Does emotion shape language?

Out, C.

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Does emotion shape language?

Studies on the influence of affective state on interactive language production

Charlotte Out

Does emotion shape language? Studies on the influence of affective state on interactive language production

Charlotte Out

Tilburg center for Cognition and Communication (TiCC)
Department of Communication and Cognition
Tilburg School of Humanities and Digital Sciences
Tilburg University

Does emotion shape language?
Studies on the influence of affective state on interactive language production

Charlotte Out
PhD thesis
Tilburg University, 2021

TiCC PhD Series Number 77
ISBN: 978-94-6416-924-9
Printing: Ridderprint | www.ridderprint.nl
Layout and design: Michiel Huige, persoonlijkproefschrift.nl
Cover design by Stasia Burrington, www.stasiaburringtonart.com

Does emotion shape language? Studies on the influence of affective state on interactive language production

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University

op gezag van de rector magnificus, prof. dr. W. B. H. J. van de Donk,

in het openbaar te verdedigen ten overstaan van een door het college voor promoties

aangewezen commissie in de Aula van de Universiteit

op donderdag 16 december 2021 om 13.30 uur

door

Charlotte Out

geboren te Heemstede

Promotor

Prof. dr. Emiel Krahmer	Tilburg University

Copromotor

Dr. Martijn Goudbeek	Tilburg University

Promotiecommissie

Prof. dr. Vera Kempe	Abertay University
Prof. dr. Jos van Berkum	Utrecht University
Dr. Mariët Theune	University of Twente
Dr. Mirjam Broersma	Radboud University
Prof. dr. Fons Maes	Tilburg University

I'm a woman of my word, now haven't you heard?

My word's the only thing I've ever needed

I'm a woman of my word, now you have heard

My word's the only thing I truly need

Cat Power - Woman

CONTENTS

1

Introduction

1.1. DOES EMOTION INFLUENCE LANGUAGE PRODUCTION?

Humans are emotional beings. Throughout the day, we experience a myriad of feelings: reluctance to get out of bed, pleasure when drinking a good cup of coffee, ambivalence about going to work or to the gym, and excitement to meet a friend later that evening. In contrast to animals, we can express these feelings: non-verbally, with facial expressions and gestures, but also verbally, using our language. We can write about our fun (or disappointing) evening with our friend in our diary, or share our story with a colleague, partner, or neighbor in a conversation. Even when we do not verbalize our emotions directly, they arguably influence how we act and how we express ourselves. For example, when we are engaging in a conversation, we will probably talk faster and louder when we are happy (Scherer, 2003), and, in our excitement, maybe even rant for a while. When we are sad, however, we will probably talk slower, using a soft voice (Scherer, 2003) and pay more attention to our conversation partner (see, e.g., Converse et al., 2008; Forgas, 2013), for example, to check whether they understand or listen to us. Even when we express ourselves solitary, as when writing a diary entry or poetry, emotions influence our writing style and the words we tend to use (e.g., Pennebaker et al., 2003).

As the examples above illustrate, and ample research confirms, there is a pivotal relationship between language production and emotions, but the specifics remain unclear. In this dissertation, we study the influence of emotions on spoken language production, both alone and in interactions, across four different experimental studies, each targeting a specific aspect of language production. Each study builds on established paradigms for studying language production, and asks what impact (if any) our emotional state has on this particular aspect. In this first, introductory chapter, we will introduce emotion and affect, language production, and the relationship between the two, followed by a section on replication research. This chapter is concluded with an overview of the studies presented in this dissertation, accompanied by the research questions they aim to answer.

1.2. Emotion

Although there is no general agreement on how affective feelings should be defined or distinguished (see, e.g., Engelen & Mennella, 2020; Izard, 2010; James, 1890), a common distinction is made between mood, emotion, and affect. A mood is usually defined as a diffuse, general feeling, that is only weakly linked to specific situations (Ekkekakis, 2012; Rottenberg, 2005). For example, someone can be in an anxious mood because they worry about their future in general. Moods typically lasts for hours or days (Rottenberg, 2005).
Emotions are usually defined as quick reactions to meaningful events, organisms (such as spiders or specific persons), objects, and the like, lasting only a few seconds or minutes (Rottenberg, 2005). A distinction can be made between various different, discrete emotions, for example, by amusement and disgust. Each emotion can be characterized, and distinguished, in several ways. For example, by the typical situations that trigger the emotion: usually, people tend to feel amused when they hear a funny joke, but disgusted

when they smell rotten fish (and not vice versa). Other ways researchers have tried to characterize emotions have involved, among other things, causes such as physiological and expressive phenomena such as changes in heart rate and blood pressure (see, e.g., Cacioppo et al., 2000) and facial expressions (e.g., Ekman et al., 1980), as well as effects for cognitive processes such as changes in perspective taking (e.g., Todd et al., 2015; Yip & Schweitzer, 2019) and the broadening- or narrowing of attentional scope (Fredrickson & Branigan, 2005; Gable & Harmon-Jones, 2010).

Finally, *affect* can be described as a basic sense of feeling (Barrett, 2017) and is commonly used as an umbrella term, encompassing more specific terms as moods and emotions (see, e.g., Ekkekakis, 2012; Munezero et al., 2014). Many scholars take a dimensional approach to affect, using, sometimes among other dimensions, valence (i.e., positive-negative or pleasant-unpleasant) and arousal (high or low sense of activation) to describe affective states (e.g., Barrett, 2017; Gillioz et al., 2016; Posner et al., 2005; Russell, 1980; but see Kuppens et al., 2017). In the current dissertation, we will mainly use the term 'affect', unless we are referring to a specific emotion or mood.

To study the influence of affective state in an experimental setting, researchers usually try to modify the affective states of participants, using so-called affect induction methods. In the present dissertation, we used various well-known and effective affect induction methods including recalling an affective, autobiographical memory (e.g., Jallais & Gilet, 2010; Schaeffer & Philippot, 2005), viewing affective film clips (e.g., Fernández-Aguilar et al., 2019; Rottenberg et al., 2007; Schaeffer et al., 2010; Uhrig et al., 2016) and viewing affective images (e.g., Uhrig et al., 2016), for example, pictures from the International Affective Pictures System (IAPS; Bradley & Lang, 2017; Lang et al., 1997).

As we briefly touched upon, affective states are known to influence a wide range of cognitive processes. In this dissertation, we zoom in on one particular process, spoken language production, which we introduce in the following section.

1.3. Spoken language production

Most individuals produce language every day, by writing, by signing, but mostly by talking: sometimes to ourselves, sometimes in front of an audience, but most of the time within a conversation. According to most researchers in the field, speaking involves multiple stages, that happen either successive, or in parallel to each other (Dell, 1986; Garrett, 1975; Griffin & Ferreira, 2006; Levelt, 1989; Levelt et al., 1999; Van Gompel et al., 2019; Vigliocco & Hartsuiker, 2002). Many of these researchers differentiate between three distinct speech production stages. First, a speaker has to conceptualize their message, or decide *what* they want to say ('content selection'). Second, they have to decide *how* they are going to say it ('message formulation'). Third, and final, they have to articulate the message, resulting in audible speech ('phonological encoding' and actual articulation; see also Griffin & Ferreira, 2006).

Indeed, speaking is a complex task that involves many of our cognitive abilities (Goldrick et al., 2014). Luckily, much of the spoken language production process happens

unconsciously: we are often not aware of the exact words we are planning to use, or how fast or slow we want to talk, even when we are talking directly to another person. In fact, to most people, speaking with someone in a conversation feels easy (Garrod & Pickering, 2004). This might seem counterintuitive, especially when keeping in mind that a conversation is not as straightforward as choosing the right words to get a message across. In a conversation, we have to consider many other things, for example, to which degree we share the same 'common ground' (Stalnaker, 1978), that is, to what extent we have the same information and beliefs as our conversation partner. For example, when a fellow psycholinguistics researcher asks about the topic of this dissertation, I expect that we share, to a certain degree, mutual knowledge about psychology and linguistics. As a result, I will aim to answer accurately, by, among other things, using research jargon, saying 'affect and content selection in conversation'. However, if a stranger on the bus would ask the same question, I expect that he does not have knowledge of psycholinguistics, academia or research. Therefore, I will answer more generally, saying 'the influence of what we feel on what we say'.

When we continue our conversation, the fellow psycholinguistics researcher, or stranger on the bus, will likely adjust their words as well, based on their perceived level of common ground with me. In this sense conversation can be seen as a joint activity in which all parties continuously adapt to each other (e.g., De Looze et al., 2014; Garrod & Pickering, 2004). Garrod and Pickering (2004) argue that one of the ways in which this happens, is via alignment: conversation partners start to 'mimic' each other in various ways, including using similar words and phrases. For example, Goudbeek and Krahmer (2012) found that speakers who listened to a pre-recorded voice (as a proxy for a conversation partner) describing various objects of furniture, tend to start mimicking the attributes used in the referential expressions (color or orientation, e.g., 'blue', 'front') they heard, when they were asked to describe objects of furniture themselves. So, when speakers described similar depictions of furniture themselves, they adapted their word use to their hypothetical conversation partner.

Another important way in which conversation partners can adapt to each other can be seen when looking at repeated, interactive references. When speakers repeatedly refer to an object, they typically start by describing the object (e.g., referring to a cute looking kitten in an animal shelter as 'the sweet kitten with the fluffy tail'), gradually converging on a shared referential expression, e.g., 'the sweetheart' for the aforementioned cute looking kitten (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986). When conversation partners agree on this temporary, shared conceptualization, a conceptual pact has been formed, which can be used in later references, e.g., to this specific kitten (Brennan & Clark, 1996). As a result, speakers need less words, time, and effort to refer to the object and mutually understand each other (Clark & Wilkes-Gibbs, 1986). In other words, the formation and use of (successful) conceptual pacts makes conversation more efficient.

While common ground, alignment and conceptual pact formation promote effective communication by allowing conversation partners to adapt to each other, it is important to keep in mind that conversations occur within an intra- and interpersonal context. Indeed,

even when interlocutors are focused on a practical goal, for example, a soccer trainer and captain discussing tactics to win the match, there are many aspects influencing their conversation. For example, how they are feeling (e.g., Morse & Afifi, 2015) and how well they know each other (Brown & Levinson, 1987; Brown, 2015) might influence, for example, how polite they are towards each other. In addition, their interaction might also look differently when one of them, or both, experience difficulties understanding and using social communication norms as politeness, which is a core characteristic in individuals with autism (e.g., Cummins et al., 2020; Mathersul et al., 2013).

So, language production in conversation involves social interactions in which both partners (commonly) aim for smooth conversation, by adjusting their speech, mimicking each other's expressions, and adhering to social norms as formulating requests using an appropriate degree of politeness. The core question of this dissertation is how these processes are impacted by the affective state of speakers.

1.4. Affect and language production

There is an important relationship between affect and language. For instance, it is often easy to indicate the affective state of a person, based on the acoustic characteristics of their speaking voice; an angry speaker tends to speak with a louder tone of voice and a higher pitch, while sad speakers speak softer and lower (Bachorowski & Owren, 1995; Scherer, 2003; for an overview, see, e.g., Scherer, 2019). In a similar vein, there is increasing support that affective states influence word use. For example, in their poetry, suicidal poets tend to use relatively more first-person singular pronouns, fewer references to other individuals, and more words referring to death, compared to their non-suicidal peers (Pająk & Trzebiński, 2014; Stirman & Pennebaker, 2001).

Moreover, the words used by a speaker might influence their affective state as well. One example that might come to mind, is that some individuals use self-directed verbalizations ('I can do this!'), or 'self-talk', to alter their affective state or encourage themselves (see, e.g., Van Raalte et al., 2016). In fact, this phenomenon is the foundation of the 'Velten method', a well-known affect induction method in affective science. When using the Velten method, participants read out loud an array of self-directed statements that gradually increase in positive content (e.g., 'God, I feel great!') or negative content (e.g., 'I want to go to bed and never wake up'), resulting in the corresponding moods in the speakers (Velten, 1967; 1968; Wilting et al., 2006).

Another example of the relationship between affect and language is (repeatedly) writing in a diary to vent negative emotions, thoughts and experiences. Previous research has shown that, indeed, so-called 'expressive writing' (e.g., Pennebaker, 1993; 1997) is a successful technique to reduce distress in the long run, and even improve the writer's physical and mental health (e.g., Baikie & Wilhelm, 2005; Soliday et al., 2004; see also Pennebaker & Chung, 2007).

As described above, the effects of affect on voice characteristics and word use are well established. However, the impact of affect on other aspects in language production, for

example, linguistic alignment between conversation partners, remains limited. In general, to date, a more systematic analysis on the relationship between affect and language production is still lacking. As a consequence, the specific details of the nature of this relationship, i.e., causal or correlational, also remains unclear. For example, as mentioned above, some findings support that language production influences affect (e.g., Velten, 1967; 1968), while other studies suggest that affect influences language production (e.g., Forgas, 1999a; 1999b; Pennebaker, 1997). Additionally, some researchers argue that affect and language production might influence *each other* (e.g., Barrett, 2017; Lindquist, 2017). While this is a very interesting and important topic in the fields of psycholinguistics and affective science, we do not aim to answer this question, but acknowledge all three perspectives.

In contrast, there is an extensive body of literature showing the influence of affect on various cognitive processes that are essential for (spoken) language production, e.g., attention (e.g., Charash et al., 2006; Fredrickson & Branigan, 2005; Gable & Harmon-Jones, 2010), perspective taking (e.g., Converse et al., 2008; Kempe et al., 2013; Todd et al., 2015) and processing style (e.g., Beukeboom & Semin, 2006). Based on the results of these previous studies, we conjecture that the influence of affect on language production extends beyond the characteristics of articulation and word use. Therefore, using four experimental studies, we aimed to investigate how affect influences various aspects of language production. As a starting point, we used well-known experimental set-ups, paradigms and/or findings in the field of psycholinguistic research. In other words: our experiments are replications of original studies, with an (added) component of affect.

1.5. On the importance of replication in research

Like all scientific work, this thesis is standing on the shoulders of giants: my work is inspired by and based on previous studies, conducted by researchers who studied (affective) language production before I had the opportunity to do so. Together, these researchers build the foundation for the scientific field of affective language production, creating budding support for the important, but still not well-understood relationship between affect and language.

What can be considered redundant information to fellow researchers, but what might not automatically come to mind to the non-academic reader, is that each experiment is, to some degree, unique (Anderson et al., 2016). One of the reasons is that experiments include not only predefined, controlled variables that are of interest to the researcher, but contextual variables as well (Van Bavel et al., 2016). For example, in Chapter 2, our Study 1 took place at Tilburg University, was conducted by me as the experiment leader, testing young, Dutch, mostly female college students living in the south of the Netherlands. Now, imagine that another group of researchers, or even ourselves, would aim to replicate this study, and find different results. This might be due to several reasons, for example, the effect of the original study was found by chance and therefore, the replication did not show the same results. However, another reason might be that the results in the original study were (partly) due to one or more of the unique characteristics of the study, or circumstances in which the

study was conducted. For example, if it took place a few days before the exam week, the participants might be slightly tense and in a hurry to finish the experiment, wanting to go back to the library to study. As a result, they might report more negative affect or distress before and after describing affective pictures, compared to, say, right after the exam week. Indeed, as Anderson and colleagues (2016) argue, '(…) all replications differ in innumerable ways from original studies. They are conducted in different facilities, in different weather, with different experimenters, with different computers and displays, in different languages, at different points in history, and so on.'

Given that scientific studies are usually conducted to gain information about the world in general, opposed to, for example, the specific situation in a laboratory (Kaplan, 1964; Rosenthal & Rosnow, 1984), it is important to perform replications, to 'help identify, diagnose, and minimize many concerns about the (…) reproducibility of research' (Makel et al., 2012, p. 537). Indeed, when studies are replicated and researchers find similar results to the original study, it means that it is more likely that these results can be generalized to, for example, different situations or a larger population (Schmidt, 2009).

Research can be replicated in various ways. One of the important aspects of replications is the degree in which the replication study remains 'true' to the original study, which is based on the aim of the replication. Therefore, replications are commonly (and roughly) classified in two categories: direct replication and conceptual replication (see, e.g., Schmidt, 2009; Van Berkel & Crandall, 2018; Zwaan et al., 2018). In a *direction replication*, researchers usually aim to find out whether they can get the (exact) same results as the original study, using the exact same methods (but bearing in mind the earlier remark by Anderson and colleagues). To reach this goal, the same experimental procedure as the original study is used, i.e., the same population, questionnaires, analyzing software, statistical analysis, etcetera. In a *conceptual replication*, the original research methods are (slightly) altered, for example, by using different populations, materials and/or experimental settings. In a conceptual replication, the aim is to re-test and extend the hypothesis or the underlying theory of the research. For example, in Chapter 5, we aimed to find out whether not only neurotypical students, but also autistic individuals, tend to be more polite towards their conversation partner if they are in a negative, compared to a more positive affective state. This is a conceptual replication of Forgas (1999a), who found that (presumably neurotypical) students tend to formulate more polite requests when they were presented with hypothetical, difficult situations in need of a request.

In the current dissertation, we performed conceptual replications. We did this by adding affect inductions to our experiments (Chapter 2, 3 and 4), by adjusting the methods (Chapter 3 and 5), the stimuli (Chapter 3 and 4) and the population sample (Chapter 5).

1.6. Research questions and overview of studies

In this dissertation, we aim to answer the following main research question:

- *To which degree do affective states influence (spoken) language production in an interactive setting?*

This question is examined in four individual experimental studies, reported in Chapter 2 to 5. Each chapter is based on an individual study that either has been published (Chapter 2 and 3), has been submitted for publication (Chapter 4) or is under review (Chapter 5) as a paper in an international peer-reviewed journal. All chapters are individual products in the sense that they consist of an abstract, introduction to the study (including a theoretical framework), method- and results section, and discussion. Given that the papers are connected to one another, the content of the chapters occasionally overlap.

In **Chapter 2**, we aim to answer the following specific research question: 'Can verbalizing affective pictures induce affective states?'. In this chapter, we were inspired by the affect induction method by Velten (1967; 1968), who found that verbalizing affective self-directed statements gradually increasing in affective content can induce the corresponding mood in the speaker. We conjecture that verbalizing affective content, in specific, describing affective pictures, might evoke corresponding affective states in the speakers as well. To study this, we select affective pictures from the IAPS (Bradley & Lang, 2017; Lang et al., 1997), a large set of various pictures (e.g., portraits, sports, mutilated bodies) widely used to induce affect (see e.g., Uhrig et al., 2016). We create a task in which individuals viewed and verbalized (Study 1) or merely viewed (Study 2) a set of pictures increasing in positive or negative content, or remaining neutral, measuring affective state before and after exposure to the pictures.

Additionally, we explore whether the verbal picture descriptions contain increasingly more affectively laden words in the expected direction. In order to study this, the verbalized picture descriptions in Study 1 were audio recorded and transcribed for analyses. Subsequently, the frequency of affective word use was calculated using the well-known and popular software Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2001), a text analysis software program for counting (percentages of) words grouped into various (affect) categories, e.g., 'positive emotion words'.

In Chapter 2, we study affective language production of individuals in a solitary setting. In the remaining chapters of this dissertation, we study this in a social interaction, starting with **Chapter 3**. In this chapter, we aim to answer the research question: 'Do emotions influence alignment between conversation partners, in specific, alignment in referential expressions?'. We replicate and extend Goudbeek and Krahmer (2012), who found that speakers listening to a pre-recorded voice describing furniture objects tend to use the same preferred (color, e.g., 'blue') *and* dispreferred (orientation, 'front') attributes in referential expressions when describing the objects themselves. The authors concluded that speakers align to their (hypothetical) conversation partner, by adjusting the content of their language, in this case, their referential expressions.

We aim to replicate this finding in a more naturalistic setting, by letting individuals describe objects of furniture to each other, instead of listening to the descriptions of a pre-recorded voice and describing the objects in private. In our experiment, participants were exposed to amusing and disgust-inducing film clips, before they were collected to be seated in front of another participant, forming dyads to engage in a director-matcher task. Using the same stimuli as Goudbeek and Krahmer (2012), both participants view pictures

of furniture, differing in size (small or large) and color, on their own computer screen. In each trial, participants view one target picture and two distractor pictures. Switching their role between trials, the director describes the target picture (which is, only for the director, framed by a red border) to the matcher, who has to select the correct picture. The interaction was audio recorded and transcribed. We investigate to which extent the two individuals align with each other during the task, in that they start using the same referential expressions – color, size, or both, and, subsequently, whether the emotional state of the speakers influenced their referential communication.

In **Chapter 4**, we studied another type of collaboration in conversation, namely, conceptual pact formation. We replicate and extent the seminal study of Clark and Wilkes-Gibbs (1986), who found that individuals repeatedly describing highly similar and complex abstract figures to each other, gradually and efficiently create temporary agreements, or conceptual pacts, referring to specific figures. As a result, over time, they need fewer words and speaking turns to describe the figures. In Chapter 4, we aim to investigate the role of affect in conceptual pact formation, answering the question: 'Do affective pictures, or affective states, influence the production of conceptual pacts in reference in dyads?'. We did this by a) studying whether dyads also form conceptual pacts when repeatedly describing naturalistic, affectively laden and highly similar (IAPS) pictures and b) assessing the influence of verbalizing affective content on both affective state of the speakers and conceptual pact formation within dyads. In order to research these two affairs, we created a paper-based director-matcher task, in which dyads described naturalistic, affectively laden pictures, varying in high or low (un)pleasant and (un)arousing content. Contrary to Chapter 2, participants were given a fixed role, director *or* matcher, and presented with identical sets of pictures, presented in different, fixed orders, on six sheets (director) or as a stack of separate cards, accompanied by a sheet depicting empty boxes (matcher). The director reported to the matcher which picture occupied which position in each particular sequence; the matcher placed the pictures in the correct order, in the boxes on their sheet. This process was repeated until the dyad ran through all six trials, and therefore, describing each picture six times. To analyze the number of words used, and turns taken to describe the pictures, the experiment was audio recorded and transcribed. Before and after engaging in the director-matcher task, participants indicated their affective state, based on their self-reported degree of pleasantness and arousal, by filling in the Self-Assessment Manikin (Bradley & Lang, 1994).

In **Chapter 5**, we aim to answer our question: 'Do affective states of speakers influence the production of (im)polite language in a conversation?'. To investigate this, we replicated and extended Forgas (1999a), who found that individuals presented with hypothetical difficult situations tend to formulate more polite requests when they are in a negative mood, compared to those in a positive mood.

We aimed to replicate this finding in a more naturalistic setting, in which dyads formulated requests in a real-life social situation, namely, a conversation. Given that autistic individuals tend to struggle with understanding and implementing social communication norms as politeness (e.g., Cummins et al., 2020; Mathersul et al., 2013), we studied the

influence of affect on politeness in non-autistic students (Study 1), as well as autistic individuals (Study 2), using (roughly) the same study design.

After an affect induction procedure to induce a happy, neutral or sad affective state, participants formed dyads to engage in a conversation. Before the onset of the conversation, they were presented with a booklet, featuring sensitive, personal affairs as bullying and infidelity. Participants were instructed to take turns asking each other about their personal experiences with the topics presented in the booklet. In order to investigate the (im)polite nature of the formulated requests for information, conversations were audio recorded and transcribed. Subsequently, requests were scored on the original request characteristics, including politeness, by Forgas (1999a), as well as politeness strategies (Brown & Levinson, 1987), to indicate the degree and type of politeness detected in the requests.

In **Chapter 6**, we provide a summary of the study findings and conclusions from the preceding chapters. Hereafter, a general discussion is given, in which we discuss the implications of our findings for the field of affective language production. Subsequently, we discuss to which degree we successfully replicated the findings of the original studies. The chapter is concluded with theoretical implications and a general conclusion.

2

Gradual positive and negative affect induction: The effect of verbalizing affective content

This chapter is based on:
Out, C., Goudbeek, M., & Krahmer, E. (2020). Gradual positive and negative affect induction: The effect of verbalizing affective content. ***PloS ONE***, ***15***(5), Article e0233592. https://doi.org/10.1371/journal.pone.0233592

ABSTRACT

In this chapter, we study the effect of verbalizing affective pictures on affective state and language production. Individuals describe (Study 1: Spoken descriptions of pictures) or passively view (Study 2: Passively viewing pictures) 40 pictures from the International Affective Picture System (IAPS) that gradually increase from neutral to either positive or negative content. We expected that both methods would result in successful affect induction, and that the effect would be stronger for verbally describing pictures than for passively viewing them. Results indicate that speakers indeed felt more negative after describing negative pictures, but that describing positive (compared to neutral) pictures did not result in a more positive state. Contrary to our hypothesis, no differences were found between describing and passively viewing the pictures. Furthermore, we analyzed the verbal picture descriptions produced by participants on various dimensions. Results indicate that positive and negative pictures were indeed described with increasingly more affective language in the expected directions. In addition to informing our understanding of the relationship between (spoken) language production and affect, these results also potentially pave the way for a new method of affect induction that uses free expression.

2.1. Introduction

Speaking about emotionally meaningful events is not a neutral act: it affects us. When we experience something positive, like getting a raise, we feel happy, and, conversely, explaining that we did not get a promotion makes us feel sad. This relationship between speaking and feeling is partly there because the event we talk about is inherently affective, but it might also be because verbalizing an emotionally meaningful fact ('I got a raise. I finally got a raise!') amplifies or even induces the emotions we experience.

In this chapter, we investigate the relationship between language production and affect. Inspired by the Velten method (Velten, 1968), we present a procedure where participants are exposed to pictures that increase or decrease in valence as the experiment progresses. In contrast to the Velten method that relies on reading aloud affective fixed self-referential statements, we use spontaneous spoken descriptions of affectively charged pictures. To elucidate the relationship between affective state and language production, we explore the content of the verbal descriptions, using the affective categories of the Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2001). To gauge the role of verbal description in eliciting affective reactions, we explore the difference between describing these pictures and silently viewing them.

2.1.1. The Velten method

The theoretical aim of Velten's (Velten, 1968) study was to find evidence for the efficacy of a type of cognitive therapy, focusing on making the patient aware of how their own verbal interpretations of events influence their affective responses. Predicting that affective phrases would indeed elicit corresponding responses, Velten created an affect induction method to elicit positive (elation) or negative (depressive) affect by having participants read out loud 60 sentences that gradually increased in affective content. The first sentence in both conditions was 'Today is neither better nor worse than any other day'. In the positive condition, it was followed by sentences such as 'Things look good, things look great!' with the final sentence being 'God, I feel great!'. In the negative condition, it was followed by sentences such as 'It often seems that no matter how hard I try, things still go wrong', with the final sentence being 'I want to go to sleep and never wake up'. Velten also included a neutral condition, containing sentences as 'The review is concerned with the first three volumes' and 'West Samoa gained its independence in 1965'. After reading the statements, various measures were obtained, including several cognitive and behavioral tasks. In one of the tasks, participants were asked to choose from a long list of adjectives which adjectives applied to them (Multiple Affect Adjective Check List, Today Form (MAACL; Marcusson-Clavertz et al., 2019), and the experiment leader kept track of the number of words the participant uttered during the tasks.

As Velten predicted, participants who read the negative statements, compared to those who read the positive statements, ticked significantly more adjectives in the Depression Scale, one of the five emotion subscales of the MAACL, and in general uttered less words

(Velten, 1967). Velten concluded that these results, together with his other measurements, indicated that the induction method was effective: participants reading the negative statements felt more depressed, and participants reading positive statements felt more elated. Participants reading the neutral statements generally fell between the scores of elation and depression, implying that no effect on mood took place (Velten, 1968). After its initial development, the Velten method has been widely used in the following decades (e.g., Jennings et al., 2000), and the effects have frequently been replicated (e.g., Jennings et al., 2000; Scherer, 2013; Wilting et al., 2006).

2.1.2. The relationship between verbal expression and affective state

In general, the Velten method appears to be highly relevant for researchers studying the relationship between affect and language, which has been under scholarly debate for only about a decade (Lindquist, 2017), resulting in various hypotheses. One hypothesis is based on the psychological constructionist approach. According to this approach, language has a pivotal, although not sufficient, role in perceiving and experiencing affective states. The approach suggests that the lexicon of (affective) words at our disposal is essential to make meaning of, and therefore shape, our affective experiences, for example, by turning general, vague feelings of displeasure ('this doesn't feel right') into concrete emotions ('I feel lonely'; Lindquist, 2017). According to Barrett (2017), this categorization is learned from infancy, and depends on the social (and cultural) environment. To a certain extent, emotions are *created* by naming, and therefore categorizing (and experiencing) them.

However, the literature shows conflicting results with respect to the impact of verbalization on affective state. On the one hand, there is support that expressing affective content can attenuate the affective experiences, e.g., via 'affect labelling'. When individuals use affect labelling and put their emotions into words this can result in a decrease in the intensity of the affective, often negative, experience (see, e.g., Fan et al., 2019; Lieberman et al., 2007; Torre & Lieberman, 2018). Therefore, some consider affect labelling to be an unintentional or incidental form of affect regulation (see e.g., Burklund et al., 2014). In one study, Fan et al. (2019) studied affect labelling in naturally occurring, spontaneous emotional expression on Twitter, looking at tweets starting with 'I feel…', followed by an adjective or adverb, written by approximately 74.487 different Twitter users. A dictionary-based affect detection algorithm, VADER (Valence Aware Dictionary and Sentiment Reasoner; Hutto & Gilbert, 2014) was used to detect possible changes in affective content of the tweets six hours after (and before) the affect labelling took place. Their results showed that for most individuals, immediately after using affect labelling in a tweet, the level of affective content of their tweets decreased, before returning to baseline (Fan et al., 2019). Negative emotions returned to baseline fairly rapidly, with a decay half-life of five minutes, while for positive emotions a less rapid reduction was observed with a decay half-life of eleven minutes. The authors conclude that their findings are in line with literature on the attenuating effects of affect labelling.

In contrast to the findings summarized above, putting emotions into words can also result in an enhanced affective experience. For example, Ortner (2015) presented participants with neutral and negative pictures. First, they merely viewed 10 neutral and 10 negative pictures. Then, for the next 10 neutral and 10 negative pictures, participants were asked to either passively view, reappraise (reinterpret the pictures in a way that it no longer seemed negative) or emotionally label pictures (observe which emotions they experience and utter their labels, e.g., 'there is… anger'). The results showed that participants using affect labelling reported stronger affective states than those reappraising or only viewing the pictures. Ortner (2015) suggests that the individuals who verbally described their emotional reactions to the affective pictures in their own words created a heightened awareness of them and therefore, experienced more intense affect.

Finally, in expressive writing, the verbalizing of affect often results in an initial increase, followed by a decrease in affective intensity. Expressive writing is a well-known and successful technique to deliberately reduce (unwanted) negative affect and distress in the long run (e.g., Baikie & Wilhelm, 2005; Soliday et al., 2004). For this technique, individuals are asked to write 15-20 minutes about a traumatic or personally emotional event for several consecutive days (Pennebaker, 1997). In contrast to Fan et al. (2019), the decrease in negative emotions was not immediate: during, and immediately after writing, individuals usually reported feeling worse (e.g., Baikie & Wilhelm, 2005; Pennebaker & Beall, 1986), although at least one study reports that individuals scored higher on positive disposition shortly after the last expressive writing episode (Soliday et al., 2004). In all these studies, participants experienced a decrease in negative affect in the long run, even while feeling worse immediately after the linguistic expression of the emotional event.

2.1.3. Describing affective pictures

The findings discussed above indicate the existence of a crucial, although unclear, relationship between language production and affective state and vice versa, especially when individuals are allowed to use their own words. Asking participants to describing affectively laden pictures might be an excellent way to bring about an affective state, while simultaneously allowing the free production of linguistic content that, in turn, might affect the extent to which the pictures induce an affective state compared to a situation where participants are merely passive observers of the pictures.

For our study, we selected pictures from a well-known and validated set of affectively laden pictures, the International Affective Picture System (IAPS; Lang et al., 1997; Uhrig et al., 2016). The IAPS is a large set of pictures of varied content, including arousing and (un)pleasant (e.g., snakes and spiders, romantic couples, and extreme sports) as well as more neutral pictures (e.g., flowers, objects, and portraits). The pictures have been rated on valence or pleasure (negative/positive), arousal (low/high) and dominance (dominated/in control; Lang et al., 1997), as well as for discrete emotion categories (e.g., Mikels et al., 2005).

Most studies using the IAPS pictures to induce affective state select a number of IAPS pictures within certain ranges of valence to create subsets of positive, negative and sometimes neutral pictures (e.g., Capecelatro et al., 2013). Per the Velten method, we aimed to gradually induce positive and negative affect by exposing individuals to sets of IAPS pictures that start with neutral content, and gradually become more positive or negative, based on their valence ratings (Lang et al., 1997).

2.1.4. The current studies

In order to study the effect of language production on affect, we devised a method where individuals expressed themselves in their own (possibly affective) language as a way to induce affect. Our method is inspired by the incremental nature of the Velten method, but asks participants to describe affectively evocative pictures instead of read out loud sentences. This modification is prompted by the desire to investigate the relationship between language and affect, while simultaneously using a more natural paradigm that is less prone to demand characteristics.

To assess whether verbally describing the pictures would indeed result in an enhanced affective experience we contrast the effect of this with the effect of passively viewing pictures, which is the more conventional way of using the IAPS pictures (e.g., Capecelatro et al., 2013). To our knowledge, this is the first study that investigates the effect of describing affectively charged pictures on the affective state of the speaker. Given that previous work found that describing personal affective experiences either decreases (e.g., Torre & Lieberman, 2018), increases (e.g., Ortner, 2015), or increases and then decreases (Baikie & Wilhelm, 2005) the intensity of the experienced affect, this comparison could go either way. However, given that the verbalizing self-referential statements were effective in the Velten method, we hypothesize that, compared to passively viewing them, verbally describing affectively laden pictures will enhance the affective experience.

