wilg”)? This pattern evaluated to true if the word “wil”, “wilt” or “willen” was found and if it was surrounded either by non-word characters (\\W) or by the start/end of a sentence (^ and $).
    if(grepl(pattern="(^|\\W)wil(t|len)?(\\W|$)",x=manipulation_table[row,sprintf("X%d_Q14",loop_num)],ignore.case = TRUE))
    {
     times_verb_used <- times_verb_used + 1
    }
   }
   manipulation_table[row,"Verb_Used"] <- times_verb_used
  }
  else if (manipulation_table[row,"Group"] == "moeten")
  {
   manipulation_table[row,"X21_Q14"] <- results[row,"X21_Q14.1"]
   manipulation_table[row,"X22_Q14"] <- results[row,"X22_Q14.1"]
   manipulation_table[row,"X23_Q14"] <- results[row,"X23_Q14.1"]
   manipulation_table[row,"X24_Q14"] <- results[row,"X24_Q14.1"]
   manipulation_table[row,"X25_Q14"] <- results[row,"X25_Q14.1"]
   manipulation_table[row,"X26_Q14"] <- results[row,"X26_Q14.1"]
   manipulation_table[row,"X27_Q14"] <- results[row,"X27_Q14.1"]
   manipulation_table[row,"X28_Q14"] <- results[row,"X28_Q14.1"]
   manipulation_table[row,"X29_Q14"] <- results[row,"X29_Q14.1"]
   manipulation_table[row,"X30_Q14"] <- results[row,"X30_Q14.1"]

   times_verb_used <- 0
   for(loop_num in 21:30)
   {
    #Did the sentence contain “moet” or “moeten”?
    if(grepl(pattern="(^|\\W)moet(en)?(\\W|$)",x=manipulation_table[row,sprintf("X%d_Q14",loop_num)],ignore.case = TRUE))
    {
     times_verb_used <- times_verb_used + 1
    }
   }
   manipulation_table[row,"Verb_Used"] <- times_verb_used
  }
  else if (manipulation_table[row,"Group"] == "zullen")
  {
   manipulation_table[row,"X21_Q14"] <- results[row,"X21_Q14.2"]
   manipulation_table[row,"X22_Q14"] <- results[row,"X22_Q14.2"]
   manipulation_table[row,"X23_Q14"] <- results[row,"X23_Q14.2"]
   manipulation_table[row,"X24_Q14"] <- results[row,"X24_Q14.2"]
   manipulation_table[row,"X25_Q14"] <- results[row,"X25_Q14.2"]
   manipulation_table[row,"X26_Q14"] <- results[row,"X26_Q14.2"]
   manipulation_table[row,"X27_Q14"] <- results[row,"X27_Q14.2"]
   manipulation_table[row,"X28_Q14"] <- results[row,"X28_Q14.2"]
   manipulation_table[row,"X29_Q14"] <- results[row,"X29_Q14.2"]
   manipulation_table[row,"X30_Q14"] <- results[row,"X30_Q14.2"]

   times_verb_used <- 0
   for(loop_num in 21:30)
   {
    #Did the sentence contain “zal”, “zul”, “zult” or “zullen”?
    if(grepl(pattern="(^|\\W)zul(t|len)?(\\W|$)|(^|\\W)zal(\\W|$)",x=manipulation_table[row,sprintf("X%d_Q14",loop_num)],ignore.case = TRUE))
    {
```

```
        times_verb_used <- times_verb_used + 1
      }
    }
    manipulation_table[row,"Verb_Used"] <- times_verb_used
   }
   else if (manipulation_table[row,"Group"] == "komen")
   {

   }
   else
   {
    print("error")
   }
  }
  write.csv(manipulation_table,file="Manipulation_check.csv")
}

#Create a document listing all the participants that did not fully finish the survey, and include an automatic judgment
create_unfinished_responses_table <- function(results)
{
  results <- results[results$Finished == 0,]
  judgment <- vector(length=nrow(results))
  for(row in 1:nrow(results))
  {
   #If participants did not finish the anagram task, do not include them in any analysis
   if(is.na(results[row,"X10_Q16_Page.Submit"]))
   {
    judgment[row] <- 0
   }
   #If participants finished the anagram task but not the post-test questionnaire, we might want to include them in some analyses
   else if(is.na(results[row,"Q19B_6"]))
   {
    judgment[row] <- 1
   }
   #If participants finished the post-test questionnaire, they can be included in all analyses
   else
   {
    judgment[row] <- 2
   }
  }

  write.csv(as.data.frame(cbind(judgment,results)),file="Unfinished_responses.csv")
}

#Create a document listing all participants that can be excluded automatically
create_automatically_excluded_participants_table <- function(results)
{
  #Exclude a participant  if they either
  #- did not give consent in Q1
  #- were not a native speaker of Dutch in Q4
  #- had been diagnosed with color blindness in Q5
  results <- results[(results$Q1 == 2) %in% TRUE | (results$Q4 == 3) %in% TRUE | (results$Q5 == 1) %in% TRUE,]
  write.csv(results,file="Automatically_excluded_participants.csv")
}

#Remove excluded participants from the dataset
filter_excluded_participants <- function(results,participants_to_exclude)
{
  return(results[!results$ParticipantId %in% participants_to_exclude,])
}