In sum, to investigate our research questions, we conduct two studies investigating whether there is an additive effect of verbalizing the content of affective pictures on affective state (compared to merely viewing them). For this, we used pictures taken from the IAPS, gradually increasing in affective content (positive, negative) or remaining neutral. In Study 1 (called 'Spoken descriptions of pictures'), participants view and describe the pictures out loud. In Study 2 (called 'Passively viewing pictures'), participants passively view the pictures, and do not describe them.

Finally, to elucidate the relationship between affective state and the language that is used in the descriptions, we will explore the content of the verbal descriptions of the pictures of Study 1, comparing the frequency of affective word use in the three (affective) content categories, and word count, using LIWC (Pennebaker et al., 2001). We preregistered the methods, hypotheses, and analyses of this study at the Open Science Foundation: https://osf.io/kv8g3.

2.1.5. Hypotheses

With respect to our gradual affect induction procedures, we have the following hypotheses:

(H1) Irrespective of whether they describe the pictures out loud, participants in the condition with positive pictures will report higher levels of pleasant affect. In the condition with negative pictures, they will report higher levels of unpleasant affect. No differential effect is expected for the neutral pictures.

(H2) We expect that *describing* affective pictures will enhance the effect on affective state compared to passively viewing the pictures. Specifically, we predict that participants viewing and describing positive pictures (Study 1) will report higher levels of pleasant affect than participants passively viewing positive pictures (Study 2), and participants viewing and describing negative pictures (Study 1) will report higher levels of unpleasant affect than participants passively viewing negative pictures (Study 2). To determine if this hypothesis is true, the results from Study 1 and Study 2 will be compared.

2.2. Study 1: Spoken descriptions of pictures

Study 1 investigated the effect of viewing and describing (out loud) pictures gradually increasing in affective content on (self-reported) affective state.

2.2.1. Method

2.2.1.1. Design

The study had a 2 (Time: pre-test, post-test) x 3 (Condition: positive, neutral, and negative) design, with time as within-subjects variable and condition as between-subjects variable.

2.2.1.2. Participants

In total, 122 participants were recruited at a Dutch university and participated in the experiment for course credit. One participant was excluded because they did not consent to their data being used. Our final sample included N = 121 participants (41 male; M_{age} = 22.22, SD_{age} = 2.90), each randomly assigned to one of the conditions (positive condition: n = 41; neutral condition: n = 40; negative condition: n = 40).

All procedures performed were in accordance with the ethical standards of the institutional research committee, the Research Ethics and Data Management Committee of Tilburg School of Humanities and Digital Sciences, Tilburg University. All participants gave written informed consent in accordance with the Declaration of Helsinki (1964) and its later amendments or comparable ethical standards.

2.2.1.3. Materials

2.2.1.3.1. Stimuli

We used the 2008 variant of the IAPS, containing 1194 pictures (Chen et al., 2015). In order to create three conditions, we selected 40 IAPS pictures per condition based on the procedure described below.

First, two sets of 600 positive and negative pictures each were created. For the positive condition, we started from the 600 pictures with the highest valence rating (range 5.22-8.34). For the negative condition, we started from the 600 pictures with the lowest valence ratings (range 1.31-5.24). Indeed, the ranges of positive and negative pictures partly overlap. This mirrors the Velten method, which starts with the same sentence in both the positive and negative set of statements.

Next, both sets were divided into 40 bins of fifteen pictures, with each bin increasing in pleasant (positive condition) or unpleasant (negative condition) content. From each bin, one picture was randomly selected, resulting in two sequences of 40 pictures, which gradually increased in (un)pleasant content.

To create the picture set for the neutral condition, we selected 301 pictures with an average valence rating (range 4.62-5.92). Forty random bins of fifteen pictures were created, and from each bin one picture was randomly selected.

While selection of pictures from bins was random in principle, sometimes a picture was deemed inappropriate and replaced by another randomly selected picture from the same bin. Exclusion criteria were: erotic or sexually suggestive (but not non-erotic nudity), too gruesome or disgust-inducing, repetitive content, or culturally sensitive content (e.g., traditions and rituals). Based on these criteria, we excluded six pictures and replaced them with more appropriate pictures from the same bin (Supporting Information, Table S1, https://doi.org/10.1371/journal.pone.0233592). Our final sample can be found in Table S2 of the same section.

Finally, to check if the three sets of pictures did not contain any outliers that would disturb the gradual increase (positive and negative condition), or would interfere with a consistent level (neutral condition) of affective content of the pictures, the sets of pictures were inspected for their valence and arousal in two line plots (Appendix, Figure 1A and Figure 2A). As shown in Figure 1A, both the positive set, and the negative set, displayed a near perfect gradual increase in affective content in the expected directions. For the neutral set, a (very) slight decrease in pleasant content can be observed. As can be found in Figure 2A, compared to the valence ratings in Figure 1A, the arousal ratings were less distinct in their sequential direction, with a strong increase in arousal for negative pictures, and no substantial in- or decrease in arousal for both the positive pictures, as the neutral pictures. For the three sets of pictures, the range, mean (with standard deviation) and median of the valence and arousal scores can be found in Table 1.

Table 1

Statistical characteristics of the final sample of pictures

	Valence			Arousal		
Pictures	Range	*M* (*SD*)	Median	Range	*M* (*SD*)	Median
Positive	5.22-8.05	6.52 (0.77)	6.52	2.63-7.31	4.33 (1.04)	4.07
Neutral	4.95-5.22	5.08 (0.82)	5.07	2.00-6.23	3.52 (0.93)	3.22
Negative	1.51-5.22	3.53 (1.13)	3.57	1.72-7.07	4.73 (1.55)	4.97

2.2.1.3.2. Viewing IAPS pictures

In Study 1, participants were given instructions to describe each picture out loud, inspired by the MS COCO instructions (Chen et al., 2015), which is a well-established method of eliciting picture descriptions. In our study, participants were instructed to describe all the important aspects and details of the pictures, describe them in a way that another person could recognize this picture out of the set of 40 pictures, and use full sentences when describing the pictures.

After piloting with various timeframes (6, 8, and 10 seconds), 10 seconds viewing time per picture appeared to be sufficient to describe the pictures. Each participant started with two practice trials describing two neutral pictures. In order to encourage participants to actively engage in the task, we presented them with a bogus purpose of the study: memorizing the pictures. The study was introduced as a memory experiment, and participants were told that they would be asked to indicate pictures they had, and had not, seen before from a set of new and old (already seen) pictures.

2.2.1.3.3. Video- and audio recording

Audio was recorded for content analysis of the picture descriptions. In addition, we video recorded facial expressions for possible future analysis.

2.2.1.3.4. Affect questionnaire

Before and after viewing the series of pictures, participants indicated their current affective state on six 7-point Likert scales: sad/happy, unpleasant/pleasant, unsatisfied/satisfied, discontent/content, sullen/cheerful, low-spirited/in high spirits (Krahmer et al., 2004, based on Bohner et al., 1992; Mackie & Worth, 1989; English translations of Dutch originals). They were instructed to choose a number per scale; the closer the numbers were to the words, the stronger they match the feeling described the word in question. Low numbers indicated the degree of negative affect (e.g., unpleasant), high numbers indicated the degree of positive affect (e.g., pleasant). In a previous study by Krahmer et al. (2004), the internal consistency of this questionnaire was good, $\alpha = .88$. We assessed the reliability of the current scale with Cronbach's α as well. This analysis indicated that the items of our affect questionnaire had excellent internal consistency, for both Study 1 (pre-test, $\alpha = .90$; post-test, $\alpha = .94$) and Study 2 (pre-test, $\alpha = .94$; post-test, $\alpha = .95$). Based on these results, the

six items were merged into one scale, 'Affect', resulting in one pre-test and one post-test score per participant, indicating (self-reported) affective state, ranging from 1 (negative affect) to 7 (positive affect). Based on these results, the six items were merged into one scale, 'Affect', resulting in one pre-test and one post-test score per participant, indicating (self-reported) affective state, ranging from 1 (unpleasant) to 7 (pleasant).

2.2.1.4. Procedure

After participants signed the informed consent form, the experiment leader explained the procedure and turned on the camera, including audio recording. If needed, the camera was adjusted to an appropriate height to record the participant's face. Participants reported their gender and age. They then filled out the affect questionnaire for the first time (pre-test). Then, starting with two practice trials, participants were asked to view and describe 40 pictures out loud. After the task, they filled out the affect questionnaire again (post-test). Then, participants were asked to indicate which pictures they had seen before, and which ones they had not. Pictures were selected beforehand, by randomly picking three numbers between 1 and 40, using the corresponding bin to select one 'old', and one 'new' picture. Ninety percent (n = 220) of all participants (N = 245) correctly identified all six pictures as 'old' or 'new'.

After the experiment, participants in the negative condition viewed a light-hearted, short video displaying a jumping competition for bunnies (Vice, 2012). This video was shown to rise their spirits, in case participants felt especially low after the experiment. Participants in the positive and neutral condition did not watch the video. At the end, the participants were debriefed and thanked for their participation.

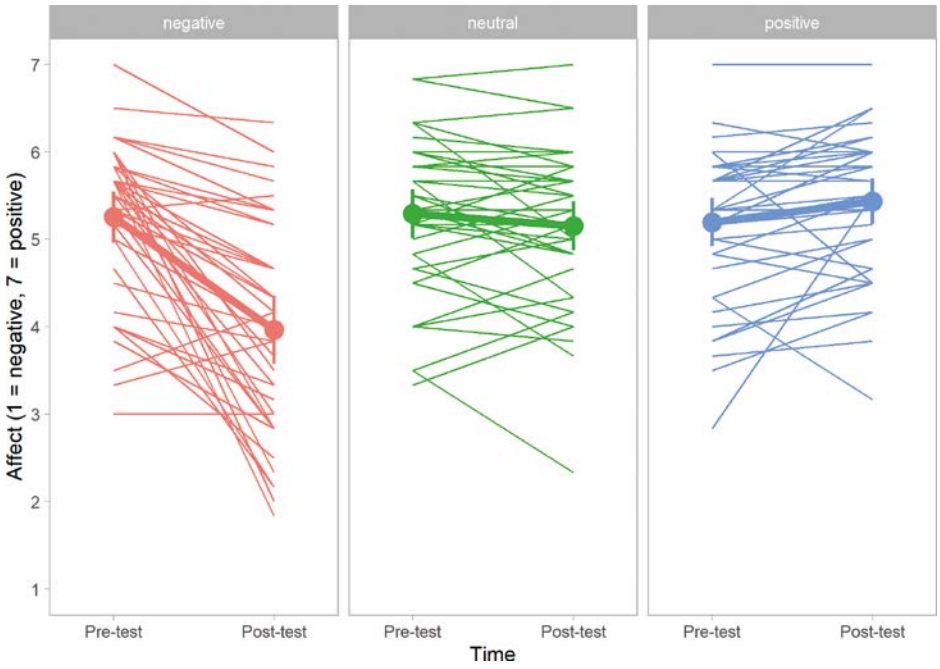
2.2.2. Results

2.2.2.1. Descriptive statistics

Figure 1 displays the individual scores on affect (y-axis) for the pre- and post-test (x-axis), sorted by condition (positive, neutral, and negative pictures). On the y-axis, lower scores indicate the degree of unpleasant affect; higher scores indicate the degree of pleasant affect. In Figure 1, the results for the three conditions show a clear pattern. Participants viewing negative pictures generally report feeling unpleasant after describing the pictures. Participants viewing positive pictures generally report feeling slightly more pleasant describing the pictures, and participants viewing the neutral pictures did not seem to report a change in affective state. In general, participants in all conditions seem to start the experiment in fairly good spirits (possibly partly explaining the limited effect in the positive condition), scoring roughly 5 to 5.5 on the 7-point Likert scale.

Figure 1

Individual affect scores for participants viewing and describing pictures



Note. The dots of the bold lines represent the mean scores.

2.2.2.2. Change in affective state

We performed a repeated measures analysis of variance with time (pre-test and post-test) as within-subjects factor, condition (positive, neutral, or negative pictures) as between-subjects factor and affective state as dependent variable. Mean scores, standard deviations, difference scores (post-test - pre-test) and range can be found in Table 2. A main effect was found for time, $F(1, 118) = 28.03$, $p < .001$, $\eta_p^2 = .19$, and for condition, $F(2, 118) = 8.05$, $p = .001$, $\eta_p^2 = .12$. However, these two main effects were qualified by a (predicted) interaction effect for time and condition, $F(2, 118) = 37.23$, $p < .001$, $\eta_p^2 = .39$. Post-hoc tests revealed that participants in the negative condition reported lower levels of pleasantness after viewing and describing the negative pictures. Participants in the positive or neutral condition did not report a significant change in affective state after viewing and describing the pictures.

Table 2

Affective state scores of participants in Study 1 and Study 2. Mean scores (with standard deviations), difference scores and range are displayed

Condition	Time	Spoken descriptions of pictures	Passively viewing pictures
		M (SD)	*M (SD)*
Positive	Pre-test	5.20 (0.88)	4.94 (1.17)
	Post-test	5.44 (0.83)	5.06 (1.08)
	Difference	0.24 (0.62)	0.12 (0.86)
	Range	-1.50, 2.67	-3.17, 2.17
Neutral	Pre-test	5.28 (0.89)	5.13 (0.92)
	Post-test	5.13 (0.89)	4.92 (0.97)
	Difference	-0.15 (0.54)	-0.21 (0.36)
	Range	-1.33, 0.67	-1.17, 0.50
Negative	Pre-test	5.25 (0.91)	4.88 (1.03)
	Post-test	3.97 (1.22)	3.58 (1.12)
	Difference	-1.28 (1.17)	-1.30 (1.04)
	Range	-4.00, 0.67	-3.83, 0.83

2.2.3. Conclusion

As predicted, participants viewing and describing negative pictures reported to experience lower levels of pleasantness, and participants viewing and describing neutral pictures did not report a change in affective state after completing the task. In contrast to our prediction (but in line with other, earlier studies reporting unsuccessful positive affect induction, as discussed below), participants viewing and describing positive pictures did not report (significantly) higher levels of positive affect after the task.

2.3. Study 2: Passively viewing pictures

Study 2 studied the effect of passively viewing, but not describing out loud, pictures gradually increasing in affective content on (self-reported) affective state. We used the same sets of pictures as in Study 1.

2.3.1. Method

2.3.1.1. Design

The design was identical to Study 1.

2.3.1.2. Participants

Participants were recruited at the same Dutch university as in Study 1. A total of 126 participants participated in the experiment for course credit; none of them participated in Study 1. Two participants did not consent to have their data published in scientific journals; therefore, we excluded their data. Our final sample included N = 124 participants (43 male; M_{age} = 23.50, SD_{age} = 4.00), again, each assigned to one of the conditions (positive condition, n = 41, neutral condition, n = 41; negative condition, n = 42). Again, all procedures were in accordance with the ethical standards of the local research committee. Written informed consent was obtained from all individual participants included in the study.

2.3.1.3. Materials

2.3.1.3.1. Stimuli

The materials we used were identical to those used in Study 1, but in contrast to Study 1, participants could do the experiment in Dutch or English, because they did not verbally describe the pictures, the language they spoke became irrelevant. Participants received informed consent, instructions and debriefing, and fill out the questionnaires, in their language of choice.

2.3.1.3.2. Viewing IAPS pictures

The viewing time per picture was identical to Study 1.

2.3.1.4. Procedure

The procedure was identical to that of Study 1, except that participants only passively viewed the pictures, instead of viewing them and describing them out loud. For this reason, no audio recording took place.

2.3.2. Results

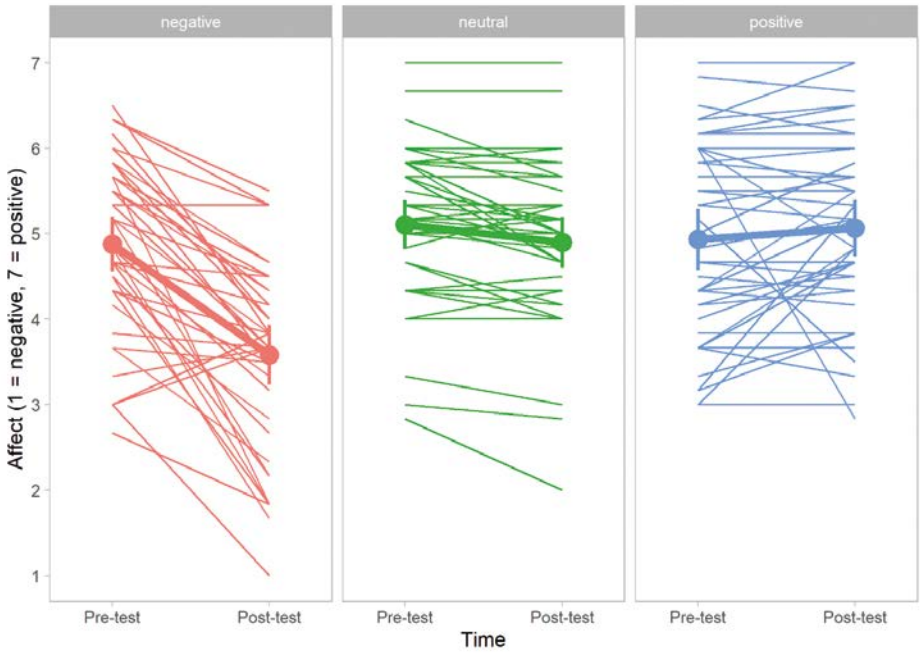
2.3.2.1. Descriptive statistics

As in Figure 1, Figure 2 displays the individual scores of affective state (y-axis) for the pre- and post-test (x-axis), sorted by condition (positive, neutral, and negative pictures). On the y-axis, lower scores indicate the degree of unpleasant affect; higher scores indicate the degrees pleasant affect.

Notice that Figure 2 looks very similar to Figure 1, showing the same pattern as described above: participants viewing negative pictures generally reported feeling unpleasant, and participants viewing positive or neutral pictures generally did not report a substantial change in affective state. Akin to the participants in Study 1, participants in Study 2 generally started the experiment in fairly good spirits, scoring roughly 5 to 5.5 on the 7-point Likert scale.

Figure 2

Individual affect scores for participants passively viewing pictures



Note. The dots of the bold lines represent the mean scores.

2.3.2.2. Change in affective state

A repeated measures analysis was performed with time (pre-test and post-test) as within-subjects factor, condition (positive, neutral or negative pictures) as between-subjects factor and affective state as dependent variable. Mean scores, standard deviations, difference scores and range can be found in Table 2. As in Study 1, a main effect was found for time, $F(1, 121) = 40.22$, $p < .001$, $\eta_p^2 = .25$, and condition, $F(2, 121) = 8.99$, $p < .001$, $\eta_p^2 = .13$. However, again, these two main effects were qualified by a (predicted) interaction effect for time and condition, $F(2, 121) = 35.16$, $p < .001$, $\eta_p^2 = .37$. Identical to Study 1, post-hoc tests revealed that participants in the negative condition reported lower levels of pleasant state after viewing the negative pictures. Again, participants in the positive or neutral condition did not report a significant difference in affective state after viewing the pictures.

2.3.3. Conclusion

Similar results to Study 1 were found: participants viewing negative pictures reported negative affect after viewing the pictures, and participants viewing positive or neutral pictures did not report a significant change in affective state.

2.4. Comparing spoken descriptions of pictures and passively viewing pictures

To determine if viewing and describing the positive and negative pictures out loud (Study 1), compared to passively viewing them (Study 2), evoked higher levels of (positive or negative, respectively) affect, the ratings from Study 1 and Study 2 were compared using an ANOVA.

2.4.1. Individual changes in affective state

To explore our dataset, we looked at the changes in affective state for all individual participants (both Study 1 and Study 2). As can be inferred from both Figure 1 and Figure 2, there is a substantial amount of variation in the effectiveness of the manipulation, with only the negative condition showing a consistent pattern for the majority of the participants.

Generally speaking, participants viewing positive pictures (n = 82) reported feeling more pleasant after the task (n = 47), albeit the change was modest (≤ 1 on a 7-point scale) for the majority of participants (n = 40). Regarding participants viewing neutral pictures (n = 81), the majority (n = 75) reported a small change in affective state, feeling more positive (≤ 1) or negative (≤ -1). Participants viewing negative pictures (n = 82) showed the same pattern, but contrary to participants exposed to the neutral pictures, the variation between participants was much larger: 74 participants reported more unpleasant affect, of which 50 individuals reported a decrease of ≥ -1 on the affect scale.

There were no large (individual) differences between the individuals describing the pictures out loud, or only passively viewing them.

2.4.2. Results

To test the hypothesis that verbally describing affective pictures, compared to only viewing them, enhances the effect on affective state, a mixed ANOVA was performed with time (pre-test and post-test) as within-subjects factor, type of study (Study 1: Spoken descriptions of pictures, or Study 2: Passively viewing pictures) and condition (positive, neutral or negative pictures) as between-subjects factors, and affective state as dependent variable.

A main effect was found for type of study, $F(1, 239) = 6.39$, $p = .012$, $\eta_p^2 = .03$ (Study 1: $M = 5.04$, $SD = .91$; Study 2: $M = 4.75$, $SD = 0.91$), indicating that affective state was overall more positive for participants in Study 1, compared to Study 2. A main effect was also found for time, $F(1, 239) = 67.50$, $p < .001$, $\eta_p^2 = .22$ (pre-test: $M = 5.11$, $SD = .97$; post-test: $M = 4.68$, $SD = 1.22$), indicating that participants experienced more negative affect after engaging in the task (reflecting the effective manipulation in the negative condition). Finally, a main effect was found for condition, $F(2, 239) = 16.82$, $p < .001$, $\eta_p^2 = .12$ (positive: $M = 5.16$, $SD = .94$; neutral: $M = 5.11$, $SD = .91$; negative: $M = 4.42$, $SD = .94$), indicating that affect was lower overall in the negative condition, both for participants that described the pictures and for those that did not. However, no three-way interaction of time, type of study, and condition

was found, $F(2, 239) = .09$, $p = .915$, indicating that describing or only viewing affective pictures did not influence affective state significantly for one or more of the conditions. Given that there was no significant change in affective state for participants after viewing positive or neutral pictures, we wanted to rule out the possibility that these null effects obscured a possible difference for the negative condition. Selecting only the negative condition, a repeated measures ANOVA was performed with time (pre-test and post-test) as within-subjects factor, type of study (Study 1 or Study 2) as between-subjects factor and affective state as dependent variable. As predicted, a main effect was found for time, $F(1, 80) = 111.95$, $p < .001$, $\eta_p^2 = .583$, with participants becoming more negative during the experiment. However, we found no effect for type of study, $F(1, 80) = 3.49$, $p = .065$, and, importantly, no interaction between time and study, $F(1, 80) = 0.01$, $p = .98$. The results of this secondary analysis again indicate that individuals experienced worse affective state after exposure to the pictures, regardless of whether they described the pictures out loud or not.

2.5. An exploratory content analysis of the picture descriptions

In order to investigate the language use of the participants, and get more insight in how individuals describe (affective) content, we explored the verbal picture descriptions of Study 1, using the word counting software LIWC (Pennebaker et al., 2001). LIWC is a text analysis software program for counting words and calculating percentages of words, grouping them in various categories, including cognitive- and affective processes. For our current analysis, we used the Dutch LIWC dictionary (Zijlstra et al., 2005) to keep track of the words in the LIWC-categories 'affective processes' (to which we will refer to as 'affective words', e.g., dirty, help), 'positive emotion' words (e.g., beautiful, hug), 'negative emotion' words (e.g., sad, cry), as well as the word count per picture description.

Verbalizations of descriptions were transcribed by five individuals outside the project. The utterances of $N = 122$ participants were transcribed, resulting in 40 x 122 = 4880 descriptions. Forty-three descriptions, less than 1% of the dataset, were missing: all descriptions from one participant (in the neutral condition), two descriptions from one participant, and one description from one participant. One participant was excluded because she did not consent to her data being used. Our final sample included 4797 picture descriptions by $n = 120$ speakers, with a mean word count of 18.89 ($SD = 6.73$) words per description.

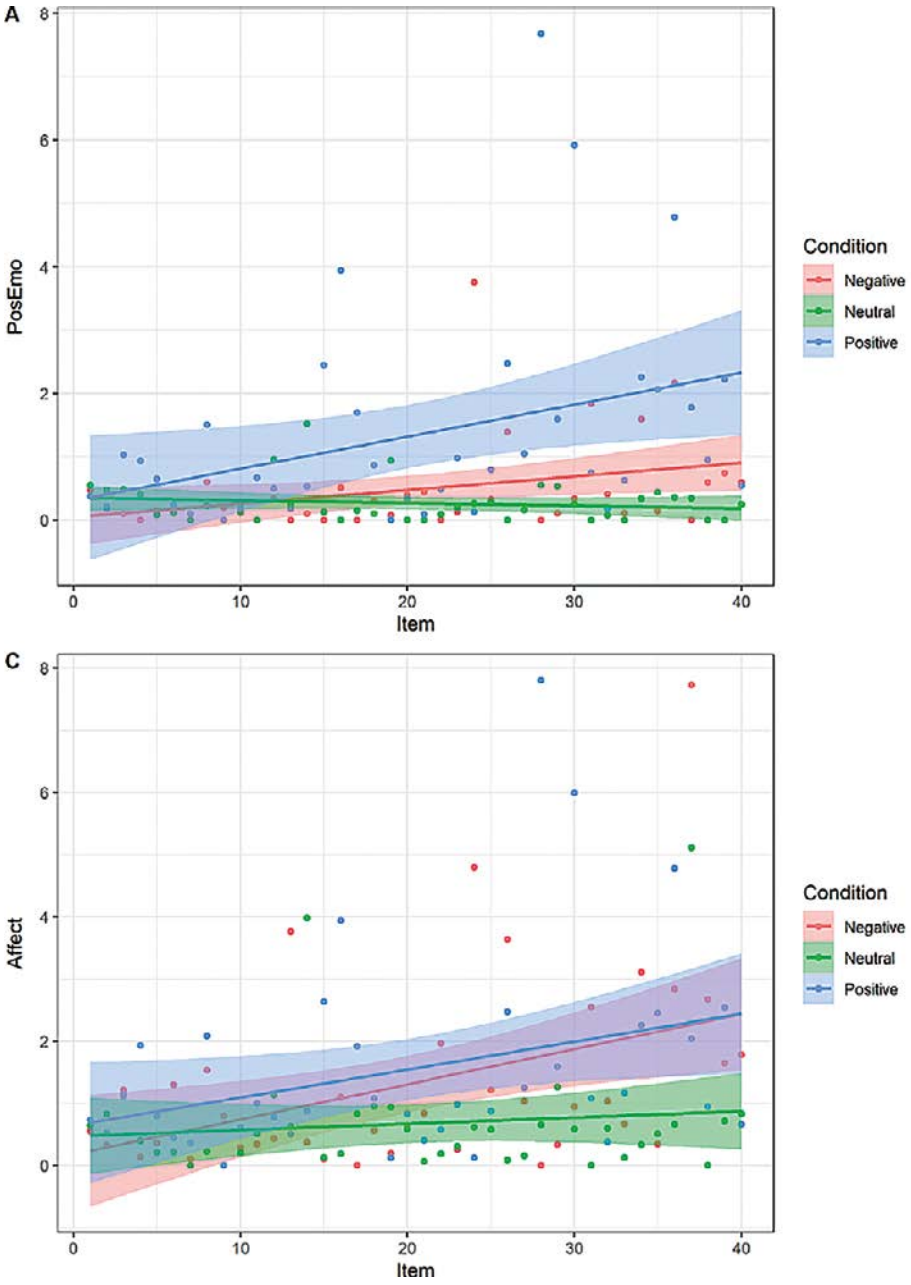
2.5.1. Descriptives

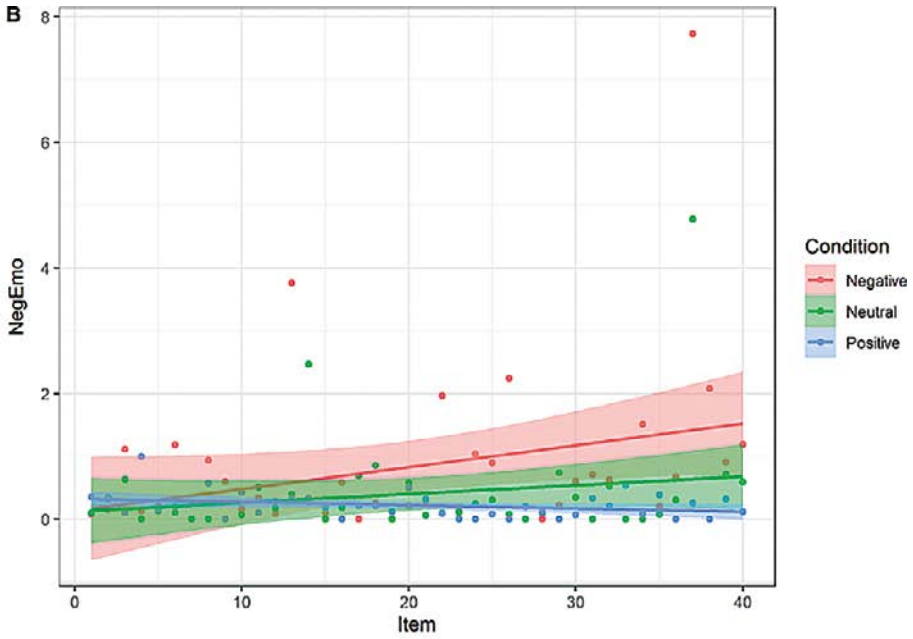
Table 3 provides the mean percentages (with standard deviations) of total words used per picture description, in the corresponding LIWC categories, per condition. Figures 3a-d depict the average scores per item in the respective LIWC category, represented by dots (the average score per item) and trend lines, including bands, representing confidence intervals. As can be seen in Figures 3a-c, most individuals tend to use no (0) or few (1, 2) affective, positive, or negative emotion words to describe a picture. Participants gradually used more positive emotion words to describe positive pictures and negative pictures, but not neutral pictures

(Figure 3a). The same pattern was found for negative emotion words, although the increase was less steep (Figure 3b). Affective word use gradually increased to describe positive and negative pictures, but not neutral pictures (Figure 3c). In all conditions, the data suggest that the number of words increases with subsequent pictures, a trend which is most clear for negative pictures (Figure 3d). However, we should be cautious interpreting this pattern since there is also substantial variation between participants.

Figure 3

Trend lines for average scores per item (represented by dots), per condition for positive emotion words, negative emotion words, affective words, and word count. Bands represent confidence intervals




B
NegEmo
Item
Condition
Negative
Neutral
Positive

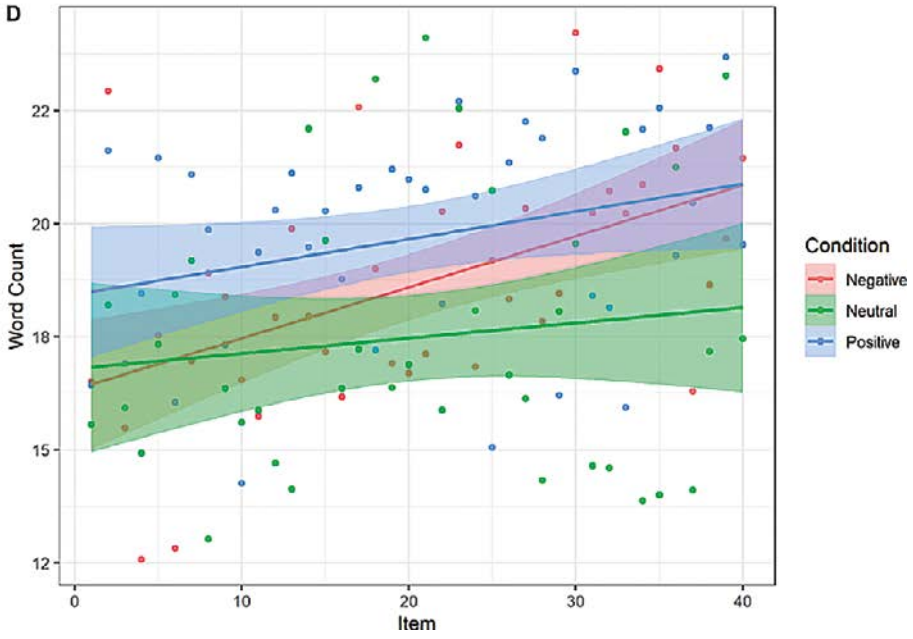

D
Word Count
Item
Condition
Negative
Neutral
Positive

Table 3

Mean scores, standard deviations and confidence intervals per condition for word count, and percentages of affective words, positive emotion words, and negative emotion words

Condition	Descriptions	Affective words		Pos. words		Neg. words		Word count	
	n	*M* (*SD*)	CI	*M* (*SD*)	CI	*M* (*SD*)	CI	*M* (*SD*)	CI
Positive	1638	1.54 (3.21)	1.39-1.70	1.26 (2.85)	1.13-1.40	0.13 (0.93)	0.09-0.18	20.16 (6.52)	19.84-20.47
Neutral	1560	0.75 (2.42)	0.63-0.87	0.45 (1.98)	0.35-0.55	0.27 (1.27)	0.21-0.33	17.58 (6.90)	17.24-17.92
Negative	1599	1.69 (3.43)	1.52-1.86	0.53 (1.70)	0.45-0.61	1.03 (2.76)	0.90-1.17	18.88 (6.53)	18.56-19.20
Total	4797	1.33 (3.08)	1.25-1.42	0.75 (2.27)	0.69-0.82	0.48 (1.87)	0.43-0.53	18.89 (6.73)	18.70-19.09

2.5.2. Results

To statistically analyze word count and affective words used in the picture descriptions, four separate one-way ANOVAs were performed, with condition as independent factor, and word count, (percentage of) affective-, positive- and negative word use, as dependent variables. Data were aggregated on individual level, combining individual scores on each picture description to one mean score for each LIWC category. We tested for homogeneity of variances using Levene's tests. Results of ANOVAs and Levene's tests can be found in Table 4. Levene's test results indicated that equal variances were assumed for word count, but not for affective-, positive-, and negative word use. Differences between the conditions were assessed with Tukey's (equal variances assumed) and Games Howell (equal variances not assumed) post hoc comparisons.