#Create a list of all unique solutions participants came up with
list_anagram_solutions <- function(results)
{
  con <- file(description="Anagram_Solutions_Participants_Came_Up_With.txt",open="w")

  anagram <- c("enoz","vnele","adar","doer","geanl","sjilt","nerda","ools","reetd","eelk")
  for(loop_num in 1:10)
  {
   anagram_solutions <- vector(mode="character")
```

```
    for(ans_num in c("A","B","C","D","E","F"))
    {
      answers <- results[,sprintf("X%d_Q17%s",loop_num,ans_num)]
      answers <- answers[!is.na(answers) & answers != ""]
      answers <- tolower(answers)
      answers <- remove_equals_and_parentheses(answers)
      answers <- remove_trailing_spaces(answers)
      anagram_solutions <- c(anagram_solutions,answers)
    }
    writeLines(sprintf("\nAnagram %d (%s):",loop_num, anagram[loop_num]),con)
    writeLines(anagram_solutions[!duplicated(anagram_solutions)],con)
  }

  close(con)
}

#Function called by the above list_anagram_solutions function
#Some participants wrote additional information next to their answers (e.g. "oslo (topographical name)")
#In order to synchronize participants' answers (and save me work from looking through non-unique solutions), this additional information was removed
remove_equals_and_parentheses <- function(strings)
{
  for(str in 1:length(strings))
  {
    strings[str] <- gsub("=.*$|\\(.*\\)|\\[.*\\]|\\?","",strings[str])
  }
  return(strings)
}

#Trailing spaces were removed as well (see the comment above)
remove_trailing_spaces <- function(strings)
{
  for(str in 1:length(strings))
  {
    strings[str] <- gsub("^ +| +$","",strings[str])
  }
  return(strings)
}

#Create a correlation matrix for the factor analysis
create_correlation_matrix <- function(results,columns_to_exclude=c())
{
  results <- results[,c("Q8A_1","Q8A_2","Q8A_3","Q8A_4","Q8A_5","Q8A_6",
                        "Q8B_1","Q8B_2","Q8B_3","Q8B_4","Q8B_5","Q8B_6",
                        "Q8C_1","Q8C_2","Q8C_3","Q8C_4","Q8C_5")]

  for(col in c("Q8A_1","Q8A_2","Q8A_4","Q8A_5","Q8C_2","Q8C_4","Q8C_5"))
  {
    results[,col] <- 8 - results[,col]
  }

  columns_to_keep = vector(length=0)
  for (col in 1:ncol(results))
  {
    if (! colnames(results)[col] %in% columns_to_exclude)
    {
      columns_to_keep <- c(columns_to_keep,col)
    }
  }
  results <- results[,columns_to_keep]

  return(cor(results))
}

#Do Principal Component Analysis on the ratings for the Intrinsic Motivation, Sense of Autonomy and Competence statements during the pre-test
do_PCA <- function(corr_matrix)
{
  fit <- principal(corr_matrix, nfactors=3, rotate="oblimin")
  print(fit)
}
```

```
#Calculate reliability of the different factors that can be extracted from the pre-test and the post-test
calculate_Cronbach_alpha <- function(results,columns_to_exclude=c(),post_test=FALSE)
{
  all_items <- list(c("Q8A_1","Q8A_6","Q8B_1","Q8B_5","Q8C_2","Q8C_3"),
                    c("Q8A_2","Q8A_5","Q8B_3","Q8B_6","Q8C_4","Q8C_5"),
                    c("Q8A_3","Q8A_4","Q8B_2","Q8B_4","Q8C_1"),
                    c("Q19A_1","Q19A_2","Q19A_3","Q19A_4","Q19B_5","Q19B_6"),
                    c("Q19A_5","Q19A_6","Q19B_1","Q19B_2","Q19B_3","Q19B_4"))

  for(col in c("Q8A_1","Q8A_2","Q8A_4","Q8A_5","Q8C_2","Q8C_4","Q8C_5","Q19A_3","Q19A_4","Q19A_5","Q19B_2","Q19B_3","Q19B_4"))
  {
    results[,col] <- 8 - results[,col]
  }

  if(post_test)
  {
    for(factor in 4:5){
      items <- results[,unlist(all_items[factor])]
      columns_to_keep = vector(length=0)
      for (col in 1:ncol(items))
      {
        if (! colnames(items)[col] %in% columns_to_exclude)
        {
          columns_to_keep <- c(columns_to_keep,col)
        }
      }
      items <- items[,columns_to_keep]
      print(cronbach.alpha(items))
    }
  }
  else
  {
    for(factor in 1:3){
      items <- results[,unlist(all_items[factor])]
      columns_to_keep = vector(length=0)
      for (col in 1:ncol(items))
      {
        if (! colnames(items)[col] %in% columns_to_exclude)
        {
          columns_to_keep <- c(columns_to_keep,col)
        }
      }
      items <- items[,columns_to_keep]
      print(cronbach.alpha(items))
    }
  }
}

#Create analysis_tbl1, which contains the information relevant for the Intrinsic Motivation and Sense of Autonomy main analyses
create_indices <- function(results,columns_to_exclude=c())
{
  factor_items <- list(c("Q8A_1","Q8A_6","Q8B_1","Q8B_5","Q8C_2","Q8C_3"),
                       c("Q8A_2","Q8A_5","Q8B_3","Q8B_6","Q8C_4","Q8C_5"),
                       c("Q8A_3","Q8A_4","Q8B_2","Q8B_4","Q8C_1"),
                       c("Q19A_1","Q19A_2","Q19A_3","Q19A_4","Q19B_5","Q19B_6"),
                       c("Q19A_5","Q19A_6","Q19B_1","Q19B_2","Q19B_3","Q19B_4"))

  factor_item_reversals <- list(c("r","","","","r",""),
                                c("r","r","","","r","r"),
                                c("","r","","",""),
                                c("","","r","r","",""),
                                c("r","","","r","r","r"))

  analysis_table <- as.data.frame(results$ParticipantId)

  for(factor in 1:5){
    items <- results[,unlist(factor_items[factor])]
    columns_to_keep = vector(length=0)
    for(col in 1:ncol(items))
    {
```

```
      if (! colnames(items)[col] %in% columns_to_exclude)
      {
        columns_to_keep <- c(columns_to_keep,col)

        if(unlist(factor_item_reversals[factor])[col] == "r")
        {
          items[,col] <- 8 - items[,col]
        }
      }
    }
    items <- items[,columns_to_keep]
    analysis_table <- cbind(analysis_table,rowSums(items) / ncol(items))
  }

  analysis_table <- cbind(analysis_table,analysis_table[,5] - analysis_table[,2],analysis_table[,6] -
analysis_table[,4],results$Group,results$Q2,results$Q3,results$Q4,results$Q5)
  colnames(analysis_table) <-
c("ParticipantId","Interest_Pre","Autonomy_Pre","Competence","Interest_Post","Autonomy_Post","Interest_Diff","Autonomy_Diff","Group","Age","Gen
der","Nativity","Color_Blindness")

  analysis_table$Group <- as.factor(analysis_table$Group)
  #Some participants wrote things such as “52 years” instead of “52” in their answer to the Age question. This gsub removes all non-digit characters from
the age field
  analysis_table$Age <- gsub(x=analysis_table$Age,pattern="\\D",replacement="")
  analysis_table$Age <- as.double(analysis_table$Age)
  analysis_table$Gender <- as.factor(analysis_table$Gender)
  analysis_table$Nativity <- as.factor(analysis_table$Nativity)
  analysis_table$Color_Blindness <- as.factor(analysis_table$Color_Blindness)

  return(analysis_table)
}

#Create analysis_tbl2, which contains the information relevant to the Persistence main analysis
calculate_persistence <- function(results,analysis_tbl1)
{
  Persistence <- ((results$X1_Q16_Page.Submit - results$X1_Q16_Last.Click) +
                  (results$X2_Q16_Page.Submit - results$X2_Q16_Last.Click) +
                  (results$X3_Q16_Page.Submit - results$X3_Q16_Last.Click) +
                  (results$X4_Q16_Page.Submit - results$X4_Q16_Last.Click) +
                  (results$X5_Q16_Page.Submit - results$X5_Q16_Last.Click) +
                  (results$X6_Q16_Page.Submit - results$X6_Q16_Last.Click) +
                  (results$X7_Q16_Page.Submit - results$X7_Q16_Last.Click) +
                  (results$X8_Q16_Page.Submit - results$X8_Q16_Last.Click) +
                  (results$X9_Q16_Page.Submit - results$X9_Q16_Last.Click) +
                  (results$X10_Q16_Page.Submit - results$X10_Q16_Last.Click)) / 10
  Persistence <- round(Persistence,4)
  ParticipantId <- results$ParticipantId
  Group <- results$Group
  Competence <- analysis_tbl1$Competence
  Age <- results$Q2
  Gender <- results$Q3
  Nativity <- results$Q4
  Color_Blindness <- results$Q5
  analysis_tbl <- as.data.frame(cbind(ParticipantId,Group,Competence,Persistence,Age,Gender,Nativity,Color_Blindness))
  analysis_tbl$Group <- as.factor(analysis_tbl$Group)
  analysis_tbl$Competence <- as.double(analysis_tbl$Competence)
  analysis_tbl$Persistence <- as.double(analysis_tbl$Persistence)
  #Remove non-digit characters from the Age field
  analysis_tbl$Age <- gsub(x=analysis_tbl$Age,pattern="\\D",replacement="")
  analysis_tbl$Age <- as.double(analysis_tbl$Age)
  analysis_tbl$Gender <- as.factor(analysis_tbl$Gender)
  analysis_tbl$Nativity <- as.factor(analysis_tbl$Nativity)
  analysis_tbl$Color_Blindness <- as.factor(analysis_tbl$Color_Blindness)
  row.names(analysis_tbl) <- analysis_tbl$ParticipantId
  return(analysis_tbl)
}

#Calculate Creative Performance and add the results to analysis_tbl2
#The reason for 2 analysis tables instead of one is that I first planned to include some participants in the Creative Performance and Persistence analyses
that I would not include in the analyses for Intrinsic Motivation and Sense of Autonomy
```