Table 4

ANOVAs and Levene's tests for condition on affective-, positive-, and negative word use, and word count

	One-way ANOVA	Levene's test
Affective word use	$F(2, 117) = 20.78$, $p < .001$	$F(2, 117) = 9.19$, $p < .001$
Positive word use	$F(2, 117) = 25.57$, $p < .001$	$F(2, 117) = 10.31$, $p < .001$
Negative word use	$F(2, 117) = 134.87$, $p < .001$	$F(2, 117) = 6.77$, $p = .002$
Word count	$F(2, 117) = 3.30$, $p = .040$	$F(2, 117) = 0.22$, $p = .801$

2.5.2.1. Affective words

Although our positive affect induction was not successful, participants describing positive pictures generally used more affective words in their descriptions, compared to participants describing neutral pictures, $p < .001$. Negative pictures were described with more affective words than neutral pictures, $p < .001$. No difference was observed for affective word use between positive and negative pictures, $p = .670$.

2.5.2.2. Positive emotion words

Positive emotion words were used significantly more when describing positive pictures, compared to negative pictures, $p < .001$, and neutral pictures, $p < .001$. There was no significant difference between negative and neutral pictures, $p = .483$.

2.5.2.3. Negative emotion words

A similar pattern was observed for negative word use: participants describing negative pictures used significantly more negative emotion words, compared to positive pictures, $p < .001$, and neutral pictures, $p < .001$. Additionally, neutral pictures were described with more negative emotion words than positive pictures, $p = .012$.

2.5.2.4. Word count

Individuals used more words describing positive pictures compared to neutral pictures, $p = .031$, but not compared to negative pictures, $p = .409$. No significant difference was found between neutral and negative pictures, $p = .405$.

2.5.3. Conclusion

The results of this exploratory content analysis are in line with what would be intuitively expected. Individuals viewing affective pictures used more affective words in their descriptions, compared to when they are describing neutral pictures. Positive pictures, compared to negative and neutral pictures, were described with more positive emotion words, and conversely, negative pictures were described with more negative emotion words than neutral pictures, and neutral pictures were described with more negative emotion words than positive pictures. Speakers used more words to describe positive pictures than neutral pictures.

2.6. General discussion

In this study, we aimed to study the effect of free verbal expression on affect induction, by investigating the effectiveness of affect induction methods, inspired by Velten (1968), where pre-defined self-referential statements are replaced with IAPS pictures, gradually increasing in affective content (positive, negative) or remaining neutral. In Study 1, 'Spoken descriptions of pictures', individuals verbalized the content of the pictures out loud. In Study 2, 'Passively viewing pictures', participants passively viewed the pictures, and did not describe them out loud. Our first hypothesis was partly confirmed: as predicted, for both studies, negative affect induction was effective, and the neutral condition did not evoke a change in affective state. However, in both studies, positive affect induction did not result in a significant enhancement of positive affective state when compared to the neutral condition. Our second hypothesis was not confirmed: describing the pictures out loud did not enhance, nor did it temper, affective state.

Additionally, the linguistic content of the verbal descriptions of the IAPS pictures was explored with LIWC (Pennebaker et al., 2001). For positive and negative pictures, we observed a gradual increase in affective word use over time. Specifically, positive pictures were described with more positive emotion words, and negative pictures were described with more negative emotion words. No effects were observed for the neutral pictures. A large variation between pictures was observed for the number of words speakers used to describe the pictures. In general, speakers used more words to describe positive pictures than neutral pictures, but not negative pictures.

2.6.1. Inducing positive and negative affect

Verbally describing and passively viewing affective pictures successfully induced negative affective states (in the negative condition), but not positive affective states (in the positive condition). Our findings did not support the hypothesis that verbally describing affective pictures would induce stronger affective states than passively viewing them. The finding that the positive affect induction turned out to be less successful than the negative affect induction is found more often (e.g., Ferrer et al., 2015; Uhrig et al., 2016; Westermann et al., 1996), and other studies using IAPS pictures (e.g., Uhrig et al., 2016) and the Velten method (e.g., Gerrards-Hesse et al., 1994; Göritz & Moser, 2003; 2006; Westermann et al., 1996) have faced this problem as well. Given the fairly positive affective state of the participants before being exposed to the pictures, a possible explanation for this lack of an effect might be that the participants' positive affect was already at ceiling (Uhrig et al., 2016; Westermann et al., 1996).

2.6.2. Affect and language

Participants reported slightly more pleasant affect after describing the positive pictures, compared to passively viewing them. However, this difference was small and not significant. As described above, the literature shows mixed results regarding the effect of affect labelling on affective (and emotional) experience. Putting emotions into words can enhance the affective experience (e.g., Ortner, 2015) or decrease it (e.g., Torre & Lieberman, 2018). However, we realize that our affect induction procedure is somewhat unique, and might not be directly comparable to affect labelling. First of all, in experiments studying affect labelling, participants are asked to describe their own affective state. In our Study 1, individuals were asked to describe the (affective) content of affective pictures, not their own affective state. Second, our participants were asked to view and describe the pictures simultaneously, making it harder to distinguish between the effect of viewing the pictures and verbalizing the content. But given that the effect of verbalizing was small, we doubt whether it would have made a significant difference to first expose individuals to the pictures to our participants, and asking them to describe them only after viewing.

To date, only a limited amount of work has been done on the description of affective content where speakers could use their own words (Castro & James, 2014; Ortner, 2015). Study 1 adds to this relatively new field that combines questions from affective science and psycholinguistics.

2.6.3. Verbal descriptions

Corresponding to the gradual increase in affective content of the positive and negative pictures, we observed a gradual increase in affective language use in the expected directions: over time, positive pictures were described with more affective and positive emotion words; negative pictures were described with more affective and negative emotion

words. The descriptions of neutral pictures were described with few affective, positive- and negative emotion words. Interestingly, individuals viewing and describing positive pictures did not (self)report enhanced positive affect, but they *did* use substantially more positive emotion words in their descriptions, compared to the negative and neutral pictures.

The usage of emotional words could be attributed to and explained by the specific (affective) content of the pictures. Therefore, we also studied a phenomenon that could not be attributed to the emotional content of pictures - the general number of words uttered to describe the pictures. Our results indicated that individuals did not use more words to describe positive than negative pictures. This is in contrast to the literature: happy individuals tend to talk faster (e.g., Laukka et al., 2005, but see also Kamiloğlu et al., 2020) and sad individuals tend to talk slower (e.g., Siegman & Boyle, 1993). Additionally, happy individuals have been found to utter more words spontaneously (Velten, 1968). However, keeping in mind that positive affect induction was not successful, this finding is not unexpected. Another explanation might be that the content of the affective pictures was more complex, compared to the neutral pictures. Indeed, many neutral pictures included depictions of objects, patterns and portraits, whereas the affective pictures often composed scenes of multiple components, e.g., individuals in various situations (e.g., plane crash, cycling), diverse backgrounds (e.g., nature, city, living room).

However, individuals used more words to describe positive pictures than neutral pictures. Assuming that affective pictures, both positive and negative, are more arousing than neutral pictures, this might explain why individuals used more words to describe affective than neutral pictures, because highly aroused speakers compared to lowly aroused speakers tend to have an increased speech rate (Goudbeek & Scherer, 2010) and thus might use more words. We found that the IAPS arousal scores were indeed positively correlated to word use, both for positive (r = .07) and negative (r = .16) pictures. However, we also found correlations between the IAPS valence scores and word use, which were systematically larger than those between arousal and word use, for both the positive (r = .12) as well as negative pictures (r = -.21). Concluding, both affective and arousing content was correlated to the number of words used to describe the pictures.

2.6.4. Strengths and limitations

Our study has a few limitations that need to be acknowledged. First, while we deliberately chose to use an incremental procedure, the incremental nature of the affect induction procedure might pose various issues. Given that order effects are at the base of our study, participants might be influenced to a lesser extent by the pictures, because they were exposed to pictures gradually increasing in affective content, instead of viewing a random selection of affective pictures (e.g., Capecelatro et al., 2013; Dhaka & Kashyap, 2017; Hot & Sequeira, 2013) that might be, on average, more positive or negative. The temporal place of a stimulus in an array of pictures can influence how the stimulus is processed in the viewer, e.g., habituation effects (e.g., Balada et al., 2014) might reduce the effectivity of the stimuli, while recency bias (e.g., Hsiao et al., 2009) might enhance

the effect of the last (few) affective pictures. However, showing affective IAPS pictures in a fixed (or non-incremental) order is not uncommon, and has been shown to effectively induce desired affective states (e.g., Carretero et al., 2020; Limonero et al., 2015; Ortner, 2015). Therefore, while we acknowledge this limitation, we do not think the incremental nature of the stimuli is responsible for the absence of an effect of, for example, the positive condition. Nevertheless, there is certainly a possibility that the last few pictures were the most effective at inducing affect, and the previous pictures' affective impact was limited. For future research, it might be interesting to compare the effectivity of exposure to highly positive or negative rated IAPS pictures, compared to exposure to pictures gradually increasing or decreasing in valence.

Based on the available evidence in the literature, predicting the precise effectivity of the incremental procedure was difficult. Hence, we were ambivalent in our predictions: the gradual increase in valence could result in a weaker effect, a stronger effect, or perhaps even no effect at all. For example, Van der Zwaag et al. (2013) compared the effectivity of gradual versus abrupt change in happy music to sad music. They found that both affect induction procedures were equally effective, lowering both valence and energy (i.e., feeling more tired, according to self-report of the participants).

Gradually increasing affective content of stimuli might have several advantages. First, as Velten argued, the gradual affect induction was favorable, 'to overcome the subjects' presumable reluctance to experience unpleasant mood' (Velten, 1967, p. 68). Indeed, recent research shows that noncompliance with an affect induction procedure is more common viewing negative videos than positive videos (Shevchenko & Bröder, 2019). Given that we started the series of negative pictures with neutral stimuli, this might prevent the initial reluctance of participants to engage in the negative affect induction procedure. Additionally, for some populations, the startle effect might be specifically unethical, because they could cause serious psychological or physiological harm, for example, to individuals with certain mental disorders (e.g., PTSD, panic disorder) or cardiovascular diseases.

Second, verbalizing the content of the pictures adds additional challenges – for example, participants likely vary in their degree of verbal skills and consequently differ in how difficult they considered the task. We did try to take this into account in the selection of our participants by excluding participants with a speech disorder or a limitation in the ability to speak fluently (e.g., stuttering). However, to check whether having Dutch as a first language had an effect on the effectivity of the affect induction of verbally describing the pictures, we repeated our analysis of Study 1, excluding the participants who did not have Dutch as their first language (n = 5), but did not find substantial differences. Additionally, given that we tested a relatively homogenous group of participants (young Dutch students), we expect that individual differences in verbal fluency, attention, and other cognitive and communicative abilities are small and randomly distributed throughout our sample.

Third, an additional benefit of our study is the collection of human, realistic verbal descriptions for the content of the subset of IAPS pictures we used. While these descriptions are not yet validated, it is a valuable first step to the possible creation of a

verbal IAPS, which might be useful in certain specific populations, e.g., visually impaired individuals.

Lastly, we conjectured that asking individuals to use their own words describing the pictures (instead of uttering pre-defined affective sentences), would reduce the awareness of the goal of the procedure (affect induction) and therefore reduce the chance of participants reporting to feel the change in affect they think they 'should' experience (e.g., social desirability or task demands), even when they do not actually experience a shift in affective states (e.g., Kenealy, 1986).

2.6.5. Future research

Affective processes often take place in a social setting, but in laboratory settings, they are generally induced in individual participants (Gilam & Hendler, 2016). Our affect induction method might be a useful, naturalistic method to induce affective states in more than one individual at the same time. For example, participants could take turns in a conversation setting describing out loud the affective IAPS pictures to each other. This might create more naturalistic opportunities in affective research to study affect induction in dyads.

Contrary to the Velten method, our method describing pictures was not self-referential in nature. Recent literature suggests that self-referencing might play a critical role in affective word processing. Soares et al. (2019) found that in a masked priming paradigm, individuals categorize positive adjectives faster when they are primed by self-related primes, compared to other-primes. In light of these findings, the Velten method might be more effective inducing positive affect than our affect induction method. Upon inspecting the verbalizations, indeed, only 26% percent of the descriptions are self-referential (e.g., 'I see…'), and less than 1% is other-referential ('Here you see…'; see Supporting Information, Table S3, https://doi.org/10.1371/journal.pone.0233592). This might be one of the reasons that our pleasant affect induction was not effective. For future research, it might be interesting to compare a condition where participants are instructed to provide a self-referential description ('I see a happy couple') to a condition where participants are instructed to provide a non-self-referential description of the pictures ('This is a picture of a happy couple').

Finally, our affect induction method is not inherently limited to valence, but also could be applied to specific emotion categories that are present in picture datasets (cf., Mikels et al., 2005, for IAPS). By replacing the current pictures with pictures that induce a specific emotion (e.g., disgust, tenderness and anger), we think our method might be able to successfully induce specific emotions and their accompanying verbal descriptions.

2.6.6. Implications

This study contributes to the sparse literature on verbalizing affective content, implying that an engaging task as verbalizing negative content, using free expression, can be an effective method to induce negative affect in a possibly more ecologically sound manner

(e.g., viewing affective videos). The results indicated that verbalizing or passively viewing affective content are equally effective methods to induce negative affective state.

We contributed to the scientific literature on the relationship between affect and language, aiming to gain understanding of the critical, but unclear relationship between language production and (un)pleasant affect.

APPENDIX

Figure 1A

Valence (y-axis) of IAPS pictures by bin (x-axis)

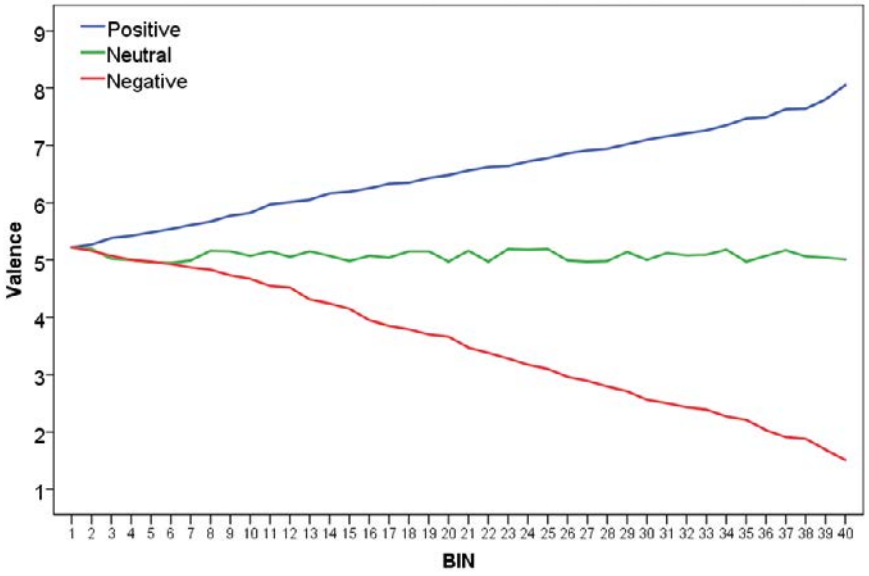
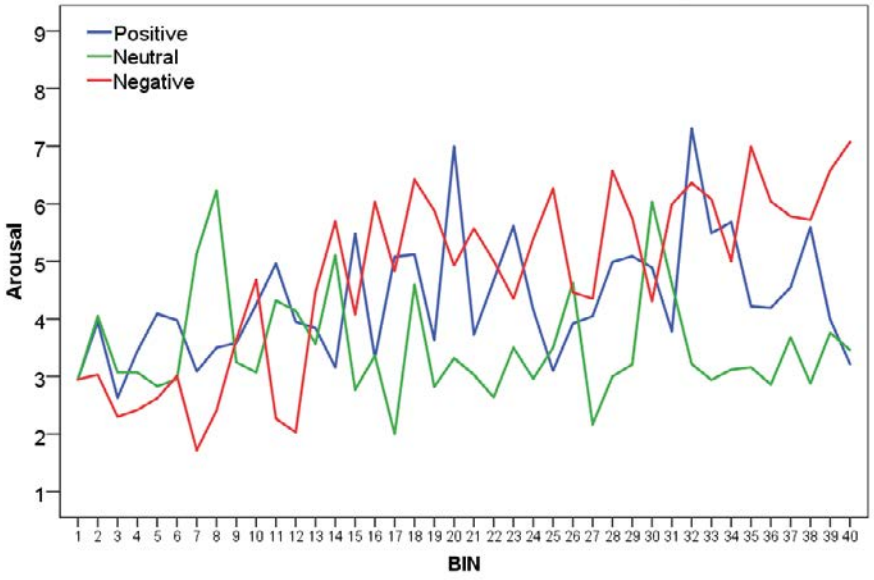


Figure 2A

Arousal (y-axis) of IAPS pictures by bin (x-axis)



3

Do speaker's emotions influence their language production? Studying the influence of disgust and amusement on alignment in interactive reference

This chapter is based on:
Out, C., Goudbeek, M., & Krahmer, E. (2020). Do Speaker's emotions influence their language production? Studying the influence of disgust and amusement on alignment in interactive reference. ***Language Sciences***, ***78***, Article 101255. https://doi.org/10.1016/j.langsci.2019.101255

ABSTRACT

The influence of affect on the early stages of spoken language production such as content selection has received little scholarly attention. During content selection in conversation, speakers often take the utterances of their conversation partners into account. For example, while speakers generally prefer to use color in their descriptions, they start to use dispreferred attributes such as orientation and size more when they are primed by a prerecorded partner using these dispreferred attributes (Goudbeek & Krahmer, 2012). The current study assessed the role of amusement and disgust in this process of conceptual alignment, while simultaneously replicating this earlier finding in a more realistic setting. Three types of alignment were analyzed: alignment of dispreferred properties (with or without additional properties), alignment of overspecified descriptions (both used by G&K), and alignment of dispreferred properties only. The results generalize the findings by Goudbeek and Krahmer (2012) to a more naturalistic conversation setting: partners indeed align with each other's attributes in the choice of their referring expressions. The effects of emotion were generally limited, but disgusted speakers do tend to align more to the dispreferred attributes (e.g., size) used by their conversation partner than amused speakers. Our findings highlight the robustness of alignment in referring expressions produced in interactive settings, and suggest that emotional state can have an impact on this process.

3.1. Introduction

The effects of affective state on voice characteristics are well established (e.g., Bachorowski & Owren, 1995; Gangamohan et al., 2016; Lee et al., 2005; Scherer, 2003; Sobin & Alpert, 1999). For example, an angry speaker is likely to talk with a higher intensity and pitch, while a sad emotional state may cause her to speak softer and lower (e.g., Bachorowski, 1999; Goudbeek & Scherer, 2010; Scherer, 2003). However, the role of affective state on other building blocks of the spoken language production process, such as conceptualizing and formulating a message (e.g., Levelt, 1989; Levelt et al., 1999) has received far less attention. Given the extensive body of literature showing effects of emotional state on a range of cognitive processes that are relevant for spoken language production, including processing style (Beukeboom & Semin, 2006), attention (Charash et al., 2006) and perspective taking (Converse et al., 2008), we conjecture that the influence of affective state on spoken language production extends beyond the established effects on articulation. Specifically, in this study, we explore whether two distinct emotions (disgust and amusement) have different effects on how speakers adapt to each other during the production of referring expressions in interactions. We do so in a newly developed interactive reference paradigm, studying reference production in a natural albeit controlled way, with speakers in which different emotions were induced. Our findings shed light on both affective language production and on reference production in interactions.

3.1.1. Spoken language production and emotion

Speaking is a highly complex cognitive activity (Goldrick et al., 2014). Speakers routinely produce 10-15 phonemes per second, the results of a spoken language production process that starts with the conceptual preparation of a message and ends in articulation. It is generally assumed that this process involves different, consecutive stages (Dell, 1986; Garrett, 1975; Griffin & Ferreira, 2006; Levelt, 1989; Levelt et al., 1999; Van Gompel et al., 2019; Vigliocco & Hartsuiker, 2002). A speaker first has to conceptualize their message, that is 'decide what to say', which results in a so-called preverbal message. A preverbal message is a conceptual representation whose expression in words and phrases will be a realization of the speaker's original intention (Levelt, 1989, p. 27). In the next stage, 'deciding how to say it', the preverbal message is converted into a linguistic representation known as the utterance plan, which involves among other things planning of the structure of the utterance and lexical access to retrieve the relevant words from memory. Finally, the resulting utterance plan is phonologically encoded and articulated, resulting in audible speech. Much of this process is unconscious: we are generally not consciously aware of selecting an attribute (whether we refer to a chair as 'large' or 'blue'), nor of retrieving a lemma, or producing a specific speech sound. Indeed, these stages of the spoken language production system have traditionally been understood as modules that receive and process input in an automatic and encapsulated way (following, for example, Fodor, 1985). However, this does not mean that external processes cannot influence them. It is well established,

for example, that the different stages can be influenced by perspective taking (Levelt, 1999) and conversation factors (Pickering & Garrod, 2013). We can consciously influence this spoken language production process, for example, by deciding to produce a particular message (say, when we want to specifically point out the size of a chair), by articulating extra loudly because of surrounding noise, and by taking into account the words and phrases used by our conversation partners when we produce our own. In general, there is an emerging consensus that the stages of the spoken language production process are cognitively permeable, like many other cognitive processes (for similar arguments, see Van Berkum et al., 2013, for language comprehension, and Vigliocco & Hartsuiker, 2002, for language production).

In this study, we concentrate on a specific external influence on the spoken language production process: the emotional state of the speaker. That affective state has an influence on spoken language production is arguably neither unexpected nor unexplored. As we have discussed above, a speaker's affective state can influence their voice characteristics, and along similar lines, there is some suggestive evidence that the affective state can also impact the words speakers use.

Forgas (1999a), for example, showed that sad speakers tend to use more polite wording than happy speakers in their requests in socially difficult situations (e.g., discussing a sensitive topic). In a somewhat similar vein, studies of the (written) word use of suicidal poets found that they relied more on first person singular pronouns, used more words about death, and made fewer references to other people in their poems, compared to non-suicidal poets (Pająk & Trzebiński, 2014; Stirman & Pennebaker, 2001). Beukeboom and Semin (2006) constitute another example; based on the idea that positive affect activates a global, and negative affect a more analytical, detail-oriented processing style (Forgas, 2012; Schwarz, 1990; Schwarz & Clore, 1988), they show that speakers in a positive mood use more abstract words (e.g., adjectives, such as 'aggressive') to describe events compared to speakers in a negative mood, who tend to use more concrete words (e.g., descriptive action-verbs, such as 'punch'). Finally, Kempe et al. (2013) looked at the effect of happiness and sadness on lexical and syntactic ambiguity in referring expressions. They found that happy speakers (compared to speakers in a neutral state) were more likely to use ambiguous bare homophones: they said 'bat', which could either refer to the mammal or a baseball bat, instead of specifying the noun with an adjective. This finding is in line with both Beukeboom and Semin (2006) and the 'affect as information theory' (Forgas, 2012; Schwarz, 1990; Schwarz & Clore, 1988), where a positive affective state leads to a more global focus and, consequently, more abstract and ambiguous language.

Our working hypothesis is that affective state can influence all stages of the spoken language production, both for deciding what to say, and deciding how to say it. In this chapter, we zoom in on one aspect of content selection not studied before: whether emotional state influences the way speakers conceptualize their descriptions — referring to an object as, say, 'the large chair' or 'the red chair' — and the extent to which they adapt to earlier references of their conversation partner when doing so.

3.1.2. Adaptation in spoken language production

In a conversation, speakers continuously adapt to each other (e.g., De Looze et al., 2014). Some argue that this adaptation is a largely automatic process (Garrod & Pickering, 2004; Pickering & Garrod, 2004), while others claim it is a more effortful, conscious process (Brennan & Clark, 1996; Wilkes-Gibbs & Clark, 1992). Garrod and Pickering (2004) argue that one way in which conversation partners can adapt to each other is via alignment: where conversation partners start to use similar representations, at representational levels ranging from sounds and words to syntactic and meaning representations. For example, when one speaker is talking in fast high-pitched voice, the other speaker is more likely to respond in kind; and when one has a preference for a particular syntactic structure, the other is increasingly likely to use this structure as well. Garrod and Pickering argue that this alignment process occurs at all levels of spoken language production, ranging from the semantic representations that drive content selection to the phonetic representations involved in articulation, and that this process helps in making conversations proceed more smoothly.

Now, when during an interaction a speaker has to produce a description to refer to some object, she can be influenced by the earlier referring expressions produced by her conversation partner. For example, if the conversation partner just described an object using size ('the large table'), she might be more likely to use size in a subsequent description as well ('the small chair'). Interestingly, using size may be at odds with another tendency that speakers have, as reported in the literature: the inherent preference of certain attributes over others (e.g., Dale & Reiter, 1995; Viethen & Dale, 2010, among many others). In particular, using color ('the red chair') is typically found to be preferred over size ('the small chair'), when both options are equally successful as descriptions (that is: when both succeed in uniquely identifying the intended referent; Pechmann, 1989). These findings have been used to develop computational models of reference production, such as Dale and Reiter's (1995) Incremental Algorithm, a computational model that uses a preference order of attributes to automatically produce human-like referential expressions. The Incremental Algorithm first uses the most preferred attribute, e.g., color, and, only when this does not identify the target object (there are multiple red chairs), it will proceed to use less preferred attributes, such as size or orientation (and produce, for example, 'the small red chair'). The algorithm adds less and less preferred attributes until the target can be properly identified. While not initially meant as an accurate and realistic model of human spoken language production (Van Deemter et al., 2012), it has substantially influenced research into the psycholinguistics of content selection (e.g., Arts et al., 2011; Frank & Goodman, 2012; Gatt et al., 2017).

However, the Incremental Algorithm might not be a good model of how speakers produce descriptions in an interactive setting, as was explored by Goudbeek and Krahmer (2012). They investigated the effect of previously mentioned (dispreferred) attributes on referential choice. In their study, participants listened to a pre-recorded voice describing one of three furniture items, using either a preferred (color) or dispreferred (orientation)

attribute. After the participants indicated which image matched the description, they then described a new target object themselves, again in the context of two other objects. The results showed that, indeed, participants preferred to use color over orientation in their descriptions. Importantly, they tended to use the same attribute they were primed with: participants primed with color used color in more than 75% of the trials, and, crucially, participants primed with orientation used orientation in more than 50% of the trials (as opposed to about 20% when primed with color). Finally, overspecified descriptions, which exposed participants to both the preferred and dispreferred attribute (i.e., 'the green left-facing chair'), also resulted in a high level of alignment (52%). The authors concluded that content selection is – at least in part – driven by alignment in conversation.

However, Goudbeek and Krahmer (2012) study alignment under artificial circumstances. As Brennan and Hanna (2009) argue, a set-up like Goudbeek and Krahmer (2012) differs markedly from spontaneous conversation: speakers did not interact with a real conversation partner, but heard a pre-recorded voice, making alignment of representations unidirectional and, arguably, irrelevant. Goudbeek and Krahmer opted for this design to be able to control what the speakers' conversation partners say. Nevertheless, in this chapter, we show that it is possible to investigate this kind of alignment in a natural setting, with participants interacting in a spontaneous yet controlled way. In addition, we study whether the emotional state of conversation partners influences their tendency to adapt to one another during referential communication. Our expectation is that it does, given that affective state has been shown to influence both people's egocentricity and their attentional focus, two aspects that are relevant for interactive referential communication, as we discuss next.

3.1.3. The influence of affective state on egocentricity and attention

To communicate effectively, conversation partners need to keep track of which information is shared (common ground) and which information is not. Previous research has shown that individuals adjust their referring expressions depending on which information is shared (Achim et al., 2017; Clark, 1996; Clark & Wilkes-Gibbs, 1986; Krauss & Fussell, 1991). These studies suggest that speakers spend considerable effort to establish common ways to refer to objects and that they do so efficiently (Brennan & Clark, 1996; Gelati & Brennan, 2010), suggesting that speakers mentally 'model' their addressee's knowledge state.

However, other studies have argued that there is an egocentric bias in spoken language production, suggesting that speakers have a tendency to focus on themselves, using their own perspective as reference point to the world (e.g., Ross & Sicoly, 1979). Some authors have argued that while speakers are able to adjust to their listener's perspective, they initially act from an egocentric perspective (Epley et al., 2004; Horton & Keysar, 1996), although others beg to differ (Bezuidenhout, 2013; Brennan & Hanna, 2009). Importantly, affective state appears to influence the amount of egocentric perspective taking, in the sense that speakers who are in a positive mood appear to be more egocentric, while speakers in a negative mood to be less egocentric (Converse et al., 2008; Kempe et al.,

2013). Clore and Hutsinger (2007), for example, argue that taking the perspective of the listener into account is more difficult for speakers that are in a positive state, because positive emotions promote more automatic responses. Conceivably, this is because people subconsciously want to extend their positive state as much as possible, and are hence less inclined to invest in activities that might negatively influence their state (like investing effort). Kempe et al. (2013) similarly suggest that the egocentric perspective is reinforced when speakers are in a positive mood: by not focusing on the addressee's perspective, less effort is spent on monitoring. This is also in line with the results of Converse et al. (2008), where individuals in a false belief task were less likely to adapt to the perspective of another person when they were in a positive mood. These results also appear consistent with studies of the influence of affective state on attention, which indicate that positive affect broadens attention (e.g., Fredrickson, 2001; Fredrickson & Branigan, 2005; Rowe et al., 2007), whereas negative affect narrows attention (e.g., Gasper & Clore, 2002). In a conversational setting, the results of these studies might translate into negative speakers being more narrowly focused on their conversation partner than positive speakers.

3.1.4. Amusement and disgust and their influence on egocentricity and attention

Additionally, recent studies find support that not only positive and negative valence, but also certain distinct emotions have an effect on egocentricity in perspective taking (e.g., disgust, Todd et al., 2015; anger, Yip & Schweitzer, 2019; guilt and shame, Yang et al., 2010). For example, Todd et al. (2015) found that disgusted, but not anxious, individuals more easily took the perspective of other individuals, in a conceptual task in which they identified the location of a green light on a screen, from their own perspective or from one of two hypothetical other individuals that appeared on a screen. Additionally, in the other-perspective trials, the more intense the individual experienced disgust, the easier it was to adapt to the perspective of the other. Other studies have found similar results: disgusted individuals, compared to individuals in a neutral condition, were found to have a narrower attentional scope (Gable & Harmon-Jones, 2010) and perform better in an emotional perspective taking task (Binyamin-Suissa et al., 2019).

In a similar vein, amused individuals have been found to have a broadened attentional scope (Fredrickson & Branigan, 2005). In a broader sense, support has been found that positive emotions that do not motivate the individual to achieve a goal, e.g., amusement while watching a funny video, broaden the attentional scope (Gable & Harmon-Jones, 2008). While on the contrary, positive emotions that are motivated by goal-achievement (e.g., desire to eat desert, after viewing images of pudding) narrow the attentional scope to shut out less important information to focus on achieving the goal (e.g., eat pudding; Gable & Harmon-Jones, 2008; for an overview, see Gable & Harmon-Jones, 2016).

3.1.5. The present study

In the current study, we induced amusement or disgust in the participants. Amusement typically occurs when a person experiences something entertaining (e.g., a joke) and feels pleasure (Tong, 2015). Disgust is elicited when a person is confronted with something repulsive, for instance certain bodily fluids (vomit, pus, urine) or (the smell of) a rotting corpse (Darwin, 1872; Seidel et al., 2010). These differences discussed above make amusement and disgust highly suitable for our purposes: to study the effect of two different emotions on conceptual alignment in a conversation.

We hypothesize that disgusted individuals will align more with their conversation partner than amused individuals. People who are in a disgusted state have been found to be less egocentric in perspective taking, compared to other negative emotions (e.g., Todd et al., 2015), and disgust has been found to narrow the attentional scope (e.g., Gable & Harmon-Jones, 2010), suggesting that disgusted people are more focused on their conversation partner and hence more inclined to align. Conversely, amused individuals have been found to have a broadened attentional scope (Fredrickson & Branigan, 2005; Gable & Harmon-Jones, 2008), which might make their attention less focused on their conversation partner. Additionally, there are indications that individuals in a positive (compared to negative) mood display more egocentric behavior and perspective taking (Clore & Hutsinger, 2007; Converse et al., 2008; Kempe et al., 2013), contributing to our hypothesis that amused individuals, compared to disgusted individuals, will align less with their conversation partner.

3.1.6. Goals of the study and hypotheses

The goal of this study is twofold. First, we replicate Goudbeek and Krahmer (2012) in a more ecologically valid set-up, investigating alignment of dispreferred attributes in reference in a more naturalistic paradigm. Second, we investigate the effect of emotion on this alignment, inducing either amusement or disgust in our participants.
Our hypotheses are the following:

(H1) We hypothesize (following Goudbeek & Krahmer, 2012) that participants will indeed align with the dispreferred attributes used by their conversation partners, given that Goudbeek and Krahmer (2012) found that individuals tended to use the same attribute they were primed with.

(H2) In addition, we hypothesize that the level of alignment will depend on the emotional state of the speaker, in particular we hypothesize that disgusted speakers will align more with their conversation partners than amused speakers will. The rationale for this hypothesis has been discussed above.