```
#However, since there was only one participant who complied with the conditions for being included in two analyses but not the other two, I decided to remove that participant entirely to create a uniform participant sample for all analyses
calculate_correct_solutions <- function(results, analysis_tbl,correct_solutions)
{
  Performance <- vector(length=nrow(analysis_tbl))
  for(row in 1:nrow(analysis_tbl))
  {
    score <- 0
    for(loop_num in 1:10)
    {
      answers <- vector(length=0)
      for(ans_num in c("A","B","C","D","E","F"))
      {
        answer <- tolower(results[row,sprintf("X%d_Q17%s",loop_num,ans_num)])
        if(!is.na(answer) & answer != "" &! answer %in% answers){
          answers <- c(answers,answer)
        }
      }
      for(answer in answers)
      {
        if(answer %in% unlist(correct_solutions[loop_num])){
          score <- score + 1
        }
      }
    }
    Performance[row] <- score
  }
  return(cbind(analysis_tbl,Performance))
}

#Descriptive statistics
show_descriptives <- function(analysis_tbl1,analysis_tbl2)
{
  print("Summary Intrinsic Motivation:")
  print(summarise(group_by(analysis_tbl1,Group),mean=mean(Interest_Diff),sd=sd(Interest_Diff)))

  print("Summary Autonomy:")
  print(summarise(group_by(analysis_tbl1,Group),mean=mean(Autonomy_Diff),sd=sd(Autonomy_Diff)))

  print("Summary Competence:")
  print(summarise(group_by(analysis_tbl1,Group),mean=mean(Competence),sd=sd(Competence)))

  print("Summary Persistence:")
  print(summarise(group_by(analysis_tbl2,Group),mean=mean(Persistence),sd=sd(Persistence)))

  print("Summary Performance:")
  print(summarise(group_by(analysis_tbl2,Group),mean=mean(Performance),sd=sd(Performance)))

  print("Gender distribution in table 1: (1=Man,2=Vrouw,3=Anders,4=Zeg ik liever niet)")
  print(table(analysis_tbl1$Gender))
  print("Nativity distribution in table 1: (1=Nederlands,2=Nederlands en een andere taal,3=Anders)")
  print(table(analysis_tbl1$Nativity))
  print("Color Blindness distribution in table 1: (1=Ja,2=Nee,3=Daar wil ik liever niets over kwijt)")
  print(table(analysis_tbl1$Color_Blindness))

  print("Gender distribution in table 2: (1=Man,2=Vrouw,3=Anders,4=Zeg ik liever niet)")
  print(table(analysis_tbl2$Gender))
  print("Nativity distribution in table 2: (1=Nederlands,2=Nederlands en een andere taal,3=Anders)")
  print(table(analysis_tbl2$Nativity))
  print("Color Blindness distribution in table 2: (1=Ja,2=Nee,3=Daar wil ik liever niets over kwijt)")
  print(table(analysis_tbl2$Color_Blindness))
}

#Create boxplots
#These are not the final boxplots that have been put in the paper, but they were here to aid the interpretation of the analysis results while the final boxplots weren't finished yet
show_boxplots <- function(analysis_tbl1,analysis_tbl2)
{
  png(filename="Boxplots/Interest_boxplot.png")
  boxplot(Interest_Diff ~ Group,
```

```
        data = analysis_tbl1,
        main = "Interest by Group",
        xlab = "Group",
        ylab = "Interest",
        col = "steelblue",
        border = "black")
  dev.off()

  png(filename="Boxplots/Autonomy_boxplot.png")
  boxplot(Autonomy_Diff ~ Group,
        data = analysis_tbl1,
        main = "Autonomy by Group",
        xlab = "Group",
        ylab = "Autonomy",
        col = "steelblue",
        border = "black")
  dev.off()

  png(filename="Boxplots/Competence_boxplot.png")
  boxplot(Competence ~ Group,
        data = analysis_tbl1,
        main = "Competence by Group",
        xlab = "Group",
        ylab = "Competence",
        col = "steelblue",
        border = "black")
  dev.off()

  png(filename="Boxplots/Persistence_boxplot.png")
  boxplot(Persistence ~ Group,
        data = analysis_tbl2,
        main = "Persistence by Group",
        xlab = "Group",
        ylab = "Persistence",
        col = "steelblue",
        border = "black")
  dev.off()

  png(filename="Boxplots/Performance_boxplot.png")
  boxplot(Performance ~ Group,
        data = analysis_tbl2,
        main = "Performance by Group",
        xlab = "Group",
        ylab = "Performance",
        col = "steelblue",
        border = "black")
  dev.off()
}

#The following six functions have been used to check various assumptions related to ANCOVA or to parametric tests in general
#This one tests whether the covariate was independent of the predictor variable
check_covariate_independent_of_predictor <- function(analysis_tbl)
{
  model <- aov(Competence ~ Group, data = analysis_tbl)
  print("Is the covariate independent of group?")
  print(summary(model))
}

#The second function checks whether there was a linear relationship between the covariate and the 4 outcome variables within each of the 3 different groups
check_linear_covariate_relationship_per_group <- function(analysis_tbl1,analysis_tbl2)
{
  for(var in c("Interest_Diff","Autonomy_Diff"))
  {
    for(group in c("willen","moeten","zullen"))
    {
      table_for_cor <- analysis_tbl1[analysis_tbl1$Group == group,c(var,"Competence")]
      print(paste0("Test for correlation between ",var," and Competence for Group ",group))
      print(cor.test(table_for_cor[,var],table_for_cor[,"Competence"]))
    }
```