3.2. Method

3.2.1. Participants

A total of 140 students (36 males) from a Dutch university, participated in pairs (forming dyads) in the experiment, either for course credit (n = 112) or in exchange for €5,- (n = 28). The mean age of the participants was 22.25 years (SD_{age} = 3.15 years, range 17-40 years). Students were randomly assigned to a conversation partner in the same condition, forming a cross-gender dyad (40 dyads), or a dyad consisting of two women (84 dyads) or two men (16 dyads). Exclusion criteria included color-blindness and having a speech disorder. No participants were excluded based on these criteria. The study was conducted in Dutch. Ethical approval for this experiment was given by the local ethics committee.

3.2.2. Materials

3.2.2.1. Emotion induction videos

Participants viewed amusement- or disgust-inducing videos. The choice of the four videos was based on existing literature (Harlé & Sanfey, 2010; Nummenmaa et al., 2012; Rottenberg et al., 2007; Schaeffer et al., 2010). The amusement inducing fragments came from 'When Harry met Sally' (1989; woman simulates having an orgasm in a restaurant) and 'There's Something About Mary' (1998; woman takes sperm from a man's ear, mistaking it for hair gel). The disgust inducing fragments came from 'Trainspotting' (1996; man dives into a dirty toilet) and 'Pink Flamingos' (1972; drag queen eats dog feces).

3.2.2.2. Emotion scales

As a manipulation check, participants rated on a 1 (not at all) to 7 (extremely) Likert scale how much they experienced amusement, pride, anger, sadness, disgust, surprise, and fear after they had viewed the video. Although we were only interested in the scores for amusement and disgust, using just two answer options could have resulted in demand characteristics. For example, participants would indicate being amused and not disgusted by viewing, for example, 'When Harry met Sally'. Hence, five other emotion scales were included as distractors, to obscure which emotions we aimed to induce.

3.2.2.3. Conversation about video

While affect induction with video clips is considered highly effective, it is also relatively transient (Bohn-Gettler & Rapp, 2014). To further enhance affect induction and transfer to the conversation setting, we asked participants to describe the video they had seen to each other. Before the conversation took place, the experiment leader asked the pair to discuss topics such as the details of the video clip, what they think the video is about, if they had seen the video before, et cetera. The last suggestion was always to discuss what they felt while viewing the video clip. This way, participants were guided to relive the emotions they experienced when viewing the video clip, and the ensuing emotional contagion (Hatfield

et al., 1993) might bolster the emotion induction. The duration of the conversation was determined by the participants themselves, and they were asked to knock on the door when finished. Usually, the conversation lasted approximately 3 minutes.
After instructions were given, the experiment leader made sure the conversation was being audio recorded, and left the room.

3.2.2.4. Stimuli

Following Goudbeek and Krahmer (2012), we used pictures from the TUNA corpus (Gatt et al., 2007) that depicted front-facing furniture items (a fan, a chair, a couch, a desk) in four different colors (blue, green, red, grey) and two sizes (large, small). Previous studies (Gatt et al., 2007; Koolen et al., 2011; 2013) revealed that participants preferred to use color when they described these pictures, in line with many other studies using different pictures (e.g., Pechmann, 1989; Sedivy, 2003). As indicated by the literature, the number of different values for the attributes has no noticeable impact on this preference, as long as values do not become too similar (e.g., light blue vs. dark blue; Viethen et al., 2017).

There were three types of trials: color, size, and filler trials. Each dyad described 60 trials in two blocks: one block consisting of 20 color trials and 10 filler trials, and one block consisting of 20 size trials and 10 filler trials. To control for order effects of the trials, four different versions were created: version 1 and version 2 (first size block, then color block), and version 3 and version 4 (first color block, then size block). Per version, each block contained a different order of trials. Trials consisted of four turns, as described below. In order to distribute the roles of primer and primee evenly across the 4 different groups, based on viewed video, block randomization was used.

3.2.2.5. Director-matcher task

Figure 1a, 1b and 1c provide an example of a size trial, a color trial and a filler trial, respectively. First, the primer was the director, describing the target picture (the picture in the middle, framed by a red border) to the primee, the matcher. Depending on the trial, the primer could use either a preferred (color) or dispreferred (size) attribute to describe the target picture. In the color trials, the primer needed to use color to distinguish the target picture from the distractors. For example, when the target picture was a large blue fan, and the distractors were a large red couch and a large red fan, she needed to use color to describe the target picture (because they all had the same size). Similarly, in size trials, only size distinguished the target picture from the distractors. For example, when the target was a large green desk, and the distractors were a small green desk and a small green fan, she had no choice but to use size to describe the target picture (see Figure 1a, panel 1).

Second, the primee, the matcher, saw the same pictures on their screen (in a different order than the primer, to prevent the use of location) with the numbers 1, 2 and 3 (from left to right) underneath them (see Figure 1a, panel 2). After listening to the description of the primer, they matched the picture by pressing the key of the corresponding number on their keyboard. After the primee's answer, the primer pressed 'Enter' and both participants continued to a new screen.

Third, the participants switched roles: the primee became the director and the primer the matcher. In contrast to the previous turns, the target picture could now be distinguished by either the preferred or dispreferred attribute. For example, the target picture was a large red couch and the distractors were a small grey chair and a small blue desk (see Figure 1a, panel 3). The primee could either use the preferred attribute ('the red couch') or use the dispreferred attribute ('the large couch') to distinguish the target picture from the distractors. In case of alignment, they would use the same attribute as the primer. However, since color is already preferred, its use does not necessarily entail alignment. Following the definition of alignment in Goudbeek and Krahmer (2012), alignment occurs only when participants have been primed with size, and use size (with or without the additional use of color). This is a strict interpretation of alignment, where using color after having been primed with color is not considered alignment, because color could have been used because it is a preferred property.
Fourth, the primer, now the matcher, selected the matching picture by pressing the key of the corresponding number on the keyboard (see Figure 1a, panel 4), after which the primee marked the end of the trial by pressing 'Enter' and a new trial would start.

Using this set-up, speakers could be primed with a dispreferred property (size) in a natural, spontaneous way, without using pre-recorded speech primes. Notice that participants were not explicitly instructed how to speak; interaction between primer and primee was free and unconstrained; only the selection of stimuli made the use of certain attributes (much) more likely than others.

3.2.2.6. Additional materials

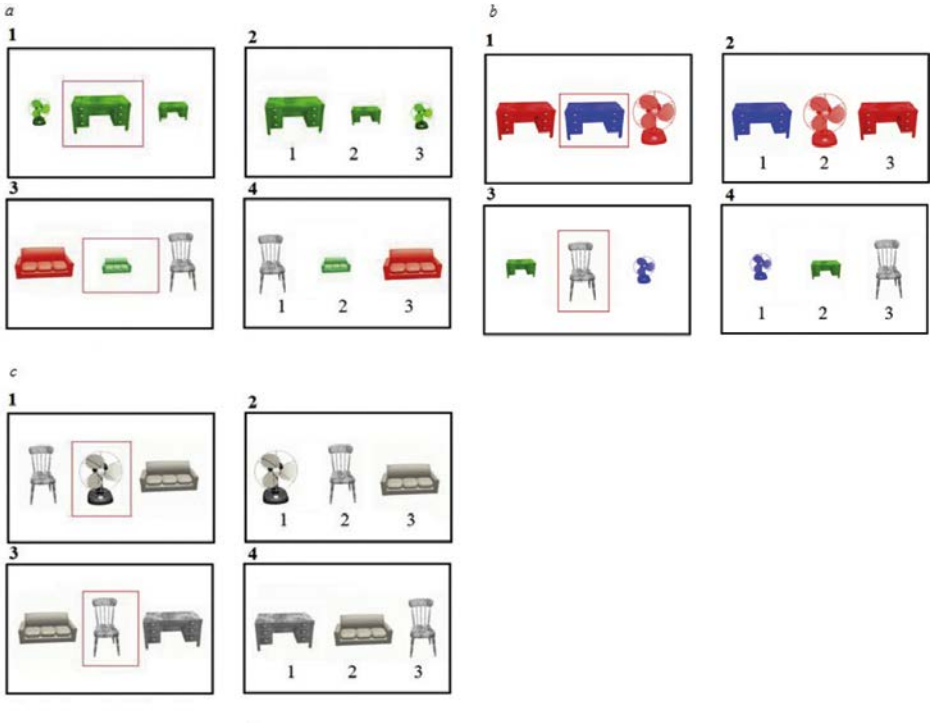
Participants rated their current mood by indicating on a 1 (very negative) to 7 (very positive) scale how much they experienced the following affective states (Dutch translations between brackets) at the time: happy/sad (gelukkig/ongelukkig), comfortable/uncomfortable (aangenaam/onaangenaam), satisfied/unsatisfied (voldaan/onvoldaan), content/discontent (tevreden/ontevreden), cheerful/sullen (vrolijk/verdrietig), in high spirits/low-spirited (opgewekt/teneergeslagen; Krahmer et al., 2004, based on Bohner et al., 1992, and Mackie & Worth, 1989; Dutch translations of English originals). Scales were filled in twice: before the emotion induction video was shown (see *3.2.2.2. Emotion scales*) and after the director-matcher task was completed (see *3.2.2.5. Director-matcher task*). Mean items scores before the mood induction ranged from M = 4.84, SD = 1.40 (comfortable/uncomfortable) to M = 5.49, SD = 1.07 (happy/sad), and after the director-matcher task from M = 5.06, SD = 1.44 (comfortable/uncomfortable) to M = 5.61, SD = 0.99 (content/discontent), indicating that participants generally felt (slightly) positive. The mean differences between the two tests indicate that while participants generally felt more positive after the director-matcher task, the differences were small, M = 0.23, range 0.02-0.32. Given that we did not have hypotheses regarding these scales, and further elaboration would not benefit the reader, we will not elaborate further on this result.

3.2.3. Procedure

Pairs of participants were welcomed in the laboratory, after which each participant was escorted to their own cubicle to perform the following tasks alone. They read and signed an informed consent form, indicating that they were going to participate in an experiment studying the effect of emotional video fragments on memory and cognitive abilities. Participants filled in the mood questionnaire and viewed an amusement- or disgust-inducing video clip. After viewing the video, they filled in the emotion scale and were asked to leave their cubicle. The dyad was escorted to an empty office. They were seated in front of each other with a desk between them, each behind a computer screen with a keyboard. They were instructed to engage in an unconstrained conversation, talking about what they had seen in the video. The experiment leader left the room, to return when they heard a knock, and give instructions to the dyad for the director-matcher task. The experiment leader stayed for the first two practice trials, to check if the participants understood how the task worked. When they were finished, they filled in the mood questionnaire, were debriefed individually on paper and thanked for their time.

Figure 1

a. Example of size trial; b. Example of color trial; c. Example of filler trial



3.3. Results

3.3.1. Manipulation check

Before inspecting the attribute use of participants in the amusement and disgust condition separately, we tested whether the emotion manipulation was effective. We performed a one-way MANOVA with emotion induction videos (amusement videos vs disgust videos) as independent variable and emotion scales (amusement vs. disgust) as dependent variable. As expected, we found a significant effect of emotion scales for the amusement videos, $F(1, 138) = 88.89$, $p < .001$, and disgust videos, $F(1, 138) = 255.47$, $p < .001$. The mean scores of the combined videos per emotion indicate that participants who viewed an amusing video reported higher levels of amusement ($M = 4.89$, $SD = 1.38$) than disgust ($M = 2.60$, $SD = 1.49$). Participants who viewed a disgusting video reported a higher level of disgust ($M = 6.26$, $SD = 1.38$) than amusement ($M = 2.51$, $SD = 1.39$). This indicates that the emotion manipulation had the desired effect. The participant's ratings on all items on the emotion scale can be found in Table A1 (Appendix).

3.3.2. Alignment

3.3.2.1. Data pre-processing

Participants in total produced 4200 descriptions — 70 dyads times 60 trials (20 color trials, 20 size trials and 20 fillers). Picture descriptions of both the primer and primee were transcribed and marked for whether they mentioned color, size, or both. Descriptions which contained both color and size were additionally classified as overspecified. Picture descriptions of the fillers were removed from the analyses, resulting in 2800 critical trials. Seventy-seven trials were excluded because of programming errors or participants not responding. Our final dataset contained 2723 trials (more than 97% of the original set).

3.3.2.2. Alignment

As mentioned, Goudbeek and Krahmer (2012) studied multiple types of alignment: alignment with dispreferred attributes (Experiment I) and alignment with overspecified primes (Experiment III). Contrary to Goudbeek and Krahmer (2012), primes in the current study were not fixed. While the stimulus characteristics clearly pointed towards the use of color or size, primers were allowed to describe the target picture with either one attribute (e.g., 'the large chair') or two attributes ('the large green chair'). This resulted in three types of produced primes: color-only, size-only and color-and-size (overspecification).

Taking this into account, in our replication attempt of Experiment I and Experiment III by Goudbeek and Krahmer (2012), we studied two situations:

1) *Alignment of dispreferred properties.* The proportion of descriptions that contain size (including size, and color-and-size) for participants that are primed with size compared to those that are primed with color.
2) *Alignment of overspecified descriptions.* The proportion of overspecified descriptions for speakers that were primed with overspecified primes compared to those that were not.

Additionally, we studied a stricter form of alignment, where overspecification (using color-and-size when primed with size) is excluded and only descriptions that contain size are considered alignment, for the following reason. Earlier studies comparable to ours, provided strong evidence that individual speakers prefer to use color over other properties in their referential expressions (see e.g., Goudbeek & Krahmer, 2012; Pechmann, 1989; Sedivy, 2003; Viethen et al., 2017), even when it is redundant (Koolen et al., 2013), resulting in overspecification (using color in addition to the defining attribute). Therefore, we reckoned that studying only the two situations described above, might give a limited, and possibly distorted representation of our results. By adding the analysis of 'alignment of dispreferred property only' (see below), we assess whether alignment with the dispreferred property was due to primees using the dispreferred property only, or due to adding color to the description.

3) *Alignment of dispreferred property only.* The proportion of size only descriptions for participants that are primed with size compared to those that are primed with color.

3.3.2.3. Design

The study had a 2 x 2 x 2 design. The dependent variable was Alignment (conceptualized and described above) with two values: no alignment (0) and alignment (1). Independent (fixed) variables were Emotion (amusement or disgust), type of Prime (color-only and size-only) and Overspecification (overspecified prime or not). The latter independent variable was created to be able to differentiate trials in which the primer used minimal specification (e.g., 'the small chair' when size is sufficient to identify the target) from those in which the primer used both attributes to describe the object, hence therefore overspecified (e.g., 'the small green chair' when size would be sufficient to identify the target).

3.3.2.4. Descriptives

Before analyzing our data, we inspected the raw numbers of observations of attribute use by primee, sorted by emotion and prime. The detailed statistical analysis is presented after this subsection (*3.3.2.5. Model testing*).

Table 1 displays which properties primees use to describe the target picture (Figure 1, panel 3). In both the amusement as disgust condition, the primee described the target pictures mostly with color-only descriptions (amusement: n = 804 descriptions; disgust: n = 739 descriptions), followed by color-and-size descriptions (amusement: n = 327 descriptions; disgust: n = 342 descriptions) and size-only descriptions (amusement: n = 179 descriptions; disgust: n = 335 descriptions). Table 1 shows that, for the size-only trials, size-only descriptions are used substantially more in the disgust condition than the amusement condition. Out of the 459 size primes in the disgust condition, 217 descriptions (47%) were size-only. In the amusement condition, however, size-only trials were described with size-only descriptions less frequently. Out of the 367 size primes in the amusement condition, only 110 descriptions (30%) were size-only.

In a similar fashion, in color-only trials, primees use size more in the disgust condition (112 descriptions out of 676; 17%) than the amusement condition (60 descriptions out of 645; 9%). In contrast, the effects of emotion on overspecification use seem to be small. Based on Table 1, and as described above, we expect an interaction between emotion and prime for size use.

3.3.2.5. Model testing

Given that we collected multiple trials for each participant, we used logit mixed modeling (Jaeger, 2008) to examine our data, enabling us to take random effects of participants and items (in our case, trials) into account when analyzing a repeated measures data set (Baayen et al., 2008). Following Barr et al. (2013), all our models start maximal with random slopes and intercepts for participants and items. When they failed to converge, we excluded random slopes (first for item, then for participant) to simplify our random effects structure, as suggested by Barr et al. (2013). Random intercepts for participant and item were kept in

all cases. We further addressed convergence failures by: a) centering the data, b) using a different optimizer (bobyqua) and c) increasing the number of iterations to infinite ('2e4'). Table 2 contains the details of the models discussed below.

Table 1

Primee attribute use by Prime and Emotion

Prime	**Emotion**	**Primee attribute use**			
		Color	Size	Color + Size	Totals
Color	Amusement	486	60	99	645
	Disgust	454	112	110	676
Size	Amusement	179	110	78	367
	Disgust	172	217	70	459
Color + Size	Amusement	139	6	150	295
	Disgust	113	6	162	281
Totals		1543	511	669	2723

Note. Amusement (***n*** = 1307), Disgust (***n*** = 1416).

Table 2

Parameter estimates of the optimal models to estimate alignment of Model 1, Model 2, and Model 3

	Fixed effects				RE participants		RE items	
Predictor (centralized)	*β*	*SE*	*Wald Z*	*p* ≤	s^2	*SD*	s^2	*SD*
Model I: Alignment of dispreferred properties, *N* = 2147, loglikelihood = -887.2								
Intercept	-0.63	0.37	-1.72	.086	5.72	2.39	1.80	1.34
Emotion	0.94	0.60	1.56	.120				
Prime	2.24	0.53	4.23	.001	4.41	2.10		
Emotion * Prime	0.54	0.62	0.86	.389				
Model II: Alignment of overspecified descriptions, *N* = 2723, loglikelihood = -1011.1								
Intercept	-2.08	0.30	-6.99	.001	4.03	2.01	0.76	0.87
Emotion	-0.25	0.51	-0.49	.626				
Overspecification	2.11	0.41	5.15	.001	2.54	1.59	0.99	1.00
Emotion * Overspecification	-0.03	0.51	-0.05	.962				
Model III: Alignment of dispreferred properties only, *N* = 2147, loglikelihood = -647.2								
Intercept	-3.18	0.48	-6.67	.001	8.32	2.89	1.64	1.28
Emotion	1.54	0.78	1.98	.048				
Prime	2.60	0.63	4.13	.001	3.16	1.78		
Emotion * Prime	0.58	0.67	0.86	.391				

Note. RE = random effect.

3.3.2.5.1. Model I: Alignment of dispreferred properties

Type of prime significantly predicted use of size, β = 2.24, *SE* = 0.53, *z* = 4.23, *p* < .001; when speakers get primed by size (as opposed to color), they were more likely to mention size (whether together with color or alone). No main effect was found for Emotion, β = 0.94, *SE* = 0.60, *z* = 1.56, *p* = .12, nor was the interaction between Emotion and Prime significant, β = 0.54, *SE* = 0.62, *z* = 0.86, *p* = .389.

3.3.2.5.2. Model II: Alignment of overspecified descriptions

Overspecified primes resulted in significantly more overspecified descriptions than primes consisting of only color or only size, β = 2.11, *SE* = 0.41, *z* = 5.15, *p* < .001. No main effect of Emotion, β = -0.25, *SE* = 0.51, *z* = -0.49, *p* = .626, or interaction between Emotion and Overspecification was found, β = -0.03, *SE* = 0.51, *z* = -0.05, *p* = .962.

3.3.2.5.3. Model III: Alignment of dispreferred property only

Type of prime significantly predicted use of size only, β = 2.60, *SE* = 0.63, *z* = 4.13, *p* < .001, indicating that when speakers get primed by size (as opposed to color), they were

more likely to mention only size (without mentioning color). A main effect for Emotion was found, β = 1.54, *SE* = 0.78, *z* = 1.98, *p* = .048, indicating that disgusted speakers, compared to amused speakers, were (slightly) more likely to use only size when they were primed with size only. No interaction between Emotion and Prime was found, β = 0.58, *SE* = 0.67, *z* = 0.86, *p* = .391.

3.4. General discussion and conclusion

This study had two goals. First, we wanted to replicate the study by Goudbeek and Krahmer (2012) in a more naturalistic (but still controlled) version, by letting two individuals interact in a referential conversation setting, and asking to what extent they aligned with each other in the properties that they used in their conceptualizations of their referring expressions. Second, we wanted to investigate the effect of two specific emotions, amusement and disgust, on alignment in interactive reference production, where we predicted that disgusted speakers would align more with their conversation partners than amused speakers. Three types of alignment were studied: alignment of dispreferred properties (primee uses size-only or overspecification), alignment of overspecified descriptions (primee uses overspecification), and alignment of dispreferred property only (primee uses size-only).

3.4.1. A naturalistic replication of conceptual priming

Goudbeek and Krahmer (2012) studied how speakers describe an object to a conversation partner, using either a preferred or a dispreferred property, when in the prior context either the preferred or the dispreferred variant was primed. Our results replicate the findings of Goudbeek and Krahmer: primed with a description containing size, participants are more likely to use size themselves, with or without adding color in their descriptions. This replicates the results of the first experiment in Goudbeek and Krahmer (2012), with two important modifications. The first is that the interaction in their study was much more artificial than in our current study. Instead of 'interacting' with a computer and being primed by a pre-recorded voice, participants in the current study interacted in dyads in a natural setting, without restrictions. This was achieved by carefully designing trials in such a way that speakers were likely to produce particular descriptions. The second modification concerns the analyses. Goudbeek and Krahmer compared the proportion of size descriptions to a somewhat unrealistic baseline of zero, because Dale and Reiter's (1995) Incremental Algorithm predict this. In contrast, the current analysis compared the proportion of size descriptions in situations where participants are primed with size to a baseline where participants are primed with color, which offers a more realistic (and more stringent) baseline.

In addition, we extended the findings of Goudbeek and Krahmer by investigating a stricter interpretation of alignment – instead of including overspecification in the definition, we also looked at alignment of dispreferred property only. We found that when primed with size, participants were more likely to use size alone in their descriptions. This is

valuable new information, because Goudbeek and Krahmer's definition of alignment did not distinguish between mentioning, and using only, the attribute they were primed with. This result strengthens the evidence for alignment during interactive reference production: when primed to use a dispreferred property, individuals do not only use the dispreferred property more, they also use it more exclusively, without combining it with the preferred property color.

Finally, analyzing alignment of overspecified descriptions, our results also showed that when their conversation partner produced overspecified descriptions ('the large red chair'), participants tended to over specify their own descriptions more often as well, compared to when they were primed with minimally specified descriptions. The proportions we found in the current experiment were very similar to those reported in Goudbeek and Krahmer (2012): overspecified primes triggered overspecification in half of the trials (47%, G&K: 52%), while for single primes this was less prevalent (13%, G&K: 11%). Again, this replicates the result reported by Goudbeek and Krahmer (their Experiment III) in a more naturalistic paradigm.

3.4.2. The influence of emotion

Our other goal was to study the effect of emotion on reference production. Overall, we found only one (small) effect of emotion on alignment of dispreferred property only: disgusted (but not amused) speakers primed with size used the dispreferred attribute (only) more than the preferred attribute, indicating that disgusted speakers have a stronger tendency to align compared to amused individuals, which we are inclined to interpret as a positive side-effect of a negative emotion.

This finding is in line with both attentional bias and the egocentricity accounts. The attentional bias account implies that disgust in specific (Gable & Harmon-Jones, 2010), and negative mood in general (Beukeboom & Semin, 2006) lead to a narrower scope of attention, which in turn might have led the disgusted individuals to focus more on the words of their conversation partner, and hence potentially align more. On the other hand, amused individuals tend to have a broadened attentional scope (Fredrickson & Branigan, 2005), as positive emotions low on goal-achieving motivation, as amusement, tend to evoke (Gable & Harmon-Jones, 2008), which might have resulted the amused individuals to focus less on their conversation partner.

Alternatively, the egocentricity account states that a positive mood in general leads to a more egocentric reference frame (Converse et al., 2008; Kempe et al., 2013). Although, to our knowledge, there is no literature on amusement in specific on egocentricity to this date, we deduce that amused, and therefore positive and egocentric, speakers rely more on their own perspective, using the a priori preferred attribute color, regardless of whether their conversation partner used size or not. Disgust has been found to decrease egocentric perspective taking, and the stronger the disgust, the easier it was to adopt to the perspective of the other (Todd et al., 2015).

Interestingly, however, the effects of emotion on alignment disappeared when we included overspecification for alignment of dispreferred attributes. This might be due to a combination of reasons. First, when inspecting our data (*3.3.2.1. Data pre-processing*), the effects of emotion on overspecification seemed to be small, or non-existent. This is in line with a study by Vonk et al. (2021). In a study similar to ours, individuals were subjected to an emotion induction (happy, sad, or neutral) before individuals engaged in a director-matcher task, describing faces. Contrary to their hypothesis, sad individuals did not over specify significantly less than happy (and neutral) individuals. Therefore, by including overspecification for alignment of dispreferred attributes (alignment of dispreferred property, instead of alignment of dispreferred property only), the effect of emotion is reduced and no longer statistically significant. Second, a possible (small) effect might be obscured due to the small number of observations, resulting in a false negative finding. As can be seen in Table 1, primees do not over specify often when primed with a single attribute: overspecification was used in only 18% (148 descriptions) of all size-only primes (826 primes), and 16% (209 descriptions) in all color-only primes (1321 primes).

3.4.3. Limitations and future research

In the introduction, we formulated our working hypothesis that the affective state of a speaker can conceivably influence every aspect of language production, in line with findings from different domains of research showing that affect can influence many cognitive processes. In the current chapter, we experimentally studied the role of emotional state on various types of alignment during spoken interactions. While we found some evidence for an influence of emotion (disgust) on referential alignment (of dispreferred property only), the effect was not overwhelmingly strong. One possible explanation might be that the induction of amusement was not sufficiently strong, certainly when compared to the induction of disgust. The median (65%) of our disgust condition was 7, on a seven-point scale, indicating that participants generally were 'extremely disgusted'. For amusement, the median (43%) was 5, indicating that most participants felt only 'somewhat amused'. These results are common in studies that use emotion induction: negative emotions are more easily, and stronger, elicited than positive emotions (see, e.g., Boyes et al., 2020; Ferrer et al., 2015; Gasper, 2004; Göritz & Moser, 2006). On the contrary, given that our disgust induction induced high levels of disgust (the mean score was 6.26 on a 7-point Likert scale), we found ourselves surprised not finding a stronger effect for alignment (with dispreferred property only) of disgusted speakers.

In future research, we plan to develop and evaluate stronger methods to induce positive (as well as negative) emotions. Additionally, a potential disadvantage of the method used in this study (and in many others) is that the emotion induction phase (watching a video) is not inherently coupled to the actual experimental task (the director-matcher task). This, too, is something we plan to address in future research, by more strongly integrating the emotion induction phase and the experimental phase. Finally, we compared the effects of disgust and amusement, but these are just two examples out of a large set of possible

emotions. In future research, it would be interesting to compare the effects of multiple negative (for example, anger and disgust) and positive emotions (amusement and pride, say) to make sure that the (small) effects we observe of emotional state can indeed be attributed to the specific emotions, rather than their general valence.

3.4.4. Conclusion

In this chapter, we used an improved version of the referential alignment paradigm of Goudbeek and Krahmer (2012) to study the impact of emotion on alignment in reference production. Our results revealed a clear replication of the earlier findings of Goudbeek and Krahmer, but using a novel experimental paradigm, which is considerably more naturalistic than theirs is. Additionally, we found a small but reliable effect of emotion on alignment, in the sense that disgusted speakers were more inclined to align their description to their conversation partner, but only by using minimally specified descriptions. This suggests that emotional state can indeed influence the way speakers conceptualize their descriptions, and hopefully offers a stepping stone towards a more extensive study of how a variety of emotional states influence different aspects of the language production process.

APPENDIX

Ratings on emotion scales

As expected, we also found significant effects for the other emotions. A one-way MANOVA was performed with Emotion induction videos (amusement and disgust) as independent between-subjects variable and Emotion scales (Anger, Surprise, Fear, Pride and Sadness) as dependent within-subjects variables. The results can be found in Table 1A (see below). All differences were statistically significant ($p < .001$). However, note that generally, the general means for the Emotion scales were on the low side of the scales, indicating that the intensity of the emotions experienced were modest. As intended, the largest differences in general means are on amusement and disgust (see *3.3.1. Manipulation check*).

Table 1A

General means and standard deviations of Emotion scales for Emotion induction videos

Emotion scale	Emotion Induction video	
	Amusement	**Disgust**
	M (SD)	*M (SD)*
Anger	1.59 (1.01)	2.60 (1.47)
Surprise	3.59 (1.70)	5.00 (1.40)
Fear	1.43 (0.83)	2.64 (1.50)
Pride	2.29 (1.37)	1.60 (0.89)
Sadness	1.39 (0.75)	2.34 (1.36)

4

The influence of affective pictures on conceptual pact formation within social interactions

This chapter is based on:
Out, C., Goudbeek, M., & Krahmer, E. (2021). *Is there a role for affective state in conceptual pact formation*? [Manuscript submitted for publication].

ABSTRACT

In conversation, interlocutors are mutually responsible to communicate timely, efficiently and successfully. To account for this observation, Clark and Wilkes-Gibbs (1986) proposed their seminal collaborative model, based on the finding that speakers in a referential matching task gradually build up conceptual pacts, leading to fewer words and speaking turns to describe abstract figures repeatedly. We aimed to replicate and extend their study with more naturalistic stimuli, namely, affective pictures varying in (un)pleasant and (un) arousing content. We assessed the influence of these pictures on both affective state of the interlocutors and conceptual pact formation between dyads. Results support the generalizability of the collaborative model: dyads formed conceptual pacts successfully, resulting in a decline in word use and turn taking during the interactions. Affective state of participants generally improved during the experiment, regardless of type of affective content discussed. Overall, only limited effects of affective content on conceptual pact formation were observed.

4.1. Introduction

To communicate efficiently, speakers in a conversation continuously adapt to each other, using various strategies (for an overview, see e.g., De Looze et al., 2014). One such strategy is keeping track of which information and beliefs are shared and salient, a phenomenon known as maintaining common ground (Stalnaker, 1978), which allows individuals to adjust their utterances to this common ground. For example, when repeatedly describing an object (such as a cute looking kitten), interlocutors can adopt the referential expressions of their conversation partner, converging on the same semantic expressions (for example 'the sweetheart'), a process referred to as lexical entrainment (Brennan & Clark, 1996). When interlocutors agree on this temporary, shared conceptualization, a conceptual pact has been formed, which can be used in later references, e.g., to this specific kitten (Brennan & Clark, 1996).

Given that many cognitive processes can be influenced by affective state, including processing style (Beukeboom & Semin, 2006), attentional scope (Charash et al., 2006) and perspective taking (Converse et al., 2008), it is reasonable to assume that the affective state of interlocutors might influence adaptation processes in conversation as well, which might affect how fast and efficient interlocutors adapt to each other. To the best of our knowledge, no prior study has looked at the impact of affective state, or affective content, on conceptual pact formation. Therefore, in this chapter, we study whether distinct affective states of conversation partners influence their tendency to adapt to each other during referential communication, inspired by the seminal work of Clark and Wilkes-Gibbs (1986).

4.1.1. Conceptual pact formation

Clark and Wilkes-Gibbs (hereafter, C&WG) were interested in studying speaking and understanding in conversation. They investigated this in a referential communication task (based on Krauss & Weinheimer, 1964) with pairs of participants (dyads) that engaged in conversation. In the original referential communication task, one individual, the director, sees twelve highly similar and complex abstract figures, taken from the ancient Chinese game of Tangram. The other half of the dyad, the matcher, sees the same figures the director sees, on cards, arranged in a different order. The director has to describe the pictures in such a way that the matcher can arrange the figures in the described ordering, or array. This task is repeated six times. C&WG found support for their collaborative model: indeed, the dyads created conceptual pacts over the six trials that they used for distinct figures. When directors described a specific picture for the first time, they provided multiple details, aiming to create a mutual agreement on how to refer to this specific figure. For example, in trial 1, the director could say: 'The next one looks like a person who's ice skating, except they're sticking two arms out in front' (C&WG, 1986, p. 12). When the matcher confirmed that he understood her correctly, a mutual agreement, or conceptual pact (Brennan & Clark, 1996), was created. As common ground between the interlocutors builds over time, the references generally became shorter. For example, in trial 4, the director

says: 'The next one's the ice skater', and in trial 6 merely 'the ice skater' (C&WG, 1986, p. 12). This reduction extended to the length of the utterances per trial: in general, directors used substantially less words to describe figures in trial 6, compared to trial 1. Additionally, directors generally needed more turns to reach mutual agreement upon encountering a picture for the first time, compared to later trials. As a result, the average number of speaking turns per figure also declined over trials.

C&WG's findings have been replicated, at least conceptually, numerous times (e.g., Arbuckle et al., 2000; Branigan et al., 2011; Derksen et al., 2015; Yoon & Brown-Schmidt, 2019). Many authors studied additional factors, e.g., gestures (De Ruiter et al., 2012), or replaced the abstract figures with other pictures, e.g., of well-known objects such as an apple or a basketball (Yoon & Brown-Schmidt, 2019). However, to our knowledge, no study has used affective stimuli in these tasks, or looked at the influence of affective state of the speakers on the conceptual pact formation between interlocutors. Previous research has shown that verbalizing affective content (e.g., self-referral sentences; Velten, 1968), including affective pictures (Ortner, 2015; Out et al., 2020a) might induce the corresponding affective state in speakers. However, other studies suggest, in contrast, that verbalizing affective state can also successfully reduce (often negative) affective states (e.g., Baikie & Wilhelm, 2005; Fan et al., 2019). In the present study, we aim to find out whether conversing about affective pictures induces the corresponding affective state in interlocutors engaging in a referential communication task. Additionally, we ask whether affective content influences the tendency to align and therefore, how fast and efficient interlocutors create conceptual pacts. This is based on the findings that, as discussed above, a speaker's affective state can influence various cognitive processes relevant for referential communication, as will be discussed next.