```
  }
  for(var in c("Performance","Persistence"))
  {
    for(group in c("willen","moeten","zullen"))
    {
      table_for_cor <- analysis_tbl2[analysis_tbl2$Group == group,c(var,"Competence")]
      print(paste0("Test for correlation between ",var," and Competence for Group ",group))
      print(cor.test(table_for_cor[,var],table_for_cor[,"Competence"]))
    }
  }
}

#The third function checks for each outcome variable whether the regression slopes for the different groups were similar
check_homogeneity_regression_slopes <- function(analysis_tbl1,analysis_tbl2)
{
  print("Homogeneity Interest regression slopes:")
  aov(Interest_Diff ~ Group*Competence, data = analysis_tbl1) %>% Anova(type="III") %>% print

  print("Homogeneity Autonomy regression slopes:")
  aov(Autonomy_Diff ~ Group*Competence, data = analysis_tbl1) %>% Anova(type="III") %>% print

  print("Homogeneity Persistence regression slopes:")
  aov(Persistence ~ Group*Competence, data = analysis_tbl2) %>% Anova(type="III") %>% print

  print("Homogeneity Performance regression slopes:")
  aov(Performance ~ Group*Competence, data = analysis_tbl2) %>% Anova(type="III") %>% print
}

#The fourth function checks whether the residuals were normally distributed
#It does so both in a statistical way (shapiro-wilk test) and in a visual way (density plots)
check_normality_of_residuals <- function(analysis_tbl1,analysis_tbl2)
{
  model <- lm(Interest_Diff ~ Group, data = analysis_tbl1)
  model.metrics <- augment(model)
  print(head(model.metrics, 5))
  print(shapiro_test(model.metrics$.resid))

  model <- lm(Autonomy_Diff ~ Competence + Group, data = analysis_tbl1)
  model.metrics <- augment(model)
  print(head(model.metrics, 5))
  print(shapiro_test(model.metrics$.resid))

  model <- lm(Persistence ~ Group, data = analysis_tbl2)
  model.metrics <- augment(model)
  print(head(model.metrics, 5))
  print(shapiro_test(model.metrics$.resid))

  model <- lm(Performance ~ Group, data = analysis_tbl2)
  model.metrics <- augment(model)
  print(head(model.metrics, 5))
  print(shapiro_test(model.metrics$.resid))

  ggdensity(analysis_tbl1$Interest_Diff, fill = "lightgray") %>%
    ggexport(filename="Density/Interest_Normality.png")
  ggdensity(analysis_tbl1$Autonomy_Diff, fill = "lightgray") %>%
    ggexport(filename="Density/Autonomy_Normality.png")
  ggdensity(analysis_tbl2$Persistence, fill = "lightgray") %>%
    ggexport(filename="Density/Persistence_Normality.png")
  ggdensity(analysis_tbl2$Performance, fill = "lightgray") %>%
    ggexport(filename="Density/Performance_Normality.png")
}

#The fifth function checks whether the residuals within each group had similar variances
check_levenes <- function(analysis_tbl1,analysis_tbl2)
{
  lm(Interest_Diff ~ Group, data = analysis_tbl1) %>%
    augment(.) %>%
    levene_test(.resid~Group) %>%
    print
```

```
  lm(Autonomy_Diff ~ Competence + Group, data = analysis_tbl1) %>%
    augment(.) %>%
    levene_test(.resid~Group) %>%
    print

  lm(Persistence ~ Group, data = analysis_tbl2) %>%
    augment(.) %>%
    levene_test(.resid~Group) %>%
    print

  lm(Performance ~ Group, data = analysis_tbl2) %>%
    augment(.) %>%
    levene_test(.resid~Group) %>%
    print
}

#The final assumption checking function printed the outliers associated with each variable
check_outliers <- function(analysis_tbl1,analysis_tbl2)
{
  lm(Interest_Diff ~ Group, data = analysis_tbl1) %>%
    augment(.) %>%
    filter(abs(.std.resid) > 3) %>%
    as.data.frame() %>%
    print

  lm(Autonomy_Diff ~ Competence + Group, data = analysis_tbl1) %>%
    augment(.) %>%
    filter(abs(.std.resid) > 3) %>%
    as.data.frame() %>%
    print

  lm(Persistence ~ Group, data = analysis_tbl2) %>%
    augment(.) %>%
    filter(abs(.std.resid) > 3) %>%
    as.data.frame() %>%
    print

  lm(Performance ~ Group, data = analysis_tbl2) %>%
    augment(.) %>%
    filter(abs(.std.resid) > 3) %>%
    as.data.frame() %>%
    print
}

#Do 4 AN(C)OVAs to test if there was an effect of Group on any of the 4 outcome variables
do_main_analysis <- function(analysis_tbl1,analysis_tbl2)
{
  #Analysis 1: the effect of Group on Intrinsic Motivation
  #Since the residuals were not normally distributed, this one uses bootstrapping
  #It also leaves out the covariate (Competence) since it did not show a linear relationship with Intrinsic Motivation within each of the different Groups
  set.seed(0)
  reps <- boot(data=analysis_tbl1, statistic=calculate_ancova, R=1000, formula=Interest_Diff~Group)
  print("Intrinsic Movation ANOVA results: (without covariate)")
  print(paste0("F is ",reps$t0[1]))
  print(paste0("p is ",reps$t0[2]))
  print(paste0("η² is ",reps$t0[3]))
  effect_size <- sqrt(reps$t0[3]/(1-reps$t0[3]))
  print(paste0("f is ",effect_size))