4.1.2. Influence of affective state on adaptation in spoken language production

One way in which affective states can be conceptualized is within a two-dimensional model spanning pleasantness or valence on one dimension, and level of activity or arousal on the other (e.g., Barrett, 2017; Gillioz et al., 2016; Posner et al., 2005; Russell, 1980). This characterization is not without its critics (see, e.g., Fontaine et al., 2007; Kuppens et al., 2017), however, it can be helpful to characterize different affective states in this way.

Based on valence and arousal, different affective states can influence cognitive processes differently, which might influence conceptual pact formation within dyads. For example, previous research has shown that speakers are able to adjust, automatically or purposefully, to the knowledge of their conversation partner (e.g., Bezuidenhout, 2013; Epley et al., 2004). It has been suggested that affective states may influence this form of perspective taking, with interlocutors in a negative mood taking the perspective of their conversation partner more easily into account than interlocutors in a positive mood (Clore & Huntsinger, 2007; Converse et al., 2008). One possible explanation might be that speakers in an unpleasant affective state tend to have a narrower attentional scope (e.g., Gasper & Clore, 2002), compared to the more broadened attentional scope associated

with a pleasant affective state (Fredrickson, 2001; Fredrickson & Branigan, 2005; Rowe et al., 2007). In a conversational setting, these results might translate into speakers in an unpleasant state focusing more on their conversation partner, compared to speakers in a pleasant state, which conceivably could lead to more efficient and quicker conceptual pact formation within dyads. Along similar lines, there is ample evidence that affective state influences the words that speakers use. For example, Beukeboom and Semin (2006) show that speakers in a positive mood tend to use more abstract words (e.g., adjectives, such as 'aggressive') to describe events compared to speakers in a negative mood, who tend to use more concrete words (e.g., descriptive action-verbs like 'punch'). In line with this finding, Kempe et al. (2013) found that speakers in a pleasant mood, compared to speakers in a neutral mood, were more likely to use ambiguous language, for example, saying 'bat' which could either refer to the animal or a baseball bat, instead of specifying the noun with an adjective (e.g., 'baseball bat'). Based on these findings, we might predict that speakers in a pleasant affective state might focus less on their conversation partner in a conversational setting, and use more abstract wording than their peers in an unpleasant affective state, hindering cooperation and the successful and timely formation of conceptual pacts, resulting in more debate before reaching mutual agreement, and therefore, more words used by the director.

However, this prediction is primarily based on differences in pleasantness, which is only one aspect of affective state. Recent studies suggest that also certain categorical emotions, differing on characteristics beyond valence, may have different effects on cognitive aspects. For example, anxiety, disgust and anger are all characterized by negative valence, but differ in their level of arousal and the degree of (un)certainty associated with these emotions (Todd et al., 2015). Anxious, but not angry or disgusted individuals, experience increased feelings of uncertainty, enhancing reliance on an egocentric perspective, at the expense of understanding the viewpoints of others (Todd et al., 2015). In a similar vein, while pleasant affective state might be associated with broadened attentional scope, this might be limited to non-goal-oriented emotions such as amusement, as opposed to goal-oriented states such as appetite for food, which seems to evoke a narrower attentional scope (Gable & Harmon-Jones, 2008).

As far as we can tell, the role of arousal on adaptation in communicative interaction has barely been studied yet. Based on the above, we argue that the level of arousal might influence conceptual pact formation as well. However, we have to keep in mind that, although evidence for the relationship between affective state and cognitive processes is growing, it is not well understood yet (see e.g., Forgas & Matovic, 2020; Lindquist, 2017). Thereby, we surmise that the formation of conceptual pacts might be, either positively or negatively, influenced by the affective state of the interlocutor.

4.1.3. The present study

In the present study, we first of all aim to replicate the original findings by C&WG (1986): do speakers in a referential matching task gradually build up conceptual pacts, resulting in

references becoming shorter over time, as represented by a decrease in number of words uttered, and speaking turns taken, by the director? In addition, we extend their study, asking whether conversing about affective pictures would induce the corresponding affective state in the speaker, and whether this influences the creation of conceptual pacts. For example, participants exposed to pleasant, highly arousing pictures, might report feeling more positive and more aroused after completing the referential communication task. We conjecture that the affect induction might be different for participants, based on their role in the referential communication task. For example, given that directors will play a more active role by describing the pictures, they might spend more time looking attentively at the pictures than the matchers, possibly resulting in stronger affect induction, compared to the matchers.

To answer these questions in the present study, we replaced the abstract Tangram figures used by C&WG with affective pictures, while simultaneously staying true to the original procedure. Participants are involved in a referential communication task with one of four types of pictures: pleasant and high arousal (jumping individuals); pleasant and low arousal (kittens); unpleasant and high arousal (spiders); unpleasant and low arousal (funerals). We assessed the change in affective state with the Self-Assessment Manikin (SAM; Bradley & Lang, 1994). Methods and hypotheses were preregistered at the Open Science Foundation: https://osf.io/3gf5n

4.1.4. Hypotheses

First, we predict that:

(H1) Participants exposed to the four categories of pictures (pleasant or unpleasant, high or low on arousal), will report heightened or reduced levels of pleasantness and arousal, congruent with the respective categories.

As a research question, we added:

(RQ1) The affect induction effect of the referential communication task (cf. H1) might influence directors differently than matchers.

Second, we predict that we will replicate the results of C&WG, finding support for conceptual pact formation:

(H2) The number of words used by the director will decline from trial 1 to trial 6;

(H3) The number of words the director used to describe the pictures will decline over the course of the array, i.e., from picture 1 to picture 12. This decline will be larger for earlier trials, compared to later trials[1].

(H4) The number of speaking turns used by the director will decline from trial 1 to trial 6.

Regarding the effect of affective state, the following research question was added:

(RQ2) Does affective state of the participants influence the number of words uttered per trial (H2), words uttered per picture over trial (H3), and speech turns taken per trial (H4)?

4.2. Method

4.2.1. Design

The study had a 2 (valence: pleasant or unpleasant) by 2 (arousal: high or low) factorial design. Dependent variables were affective state (measured by valence and arousal levels, before and after the task) and conceptual pact formation (measured by the number of words uttered and turns taken by the director, indicating (successful) conceptual pact formation). To support readability, we abbreviate the conditions as +valence/+arousal (+V/+A), +valence/-arousal (+V/-A), -valence/+arousal (-V/+A), and -valence/-arousal (-V/-A), respectively.

4.2.2. Participants

A total of 134 students from Tilburg University participated in pairs (forming dyads) in the experiment for course credit. Our sample included 52 men, 81 women and 1 participant who did not disclose their gender. The range of age was 18 to 27 years (M_{age} = 21.28 years, SD_{age} = 2.31 years). Each participant was assigned to one of the four conditions: +valence/+arousal (n = 34); +valence/-arousal (n = 34); -valence/+arousal (n = 34); or -valence/-arousal (n = 32). Students were randomly assigned to a conversation partner in the same condition, forming a cross-gender dyad (32 dyads), a dyad consisting of two women (24 dyads) or two men (10 dyads). One dyad consisted of one woman and the participant who did not disclosure their gender.

The inclusion criterion was the ability to speak Dutch fluently. No participants were excluded based on this criterion.

The sample size was based on a statistical power analysis in G*Power 3.9.1.4 (Faul et al., 2007), measuring conceptual pact formation (in terms of the difference in number of words, and turn taking, on two separate moments). Our goal was to obtain .95 power

1 This hypothesis was not preregistered by us, but is part of the original study.

to detect a medium effect size of *f* = .25 (Cohen, 1988). We settled for one group, and two moments of testing. Given that there was no straightforward option in G*Power to calculate the power for our analysis of choice, (generalized) linear mixed modelling, we opted for the rudimentary form of the linear mixed effects model, the repeated measures analysis of variance (RM ANOVA; Quené & Van den Bergh, 2004). With the standard alpha error probability of = .05, G*Power suggested *n* = 54 for our group. Since we have four affective conditions, and 54 is not devisable by 4 we settled for *n* = 56. Given that we will only measure the data of the director, we doubled that number, resulting in *N* = 112. Thus, our final sample size of *N* = 134 should be adequate for the main objective of the study (replicating the study of C&WG) and should also allow for the additional objective of controlling for affect.

4.2.3. Consent

All procedures performed were in accordance with the ethical standards of the institutional research committee, the Research Ethics and Data Management Committee of Tilburg School of Humanities and Digital Sciences, Tilburg University (REC#2017/26). All participants gave written informed consent in accordance with the Declaration of Helsinki (1964) and its later amendments or comparable ethical standards.

4.2.4. Materials

4.2.4.1. Assessment and selection of pictures

To select our stimulus material, we started by selecting pictures from the well-known and validated database of affective pictures, the International Affective Picture System (IAPS; Lang et al., 1997; Uhrig et al., 2016*)*. The IAPS contains pictures with varied content, ranging from flowers to mutilated bodies, and have been rated on valence or pleasantness (negative/positive), arousal (low/high) and dominance (dominated/in control; Lang et al., 1997). IAPS pictures have often been successful in inducing affective state (e.g., Lench et al., 2011; Uhrig et al., 2016), influencing both levels of pleasantness and arousal in individuals (Mirandola & Toffalini, 2016; Ritz & Thöns, 2006).

Based on valence and arousal ratings, we selected the most suitable pictures for our experiment, choosing two content categories per affective dimension: friends and jumping individuals (positive valence, high arousal), kittens and babies (positive valence, low arousal), snakes and spiders (negative valence, high arousal), and graveyards and funerals (negative valence, low arousal; see Table 1). Since the IAPS did not contain a sufficient number of different pictures in the relevant categories, we expanded our pool of picture options using Google Picture Search, searching for pictures in the same content categories as those in the targeted picture sets (e.g., kittens). We made sure that the final pictures of each set were sufficiently similar to each other, that they could only be distinguished from one another by elaborate descriptions, stimulating our participants to go beyond simple descriptions ('a cat'), thereby facilitating conceptual pact formation.

Our final selection contained 104 pictures, which we pre-tested through an online questionnaire in Qualtrics. Eighty-one students (all different from the 134 participants of the main study; 21 men; 60 women; age range: 18-40 years, M_{age} = 21.88 years, SD_{age} = 3.32 years) participated for course credit, rating pictures of one content category on pleasantness and arousal, using the 9-point SAM. For pleasantness, the scale ranged from a frowning, unhappy manikin (1) to a smiling, happy manikin (9). For arousal, the scale ranged from a relaxed, sleepy manikin with a dot in their abdomen, representing low arousal (1) to an excited, wide-eyed manikin with an explosion in their abdomen area, representing high arousal (9). Each scale contained five manikins, representing the gradual increase in pleasantness or arousal. Participants chose one number per scale; the closer the numbers were to the words, the stronger they match the feeling described. Mean and standard deviations of all content picture categories can be found in Table 1.

Table 1

Pilot-test results for the eight picture categories, based on SAM scores

Valence	Arousal	Content	*n*	Valence *M* (*SD*)	Arousal *M* (*SD*)
Positive	High	Friends	11	6.61 (1.22)	4.25 (1.89)
Positive	**High**	**Jump**	**5**	**6.80 (0.79)**	**4.78 (2.18)**
Positive	**Low**	**Kittens**	**11**	**6.36 (1.10)**	**3.80 (2.06)**
Positive	Low	Babies	9	6.05 (0.79)	4.22 (1.69)
Negative	High	Snakes	12	3.78 (1.04)	4.50 (2.12)
Negative	**High**	**Spiders**	**7**	**3.07 (1.98)**	**6.18 (2.15)**
Negative	Low	Graveyards	20	3.78 (0.95)	3.44 (1.07)
Negative	**Low**	**Funerals**	**6**	**3.04 (0.75)**	**4.11 (0.82)**

Note. Boldfaced categories represent our final selection.

4.2.4.1.1. Final selection

Out of the two affective dimension categories, we chose the best content category, i.e., scoring most in line with the desired valence- and arousal ratings. Our final selection included 1 content picture set per condition, containing 12 distinct pictures of jumping individuals (positive valence, high arousal), kittens (positive valence, low arousal), spiders (negative valence, high arousal) and funerals (negative valence, low arousal).

4.2.4.2. Director's sheets

Directors were provided with six sheets of A2 paper in landscape format. For each condition, all 12 corresponding pictures (10 x 12.5 cm each) were printed on these sheets. Following C&WG, each sheet contained a different array of pictures per category (see Figure 1a-d).

Figure 1a

Director's sheet: Example trial of +valence/+arousal

Figure 1b

Director's sheet: Example trial of +valence/-arousal

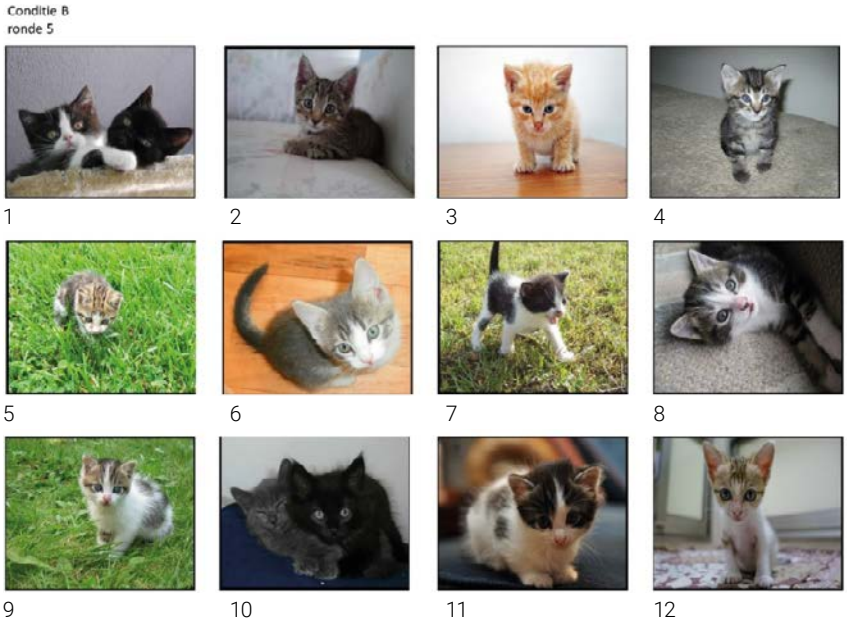


Figure 1c

Director's sheet: Example trial of -valence/+arousal

Figure 1d

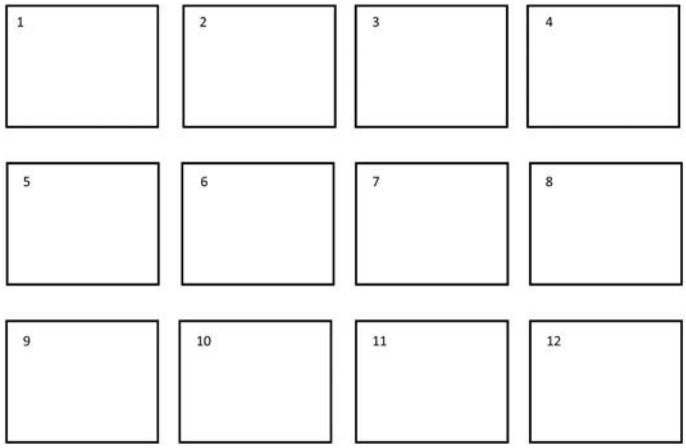
Director's sheet: Example trial of -valence/-arousal

4.2.4.3. Matcher's sheet and envelope with pictures

For the matcher, one sheet of identical size was created, containing twelve empty boxes of 10 x 12.5 cm each (see Figure 1e). The sheet was accompanied with an envelope, containing pictures of the corresponding condition on separate cards.

Figure 1e

Matcher's sheet



4.2.4.4. Referential communication task

The first participant entering the room was given the role of director, and seated on the left end of the table. The second became the matcher and sat on the right end of the table. An opaque screen in the middle of the table obscured the ability to view the other person's activities, but did enable them to see each other's faces. The director received their sheets, and was asked to turn the stack around, facing the sheet of the first trial. The matcher received their sheet and envelope with pictures. Both participants were given a minute to view the pictures, and arrange the materials to their liking.

Participants were informed that they were presented with identical sets of pictures, and were going to engage in the task together, before describing the respective roles they were assigned to. The director's task was to report to the matcher which picture occupied which position in each particular array. The matcher's task was to put the pictures in the correct order, in the blank boxes on their sheet. For each trial, the director moved the top sheet of paper to the back of the stack, revealing a new sheet with the same set of pictures but in a different array. This process was repeated a number of times, resulting in six trials per picture set. After the completion of each trial, participants indicated when they were done describing and placing the pictures. The experiment leader checked the accuracy for each trial; pointing out errors when necessary.

Directors received explicit instructions to perform the task in order (i.e., first describe the picture on position 1, then the picture on the second position, etc.), and were prohibited from showing matchers the pictures, because this would obviously undermine the purpose of the study. Additionally, participants were informed that they were allowed to converse about the pictures and ask each other questions, encouraging them to interact naturally, using conversational speech. When both participants understood what was expected of them, the task began. Each of the 67 dyads discussed twelve pictures per six trials, resulting in a total of 72 picture descriptions per dyad. All speech during the task was recorded.

4.2.4.5. SAM

To measure whether describing the pictures changed pleasantness and arousal levels in participants, participants filled in the pen-and-paper version of the SAM before and after the referential communication task, indicating their current affective state on 9-point scales as described above.

4.2.4.6. Familiarity question

Participants indicated on a 7-point Likert scale to which extent they knew their conversation partner, with a 1 indicating 'not at all' and a 7 'very well'. The vast majority (70.9%) reported 1 ('not at all', n = 77) or 2 (n = 18), indicating that they did not know their conversation partner. 14.9% reported 6 (n = 18) or 7 ('very well'; n = 2). This question was asked after the referential communication task took place, to make sure that this question would not trigger (additional) speculation about the purpose of the task and the research in general.

4.2.5. Procedure

We kept the set-up of the study as similar as possible to C&WG (1986). Participants read and signed an informed consent form, reported their student number (or name), age, gender and role (director or matcher). Participants first filled in the SAM, engaged in the referential communication task, and then filled in the SAM for the second time. Finally, they indicated how well they knew their conversation partner, wrote down what they thought the experiment was about, and were debriefed individually on paper.

4.2.6. Data pre-processing

4.2.6.1. Transcriptions

Two student assistants, naïve to the hypotheses and research questions of the project, manually transcribed the audio recordings of the referential communication tasks. One conversation, in the -valence/+arousal condition, was missing due to a technical glitch. The final sample included 66 transcribed conversations, which were used for further analysis. Excluded from the transcriptions were incomprehensible utterances (these were tagged as incomprehensible), laughter, false starts (e.g., 'Twe-' meaning twelve, continuing with 'eleven'), and half-words, including speech disfluency and correcting (e.g., 'fune-' meaning

funeral, continuing with 'ceremony'). Included were repetitions of complete words (e.g., 'picture four, picture four'), and (self) repairs in complete words (e.g., 'one, no, sorry, two').

Data was structured in such a way that each director described twelve pictures for six trials. When one or more pictures were described repeatedly in one trial (50 cases), they were added to the picture description moment before the repetition took place. For example, a director describes picture 8, and then describes picture 4 again. The second description of picture 4 is added to the picture description moment of picture 8 in the array. Three trials were described non-chronologically. We treated them as if they were chronological. For example, the first picture which should be described in trial 1 is picture 5. However, this picture was not described first. Then, the description of picture 5 is nevertheless considered to be the first picture description for trial 1.

In 61 cases, the matcher described, or aimed to describe, a picture before the director did so. In 54 of these 61 cases (88.5%), this happened on the last picture in the array of a trial. Given that we only focused on the utterances of the director, we analyzed the words used by the director in these moments, as usual.

4.2.6.2. Word count

Words were defined as complete words, including fillers (e.g., 'uh'), non-Dutch, dialect and slang words (e.g., English words, 'nummero' instead of 'number') and names (e.g., 'Medoza', 'Walker'). Word count was defined as the number of words used by the director to describe each picture; including answers to questions asked by the matcher.

Based on the above, 37 descriptions were added to their corresponding picture description moments, resulting in a slight reduction of picture descriptions. The final dataset included 4752 descriptions: 1224 for +valence/+arousal and +valence/-arousal, and 1152 for -valence/+arousal and -valence/-arousal.

4.2.6.3. Turn taking

Turn taking was defined as every speaking turn the director takes, containing utterances that include words (e.g., 'Indeed', 'The fifth picture is the boy on the yellow trampoline'), sounds of approval, listening or backchanneling (e.g., 'uhuh', 'uh', 'hm', etc.), and/or aiming to produce content words (incomprehensible utterances). Common turns by the director were descriptions of pictures, answering a question by the matcher, adding information to a description, and requests for confirmation (e.g., 'Do you understand?'). Utterances including only non-linguistic sounds (e.g., laughter, sneezing, coughing) were not considered as separate turns and therefore, we considered the current speaker to continue their turn (see Example 1 in the Appendix).

When the director was talking through the utterances of the matcher by adding new information, repeating information, confirming, interrupting, or backchanneling, this was considered a separate turn (see Example 2 in the Appendix). When the matcher and director spoke at the same time, the initiator of the speech was considered to have the turn.

4.2.7. Missing values and error rate

4.2.7.1. Missing values

In the experiment, missing values in SAM scores occurred in three (out of a total of *N* = 134) participants. Missing values were less than 1% of the SAM scores, including five values: pre-test valence score and pre-test arousal score (*n* = 1), pre-test arousal score and post-test arousal score (*n* = 1), and post-test arousal score (*n* = 1). All missing values occurred in the -valence/+arousal condition.

4.2.7.2. Error rate

While engaging in the referential communication task, seven dyads (out of *N* = 67 dyads), 10.5% percent, made mistakes, by mixing up pictures. In all 402 trials, the error rate was 1.7%. Six dyads made one error in one trial; one dyad made one error in two trials. Given the small error rate, errors were not excluded from the dataset.

4.2.8. Statistical analyses

To test our first hypothesis (H1) and explore our first research question (RQ1), data was analyzed using IBM SPSS statistics software (version 26). Before our manipulation check, we checked whether difference scores of pleasantness and arousal (post-test minus pre-test) were approximately normally distributed by visually inspecting QQ-plots and histograms, as well as using Shapiro Wilk's test. If approximate normality could be assumed, effectivity of talking about affective pictures on affective state (H1) was measured by performing two RM ANOVAs with time (pre-test and post-test) as within-subjects factor, condition as between-subjects factor, and self-reported pleasantness and arousal scores as dependent variables. Bonferroni corrected posthoc tests were included. If approximate normality could not be assumed, indicated by (severe) deviation from normality according to our measures, a non-parametric Wilcoxon signed rank test was performed.

To answer our first research question (RQ1) that the affect induction effect of the referential communication task might influence directors differently than matchers, we inspected the scores on pleasantness and arousal of the directors and matchers separately. Difference scores were obtained in the same fashion as described above.

To test our second, third and fourth hypothesis (H2, H3, H4), data was analyzed using the statistical program R 3.6.3. (R Core Team, 2020). Given that we collected multiple trials for each director, generalized linear mixed model analyses were used, enabling us to take random effects of participant and items (unique pictures) into account when analyzing the number of words uttered, and speaking turns taken by the director in the repeated measures data set. For this purpose, we used the lme4 package (Bates et al., 2014), with p-values provided by the lmerTest package (Kuznetsova et al., 2017). Visual inspection of our data revealed a right-skewed distribution of the data; hence, a log link function was added to the Poisson probability distribution.

Four models were created, starting with the fixed effect(s) only, and adding interactions when appropriate. Then, we added the by-participant random intercept, by-item random intercept, random intercept of trial (when trial was not selected as a fixed effect), and by-item random slope for trial, respectively. When the model did not converge after these steps were taken, we removed the random slope, random intercept for trial, by-item random intercept, and by-participant random intercept, respectively. The relevant fixed effects remained unchanged, irrespective of the exact converging or non-converging model: (non) significant effects remained the same, and betas remained similar in size. The most complicated model that still converged was selected as the best model.

To explore our second research question (RQ2), four additional models were created, by selecting the final models and adding the affective picture dimensions, valence and arousal, as fixed main effects, as well as the interaction. For clarification reasons, additional details and further explanation about these models will be provided in the results section later. All model components, based on linear fit, can be found in Table 2.

Table 2

Components used in model equations, including descriptions

Component	Description	Type of variable and values
WC	number of words uttered by director	interval: 0 - 244
Turns	number of speech turns taken by director	interval: 0 - 25
trial	fixed effect for trial (between-trial)	ordinal: trial 1[1] - 6
position	fixed effect for position of picture in the array (within-trial)	ordinal: position 1[1] to 12
valence	fixed effect for affective picture dimension valence	factor: 0[1] = negative, 1 = positive
arousal	fixed effect for affective picture dimension arousal	factor: 0[1] = low arousal, 1 = high arousal
1|item	random intercept for unique image (item)	factor: 1 to 48
1|ID	random intercept for participant	factor: 1 – 131 (only uneven numbers, indicating directors)
1|trial	random intercept for trial	
trial|item	by-item random slope for trial	

Note. DV = dependent variable.
[1]Reference level.

4.3. Results

4.3.1. Affect induction

4.3.1.1. Descriptives

In general, participants were in fairly good spirits, with the vast majority scoring a 7 or higher on valence, both before (*n* = 104, 77.6%) and after viewing the pictures (*n* = 116, 86.6%), indicating high levels of pleasantness. Regarding arousal, the results were more mixed. The majority scored between 2 and 5, both before (*n* = 106, 79.1%, mode = 3) and after viewing the pictures (*n* = 96, 71.6%, mode = 3), indicating that that participants tended to experience fairly low levels of arousal. Means, standard deviations and difference scores can be found in Table 3 (left). Difference scores were obtained by subtracting each individual post-test from pre-test score, with positive numbers indicating an increase in pleasantness and level of arousal.

4.3.1.2. Manipulation check

4.3.1.2.1. Assumption checking

For pleasantness, Shapiro Wilk's test showed significant departure from normality for all conditions (all $p \leq .005$). For arousal, Shapiro Wilk's test indicated that all conditions met the assumption of normality (p-values ranging between $p = .063$ and $p = .245$), except +valence/-arousal ($p = .005$). Inspection of QQ-plots and histograms indicated that the difference scores of both valence and arousal were approximately, although not perfectly, normally distributed. Therefore, a parametric RM ANOVA was performed (see Table 4). However, given the slight deviance from normality, we were somewhat more careful interpreting our results; confidence intervals (CI) are provided.

Table 3

Means, standard deviations and difference scores (Diff) on pre- and post-test valence- and arousal scores for directors, matchers and combined

Valence													
	Total				Directors				Matchers				D-M
Condition	*n*	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	Diff *M*	*n*	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	Diff *M*	*n*	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	Diff *M*	Diff *M*
+V/+A	34	6.56 (1.37)	7.15 (1.16)	0.59	17	6.59 (1.00)	7.29 (0.69)	0.70	17	6.53 (1.70)	7.00 (1.50)	0.47	0.23
+V/-A	34	6.88 (0.98)	7.38 (0.78)	0.50	17	6.71 (1.05)	7.24 (0.83)	0.53	17	7.06 (0.90)	7.53 (0.72)	0.47	0.06
-V/+A	33	7.00 (0.56)	7.27 (0.94)	0.27	17	6.88 (0.60)	7.41 (1.06)	0.53	16	7.13 (0.50)	7.13 (0.81)	0.00	0.53
-V/-A	32	7.03 (1.00)	7.47 (1.24)	0.44	16	7.00 (1.03)	7.56 (1.31)	0.56	16	7.06 (1.00)	7.38 (1.20)	0.32	0.24
Total	133	6.86 (1.03)	7.32 (1.04)	0.46	67	6.79 (0.93)	7.37 (0.98)	0.58	66	6.94 (1.12)	7.26 (1.10)	0.32	0.26
Arousal													
	Total				Directors				Matchers				D-M
Condition	*n*	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	Diff *M*	*n*	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	Diff *M*	*n*	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	Diff *M*	Diff *M*
+V/+A	34	3.00 (1.43)	3.56 (1.50)	0.56	17	2.88 (1.65)	4.24 (1.39)	1.36	17	3.12 (1.22)	2.88 (1.32)	-0.24	1.60
+V/-A	34	4.03 (1.73)	4.00 (1.77)	-0.03	17	4.06 (1.56)	4.06 (2.01)	0.00	17	4.00 (1.94)	3.94 (1.56)	-0.06	0.06
-V/+A	31	3.55 (1.43)	4.52 (1.91)	0.97	15*	3.73 (1.44)	5.60 (1.80)	1.87	16	3.37 (1.45)	3.50 (1.41)	0.13	1.74
-V/-A	32	3.53 (1.54)	4.25 (2.02)	0.72	16	3.94 (1.48)	4.38 (2.36)	0.44	16	3.13 (1.54)	4.13 (1.67)	1.00	-0.56
Total	131	3.53 (1.57)	4.07 (1.82)	0.54	65	3.65 (1.58)	4.54 (1.97)	0.89	66	3.14 (1.57)	3.61 (1.54)	0.47	0.42

Note. D = Directors, M = Matchers. *n = 16 for pre-test; n = 15 for post-test.

Table 4

RM ANOVA summary for manipulation check

	Valence					Arousal				
Effect	*MS*	*df*	*F*	*p* ≤	η^2	*MS*	*df*	*F*	*p* ≤	η^2
time	13.43	1	26.83	.001	.172	20.07	1	14.61	.001	.103
condition	1.91	3	1.17	.325		8.48	3	1.98	.120	
time * condition	0.30	3	.59	.622		2.95	3	2.15	.097	
error (time)	0.50	129				1.37	127			
error (condition)	1.64	129				4.27	127			

4.3.1.2.2. Hypothesis testing

Regarding valence, a main effect was found for time, $F(1, 129) = 26.84$, $p < .001$, $\mu^2 = .172$. Participants reported to feel significantly more pleasant after engaging in the referential communication task ($M = 7.32$, $SD = 1.04$) than before ($M = 6.87$, $SD = 1.03$), with a mean difference of 0.45 (95% CI: 0.28 to 0.62). No main effect for condition was found, $F(3, 129) = 1.17$, $p = .325$, indicating that overall, participants felt equally pleasant. No interaction effect of condition with time was found either, $F(3, 129) = 0.59$, $p = .622$, indicating that participants did not report to feel more or less pleasant after being exposed to one of the four picture categories. In this case, the enhanced levels of pleasantness after the task were independent of the type of pictures participants viewed.

Regarding arousal, a main effect was found for time, $F(1, 127) = 14.62$, $p < .001$, $\mu^2 = .103$. Participants reported to experience significantly more arousal after engaging in the referential communication task ($M = 4.07$, $SD = 1.82$) than before ($M = 3.53$, $SD = 1.57$), with a mean increase of 0.55 (95% CI: 0.27 to 0.84). No main effect for condition was found, $F(3, 127) = 1.99$, $p = .120$, indicating that overall, participants felt equally aroused. More importantly, again, no interaction effect of condition with time was found, $F(3, 127) = 2.15$, $p = .097$. This finding indicates that participants did not report to feel more or less aroused after exposure to one of the four picture categories; therefore, the enhanced levels of arousal after the task could not be attributed to the type of pictures participants viewed.

Concluding, participants exposed to the four picture categories reported an increase in pleasantness and arousal, regardless of condition. Based on these results, we could not reject the null hypothesis.

4.3.2. Affect induction in directors and matchers

Means, standard deviations and difference scores for directors and matchers can be found in Table 3 (middle and right).

As can be seen in Table 3, both directors and matchers reported to feel somewhat more pleasant after viewing the pictures, regardless of the content, with the largest increase of 0.70 for positive content, and 0.53 for negative content. In general, directors reported

a higher increase in pleasantness than matchers. Viewing arousing pictures enhanced arousal levels in directors, which increased by 1.36 or by 1.87, depending on the valence condition. No such effect was observed for the matchers, who did not report a substantial change in arousal. Contrary to our expectations, matchers reported a substantial (1.00), and directors a smaller (0.44) increase in arousal after viewing the -valence/-arousal pictures. No effects were found for the +valence/-arousal condition.

To answer our first research question (RQ1), affect induction influenced directors and matchers differently. Difference scores show that directors, compared to matchers, generally reported (slightly) more pleasantness after engagement in the task, independent of type of pictures. Regarding the effect of arousal, directors reported substantially higher levels of arousal after viewing the highly (compared to low) arousing pictures. This pattern was not observed for the matchers; none of the conditions evoked a substantial in- or decrease in arousal in the expected directions. Unexpectedly, matchers did report to experience more arousal after the -valence/-arousal condition.

4.3.3. Conceptual pact formation

To test whether we could replicate the main effects found in C&WG (1986), we used both data visualization, guided by descriptive statistics, as well as statistical modelling. Means, modes, standard deviations, confidence intervals (CI) and range of the number of words and speaking turns over trials are provided in Table 5. The most complicated model that still converged was selected as the best model, details of which, including betas, are provided in Table 6.

To explore our second research question (RQ2) and subsequently assess the effect of repeatedly describing (affective) pictures on number of words used, and speaking turns taken by director, four additional models were created. Given that we were unable to find support for the effectivity of the affect induction, we did not explore whether affective state of the directors improved the abovementioned models. Instead, we investigated whether the affective picture content improved our models, predicting the number of words uttered per trial, position of picture in the array, and position of picture in the array per trial, as well as the number of speech turns used per trial. To our models, we added the affective picture dimensions valence and arousal as fixed main effects, as well as the interaction. The anova() function was used to compare the model with and without the affective picture dimensions, to obtain the F-tests with p-values, using Satterwhaite's method for denominator degrees of freedom and F-statistic. The results, and full equations of the models, are provided in Table 7.

4.3.3.1. Words per trial

To predict the number of words used per trial, we included the fixed effect for trial, random slope for unique picture per trial, and random intercepts for unique picture and participant. As can be observed in Table 3 and Figure 2, directors used more words to describe the pictures in trial 1 than in trial 6. A consecutive decline in words uttered over trials was observed, with the largest decline for the first trials.