  #Analysis 2: the effect of Group on Sense of Autonomy
  #This one also uses bootstrapping, but it does include the covariate
  set.seed(0)
  reps <- boot(data=analysis_tbl1, statistic=calculate_ancova, R=1000, formula=Autonomy_Diff~Competence + Group)
  print("Autonomy ANCOVA results:")
  print(paste0("F is ",reps$t0[1]))
  print(paste0("p is ",reps$t0[2]))
  print(paste0("η² is ",reps$t0[3]))
  effect_size <- sqrt(reps$t0[3]/(1-reps$t0[3]))
  print(paste0("f is ",effect_size))
```

```
  #Analysis 3: the effect of Group on Persistence
  #This one also uses bootstrapping, and it leaves out the covariate for similar reasons as to why it was left out in the first analysis
  set.seed(0)
  reps <- boot(data=analysis_tbl2, statistic=calculate_ancova, R=1000, formula=Persistence~Group)
  print("Persistence ANOVA results: (without covariate)")
  print(paste0("F is ",reps$t0[1]))
  print(paste0("p is ",reps$t0[2]))
  print(paste0("η² is ",reps$t0[3]))
  effect_size <- sqrt(reps$t0[3]/(1-reps$t0[3]))
  print(paste0("f is ",effect_size))

  #Analysis 4: the effect of Group on Creative Performance
  #This is the only one that does not use bootstrapping, since the residuals for Creative Performance were normally distributed
  #It does leave out the covariate, however
  model <- aov(Performance ~ Group, data = analysis_tbl2)
  print("Performance ANOVA results: (without covariate)")
  anova_results <- Anova(model,type="III")
  eta_sq <- anova_results["Group","Sum Sq"] / sum(anova_results[,"Sum Sq"])
  print(anova_results)
  print(paste0("η² is ",eta_sq))
  effect_size <- sqrt(eta_sq/(1-eta_sq))
  print(paste0("f is ",effect_size))
}

#This is a helper function to calculate AN(C)OVA
#It is used by the boot() function in those main analyses where bootstrapping was employed
calculate_ancova <- function(formula, data, indices) {
  d <- data[indices,]
  mod <- aov(formula, data = d)
  anova_results <- Anova(mod,type="III")
  return(c(anova_results["Group","F value"],anova_results["Group","Pr(>F)"],anova_results["Group","Sum Sq"] / sum(anova_results[,"Sum Sq"])))
}

#Make a visually attractive boxplot for inclusion in the thesis
create_boxplots_for_paper <- function(analysis_tbl1,analysis_tbl2)
{
  Group <- vector(length=nrow(analysis_tbl1))
  for(row in 1:nrow(analysis_tbl1))
  {
    if (analysis_tbl1[row,"Group"] == "willen")
    {
      Group[row] <- "Intrinsic"
    }
    else if (analysis_tbl1[row,"Group"] == "moeten")
    {
      Group[row] <- "Extrinsic"
    }
    else if (analysis_tbl1[row,"Group"] == "zullen")
    {
      Group[row] <- "Control"
    }
    else
    {
      Group[row] <- "Other"
    }
  }
  analysis_tbl1$Group <- factor(Group, levels = c("Intrinsic","Control","Extrinsic"))

  Group <- vector(length=nrow(analysis_tbl2))
  for(row in 1:nrow(analysis_tbl2))
  {
    if (analysis_tbl2[row,"Group"] == "willen")
    {
      Group[row] <- "Intrinsic"
    }
    else if (analysis_tbl2[row,"Group"] == "moeten")
    {
      Group[row] <- "Extrinsic"
    }
    else if (analysis_tbl2[row,"Group"] == "zullen")
```

```
    {
      Group[row] <- "Control"
    }
    else
    {
      Group[row] <- "Other"
    }
  }
  analysis_tbl2$Group <- factor(Group, levels = c("Intrinsic","Control","Extrinsic"))

  Interest_bp <- ggboxplot(data=analysis_tbl1,x="Group",y="Interest_Diff",
                           ylab="Intrinsic Motivation\n(Difference in Likert ratings)",
                           fill = "Group", palette = c("#00AFBB", "#E7B800", "#FC4E07"))
  Autonomy_bp <- ggboxplot(data=analysis_tbl1,x="Group",y="Autonomy_Diff",
                           ylab="Sense of Autonomy\n(Difference in Likert ratings)",
                           fill = "Group", palette = c("#00AFBB", "#E7B800", "#FC4E07"))
  Persistence_bp <- ggboxplot(data=analysis_tbl2,x="Group",y="Persistence",
                              ylab="Persistence\n(Seconds)",
                              fill = "Group", palette = c("#00AFBB", "#E7B800", "#FC4E07"))
  Performance_bp <- ggboxplot(data=analysis_tbl2,x="Group",y="Performance",
                              ylab="Creative Performance\n(Number of solutions)",
                              fill = "Group", palette = c("#00AFBB", "#E7B800", "#FC4E07"))
  ggarrange(Interest_bp,Autonomy_bp,Persistence_bp,Performance_bp,labels=c("A","B","C","D")) %>%
    ggexport(filename="Boxplots/BPs_for_paper.png")
}