As can be seen in Figure 3, the consecutive decline in words used over trials was found for all picture categories. Indeed, as can be seen in Table 7, adding valence and arousal to the model did not improve the model, indicating that the decrease in word use over trials was irrespective of affective picture condition.

Table 5

Number of words and speaking turns used per picture description over trial

	Words uttered					**Turns taken**				
Trial	*M*	*Mode*	*SD*	*CI*	Range	*M*	*Mode*	*SD*	*CI*	Range
1	27.22	20	18.78	25.91-28.53	1-244	2.01	1	1.62	1.90-2.13	1-25
2	15.99	10	9.74	15.31-16.67	1-124	1.37	1	0.80	1.32-1.43	1-8
3	12.17	9	8.22	11.59-12.74	1-115	1.23	1	0.85	1.17-1.29	1-18
4	10.48	9	5.66	10.08-10.87	0-60	1.13	1	0.45	1.10-1.16	0-6
5	9.41	7	4.80	9.08-9.75	1-52	1.08	1	0.33	1.06-1.10	1-4
6	9.06	7	4.35	8.76-9.36	0-28	1.08	1	0.30	1.06-1.10	0-3

Table 6

Estimated parameters of the optimal models to estimate word count and turn taking

Predictor	Fixed effects				RE ID		RE item		RE trial		Random slope trial|item	
	β	*SE*	*z*	*p* ≤	s^2	*SD*	s^2	*SD*	s^2	*SD*	s^2	*SD*
Words per trial: Effect of trial on decrease in word use												
Intercept	2.51	0.03	71.47	.001	0.04	0.20	0.03	0.17	-	-	0.09	0.30
Trial	-0.82	0.04	-18.19	.001								
Words per position: Effect of position of picture in the array on decrease in word use												
Intercept	2.51	0.16	15.74	.001	0.05	0.21	0.03	0.18	0.14	0.38	-	-
Position	-0.50	0.02	-31.65	.001								
Words per position: Effect of position of picture in the array and trial on decrease in word use												
Intercept	2.51	0.04	67.83	.001	0.04	0.21	0.03	0.18	-	-	-	-
Position	-0.43	0.02	-25.27	.001								
Trial	-0.84	0.01	-83.79	.001								
Interaction	0.56	0.04	14.22	.001								
Turns per trial: Effect of trial on decrease in turn taking												
Intercept	0.24	0.02	10.36	.001	0.01	0.10	0.01	0.10	-	-	-	-
Trial	-0.47	0.03	-15.25	.001								

Note. RE = random effect.

Table 7

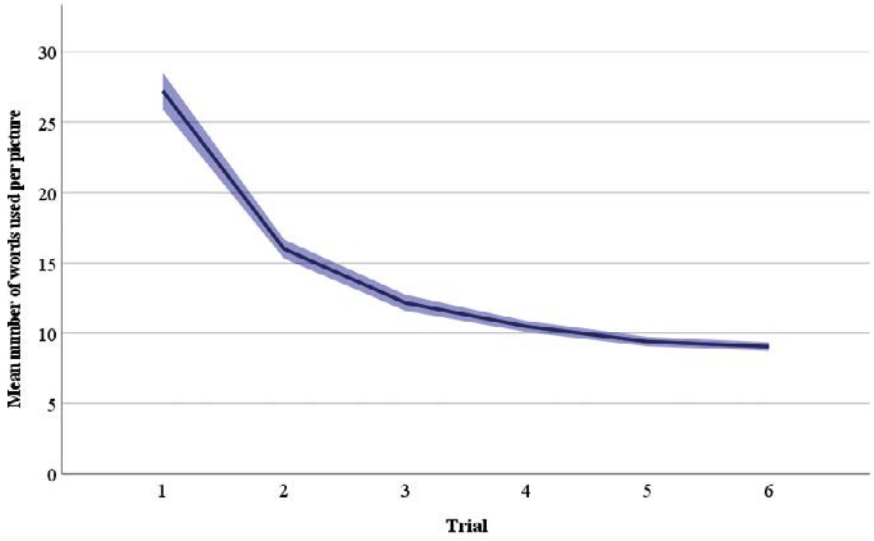
Comparisons of models with- and without affective picture dimensions valence and arousal as factor

Model	Affect dimensions	Model equation	AIC	BIC	loglike	deviance	χ^2	df	Pr(>Chisq)
M1: Words per trial	Excluded	WC ~ trial + (1+trial|item) + (1|ID)	33115	33297	-16530	33059			
	Included	WC ~ trial + valence + arousal + valence*arousal + (1+trial|item) + (1|ID)	33116	33316	-16527	33054	5.645	3	.130
M2: Words per position of picture in the array	Excluded	WC ~ position + (1|item) + (1|ID) + (1|trial)	34203	34300	-17087	34173			
	Included	WC ~ position + valence + arousal + valence*arousal + (1|item) + (1|ID) + (1|trial)	34194	34311	-17079	34158	15.317	3	.002*
M3: Words per position of picture in the array, per trial	Excluded	WC ~ position + trial + position*trial + (1|item) + (1|ID)	33682	34161	-16767	33534			
	Included	WC ~ position + trial + position*trial + valence + arousal + valence*arousal + (1|item) + (1|ID)	33674	34171	-16760	33520	14.871	3	.002*
M4: Turns per trial	Excluded	Turns ~ trial + (1|item) + (1|ID)	11593	11645	-5788.6	11577			
	Included	Turns ~ trial + valence + arousal + valence*arousal + (1|item) + (1|ID)	11573	11644	-5775.3	11551	26.598	3	.001*

Note. * = sign. at $p \leq .002$

Figure 2

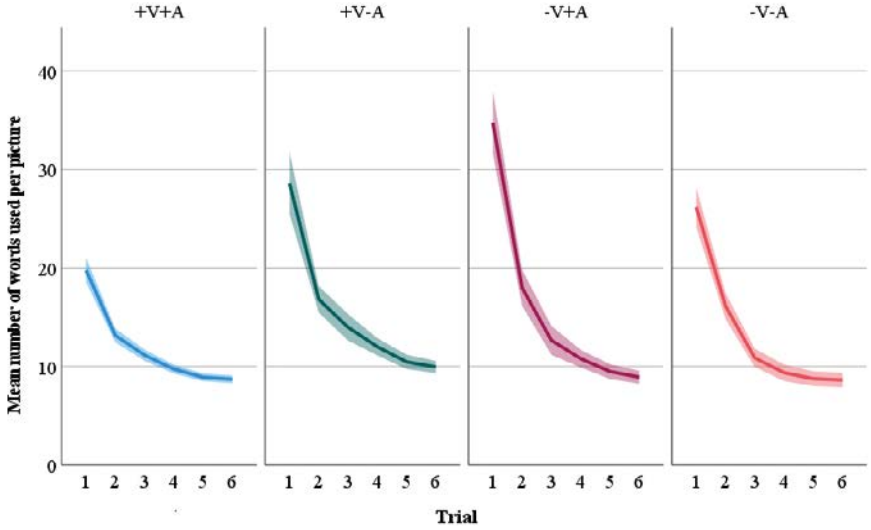
Mean number of words the director used to describe one picture (y-axis), per trial (x-axis)



Note. Band represents 95% confidence interval.

Figure 3

Mean number of words the director used to describe one picture (y-axis), per trial (y-axis), sorted by affective picture category (panels)



Note. Bands represent 95% confidence intervals.

4.3.3.2. Words per position of picture in the array

To predict the number of words used per position of picture in the array, regardless of trial, we included the fixed effect for position of picture in the array, and random intercepts for trial, unique picture and participant. As can be observed in Figure 4, directors used more words to describe the first picture (*M* = 18.06, *SD* = 16.82, CI: 16.39–19.72) than the last picture in the array of a trial (*M* = 9.77, *SD* = 8.13, CI: 8.96–10.57). The general pattern of this decline was steady, but irregular, with in- and decreases along the way. Upon inspecting the transcriptions of individual directors, we found that most directors indeed fluctuated in the number of words uttered describing pictures within one array. Table 8 provides an example of this for one director, describing +valence/-arousal pictures in trial 1.

As can be seen in Figure 5, the gradual but irregular decline in words used per position of picture in the array was observed in all categories. When we added valence and arousal to the model, as can be seen in Table 7, the model improved. Valence and arousal seemed to have the same effect, although, valence (statistically) significantly predicted the number of words directors used per position of picture in an array, β = .20, p = .043, whereas arousal did not, β = .19, p = .066 (although the β-values are clearly close together). However, more importantly, the interaction between valence and arousal was significant, β = -.49, p < .001. Based on our inspection of Figure 5, we conjecture that the decline in word use over time was gentler for the +valence/+arousal picture category, compared to the other affective picture categories. In other words, while directors did use less words over time to describe the depicted jumping individuals within one trial, this decline was not as prominent as for directors describing the pictures in the other sets, depicting kittens, spiders, or funerals.

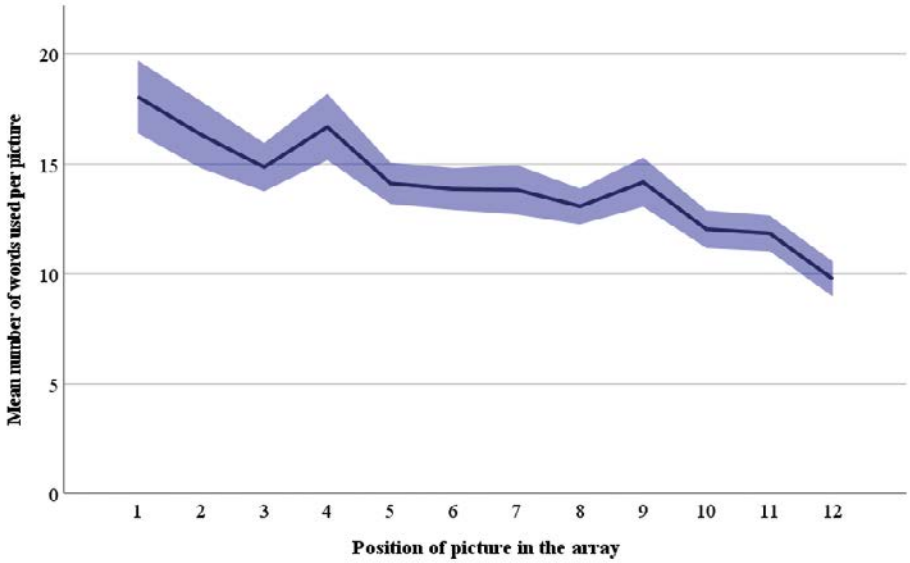
4.3.3.3. Words of position of picture in an array, over trial

To predict the number of words used per position of picture in an array, per trial, we included the fixed effects for position of picture in the array and trial, as well as the interaction. Random intercepts for unique picture and participant were added. As can be observed in Figure 6, the decline in word use is larger for earlier trials, compared to later trials. We also observed a decrease in number of words used to describe each unique picture, indicating the result of successful conceptual pact formation. An example is provided in Table 9.

As can be seen in Figure 7, for all picture categories, the decline in word use is larger for earlier trials, compared to later trials. Again, valence and arousal seemed to have a similar effect on the model, predicting the number of words directors used per position of picture in the array, over trials; a (statistically) significant effect was found for valence, β = .20, p = .043, but not for arousal, β = .19, p = .059. More importantly, the interaction between valence and arousal was significant, β = -.49, p < .001. Based on our inspection of Figure 7, we suspect that this might be explained by the first trial, where the decline in word use over time seemed to be the sharpest for +valence/-arousal picture category, and most gentle for the +valence/+arousal picture category, compared to the other two picture categories. Regarding the remaining trials, no substantial differences were observed between the picture categories.

Figure 4

Mean number of words the director used to describe one picture (y-axis), per position of the picture in the array (x-axis)



Note. Band represents 95% confidence interval.

Table 8

Example of utterances by one director in the +V/-A condition, describing pictures in trial 1

Position of picture in the array	Utterances (Dutch)	Utterances (English translation)	WC (Dutch)
1	Oké. Nou, plaatje één is een katje in heel fel groen gras, die naar beneden kijkt.	Okay. Well, picture one is a kitten in very bright green grass, looking down.	16
2	Het tweede is een grijs katje en een zwart katje, die samen-	The second is a grey kitten and a black kitten, which together-	12
3	De derde is een katje die recht in de camera kijkt op een witte bank met een soort van bloemenpatroon. Met blauwe ogen, beetje tijger-achtig.	The third is a kitten looking straight into the camera on a white sofa with some sort of floral pattern. With blue eyes, a bit tiger-like.	26
4	De vierde is een heel dun katje, die op een rood-achtig kleed zit. Witte pootjes, blauwe ogen, rood hoofdje. Of ja, oranje-achtig. Ik denk de smalste van allemaal. Heel smal, zijn hoofd is groter dan uh het lichaam, lijkt wel. En het lichaam is heel wit en z'n hoofd is dan een beetje rossig. En hij heeft een heel driehoekig hoofdje. Best wel grote oren. En dan wat op de vloer ligt is een rood tapijt met witte figuurtjes.	The fourth is a very thin kitten, sitting on a reddish carpet. White legs, blue eyes, red head. Or well, orangey. I think the smallest of them all. Very small, his head is bigger than uh the body, it seems. And the body is very white and his head is a bit ginger. And he has a very triangular head. Pretty big ears. And then what lies on the floor is a red carpet with white figures.	81
5	En dan heb je een hele kleine zwart-witte en die kijkt heel zielig en die zit op de grond. Er is ook een andere zwart-witte maar die loopt, deze zit.	And then you have a very small black-and-white one and it looks very sad and it sits on the ground. There is also another black-and-white one but they walk, this one is sitting.	32
6	En dan een grijs katje die op een houten vloer kijkt en omhoog kijkt. Die is ook van boven, boven gefotografeerd.	And then a grey kitten looking to a wooden floor and looking up. This one is also photographed from above, above.	21
7	En dan een rossig katje.	And then a ginger kitten.	5
8	En dan de andere katjes die met z'n tweeën op de foto staan.	And then the other kittens that are in the picture together.	13
9	En dan het zwart-witte katje dat op het gras loopt.	And then the black-and-white kitten walking on the grass.	11
10	En dan, uh, ja wat is dit? Hij ligt op een bank denk ik en het licht- En die kijkt zo omhoog.	And then, uh, yes what is this? He's on a couch I think, and the light – And he looks up like that.	22
11	En dan heb je eentje met hele blauwe ogen, die op een beige-achtig tapijt zit.	And then you have one with very blue eyes, sitting on a beige-like carpet.	16
12	En dan eentje die in het gras zit. Oké.	And then one sitting in the grass. Okay.	9

Table 9

Example of conceptual pact formation for 1 unique picture, over trials

Trial	Utterances (Dutch)	Utterances (English translation)	WC
1	Uh, derde foto is een vrouw die ook in de lucht springt en op de achtergrond zie je een stad. Uh, ja heel veel gebouwen, uh, en aan de rechtere kant zie je een lantaarnpaal.	Uh, third photo is a woman who also jumps in the air and in the background you can see a city. Uh, yes a lot of buildings, uh, and on the right side you see a lamppost.	35
2	Uh, derde vrouw, uh, foto is van die vrouw die in de lucht springt met, uh, stad op de achtergrond en lantaarnpaal, uh, rechts.	Uh, third woman, uh, photo is of that woman jumping in the air with, uh, city in the background and lamppost, uh, on the right.	24
3	Vierde foto is van de vrouw, uh met de stad op de achtergrond, de lantaarnpaal rechts.	Fourth photo is of the woman, uh, with the city in the background, the lamppost on the right.	16
4	Uh, elfde is vrouw, uh, met de stad op de achtergrond.	Uh, eleventh is woman, uh, with the city in the background.	11
5	Zesde is de vrouw met de stad op de achtergrond.	Sixth is the woman with the city in the background.	10
6	Zevende is de vrouw, uh, met de stad op de achtergrond.	Seventh is the woman, uh, with the city in the background.	11

Note. Naturally, the position of the picture in the array varies per trial.

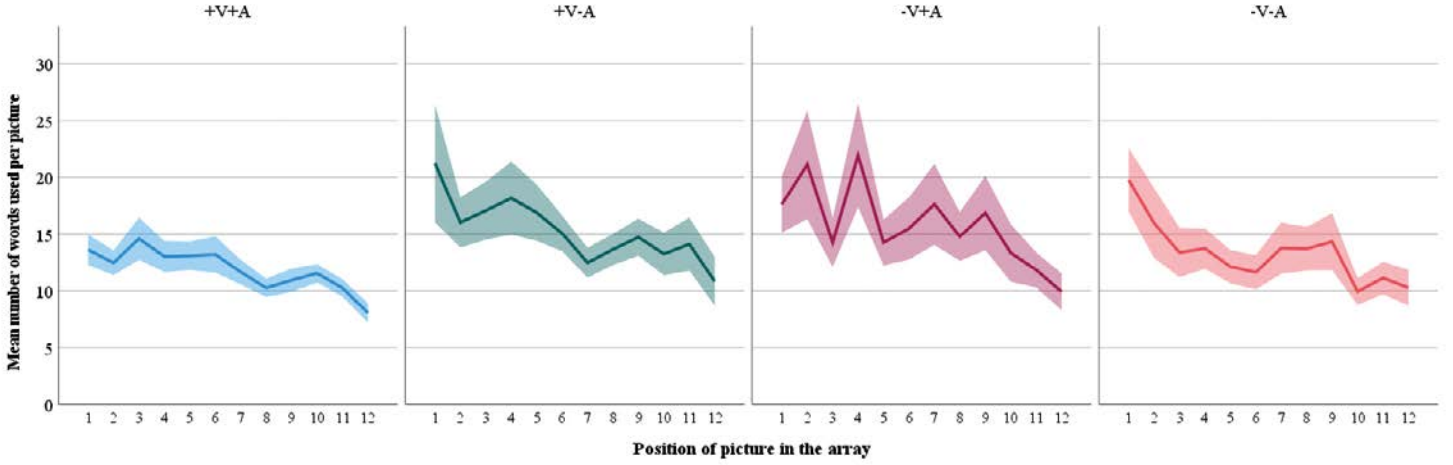
4.3.3.4. Turns per trial

To predict the number of turns used per trial, we included the fixed effect for trial and random intercepts for unique picture and participant. As can be seen in Figure 8, directors used most speaking turns in trial 1; a consecutive decline in turn taking over trials was observed, again, with the largest decline at the beginning of the array of trials.

As can be seen in Figure 9, the consecutive decline in turns taken over trials was found for all picture categories, with the largest decline at the beginning of the array of trials, especially in the -valence/+arousal category. No significant effect was found for valence, $\beta = .06$, $p = .261$, indicating that frequency of turn taking did not depend on the positive or negative valence of the affective picture category. A significant effect was found for arousal, $\beta = .18$, $p < .001$, indicating that directors used more turns per trial to describe the highly arousing, compared to the lowly arousing pictures. The interaction was significant as well, $\beta = -.36$, $p < .001$. Based on our inspection of Figure 9, we think that this can be explained by the steep decline in turn taking by directors in the -valence/+arousal category, compared to the other categories.

Figure 5

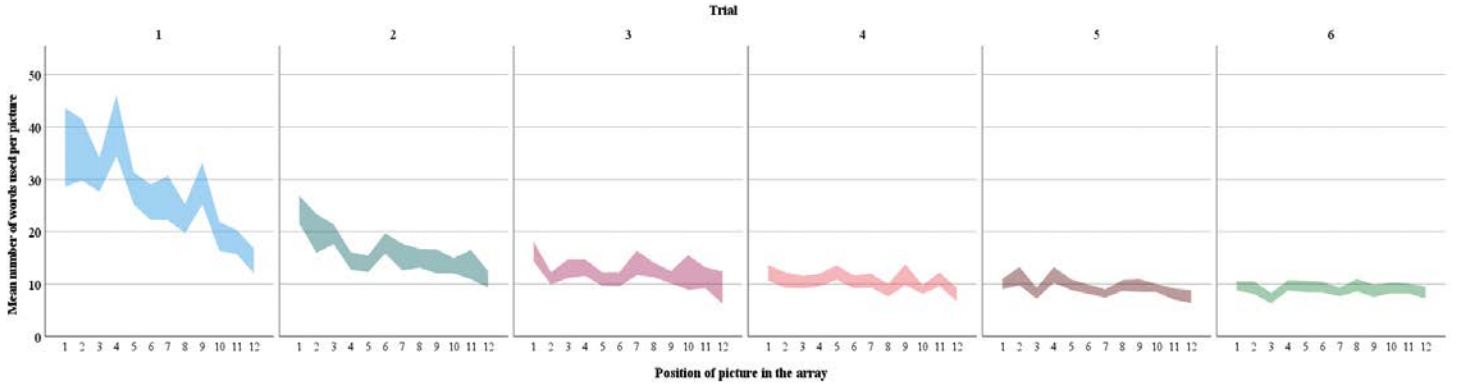
Mean number of words the director used to describe one picture (y-axis), per position of the picture in the array (x-axis), sorted by affective picture category (panels)



Note. Bands represent 95% confidence intervals.

Figure 6

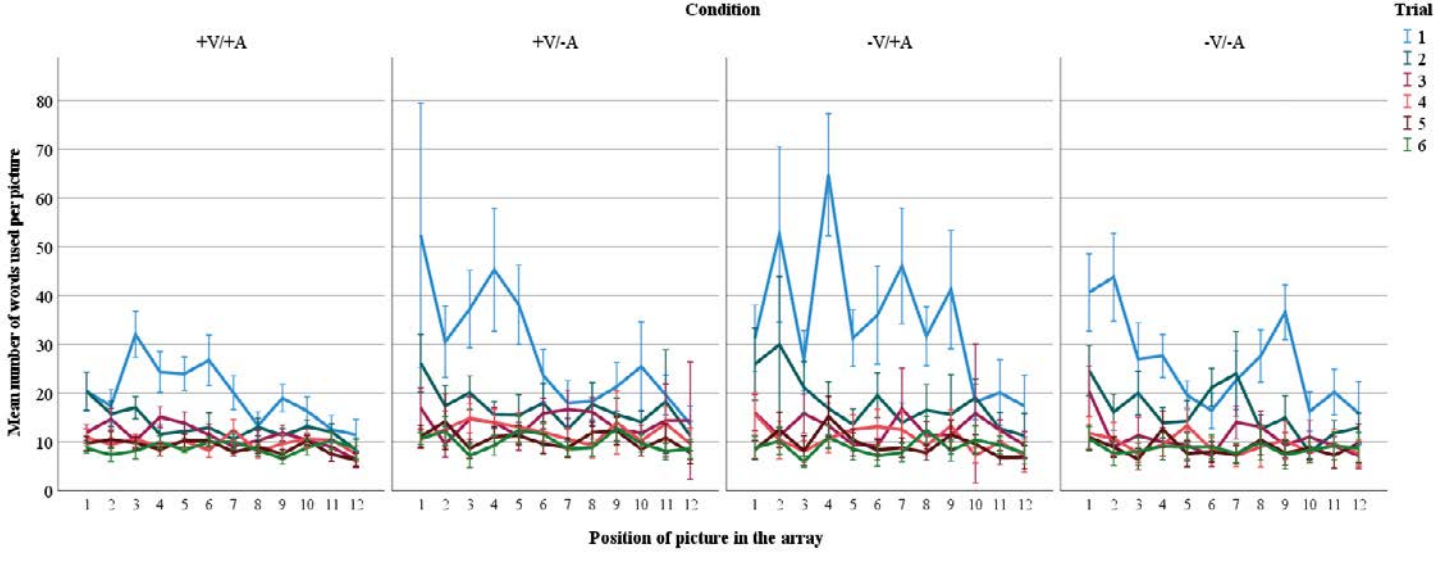
Mean number of words the director used to describe one picture (y-axis), per position of the picture in the array (x-axis), sorted by trial number (panels)



Note. Bands represent 95% confidence intervals.

Figure 7

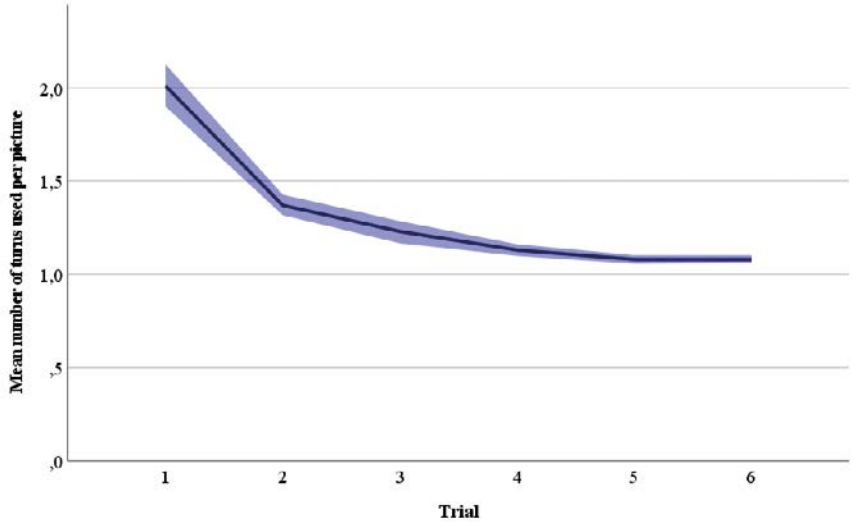
Mean number of words the director used to describe one picture (y-axis), per position of the picture in the array (x-axis), sorted by affective picture category (panels). Lines represent the trial number



Note. Error bars represent 95% confidence intervals.

Figure 8

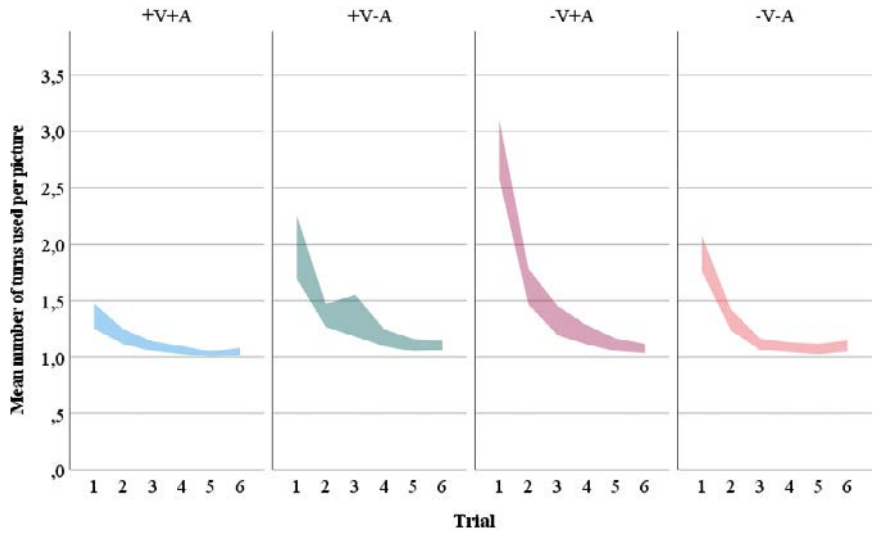
Mean number of speech turns the director used to describe one picture (y-axis), per trial (x-axis)



Note. Band represents 95% confidence interval.

Figure 9

Mean number of speech turns the director used to describe one picture (y-axis), per trial (x-axis), sorted by affective picture category (panels)



Note. Bands represent confidence intervals.

4.4. Discussion

The purpose of this study was to replicate C&WG (1986) findings that interlocutors repeatedly discussing identical abstract figures will require less words and speech turns over time to correctly identify them, a phenomenon attributed to conceptual pact formation (Brennan & Clark, 1996). In the present study, we extended their original referential communication task by using more naturalistic, and affectively-laden (IAPS) pictures, either pleasant or unpleasant, and high or low in arousal, wondering whether conversing about affective pictures would influence affective state, as well as the creation of conceptual pacts.

Our first hypothesis was not confirmed: in general, our affective pictures did not induce the expected levels of pleasantness and arousal. Answering our first research question, we found that although the differences between directors and matchers were small, after exposure to the highly arousing pictures, directors, but not matchers, reported substantially enhanced levels of arousal. Using linear mixed models, we replicated C&WG's findings, confirming the rest of our hypotheses: over time, directors used fewer words to describe the pictures over trials (H2), fewer words to describe pictures that were positioned later, compared to earlier in an array (within-trial), as well as the interaction (between-trial; H3). Additionally, directors used fewer speaking turns to describe the pictures over trials (H4). These results support the collaborative model.

Given that we were unable to find support for the effectivity of the affect induction, instead of investigating the influence of affective state of the directors on word use and turn taking, we investigated whether the different affective picture categories, represented by their dimensions high or low in levels of valence and arousal, improved our models. Answering our second research question, we found that the strength of the decline in word use and turn taking was influenced by the affective picture category. Within the arrays of trials, from picture 1 to picture 12, directors describing jumping individuals in the +valence/+arousal category showed the strongest decline in word use. However, when we also took the number of trial (trial 1 to trial 6) into consideration, examining the events occurring between trials as well, the most prominent decline in word use was found in the +valence/-arousal category. The frequency of turn taking showed the strongest decrease for directors describing spiders in the -valence/+arousal category. Finally, the decline in number of words used to describe pictures over trials seemed to be independent of the affective picture category.

4.4.1. Affect induction

Generally, participants reported heightened levels of pleasantness after engaging in the referential communication task, regardless of the content of pictures. We conjecture this might be due to several reasons. First and foremost, the task was fun: most participants displayed joy while engaging in the task. Second, the participants were already in high spirits before exposure to the pictures took place; a possible ceiling effect might explain

pictures. The combination of highly arousing pictures, actively engaging in a task and (perceived) high responsibility to perform well, might have resulted in a substantial increase in arousal.

In general, the differences between directors and matchers were modest. However, when observing individual scores, larger difference scores were found for directors than matchers, with more in- and decreases in affective state of >1 (as measured by difference scores in valence and arousal) for directors (45 cases) than matchers (24 cases), indicating that affect induction indeed had a stronger effect for directors than for matchers.

4.4.3. Conceptual pact formation

As predicted, the number of words used by the director to describe pictures declined between and within trials, and the number of speaking turns declined between trials. The general patterns are highly similar to C&WG (1986) and in line with other, previous findings concerning conceptual pact formation in referential communication tasks (e.g., Garrod & Pickering, 2004; Hupet et al., 1991). An additional benefit of the study is the external validity of these results, generalizing the collaborative model, given our choice of picture sets. To our knowledge, most referential communication tasks using pictures to study conceptual pact formation, used Tangram figures (e.g., Arbuckle et al., 2000; Branigan et al., 2011; Brannan & Clark, 1996; Hupet et al., 1991; Yoon & Brown-Schmidt, 2019) or depictions of everyday objects (e.g., fruit, house, shoe) against a neutral background (e.g., Brannan & Clark, 1996; Carbary & Tanenhaus, 2011; Yoon & Brown-Schmidt, 2019). Whereas our study included two picture categories more or less similar to the latter category (kittens and spiders), we also included two categories containing content that is substantially more complex and naturalistic (jumping individuals and funerals). To our knowledge, this not very common (although one exception might be found in Bortfeld et al., 2001, who compared Tangram figures to pictures of children). Given that our results show that interlocutors tend to form conceptual pacts across all four different categories of pictures, describing and matching relatively simple pictures high in similarity, as well as sets of more discriminatory, complex pictures, this allows the conclusion that our results support the generalizability of the collaborative model.

4.4.4. Adaptation in conversation and affective content

Although the different categories of pictures did not predict the changes in affective state, they did influence the number of words used, and turns taken, over time. The largest effects were seemed to be found in two categories, with sharper declines in word use for pictures over time (both between and within trials) for +valence/-arousal, and sharper decline in turns taken over trial for the -valence/+arousal condition. One possible reason might be that, as already mentioned above, the content of these pictures was relatively uniform and plain, featuring one or two animals (kittens or spiders) against a neutral background (e.g., white wall, grass). It might have been more challenging to distinguish the unique pictures

why the pleasant pictures failed to enhance levels of pleasant affect in participants (Uhrig et al., 2016; Westermann et al., 1996). Third, because the dyads performed well (less than 2% of all trials included errors), they received predominantly positive feedback after each trial. Some participants expressed delight when receiving these affirmations; this positive feeling might have contributed to the overall increase in self-reported levels of positive affect.

An alternative, fourth explanation might be that, although the pre-testing of affective pictures suggested otherwise, the content of the pictures was not sufficient to induce the targeted affective states successfully, or a possible habituation hindered the effect (although persistent effects of prolonged exposure to negative IAPS pictures have been found; Smith et al., 2005). Lastly, another possible explanation might be that the SAM was not the right instrument for our participants to measure their self-assessed affective state. Although the SAM is a widely used and validated tool (e.g., Bradley & Lang, 1994; Bynion & Feldner, 2017) with a high test-retest reliability (e.g., Kanske & Kotz, 2010), a number of authors have acknowledged its possible limitations as well. Some scholars question the intuitiveness of the scale, reporting that participants misinterpreted or misunderstood the depicted affect dimensions (especially arousal; Hayashi et al., 2016), or frequently asked for clarification, even after they were informed what the manikins represented (Betella & Verschure, 2016). While misinterpretation is always a possibility, we aimed to minimize the odds by explicitly giving participants instructions to indicate the degree of arousal, with the left manikin representing feeling 'very relaxed' and the right manikin 'very excited'. Given that we did not receive any questions or comments about this, we assumed that participants understood the task correctly.

However, as Broekens and Brinkman (2013) argue, the SAM represents only very clearly expressed emotions, no 'spontaneous, more subtle and varied expressions' (p. 2). In this line, one could conceivably argue that the SAM might not have been sensitive enough as a self-report affective state instrument, because individuals probably did not relate to the stereotypical, strongly expressed affective states depicted in the manikins. Given that we found only a subtle effect of affective pictures on affect induction, it might be the case that affect induction was successful, but too subtle to be detected by the SAM.