#Do an alternative analysis to check for the effect of Group on a fifth outcome variable: time used to come up with sentences containing either "willen", "moeten" or "zullen" during the sentence generation task
do_alternative_analysis <- function(results,analysis_tbl2)
{
  #Create table
  X21_Q13_Page.Submit <- vector(mode="character",length=nrow(results))
  X22_Q13_Page.Submit <- X21_Q13_Page.Submit
  X23_Q13_Page.Submit <- X21_Q13_Page.Submit
  X24_Q13_Page.Submit <- X21_Q13_Page.Submit
  X25_Q13_Page.Submit <- X21_Q13_Page.Submit
  X26_Q13_Page.Submit <- X21_Q13_Page.Submit
  X27_Q13_Page.Submit <- X21_Q13_Page.Submit
  X28_Q13_Page.Submit <- X21_Q13_Page.Submit
  X29_Q13_Page.Submit <- X21_Q13_Page.Submit
  X30_Q13_Page.Submit <- X21_Q13_Page.Submit
  analysis_tbl3 <<-
cbind(results[,c("ParticipantId","Group")],X21_Q13_Page.Submit,X22_Q13_Page.Submit,X23_Q13_Page.Submit,X24_Q13_Page.Submit,X25_Q13_Page.Submit,X26_Q13_Page.Submit,X27_Q13_Page.Submit,X28_Q13_Page.Submit,X29_Q13_Page.Submit,X30_Q13_Page.Submit)
  for (row in 1:nrow(analysis_tbl3))
  {
    if (analysis_tbl3[row,"Group"] == "willen")
    {
      analysis_tbl3[row,"X21_Q13_Page.Submit"] <<- results[row,"X21_Q13_Page.Submit"]
      analysis_tbl3[row,"X22_Q13_Page.Submit"] <<- results[row,"X22_Q13_Page.Submit"]
      analysis_tbl3[row,"X23_Q13_Page.Submit"] <<- results[row,"X23_Q13_Page.Submit"]
      analysis_tbl3[row,"X24_Q13_Page.Submit"] <<- results[row,"X24_Q13_Page.Submit"]
      analysis_tbl3[row,"X25_Q13_Page.Submit"] <<- results[row,"X25_Q13_Page.Submit"]
      analysis_tbl3[row,"X26_Q13_Page.Submit"] <<- results[row,"X26_Q13_Page.Submit"]
      analysis_tbl3[row,"X27_Q13_Page.Submit"] <<- results[row,"X27_Q13_Page.Submit"]
      analysis_tbl3[row,"X28_Q13_Page.Submit"] <<- results[row,"X28_Q13_Page.Submit"]
      analysis_tbl3[row,"X29_Q13_Page.Submit"] <<- results[row,"X29_Q13_Page.Submit"]
      analysis_tbl3[row,"X30_Q13_Page.Submit"] <<- results[row,"X30_Q13_Page.Submit"]
    }
    else if (analysis_tbl3[row,"Group"] == "moeten")
    {
      analysis_tbl3[row,"X21_Q13_Page.Submit"] <<- results[row,"X21_Q13_Page.Submit.1"]
      analysis_tbl3[row,"X22_Q13_Page.Submit"] <<- results[row,"X22_Q13_Page.Submit.1"]
      analysis_tbl3[row,"X23_Q13_Page.Submit"] <<- results[row,"X23_Q13_Page.Submit.1"]
      analysis_tbl3[row,"X24_Q13_Page.Submit"] <<- results[row,"X24_Q13_Page.Submit.1"]
      analysis_tbl3[row,"X25_Q13_Page.Submit"] <<- results[row,"X25_Q13_Page.Submit.1"]
      analysis_tbl3[row,"X26_Q13_Page.Submit"] <<- results[row,"X26_Q13_Page.Submit.1"]
      analysis_tbl3[row,"X27_Q13_Page.Submit"] <<- results[row,"X27_Q13_Page.Submit.1"]
      analysis_tbl3[row,"X28_Q13_Page.Submit"] <<- results[row,"X28_Q13_Page.Submit.1"]
      analysis_tbl3[row,"X29_Q13_Page.Submit"] <<- results[row,"X29_Q13_Page.Submit.1"]
```

```
    analysis_tbl3[row,"X30_Q13_Page.Submit"] <<- results[row,"X30_Q13_Page.Submit.1"]
   }
   else if (analysis_tbl3[row,"Group"] == "zullen")
   {
    analysis_tbl3[row,"X21_Q13_Page.Submit"] <<- results[row,"X21_Q13_Page.Submit.2"]
    analysis_tbl3[row,"X22_Q13_Page.Submit"] <<- results[row,"X22_Q13_Page.Submit.2"]
    analysis_tbl3[row,"X23_Q13_Page.Submit"] <<- results[row,"X23_Q13_Page.Submit.2"]
    analysis_tbl3[row,"X24_Q13_Page.Submit"] <<- results[row,"X24_Q13_Page.Submit.2"]
    analysis_tbl3[row,"X25_Q13_Page.Submit"] <<- results[row,"X25_Q13_Page.Submit.2"]
    analysis_tbl3[row,"X26_Q13_Page.Submit"] <<- results[row,"X26_Q13_Page.Submit.2"]
    analysis_tbl3[row,"X27_Q13_Page.Submit"] <<- results[row,"X27_Q13_Page.Submit.2"]
    analysis_tbl3[row,"X28_Q13_Page.Submit"] <<- results[row,"X28_Q13_Page.Submit.2"]
    analysis_tbl3[row,"X29_Q13_Page.Submit"] <<- results[row,"X29_Q13_Page.Submit.2"]
    analysis_tbl3[row,"X30_Q13_Page.Submit"] <<- results[row,"X30_Q13_Page.Submit.2"]
   }
   else
   {
    print("huh")
   }
  }
  analysis_tbl3$Group <<- as.factor(analysis_tbl3$Group)
  analysis_tbl3$X21_Q13_Page.Submit <<- as.double(analysis_tbl3$X21_Q13_Page.Submit)
  analysis_tbl3$X22_Q13_Page.Submit <<- as.double(analysis_tbl3$X22_Q13_Page.Submit)
  analysis_tbl3$X23_Q13_Page.Submit <<- as.double(analysis_tbl3$X23_Q13_Page.Submit)
  analysis_tbl3$X24_Q13_Page.Submit <<- as.double(analysis_tbl3$X24_Q13_Page.Submit)
  analysis_tbl3$X25_Q13_Page.Submit <<- as.double(analysis_tbl3$X25_Q13_Page.Submit)
  analysis_tbl3$X26_Q13_Page.Submit <<- as.double(analysis_tbl3$X26_Q13_Page.Submit)
  analysis_tbl3$X27_Q13_Page.Submit <<- as.double(analysis_tbl3$X27_Q13_Page.Submit)
  analysis_tbl3$X28_Q13_Page.Submit <<- as.double(analysis_tbl3$X28_Q13_Page.Submit)
  analysis_tbl3$X29_Q13_Page.Submit <<- as.double(analysis_tbl3$X29_Q13_Page.Submit)
  analysis_tbl3$X30_Q13_Page.Submit <<- as.double(analysis_tbl3$X30_Q13_Page.Submit)
  Average_Time <- (analysis_tbl3$X21_Q13_Page.Submit + analysis_tbl3$X22_Q13_Page.Submit + analysis_tbl3$X23_Q13_Page.Submit +
analysis_tbl3$X24_Q13_Page.Submit + analysis_tbl3$X25_Q13_Page.Submit + analysis_tbl3$X26_Q13_Page.Submit +
analysis_tbl3$X27_Q13_Page.Submit + analysis_tbl3$X28_Q13_Page.Submit + analysis_tbl3$X29_Q13_Page.Submit +
analysis_tbl3$X30_Q13_Page.Submit) / 10
  Competence <- analysis_tbl2$Competence
  analysis_tbl3 <<- cbind(analysis_tbl3,Average_Time,Competence)

  #Descriptive analysis
  print(summary(analysis_tbl3))
  print(summarise(group_by(analysis_tbl3,Group),mean=mean(Average_Time),sd=sd(Average_Time)))
  png(filename="Boxplots/Sentence_Time_Boxplot.png")
  boxplot(Average_Time ~ Group,
        data = analysis_tbl3,
        main = "Average Time by Group",
        xlab = "Group",
        ylab = "Average Time",
        col = "steelblue",
        border = "black")
  dev.off()

  #Check assumptions
  #Linear relationship?
  for(group in c("willen","moeten","zullen"))
  {
   table_for_cor <- analysis_tbl3[analysis_tbl3$Group == group,c("Average_Time","Competence")]
   print(paste0("Test for correlation between Average Time and Competence for Group ",group))
   print(cor.test(table_for_cor[,"Average_Time"],table_for_cor[,"Competence"]))
  }

  #Normal residuals?
  model <- lm(Average_Time ~ Group, data = analysis_tbl3)
  model.metrics <- augment(model)
  print(head(model.metrics, 5))
  print(shapiro_test(model.metrics$.resid))

  ggdensity(analysis_tbl3$Average_Time, fill = "lightgray") %>%
   ggexport(filename="Density/Sentence_Time_Normality.png")

  #Equal variances?
```

```
 lm(Average_Time ~ Group, data = analysis_tbl2) %>%
   augment(.) %>%
   levene_test(.resid~Group) %>%
   print

  #Outliers?
  lm(Average_Time ~ Group, data = analysis_tbl3) %>%
   augment(.) %>%
   filter(abs(.std.resid) > 3) %>%
   as.data.frame() %>%
   print

  #Analysis time
  set.seed(0)
  reps <- boot(data=analysis_tbl3, statistic=calculate_ancova, R=1000, formula=Average_Time~Group)
  print("Average Time ANOVA results: (without covariate)")
  print(paste0("F is ",reps$t0[1]))
  print(paste0("p is ",reps$t0[2]))
  print(paste0("η² is ",reps$t0[3]))
  effect_size <- sqrt(reps$t0[3]/(1-reps$t0[3]))
  print(paste0("f is ",effect_size))
}

#Create a box plot showing how long participants took to take the full survey
investigate_survey_time <- function(results)
{
  colnames(results)[6] <- "Duration"
  results$Same <- vector(length=nrow(results))

  png(filename="Boxplots/Survey_time_boxplot.png")
  boxplot(x = results$Duration / 60 / 60,
          xlab = "Participants",
          ylab = "Total Time (hours)",
          col = "steelblue",
          border = "black")
  dev.off()

  lm(Duration ~ Same, data = results) %>%
   augment(.) %>%
   filter(abs(.std.resid) > 3) %>%
   as.data.frame() %>%
   print
}
```