4.4.2. Affective states of directors vs. matchers

In general, directors, compared to matchers, reported more pleasant affect and higher levels of arousal after the task. Given that directors played a more active role by describing the pictures, opposed to the more passive role of the matchers of ordering the pictures correctly, we conjecture that the role of director might have been more interesting, engaging and activating, resulting in heightened levels of both pleasantness and arousal. Additionally, one could argue that within a dyad, the role of director is more cognitively demanding than the role of matcher, as they typically need more time to plan their utterances, taking the initiative to describe the content of the pictures, and compare them to the other pictures (Bortfeld et al., 2001). Given this responsibility, we can presume they spend more time looking (intensely) at the pictures, resulting in enhanced effectivity of the highly arousing

from each other, compared to the pictures in the other two categories which featured more complex content, e.g., one or more individuals with clear distinctive characteristics (e.g., having brown hair, wearing white shorts) and set against varied backgrounds (e.g., featuring a lighthouse or snowy mountains). This is in line with Hupet et al. (1991), who studied collaboration between dyads engaging in a referential communication task, finding that interlocutors needed more words and turns per figure to reach mutual agreement for the Tangram figures that were harder to differentiate from each other, compared to the figures that were more distinct. In a similar vein, Bortfeld et al. (2001) found that after a referential communication task, interlocutors reported that they found it easier to describe and match pictures of children, compared to Tangram figures, arguing that 'unusual objects need more planning than common objects' (p. 134-135). Indeed, we found 37 cases in which directors in the +valence/-arousal and –valence/+arousal categories described the same picture for a second time within the array of a trial (excluding the cases in which this was due to the dyad not understanding the task instructions correctly, therefore restarting the trial). This only happened once for the -valence/-arousal pictures, and never for the +valence/+arousal pictures.

The number of words used by the director to describe pictures over trials was strongly predicted by trial number. Adding valence and arousal dimensions did not improve the model, indicating that the decline in word use was irrespective of picture category. We conjecture our model was already saturated.

4.4.5. Conclusion

In this chapter, we replicated the referential communication task by C&WG, finding support for the collaborative model by using affectively-laden pictures, while simultaneously studying the effect of these affective pictures on affective state in the interlocutors. While we replicated C&WG, no effect of affectively-laden pictures on affective state could be detected, which, as explained above, might be due to several reasons. This study contributed to the sparse literature on the relationship between affective content and spoken language production in naturalistic settings as interactions.

APPENDIX

Example 1

turn 1 Director: For two, shoes are flying into the air. <laughter>
Matcher: <laughter>
Director: The sixth one is a <uh> woman in a (continues)

Example 2

turn 1 Director: Boy in orange shirt, center on one.
Matcher: Have to have a look, have to
turn 2 Director: Oh sure
Matcher: put down so I can see them all. Oh yes, with those three kids, okay.
turn 3 Director: Yes.

5

An intimate conversation: The influence of affective state on (im)polite request formulations of neurotypical and autistic speakers in dyads

This chapter is based on:
Out, C., Goudbeek, M., & Krahmer, E. (2021). *An intimate conversation: The influence of affective state on (im)polite request formulations of neurotypical and autistic speakers in dyads.* [Manuscript under revision].

ABSTRACT

Affective state has a crucial, yet unclear relationship with verbal communication. For example, when people are presented with hypothetical interpersonal scenarios in which they need to formulate a request (e.g., asking money back that an acquaintance had previously borrowed), individuals in a negative (compared to positive) affective state tend to formulate these requests more politely, based on six different request characteristics, including politeness (Forgas, 1999a). In the current study, we aimed to replicate and extend this finding in a more realistic setting, in both neurotypical (Study 1, *N* = 166) as well as a small group of autistic individuals (Study 2, *N* = 29). After an affect induction procedure, dyads asked each other about their personal experience with sensitive affairs such as bullying. Transcripts of the recorded conversations were scored on request characteristics related to politeness (Forgas, 1999a) by both neurotypical and autistic raters. Although affect induction was successful in both studies, surprisingly, only an effect of affective state on politeness was found for autistic but not neurotypical speakers, with autistic speakers in a neutral or negative state formulating more elaborate, hedged, indirect and complex requests, compared to their peers in a positive affective state. Furthermore, complementary ratings on politeness strategies (Brown & Levinson, 1987) by two neurotypical raters indicated no effect of affective state for either group of speakers. Neurotypical and autistic speakers, regardless of affective state, did not produce substantially different requests, regarding the degree of politeness. Although we did not replicate the original findings by Forgas in our neurotypical sample, our findings show a limited effect of affect on request characteristics formed in spontaneous, free conversation in autistic speakers. This suggests that the influence of affective state on verbal communication might extend beyond the neurotypical population.

5.1. Introduction

When you ask another person to do something, you usually will need to formulate your request in a specific manner, in order for the other person to grant you the favor. On the one hand, you want to be sufficiently direct to achieve your goal; on the other hand, you want to be sufficiently polite to not give offence. In other words, you need to be careful in selecting the appropriate words to create the right balance between politeness and directness. Say, you want to ask your friend Mark to help you move on Sunday, knowing that he would rather not because he wants to play a soccer match that morning. A (very) polite request would be: 'I know you would really like to play your soccer match on Sunday morning, but maybe you could help me move for just two hours? That would be amazing.' This request is optimistic, friendly of tone and acknowledges Mark's needs. Chances are higher that this request will be granted than if you are more impolite, demanding and direct: 'I will move this Sunday morning, and I need you to be there to help me.' However, being too polite might not be successful either, e.g., when you merely implicate your needs, instead of asking Mark actually for help: 'This Sunday morning is going to be a rough, exhausting morning.' As illustrated by the example above, formulating a request in order to create a maximal chance of compliance by the other party, without offending them, must be handled with finesse. This process can be influenced by various factors, including (most relevant for the current chapter) the affective state of the speaker. Like many demanding cognitive tasks involved in interpersonal verbal communication which can be influenced by affect (e.g., processing style, Beukeboom & Semin, 2006; perspective taking, Converse et al., 2008), politeness, too, can be promoted or hampered by the affective state of the speaker with happier speakers (or the person formulating the request) being less polite (Forgas, 1999a). In the current studies, we aimed to replicate and extend Forgas (1999a), investigating the relationship between affective state and politeness.

Forgas (1999a) conducted three studies exploring the effect of affective state (more specifically, positive and negative mood) on verbal communication, focusing mainly on politeness in request formulation. In his Experiment I and II, participants were first asked to recall a situation in which they felt very happy (or sad), and write about it in 10-12 minutes. Then, they were presented with a vignette describing two situations in which they were to request something from an acquaintance or friend: an easy situation (asking back 20 dollars that an acquaintance had previously borrowed) and a more difficult situation (asking for ketchup with the meal that their friend, proud of their cooking, made for them). In Experiment I, the participants were presented with 5 response options per situation, containing a request ranging from very polite and indirect (the easy situation: 'I wonder if I have any money on me today?'; the difficult situation: 'This meal might need something…') to relatively impolite and very direct ('I want you to repay my twenty dollars'; 'I want some ketchup'), and were asked to select the options they preferred to use. In Experiment II, participants were asked to write down how they would phrase their request in their own words. Experiment III extended of the first two studies, with a neutral mood condition added and a different affect induction technique (affective videos). In Experiment III, participants

were presented with descriptions of 3 easy and 3 difficult situations, and, as in Experiment II, wrote down for each situation how they would phrase their requests in that particular situation. Two individuals, blind to the mood conditions, used 7-point bipolar scales to rate the requests of Experiment II and III on the degree of politeness, directness, friendliness, among other request characteristics.

Results indicated that sad individuals selected (Experiment I) and formulated (Experiment II and III) more polite requests than happy individuals, indicating that sad individuals, compared to happy individuals, prefer to use more polite requests. This was in line with Forgas' prediction, given that his earlier work had shown that individuals in a positive mood are more optimistic and confident than individuals in a negative mood (Forgas, 1995). Experiment I, II and III showed that this effect was stronger in more difficult situations, compared to relatively easy situations. In Experiment III, politeness and directness scores of individuals in the neutral condition generally fell between those of the happy and sad participants. Overall, requests were more polite and less direct in difficult situations than in easy situations, suggesting that more demanding tasks are likely to trigger more elaborate processing strategies in individuals. This finding is in line with the Affect Infusion Model (Forgas, 1995), which states that mood and affect influence information processing, and therefore our (verbal) actions. Articulating requests in more difficult (compared to easier) situations demands relatively more elaborate processing, and therefore allows more space for affectively primed thoughts. This implies that the type of situation is relevant for the effect of affective state on politeness.

Forgas (1999a) studied the effects of affective state on politeness in a somewhat artificial setting. The participants did not interact with another individual, raising the question whether the findings are generalizable to real-life interactions. Therefore, in the present study, verbal politeness is studied in a more naturalistic setting, in which participants engage in a conversation. To our knowledge, research on verbal politeness in an experimental, face-to-face setting with another human conversation partner is scarce, although some notable examples do occur. For example, in a different study by Forgas (1999b), after participants (N = 78) were exposed to a mood induction procedure (positive, neutral or negative, n = 26 per condition), the experiment leader asked the participant to go to the next office, and ask the person there to give them the stimulus file from the filing cabinet, that the experiment leader needed to set up the next part of the experiment. The results were in line with earlier studies: sad individuals used more polite and indirect requests for the stimulus file compared to individuals in a neutral or happy mood, and individuals in a happy mood used more impolite and direct requests compared to individuals in a neutral mood.

In a more recent study, Morse and Afifi (2015) also conducted a more ecologically valid study on verbal politeness, studying the language use of individuals aiming to persuade their opponent (a confederate of the study) in a debate. Participants were exposed to a positive, negative or no mood induction procedure (n = 33 per condition); manipulation checks indicated that mood induction was generally successful, i.e., participants in the positive condition reported more positive mood than those in the negative condition,

but not more than those in the neutral condition. Participants in the positive condition, compared to the negative or control condition, displayed more verbal impoliteness, based on, among other things, manual ratings of verbal aggressiveness (e.g., insulting the other person) and statements preventing the other person from having a choice (e.g., assuming their opponent agreed with them).

The findings discussed above indicate that individuals in a negative affective state, compared to a positive affective state, tend to be more polite. This is in line with previous research, giving suggestive evidence that affective state might influence the attentional scope and perspective taking of an individual. Specifically, speakers in a positive mood tend to take a more egocentric perspective, compared to speakers in a negative mood, who appear to be less egocentric (Converse et al., 2008; Kempe et al., 2013). In a similar vein, studies on the influence of affect on attention indicate that positive emotions broaden attention (e.g., Fredrickson, 2001; Fredrickson & Branigan, 2005; Rowe et al., 2007), whereas negative mood narrows the attentional scope (e.g., Gasper & Clore, 2002). In a conversation setting, these results might translate into negative speakers being more focused on their conversation partner (see also Koch et al., 2013), and therefore, formulating more polite requests than speakers in a positive affective state, who might be more focused on themselves, i.e., achieving their goal. Additionally, in line with Forgas (1999a), we expected the influence of affective state on politeness to be more pronounced in more difficult situations. Therefore, we aimed to create similarly difficult situations as well, in which participants wanted to request something from another individual. Thus, in the current study, participants were instructed to ask another participant to share their experiences with sensitive affairs such as financial debts, sexting, or experiences with bullying.

Regardless of affective state, what individuals perceive to be (im)polite can vary greatly from person to person (e.g., Haugh & Chang, 2019). For example, it might depend on someone's region (Schneider & Placencia, 2017), culture (e.g., Al-Duleimi et al., 2016), generation (e.g., Bella & Ogiermann, 2019) and gender (e.g., Mills, 2003). Verbal politeness behavior might also depend on an individual's social interaction skills, e.g., adhering to implicit social norms as being sufficiently polite. Persistent difficulties in social communication, as well as restrictive, repetitive behaviors, is one of the two core characteristics of an autism spectrum disorder (ASD; American Psychiatric Association, 2013). Autistic individuals[2] are known to struggle with implementing and understanding social communication norms (e.g., Cummins et al., 2020; Mathersul et al., 2013), which are recognizable in their conversational speech (Paul et al., 2009). For example, they might use overly formal speech, abruptly switch topics, or barely respond to their conversation partner (De Villiers et al., 2007). As a result, autistic individuals are sometimes considered to communicate in a socially unaccepted or 'inappropriate' manner (Thomas & Mambara,

2 In this chapter, we use 'identity-first' (e.g., 'autistic individual') instead of 'person-first' (e.g., 'individual with ASD') language. This decision is based on the findings of a recent study on (British) autistic individuals, showing that the majority of the community preferred to use identity-first language when communicating about autism (e.g., 'autistic person'; Kenny et al., 2016).

2020). In the present study, we ask whether autistic individuals also struggle to use another aspect of social communication, namely, politeness. While anecdotal evidence suggests that autistic individuals have difficulties with being polite, understanding politeness, or both (Belek, 2018; Manett, 2020; Sterponi, 2004), experimental research shows inconsistent results with respect to this matter: some published studies are in line with these anecdotes (e.g., Zalla et al., 2014), while other studies do not find substantial differences in verbal politeness between autistic individuals and non-autistic individuals, or neurotypicals (Volden & Sorenson 2009; Yang et al., 2020).

In a similar vein, although currently not included as a core symptom of ASD (American Psychiatric Association, 2013), intra- and interpersonal affective impairments, or atypicalities, are common in autistic individuals[3], for example difficulties in emotion regulation (e.g., Cai et al., 2018), in recognizing emotions in other individuals (e.g., Velikonja et al., 2019), but possibly also *experiencing* emotions differently than neurotypicals (Bölte et al., 2008). Therefore, it is possible that affective state might have a different influence on the cognitive processes involved in politeness behavior (e.g., attentional scope, egocentricity) of autistic individuals, compared to neurotypicals. Hence, in this chapter we not only intend to replicate the study by Forgas (1999a) with neurotypicals, but also to explore how autistic individuals fare in this set-up. We tested this in same-diagnosis (both neurotypical or both autistic) instead of mixed-diagnosis (neurotypical and autistic) dyads, for the following reason. Recent studies show that neurotypicals tend to form negative first impressions of autistic individuals, subsequently reporting low interest to socially engage with them in the future (e.g., Grossman et al., 2019; Sasson et al., 2017). Given that we thought that this would be an unfortunate situation, and the fact that many studies report that autistic individuals tend to be interested to socially interact with both neurotypicals *and* other autistic people (e.g., DeBrabander et al., 2019; Morrison et al., 2020), we made this decision.

The goal of this study is threefold. First, in Study 1, we (conceptually) replicate Forgas (1999a) in an interactive and therefore, more ecologically valid set-up, by investigating the effect of (positive, neutral or negative) affective state on politeness in asking verbal requests in a conversation between two neurotypicals.

Second, in addition to measuring politeness using Forgas' original scales to assess, among other things, the degree of politeness of the requests, we explored whether affective state of the speaker also influenced which type of politeness they use when formulating a request. Brown and Levinson (1987), inspired by Goffman (1967), describe four types of politeness strategies: positive politeness, negative politeness, bald on-record, and the off-record strategy. Given that Brown and Levinson's politeness theory is, by many scholars, considered to be one of the most influential politeness models to date (e.g., Feng, 2015; Leech, 2005; albeit highly debated, see, e.g., Al-Hindawi & Alkhazaali, 2016; Mills, 2003), we also analyze the collected requests in terms of their politeness strategies.

---

3 However, research on the emotional deficits within the autistic population shows inconsistent results with respect to global emotional difficulties (Bird & Cook, 2013).

Third, because we aim to extend the findings of Forgas (1999a) by studying politeness in a non-student population, while simultaneously contributing to the sparse literature on autism and politeness, we replicated Forgas (1999a) in a sample of autistic individuals in the additional, exploratory Study 2.

5.1.1. Hypothesis and research questions

With respect to Study 1, we have the following hypothesis and research question:

(H1) Following Forgas (1999a), we hypothesize that neurotypicals in a positive affective state will be more impolite, compared to neurotypicals in a negative affective state, who will be more polite. Neurotypicals in a neutral affective state will fall in between.
(RQ1) In addition to H1, we will explore whether there is a relation between affective state and the implementation of different politeness strategies (Brown & Levinson, 1987).

Regarding Study 2, we have the following research questions:

(RQ2) Does affective state have an influence on politeness behavior in autistic individuals, as scored by Forgas' request characteristics related to politeness?
(RQ3) Is there a relation between affective state of autistic speakers and the implementation of different politeness strategies?

Finally, we will compare Study 1 and Study 2:

(RQ4) Regardless of affective state, are autistic individuals more, less, or equally polite as neurotypicals?

We preregistered our methods and hypotheses at the Open Science Foundation (OSF; https://osf.io/z2arj). As it was difficult to recruit autistic participants, we did not reach our preregistered desired sample size for this group (N = 66). As a result, the samples of Study 1 and Study 2 were rather unequally sized (N = 166 neurotypicals; N = 29 autistic individuals), making us hesitant to statistically compare the two groups. Therefore, we explored the results of Study 1 and Study 2 separately when testing our hypothesis and research questions.

5.2. Study 1. Affective state and politeness in neurotypicals

5.2.1. Method

5.2.1.1. Design

The study had a between subjects design with 3 conditions (positive, neutral and negative), with the six request characteristics scales by Forgas (polite-impolite, direct-indirect,

friendly-unfriendly, elaborate-simple, hedging-not hedging, and simple-complex) and the four Brown and Levinson's politeness strategies (bald on record, positive politeness, negative politeness and off-record) as dependent variables.

5.2.1.2. Participants and data collection

In total, 166 neurotypical (NT) participants (aged 17-40, M_{age} = 21.67 years, SD_{age} = 2.89 years) from a Dutch university participated in pairs (forming dyads) in the experiment for course credit. The sample included 59 men, 106 women, and 1 non-binary individual. Students were randomly assigned to a conversation partner, forming a cross-gender dyad (36 dyads), a dyad consisting of two women (35 dyads) or two men (11 dyads). One dyad consisted of one man and one non-binary individual. Each dyad was assigned to one of three conditions: positive, neutral or negative. Data was collected in two phases. The first phase was carried out by a research assistant, in exchange for course credit, testing participants randomly assigned to the positive (n = 48) or negative affect condition (n = 48). The second phase was carried out by the first author of the present chapter (CO), testing participants in the neutral condition (n = 70).

The inclusion criterion was the ability to speak Dutch fluently; the exclusion criterion was having an ASD. After completion of the experiment, one participant reported that she was, in fact, autistic. Given the size of the sample, this one participant has a negligible effect on the results, hence we decided to leave her in the sample.

5.2.1.3. Consent

All participants consented to have their anonymized audio recordings, as well as their responses to the questionnaires, used in analyses and publications. Two participants did not grant their permission for us to quote their utterances, i.e., use (fragments of) the audio recording of their speech, both in original and transcribed form, in a presentation or a publication. All procedures performed were in accordance with the ethical standards of the institutional research committee, the Research Ethics and Data Management Committee of Tilburg School of Humanities and Digital Sciences, Tilburg University (REC#2018/42). All participants were clearly informed of their right to stop at any time without consequences, and gave written informed consent in accordance with the Declaration of Helsinki (1964) and its later amendments or comparable ethical standards.

5.2.1.4. Materials of experiment

5.2.1.4.1. Autobiographical memory task

Following Forgas (1999a), we used an autobiographical memory task to induce positive and negative affective states. In this task, participants reimagine a specific situation in their social life which made them very happy (positive affective state) or very sad (negative affective state). Participants are given 10 minutes to imagine (as lively as possible) and write down their chosen positive or negative experience. They are asked to focus on the details of the situation, and encouraged to re-experience the same feelings they felt in

their chosen situation. Given that the original task did not include a neutral condition, we created a neutral version of the autobiographical memory task ourselves, asking participants to imagine and write down the specific details of a regular day in their lives (including their morning routine, school or work, and evening activities), which is a well-known and common method to induce neutral mood (e.g., Young et al., 2011). To make sure that participants spend 10 minutes on this task, the screen was time locked for this duration; they were unable to proceed with the study until 10 minutes had passed. For the exact autobiographical memory task instructions, see Supplementary Table 1.

5.2.1.4.2. Affect questionnaire

In line with Forgas (1999a), participants indicated on four 7-point Likert scales to which extent describing their memory made them feel sad (1) or happy (7); active (1) or not active (7); bad (1) or good (7); and relaxed (1) or tense (7). Participants were instructed to choose a number per scale; the closer the numbers were to the ends of the scale, the stronger they match the feeling described by the word in question. To conceal the purpose of the affect manipulation, and following Forgas (1999a), four fillers, related to the affect induction task, accompanied the affect questionnaire (e.g., indicate to which extent the described situation was nice or troublesome). The scores on the filler questions are not further analyzed.

Before our analyses, we re-coded our sad-happy and bad-good scales to happy (1) and sad (7), and good (1) and bad (7). This was done to facilitate comparison with Forgas' scales of which some were reverse coded.

5.2.1.4.3. AQ-Short

Given that we aimed to study politeness in autistic individuals as well (Study 2), we checked whether our sample of (presumably) neurotypicals included individuals scoring high on autistic traits. Participants filled in the 28-item self-report questionnaire AQ-Short (Hoekstra et al., 2011), an abridged, validated version of the Autism Quotient (AQ; Baron-Cohen et al., 2001). More information on the AQ-Short can be found in the Appendix.

5.2.1.4.4. Booklet for conversation

Before the conversation, the dyad was offered a booklet, consisting of 12 pages. The front page requested the participant to turn the page when their time had started. The booklet featured 6 pre-set conversation topics: debts, bullying, sexting, cheating (i.e., infidelity), excessive use of alcohol and sexual intimidation. For each topic, the participant was instructed to ask their conversation partner about his or her experiences with one of the six topics (p. 2, 4, 6, 8, 10 and 12), followed by the instruction to hand the booklet to their conversation partner, and not turn the page themselves (p. 3, 5, 7, 9 and 11).

5.2.1.4.5. Familiarity question

Given that individuals tend to be more polite towards persons they do not know (Brown & Levinson, 1987, hereafter B&L, 1987), participants indicated on a 7-point Likert scale to which extent they knew their conversation partner (before their interaction), with a 1

indicating 'not at all' and a 7 'very well'. This question was asked after the conversation took place, to make sure that this question would not trigger (additional) speculation about the purpose of the conversation in specific and research in general.

5.2.1.4.6. Opinion questions

Participants were asked to give their opinion about the 3 topics they asked their conversation partner about. On three 7-point Likert scales, they indicated to which extent they felt uneasy (1) or comfortable (7) with the topic; how personal (1) or impersonal (7), and how difficult (1) or easy (7) the topic was to them.

5.2.1.5. Procedure

Pairs of participants were welcomed to the laboratory. The first participant entering the room was guided to a separate booth inside the laboratory and seated behind a desktop computer; the second participant was seated behind a desk with a laptop. Both participants were asked to read and sign the informed consent form, asked to report their age, gender, and, in phase two, (possible) psychiatric diagnoses. They then engaged in the autobiographical memory task, filled in the affect questionnaire and the AQ-Short.

After these solitary tasks, the experiment leader explained the procedure of the conversation that the participants would engage in. First, the dyad was presented with the booklet. They were told that they were going to get to know each other better, by asking each other questions about their personal experiences with 6 pre-set topics, which were featured in the booklet. Each participant asked questions about 3 of the 6 topics, taking turns per topic, thus the role of requesting switches between the participants for each topic. They were instructed to turn around one page at the time, to make sure that the content of the topics was unknown to both. Lastly, they were given a stopwatch, set to 1.5 minutes, and instructed to press the button each time they proceeded to the next topic. After each countdown, the alarm went off, which indicated that the participant should hand the booklet over to their partner, to have them ask the question about the next topic. This continued until the booklet was finished. The experiment leader asked if the instructions were clear, answering any questions that may arise. When both participants understood the task, the experiment leader informed them that audio recording would take place, and asked them to knock on the door when they had finished the task. The experiment leader pressed 'play' on the audio recorder and left the room. When one of the participants knocked, the experiment leader re-entered the room, pressed 'pause' on the audio recorder and asked the participants kindly to proceed with the last, individual part of the experiment, behind their computer or laptop. They answered the familiarity question, filled in the affect questionnaire again (without fillers), and answered the opinion questions. After this, they were informed that they had completed the study; they were debriefed and thanked for their participation.

5.2.1.6. Rating materials

5.2.1.6.1. Forgas' request characteristics

Requests were rated on six 7-point bipolar scales, derived from Forgas (1999a): polite (1) – impolite (7), direct (1) – indirect (7), friendly (1) - unfriendly (7), elaborate (1) – simple (7), hedging (1) - not hedging (7), and simple (1) – complex (7).

In the original article by Forgas (1999a), politeness and directness are used more or less interchangeably, based on the rationale that they are opposite and strongly related constructs (see Forgas, 1999a, p. 929). Although the work of Brown and Levinson (1987) was mentioned, no (clear) definition was provided for the concept of politeness. Directness 'referred to the degree of immediacy in the syntactic and semantic formulation of the request' (p. 928-929). The hedging and friendliness scales were included without providing a clear rationale.

Given that Forgas (1999a) aimed to not only study the influence of mood on politeness, but *also* on the degree of elaboration of the formulated requests, the scales measuring elaboration and complexity were included in the original study. Although we were not necessarily interested in these aspects of the requests, given that we aim to replicate the work by Forgas, all six scales were included.

5.2.1.6.2. Brown and Levinson's politeness strategies

Requests were subjected to the classification into the four different politeness strategies (B&L, 1987):

Bald on-record. Request was rated as bald on-record when the message was direct, clear, and to the point. This example, and the following examples are taken from the interactions collected in this study: 'Do you have experience with alcohol abuse?'.

Positive politeness. Request was rated as positive politeness when the speaker aimed to support and protect the listener's positive self-image, or showed that they cared about the listener's need to be recognized, understood and liked. For example, they approached the listener in a positive manner, using humor and asking their opinion: 'Excessive use of alcohol, yes you are doing that! (laughter) Drinking alcohol, yes, what is your opinion about that?'.

Negative politeness. Request was rated as negative politeness when the speaker aims to avoid imposition on the listener, showing respect for the listener's need for autonomy. For example, the speaker waters down their request, or uses words as 'maybe' and 'perhaps', suggesting that the listener does not have to answer the question if they do not want to. For example: 'What are your experiences with uh, excessive alcohol consumption? Do you have experiences?'.

Off-record. Request was rated as off-record when the request could be interpreted in multiple ways. For example, the speaker tries to avoid the possible negative consequences of asking a question that might be too personal, therefore being ambiguous in their message. For example, when asking about excessive use of alcohol, they might ask 'Did

you celebrate carnival?', referring to a regional celebration involving, among other things, (excessive) drinking in pubs.

5.2.1.7. Data pre-processing (Study 1 and Study 2)

For reasons of clarity and economy, data pre-processing of both Study 1 *and* Study 2 will be discussed in the present section.

5.2.1.7.1. Requests

Recorded conversations were transcribed verbatim by CO and a research assistant. Excluded from the transcriptions were questions asked that were misinterpretations of the presented topic[4]. Included and annotated between angle brackets were half-words (<mic-> meaning 'microphone'), speech disfluencies (<uh>, <uhm>), corrections and stuttering (<intimida-> <uh> <in->), laughter (<laughter>), beeping sound of the stop watch (<beep>) and unintelligible speech <unintelligible>. In line with this annotation style, an example of two conversations (one example for each study) can be found in Supplementary Table 2.

For each topic discussed by the dyad, the first question asked by the speaker, as well as utterances preceding the question were selected and classified as a request (e.g., 'Excessive use of alcohol. Do you drink [alcohol]?'). In total, 585 requests were counted, including 498 by neurotypicals (Study 1) and 87 by autistic speakers (Study 2)[5].

Missing or inaccurate requests occurred in Study 1, due to technical error (4 requests), or the speaker not discussing (1 request) or misinterpreting the topic (2 requests); these 7 requests were removed from the dataset. Requests formulated by multiple (NT or autistic) speakers that were highly similar or identical in content (e.g., 'Uh, (do) you have experience with excessive use of alcohol?') were merged, resulting in a set of 535 requests, about excessive use of alcohol (90 requests), sexual intimidation (93 requests), bullying (85 requests), debt (88 requests), sexting (93 requests) and cheating (86 requests).

Aiming to create a final sample of requests by both NT and autistic speakers that could be rated on politeness, as well as further analyzed, the following steps were taken. First, our goal was to present randomized requests by speakers in each study, condition and conversation topic in a balanced manner. Therefore, we created 8 different versions, containing 24 or 23 unique requests, a number that was considered to be feasible to judge per individual rater. Second, all 84 requests by autistic speakers and 105 requests by NT speakers were selected and distributed over the versions. Requests by NT speakers were randomly selected, using a randomization formula in Excel. Details of the final sample can be found at OSF (Chapter 5, Additional materials, https://osf.io/bhqzu/).

---

4 Given that the Dutch word 'schulden' (financial debt) is almost a homonym for 'schuldig' (feeling guilty), in the first phase of data collection, participants in two dyads asked their conversation partner about the latter, instead of the first topic. To avoid this issue in the second phase of data collection, we added the word 'geld' (money) to the topic.

5 One participant who did not meet the inclusion criteria was mistakenly included in Study 2. Before our analyses took place, his data (requests, experimental data) were removed from our datasets.

5.2.1.7.2. Forgas' request characteristics coding

The final sample of 189 unique requests was rated by 17 NT students (14 women and 3 men; M_{age} = 21.94 years, SD_{age} = 2.16 years, range: 18-25 years) and 32 autistic individuals (14 women, 17 men, 1 non-binary individual; M_{age} = 43.00 years, SD_{age} = 11.47 years, range: 22-63 years). NT raters were recruited from the same Dutch university as NT participants, and compensated with course credit. ASD raters were recruited through national autism interest groups, a Dutch online message board for autistic individuals, LinkedIn, and word-of mouth. They were offered €5,- as compensation, with the added option to donate the fee to an autism research institute (Autism Research Centre). Both groups only rated the requests with the proviso they had not participated in the experimental study.

All raters were blind to the affect conditions and group of the speaker (NT or autistic). Raters were distributed randomly, until each version, or collection of unique requests, was rated independently. Each version was rated independently by 2 or 3 NT raters, and 4 autistic raters. In total, 401 requests were rated by NT raters (222 requests by neurotypicals; 179 requests by autistic individuals) and 756 by autistic raters (420 requests by neurotypicals; 336 requests by autistic individuals).

Autistic individuals are often studied from a neurotypical perspective (Heasman & Gillespie, 2019), which might give an incorrect representation of how autistic individuals experience, for example, politeness. In the current study, we also wanted to know what autistic individuals consider to be (im)polite, contributing to this gap in the literature, while simultaneously keeping in mind that spoken language of autistic individuals is, according to neurotypicals, regularly considered to be 'inappropriate' (see, e.g., Thomas & Mambara, 2020). Therefore, focusing on the verdict of peers, the 222 ratings of NT raters on requests by neurotypicals, and 336 ratings of autistic raters on requests by autistic individuals were selected for further analysis (Study 1 and Study 2, respectively).

Interrater agreement. Aiming to measure the degree of agreement among the raters, interrater reliability was measured. Krippendorffs's alpha (KALPHA) was calculated, using the 'KALPHA' macro (version 4.0) for SPSS (Hayes & Krippendorff, 2007). For NT raters, alphas ranged between 0.25 (simple-elaborate) and 0.44 (hedging-not hedging), with mean α = 0.34, indicating 34% agreement between raters. For autistic raters, alphas ranged between 0.04 (polite-impolite scale) and 0.31 (direct-indirect), with mean α = 0.15, indicating 15% agreement between raters. Based on the recommendation to strive for .80 ≥ α ≥ .667 (Krippendorff, 2004), our interrater agreement is low. KALPHA scores per scale are provided in Supplementary Table 3 (left for neurotypicals; right for autistic individuals).

5.2.1.7.3. B&L coding

The final sample of 189 unique requests was also subjected to the classification into the four different politeness strategies by B&L (1987), where we took the 'Politeness Annotation Scheme' of Imtiaz et al. (2018) as our starting point. Two raters outside of the project, one female (rater 1) and one male (rater 2), blind to the conditions, individually rated the requests on the four politeness strategies. Both raters were presented with an annotation scheme, explaining the politeness strategies, accompanied by examples from the study.

After rating the first 20 requests, the researcher met with the raters to discuss their rating decisions, estimate the level of interrater agreement (approximately 50% agreement) and check whether the annotation scheme was sufficiently clear. After editing the annotation scheme for the purpose of clarity, the remaining requests were rated. The final annotation scheme is provided at OSF (Chapter 5, Additional materials, https://osf.io/bhqzu/).

Interrater agreement. KALPHA was used to measure interrater reliability. Mean alpha's indicated low interrater agreement between the two raters, for requests by neurotypicals (α = 0.36) as well as requests by autistic individuals (α = 0.39). In case of disagreement, the first author acted as a tie-breaker, creating a 'general politeness strategy score'. In 8 cases (4 requests by neurotypicals and 4 by autistic individuals), the disagreement between raters was resolved by selecting a strategy not originally chosen by either of the raters.

5.2.1.8. Missing data

Due to a coding error, twenty-two participants (13.2% of *N* = 166) in the positive condition did not view the affect questionnaire after the autobiographical memory task. In contrast to our preregistration, we did not use multiple imputation; reasons for this decision are provided at OSF (Chapter 5, Additional materials, https://osf.io/bhqzu/). Our final sample of participants who filled in all affect questionnaires includes *n* = 26 for the positive condition, *n* = 70 for the neutral condition, and *n* = 48 for the negative condition. Due to the same coding error, four participants (2.4% of *N* = 166) did not view the familiarity question, the opinion questions, and the second affect questionnaire (presented after the conversation). For these items and questionnaires, our final sample includes *n* = 47 for the positive condition, *n* = 70 for the neutral condition, and *n* = 45 for the negative condition. Sixteen participants who filled in the AQ-Short (12.6% of *N* = 127) did not complete all 28 items, resulting in missing values for one item (*n* = 9), two items (*n* = 2), three items (*n* = 3) or four items (*n* = 2) per case. Following the criteria by R.A. Hoekstra (Hoekstra et al., 2011; personal communication, July 12, 2019), we excluded the two cases with four missing items, and performed the correction formula for the other cases (*n* = 14; for details, see Appendix). Our final sample for the AQ-Short includes *N* = 125 participants. While unfortunate, these missing data do not seriously hamper the study, given that 1) the data was missing completely at random, and 2) the sample size was still sufficient to test H1. An alternative would have been to remove the participants for which part of the data was missing, but we decided against this since we wanted to make maximum use of the data collected in the experiment.

5.2.1.9. Statistical analysis

Data analysis was performed using IBM SPPS (version 26). Furthermore, linear mixed models were created in R (version 1.2.5033).

5.2.1.9.1. Affect induction

Before our manipulation check, internal consistency of the affect questionnaires was subjected to a principal component analysis (PCA) with a Varimax rotation, using the

criterium of eigenvalue ≥1 for each detected component. Judgments on the scales loading on the respective component(s) were combined into new (aggregated) affect scale(s). Levene's test was used to check for equal variances and QQ-plots were inspected to check whether the data was approximately normally distributed. If equal variances and normality could be assumed, effectivity of the affect manipulation was measured by performing an ANOVA with condition as independent variable and aggregated new affect scale(s) as dependent variable(s). If equal variances and normality could not be assumed, a nonparametric Kruskal-Wallis test was performed. When applicable, Bonferroni corrected posthoc tests were included for both tests.

5.2.1.9.2. Internal consistency of Forgas' request characteristics scales

Following Forgas (1999a), internal consistency of the six request characteristics scales were subjected to a PCA with an Oblique rotation. The identified component(s) with eigenvalue ≥1, with the component scores as weights, were saved as component variables.

5.2.1.9.3. Linear mixed modelling

Next, we used R to perform six linear mixed models, using the lme4 package (Bates et al., 2014, version 1.1-26) with p-values provided by the lmerTest package (Kuznetsova et al., 2017). Model creation started with the fixed effect for condition, adding the random effect (RE) for by-rater random intercept, by-item random intercept and by-topic random intercept, and by-topic random slope for rater. When the model did not converge after these steps were taken, we removed the random slope, followed by the random intercept for topic, respectively. Random intercepts for item and rater were kept in all cases. Following Barr et al. (2013), the most complicated model that still converged was selected as the best, and therefore, final model. The anova() function was used for each model to obtain the F-tests with p-values, using Satterwhaite's method for denominator degrees of freedom and F-statistic. When applicable, sub setting was used, selecting ratings of NT raters on NT speakers (SubsetNT), ratings of autistic raters on requests by autistic speakers (SubsetASD), or both (SubsetAll). Reference levels were positive condition for condition, and NT for group. In order to get all the coefficients of the fixed effect for condition, Models 1-4 were run for a second time, with neutral condition as reference level. An overview of the model components is provided in Table 1.

5.2.1.9.4. B&L politeness strategies

To examine whether there is a relation between the affective state of the speaker and the (B&L) politeness strategy used for their requests, a Chi-square test of independence was performed, including frequency of each politeness strategy, per rater, as well as affective state (positive, neutral, or negative) of the speaker. In addition, a Chi-square test of independence was also performed with the general politeness strategy score.

Table 1

Components used in model equations, including descriptions

Name	Description	Type of variable and values
Component friendly NT	Dependent variable model 1	interval
Component elaboration NT	Dependent variable model 2	interval
Component friendly ASD	Dependent variable model 3	interval
Component elaboration ASD	Dependent variable model 4	interval
Component friendly all	Dependent variable model 5	interval
Component elaboration all	Dependent variable model 6	interval
condition	Fixed effect for affect condition	factor: 1 = positive[1], 2 = neutral, 3 = negative
group	Fixed effect for group of speaker	factor: 1 = NT[1], 2 = ASD
1|item	Random intercept for unique request	factor: 1 - 588
1|rater	Random intercept for unique rater	factor: 1 - 49
1|topic	Random intercept for conversation topic	factor: 1 (bullying)[1] – 6 (cheating)
rater|topic	By-topic random slope for unique rater	factor
SubsetNT	Subset: NT requests by NT raters	222 requests
SubsetASD	Subset: ASD requests by ASD raters	336 requests
SubsetAll	Subset: SubsetNT and SubsetASD combined	558 requests

Note. [1]Reference level.

5.2.2. Results

5.2.2.1. Descriptives

5.2.2.1.1. Familiarity question

Participants indicated on a 7-point Likert scale how well they knew their conversation partner. More than 72% reported 1 ('not at all', *n* = 106) or 2 (*n* = 12), indicating that they did not know their conversation partner. More than 11% reported 6 (*n* = 11) or 7 ('very well'; *n* = 7).

5.2.2.1.2. AQ-Short

The mean score on the AQ-Short was 55.87 (*SD* = 8.50, mode = 58). Five participants exceeded the cut-off score of 70 (Hoekstra et al., 2011), indicating autistic traits. Given that the AQ-Short is not a diagnostic tool (Hoekstra et al., 2011), no participants were excluded based on their AQ-Short score; all participants were included in the analyses.

5.2.2.1.3. Opinion questions

In general, participants considered the six discussion topics to be equally comfortable and uncomfortable, equally personal and neither easy nor difficult. Excessive use of alcohol was considered to be the most comfortable (*M* = 5.41, *SD* = 1.52), least personal (*M* = 3.68, *SD* = 1.78) and easiest (*M* = 5.29, *SD* = 1.69) topic. Sexting was the most uneasy (*M* = 3.78, *SD* = 1.77), personal (*M* = 2.46, *SD* = 1.74) and difficult (*M* = 3.94, *SD* = 1.79) topic.

5.2.2.2. Affect induction

To measure the internal consistency of the affect questionnaire, we subjected the four affect scales to a PCA. All four scales loaded on a single component, with an eigenvalue of 2.58, accounting for 64% of the variance. Judgements on the four scales were combined into one scale called Affect (Chronbach's α = .81), with low scores indicating positive affect, and high scores indicating negative affect. Levene's test indicated equal variances across conditions, $F(2, 141) = 0.62$, $p = .539$. A one-way ANOVA was performed with condition (positive, neutral and negative) as independent variable and Affect as dependent variable. As predicted, we found a significant effect for Affect, $F(2,141) = 126.51$, $p < .001$. The differences were in the expected direction: participants recalling a positive event reported a significantly more positive affective state (*M* = 2.29, *SD* = 0.83) compared to participants in the neutral condition, recalling a typical day (*M* = 2.77, *SD* = 0.73, $p = .023$). Participants recalling a negative event reported a significantly more negative affective state (*M* = 4.81, *SD* = 0.83) compared to participants in the neutral or positive condition (both $p < .001$). Means and standard deviations can be found in Table 2 (left).

Table 2

Means and standard deviations per condition for (aggregated) affect induction scales

	Study 1: NT		Study 2: ASD			
	Affect		HappyAffect		RelaxedTense	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Positive	2.29	0.83	3.23	1.05	4.30	2.00
Neutral	2.77	0.73	4.00	0.75	4.40	1.51
Negative	4.81	0.83	4.85	0.41	5.11	0.60
Overall	3.36	1.30	4.00	1.02	4.59	1.50

5.2.2.2.1. Internal consistency of Forgas' request characteristics

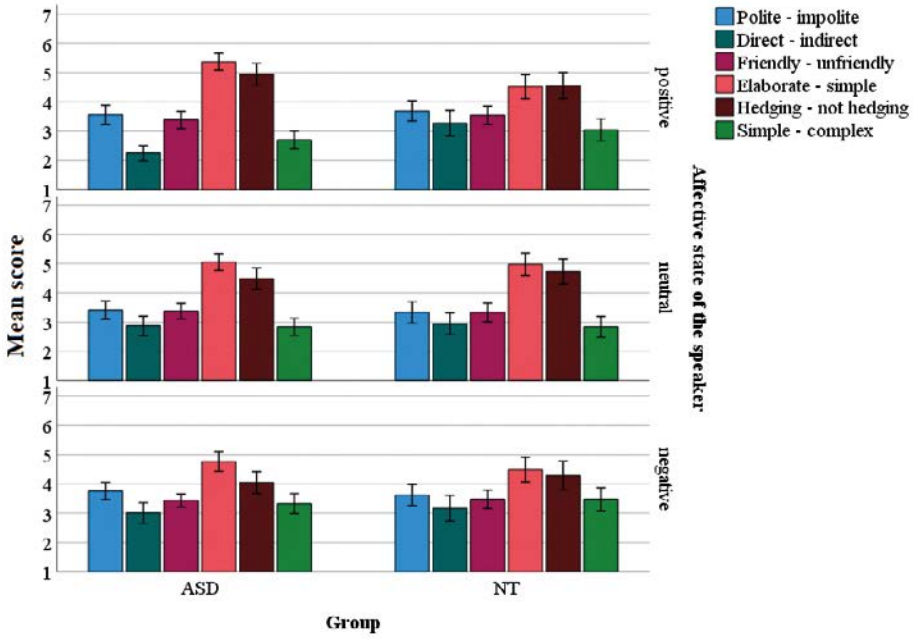
The PCA identified two components with eigenvalues ≥ 1.0 (2.66 and 1.68), accounting for 44% and 28% of the variances explained, respectively. Direct-indirect, elaborate-simple, hedging-not hedging, and simple-complex loaded on the first component, which was named *component elaboration NT*. The remaining scales (polite-impolite and friendly-unfriendly) loaded on the second component, which was named *component friendly NT*. For clarity's sake, means and standard deviations are provided for the aggravated raw scores of the request characteristics loading on component elaboration and component friendly, named *elaboration (aggr.)* and *friendly (aggr.),* respectively (see Supplementary Table 4, left).

5.2.2.3. Affective state on Forgas' request characteristics

Means and standard deviations for affect condition, for all six request characteristics scales, are provided in Supplementary Table 5 (left) and visualized in Figure 1 (right).

Figure 1

Mean scores for request characteristics scales, per condition and group (rater and request)



Note. Error bars represent 95% confidence intervals. Group represents the type of rater as well as the type of rated request. Low scores are represented by the word on the left side of the dash; high scores by the word on the right side of the dash, e.g., polite (1) – impolite (7).

5.2.2.3.1. Model testing

To test our hypothesis (H1) that neurotypicals were more impolite when they were in a positive, compared to neutral, and, subsequently, negative affective state, two linear mixed models, namely, Model 1 and Model 2, were created. To predict the degree of friendliness and politeness (*component friendly NT*) and elaboration, directness, hedging and complexity (*component elaboration NT*) of the request, we included the fixed effect for affective state of the speaker (condition), and random intercepts for item, rater, and topic, with subset SubsetNT. Affective state of the speaker did not predict the degree of friendliness or politeness, $F(2, 95.82) = 0.81$, $p = .448$, nor the degree of elaboration, directness, hedging and complexity of the request, $F(2, 91.60) = 2.26$, $p = .110$. As an additional check, we also ran the models with four additional dependent variables. First, based on our PCA, the aggregated raw scores of the scales loading on the factors (i.e., elaboration (aggr.) and friendly (aggr.)). Second, based on the PCA solution found by Forgas (1999a), the aggregated raw scores of the scales polite-impolite, direct-indirect and friendly-unfriendly (Forgas friendly (aggr.)) and elaborate-simple, hedging – no hedging and simple-

complex (Forgas elaboration (aggr.)). Again, results of the models indicated no significant contribution of condition (i.e., affect) on these models. Model equations are provided in Table 3; details are provided in Table 4.

Table 3

Model equations for each final model

Model	Subset	Model equation
Model 1	SubsetNT	Component friendly NT ~ condition + (1|item) + (1|rater) + (1|topic)
Model 2	SubsetNT	Component elaboration NT ~ condition + (1|item) + (1|rater) + (1|topic)
Model 3	SubsetASD	Component friendly ASD ~ condition + (1|item) + (1|rater) + (1|topic)
Model 4	SubsetASD	Component elaboration ASD ~ condition + (1|item) + (1|rater) + (1|topic)
Model 5	SubsetAll	Component friendly all ~ group + (1|item) + (1|rater) + (1|topic)
Model 6	SubsetAll	Component elaboration all ~ group+ (1|item) + (1|rater) + (1|topic)

Table 4

Estimated parameters of the optimal models to estimate friendliness and politeness, and elaboration, directness, hedging and complexity in NT speakers, according to NT raters

Predictor	Fixed effects					RE item		RE rater		RE topic	
Condition	β	*SE*	*df*	*t*	*p* ≤	s^2	*SD*	s^2	*SD*	s^2	*SD*
Model 1: Effect of affective state on friendliness and politeness in neurotypical speakers (component friendly NT)											
Intercept	0.09	0.09	77.97	0.92	.360	0.15	0.39	0.02	0.14	≥0.01	≥0.01
Neu-Pos	-0.16	0.13	95.34	-1.26	.212						
Neg-Pos	-0.05	0.13	96.17	-0.41	.648						
Neg-Neu	0.10	0.13	95.97	0.81	.417						
Model 2: Effect of affective state on elaboration, directness, hedging and complexity in neurotypical speakers (component elaboration NT)											
Intercept	-0.04	0.10	35.22	-0.40	.691	0.17	0.41	0.04	0.21	≥0.01	0.02
Neu-Pos	0.11	0.13	87.25	0.94	.349						
Neg-Pos	-0.01	0.13	87.07	-0.15	.882						
Neg-Neu	-0.14	0.13	88.26	-1.05	.295						

5.2.2.4. B&L politeness strategies

5.2.2.4.1. Investigating the relationship between affect and politeness strategies

Frequencies of each politeness strategy can be found in Supplementary Table 6.

To explore our first research question (RQ1) whether there is a relationship between the affective state of the speaker and the politeness strategy used in their requests, a Chi-square test of independence was performed, including frequency of assignment to each politeness strategy (in count data), per rater, as well as affective state (positive, neutral, or negative) of the speaker. Initially, we did not aggregate the frequencies for two reasons. First, the frequencies were nominal count data; transforming the data might hinder interpretation of the results. Second, interrater reliability was low; aggregating the data might misrepresent our true observations.

No significant relationship between politeness strategy and affective state was found, for both rater 1, χ^2 (6, N = 105) = 3.51, p = .742, and rater 2, χ^2 (6, N = 105) = 2.85, p = .827, indicating that affective state of the speaker did not influence the type of politeness strategy neurotypicals used for their requests. In a similar vein, no effect of affective state was found for the general politeness strategy score (see *5.2.1.7.3 B&L coding*) either, χ^2 (6, N = 105) = 4.46, p = .614.

5.2.2.5. Conclusion

Even though affect induction was successful, no significant effect was found for affective state on politeness. Neither the degree of politeness, directness, friendliness, elaboration, hedging or simplicity of the request (H1), nor the politeness strategy used to communicate the message of the request (RQ1) was significantly associated with the affective state of the NT speaker. Therefore, our hypothesis predicting that participants in the positive affective state would be less polite, compared to individuals in a negative affective state, was not supported by the data. In conclusion, our findings do not replicate the findings by Forgas (1999a).

5.3. Study 2. Affective state and politeness in autistic individuals

In order to investigate the effect of affective state on politeness in a population known to struggle with, or experience affect and politeness differently than neurotypicals, we conducted an additional experiment with a sample of autistic individuals. Methods, materials and procedure closely mirrored those of Study 1.

5.3.1. Method

5.3.1.1. Design, materials, rating materials and statistical analysis

The study design, (rating of) materials and statistical analysis was identical to Study 1.

5.3.1.2. Participants

Thirty participants were recruited through national autism interest groups, distributing flyers on campus, LinkedIn, a study-support group for autistic students (from an external university), an organization focused on autistic LGBT+ individuals, local social activities groups for autistic individuals, a Dutch online message board for autistic individuals, and word-of-mouth. Additionally, snowballing was used to extend the group of participants.

5.3.1.2.1. Inclusion and exclusion criteria

The inclusion criteria were: having ASD (e.g., autism, Asperger's syndrome, PDD-NOS); the ability to read and write Dutch fluently, and being at least 18 years old. Exclusion criteria were: having an intellectual disability (disharmonious IQ was allowed), not being able to speak Dutch fluently (e.g., stuttering) and being admitted to a ward (outpatient treatment was allowed). Given the high comorbidity of ASD with other psychiatric disorders (e.g., Gotham et al., 2015; Hofvander et al., 2009), participants with additional psychiatric disorders were not excluded[6].

One participant reported to have no official autism diagnosis, but indicated that he, as well as his friends and acquaintances, recognized signs of autism in himself. Given the struggle to find enough autistic participants, we decided to let him participate in the study. Albeit his AQ-Short score was 91, indicating autistic traits, we did not feel comfortable to include him in the dataset without an ASD diagnosis, or treatment for ASD-related symptoms. Therefore, for our analyses, we removed his data from the dataset. To avoid any confusion, we have already changed this throughout the manuscript.

5.3.1.2.2. Final sample

The final dataset included 29 participants (aged 19-58 years; M_{age} = 32.00 years, SD_{age} = 12.24 years). The sample included 19 men, 7 women, and 3 non-binary individuals, who participated in the positive (n =10), neutral (n = 10) or negative condition (n = 9). Participants reported to have Asperger's syndrome (n = 12), Pervasive Developmental Disorder, Not Otherwise Specified (PDD-NOS; n = 8), (high functioning) autism (n = 5) and 'classic autism' (n = 4). Individuals participated in pairs, and each dyad was assigned to one of three conditions. Dyads consisted of a man and a woman (4 dyads), two men (7 dyads, including the excluded participant), two women (1 dyad), one man and one non-binary individual (2 dyads) or one woman and one non-binary individual (1 dyad). All participants

6 Thirteen participants reported to (currently) have no additional psychiatric diagnoses. Sixteen participants reported that they had one or two additional diagnoses. Participants reported depression and mood disorders (***n*** = 8), ADHD/ADD (***n*** = 5), panic disorders, anxiety disorders or social phobias (***n*** = 4), psychoses and hallucinations (***n*** = 2) and other psychiatric disorders (***n*** = 3). Although not classified as a mental disorder, three participants mentioned to have language-related disorders, e.g., dyslexia, but were able to write, speak and read Dutch on a satisfactory level to participate in our study.

received monetary compensation with a voucher for a well-known Dutch online web shop, and were offered a candy bar in addition.

5.3.1.3. Consent

All participants agreed to have their *anonymized* audio recordings, as well as their responses to the questionnaires, used in analyses and publications. Five participants did not grant their permission for us to quote their utterances, i.e., use (fragments of) the audio recording of their speech, both in original and transcribed form, in a presentation or a publication. All participants were clearly informed of their right to stop at any time without consequences.

5.3.1.4. Procedure

The procedure was identical to Study 1, except for the following modifications. First, contrary to the students who participated in Study 1, most participants in the current study had little or no experience participating in scientific studies. Therefore, they received extensive explanation about the procedure, and were frequently asked whether they had questions. In addition, the experiment leader made sure that the participants felt at ease, e.g., by offering them coffee or tea upon arrival and informing them that they could always ask questions.

Second, the experiment took place on different locations, both for convenience (options to recruit and test participants on site, e.g., a networking event) and for ethical reasons (to cater to the preferences of the community, opting for sites familiar to and in the vicinity of (potential) participants, e.g., the office of the student counsellor, spare room in a church). When the experiment did not take place in the university laboratory, both participants were provided with a laptop. Other than that, the set-up from the lab was mimicked as much as possible, e.g., when engaging in the individual tasks of the experiment participants were seated apart, with their back towards each other.

Third, after the participants engaged in the autobiographical memory recall and filled in the affect scales, they filled in the AQ-Short.

5.3.1.5. Missing data

One participant had missing values, missing 2 items. Their AQ-Short score was corrected accordingly (see Appendix).

5.3.2. Results

5.3.2.1. Descriptives

5.3.2.1.1. Familiarity question

The majority of participants (n = 23) did not know their conversation partner, reporting a 1 ('not at all', n = 14) or 2 (n = 9). Only two participants knew their conversation partner fairly well, reporting a 6 (with 7 indicating 'very well').

5.3.2.1.2. AQ-Short

All participants filled in the AQ-Short. The mean score on the AQ-Short was 76.22 (*SD* = 10.45, mode = 78), with a range of 53.00 - 95.00.

5.3.2.1.3. Opinion question

Bullying was, by far, considered the most personal topic (*M* = 1.87, *SD* = 0.91), but also the easiest to talk about (*M* = 4.67, *SD* = 1.80) and one of the most comfortable topics as well (*M* = 4.80, *SD* = 1.66). Cheating was the most uncomfortable (*M* = 3.77, *SD* = 1.64) and most difficult topic (*M* = 3.62, *SD* = 1.39). The most unpersonal topic was debt (*M* = 4.57, *SD* = 1.50). The most comfortable topic was excessive use of alcohol (*M* = 4.88, *SD* = 1.71), followed closely by sexual intimidation (*M* = 4.87, *SD* = 1.64).

5.3.2.2. Affect induction

To measure the internal consistency of the affect questionnaire, we subjected the four affect scales to a PCA. Contrary to Study 1, the affect scales loaded on two components instead of one: happy-sad, good-bad and aroused-not aroused loaded on the first component, with an eigenvalue of 2.15, accounting for 54% of the variance. Judgements on the three scales of the first component were combined into one scale called HappyAffect, with low scores indicating positive affect and high arousal, and negative scores indicating negative affect and low arousal. Relaxed-Tense loaded on the second component, with an eigenvalue of 1.10, accounting for 27% of the variance. Cronbach's alpha for HappyAffect and RelaxedTense was α = 0.33.

Levene's test indicated equal variances across conditions for HappyAffect, $F(2, 26) = 3.30$, $p = .053$, but not RelaxedTense, $F(2, 26) = 6.02$, $p = .007$. Results of Shapiro Wilk's tests and inspection of QQ-plots indicated that the data was not approximately normally distributed. Therefore, a nonparametric Kruskal-Wallis test was performed. Means and standard deviations can be found in Table 2 (right). As an additional check, we performed an ANOVA as well, which provided similar results.

A significant effect was found for HappyAffect, χ^2 (2) = 11.68, $p = .003$, with a mean rank score of -5.50 for positive-neutral ($p = .435$), -13.22 for positive-negative ($p = .002$) and 7.72 for neutral-negative ($p = .139$). The results indicated that participants in the positive condition experienced more positive affect and higher arousal, compared to the participants in the negative condition. Participants in the neutral condition did not experience a significantly different affective state or arousal level than individuals in the positive or negative condition. No significant effects of condition were found for RelaxedTense, χ^2 (2) = 1.47, $p = .480$, indicating that the affect manipulation did not significantly change the level of tension or relaxation that autistic individuals experienced. However, as can be observed in Table 2 (right), mean scores indicated that participants in the negative condition generally experienced more tension, compared to individuals in the positive or neutral condition.

5.3.2.3. Internal consistency of Forgas' request characteristics

The PCA identified two components with eigenvalues ≥ 1.0 (2.43 and 1.76), accounting for 40.5% and 29.4% of the variance explained, respectively. Direct-indirect, elaborate-simple, hedging-not hedging, and simple-complex loaded on the first component, which was named *component elaboration ASD.* The remaining scales (polite-impolite and friendly-unfriendly) loaded on the second component, which was named *component friendly ASD.* For clarity's sake, means and standard deviations are provided for the aggravated raw scores of the request characteristics scales loading on component elaboration and component friendly, named *elaboration (aggr.)* and *friendly (aggr.),* respectively (see Supplementary Table 4, right).

5.3.2.4. Affective state on politeness

Means and standard deviations for affect condition, for all six request characteristics scales, are provided in Supplementary Table 5 (right) and visualized in Figure 1 (left).

5.3.2.4.1. Model testing

To explore our second research question (RQ2), investigating whether affective state influenced the degree of politeness used by autistic individuals in their requests, two linear mixed models, Model 3 and Model 4, were created. To predict the degree of friendliness and politeness (*component friendly ASD*) and elaboration, directness, hedging and complexity (*component elaboration ASD*) of the request, we included the fixed effect for affective state of the speaker (condition), random intercepts for item, rater, and topic, with SubsetASD. Affective state of the speaker did not influence the degree of friendliness or politeness, $F(2, 72.29) = 0.61$, $p = .546$, but it did influence the degree of elaboration, directness, hedging and complexity of the request, $F(2, 74.12) = 6.37$, $p = .003$. Significant differences were found between positive and neutral condition, $p = .035$, positive and negative condition, $p < .001$, but not between neutral and negative condition, $p = .142$, indicating that autistic speakers in a neutral or negative affective state used more hedging, and formulated more elaborate, indirect and complex requests, compared to positive speakers. Again, as an additional check, we also ran the models with four different dependent variables: elaboration (aggr.), friendly (aggr.), Forgas friendly (aggr.) and Forgas elaboration (aggr.). Results of the models indicated no significant contribution for condition on these models. Equations for Model 3 and Model 4 are provided in Table 3; details are provided in Table 5.

Table 5

Estimated parameters of the optimal models to estimate friendliness and politeness, and elaboration, directness, hedging and complexity in ASD speakers, according to ASD raters

Predictor	Fixed effects					RE item		RE rater		RE topic	
Condition	β	*SE*	*df*	*t*	*p* ≤	s^2	*SD*	s^2	*SD*	s^2	*SD*
Model 3: Effect of affective state on friendliness and politeness in ASD speakers (component friendly ASD)											
Intercept	0.05	0.09	49.26	0.59	.555	0.02	0.16	0.17	0.42	>0.01	0.02
Neu-Pos	-0.09	0.08	72.01	-1.10	.275						
Neg-Pos	-0.05	0.08	72.65	-0.67	.508						
Neg-Neu	-0.03	0.08	72.17	0.44	.664						
Model 4: Effect of affective state on elaboration, directness, hedging and complexity in ASD speakers (component elaboration ASD)											
Intercept	0.20	0.10	20.01	1.89	.073	0.08	0.28	0.06	0.25	0.02	0.14
Neu-Pos	-0.22	0.10	74.16	-2.14	.035						
Neg-Pos	-0.37	0.10	74.75	-3.55	.001						
Neg-Neu	-0.15	0.10	73.48	-1.48	.142						

5.3.2.5. B&L politeness strategies

5.3.2.5.1. Investigating the relationship between affect and politeness strategies

Frequencies of each politeness strategy can be found in Supplementary Table 7.

To explore our third research question (RQ3) whether there is a relationship between the affective state of the speaker and the politeness strategy used for their requests, a Chi-square test of independence was performed, including frequency of assignment to each politeness strategy (in count data), per rater, as well as affective state (positive, neutral, or negative) of the speaker. No significant relationship between these two variables were found, for both rater 1, χ^2 (6, N = 84) = 8.63, p = .195, and rater 2, χ^2 (6, N = 84) = 5.80, p = .446, indicating that affective state of the autistic speaker did not influence the type of politeness strategy they used for their requests. No effect of affective state was found for the general politeness strategy score either, χ^2 (6, N = 84) = 3.40, p = .758.

5.3.2.6. Conclusion

Given the relatively modest sample size in Study 2 (N = 29), our study might have been underpowered to detect small effects (type II error), and might be more prone to false positives (type I error) as well. Therefore, the following conclusions should be interpreted using an appropriate degree of caution.

Affect induction was generally successful: individuals in the positive condition reported more positive affective state and higher levels of arousal than individuals in the negative condition. Exploring whether affective state might have an influence on the politeness

behavior of autistic individuals (RQ2), we indeed found a significant association between the two components. Autistic individuals in a negative affective state formulated more elaborate, indirect and complex requests, and used more hedging than positive speakers, a finding similar to Forgas (1999a). However, in contrast to Forgas (1999a), affective state of autistic speakers did not influence the degree of friendliness and politeness of their requests. As in Study 1, no relationship was found between affective state of the speaker and the implementation of different politeness strategies, which answered RQ3.

5.3.2.7. Difference in politeness between NT and autistic individuals

To explore our fourth research question (RQ4) whether, regardless of affective state, autistic individuals might be more, less, or equally polite as neurotypicals, the following steps were taken.

First, using SubsetAll (i.e., the combination of SubsetNT and SubsetASD), a PCA was performed, identifying two components with eigenvalues ≥ 1.0 (2.52 and 1.72), accounting for 42% and 29% of the variances explained, respectively. Direct-indirect, elaborate-simple, hedging-not hedging, and simple-complex loaded on the first component, which was named *component elaboration all*. The remaining scales (polite-impolite and friendly-unfriendly) loaded on the second component, which was named *component friendly all*.

Second, two linear mixed models, Model 5 and Model 6, were built. Model equations are provided in Table 3; details are provided in Table 6. Results indicated that autistic speakers, compared to NT speakers, did not differ in their degree of friendliness or politeness, $F(1, 50.96) = 0.10$, $p = .753$, nor degree of elaboration, directness, elaboration and hedging in their requests, $F(1, 61.43) = 0.92$, $p = .342$. In other words, autistic speakers did not produce more or less polite requests, compared to neurotypicals, which answers our fourth research question.

Table 6

Estimated parameters of the optimal models to estimate politeness and elaboration

Predictor	Fixed effects				RE item		RE rater		RE topic	
Group	β	*SE*	*t*	*p* ≤	s^2	*SD*	s^2	*SD*	s^2	*SD*
Model 5: Effect of type of speaker (NT or ASD) on friendliness and politeness (component friendly all)										
Intercept	0.05	0.10	0.52	.607	>0.01	0.25	>0.01	0.36	>0.01	0.02
ASD – NT	-0.04	0.12	-0.32	.753						
Model 6: Effect of type of speaker (NT or ASD) on elaboration, directness, hedging and complexity (component elaboration all)										
Intercept	-0.06	0.08	-0.75	.459	>0.01	0.33	>0.01	0.24	>0.01	0.08
ASD – NT	-0.09	0.10	0.96	.342						

5.4. Discussion

In this study, we aimed to replicate and extend Forgas (1999a), who found that individuals in a negative, compared to positive affective state tend to be more polite when formulating requests in hypothetical situations in need of a request. Aiming to replicate these findings in a more naturalistic setting, namely, a face-to-face conversation, two affectively primed individuals asked each other about their experiences with sensitive topics. In Study 1, mirroring Forgas (1999a), we explored this in NT dyads. In Study 2, we extended this concept by studying the above in a relatively small sample of autistic individuals, to investigate whether the influence of affective state on verbal communication might extend beyond the NT population. Subsequently, the content of the requests was rated on language characteristics related to politeness by both NT and autistic raters, using the request characteristics scales by Forgas (1999a), selecting the ratings of peers for analysis. Additionally, requests were classified in terms of B&L's politeness strategies by two NT annotators.

Contrary to our expectations, we were unable to replicate the findings by Forgas (1999a). In Study 1, neurotypicals in a negative affective state were not more (or, for that matter, less) polite than neurotypicals in a positive or neutral affective state, based on request characteristics, as well as the classification to B&L's politeness strategies. Interestingly, affective state did seem to influence request formulation in autistic speakers. In Study 2, autistic speakers in a negative or neutral affective state formulated more elaborate, indirect, hedged and complex requests, compared to autistic speakers in a positive affective state, which is in line with Forgas' original findings. However, again, no effects of affective state were found on B&L's politeness strategies. Finally, when comparing the two groups, we found that, regardless of affective state, autistic speakers did not produce more or less polite requests, compared to neurotypical speakers.

5.4.1. Why did we not replicate Forgas' findings relating affect and politeness?

Our findings did not support the hypothesis that speakers in a negative affective state would be more polite, compared to speakers in a positive affective state. This could conceivably be due to a combination of reasons.

First, although affect induction was successful and support for the relationship between affect and spoken language production is growing steadily (e.g., Converse et al., 2008; Kempe et al., 2013; Out et al., 2020b), we did not replicate the main findings of Forgas. The changes in the experimental set up might be responsible for this. However, as discussed earlier in this chapter, the changes we made to the paradigm were motivated by our aim to enhance the ecological and external validity of the study (by conducting the experiment in real-life interactions). It is conceivable that because of this more realistic setting the effect of affective state on politeness behavior, got lost.

Second, it might be possible that the original effect is insufficiently robust to replicate. However, affective state did seem to influence request formulation in autistic speakers. In Study 2, autistic speakers in a negative or neutral affective state formulated more elaborate, indirect, hedged and complex requests, compared to autistic speakers in a positive affective state. But we must keep in mind that, given the modest sample size, we should interpret these results with caution. Yet, it is unclear why we found this result for autistic speakers, but not NT speakers in Study 1. Future studies could explore this issue further, for example, by aiming to replicate these findings, using a larger sample of autistic speakers.

5.4.2. Strengths

Our study had various strengths.

5.4.2.1. Requests were rated by peers

In the current study, requests were rated on politeness by peers, with autistic individuals rating the requests uttered by autistic speakers, and NT individuals rating the requests uttered by NT speakers. This decision was made, as described in *5.2.7.1.2. Forgas' request characteristics coding* because autistic individuals are often studied from a neurotypical point of view, using methods used to investigate neurotypical behavior, while using neurotypical norms and definitions of social behavior (Heasman & Gillespie, 2019). While this might seem fair enough, it might misrepresent how autistic individuals experience social interactions. For example, Heasman and Gillespie (2019) studied spontaneous conversations between autistic individuals playing a videogame together. They found that autistic individuals seem to have a low coordination threshold in conversations, meaning that, for example, misunderstandings between the two autistic players did not always hamper or deteriorate the conversation; they were often able to move on quickly. This is also in line with the findings of an experiment by Morrison and colleagues (2020), who let three types of dyads (NT/NT, ASD/ASD and NT/ASD) engage in a short conversation, and

asked them to subsequently evaluate their conversation partner. While both NT and autistic individuals rated their autistic conversation partner as, among other things, more awkward and less socially warm than NT conversation partners, autistic individuals did not report less interest to socially interact with them in the future. In this line, what neurotypicals consider to be polite or 'normal' social behavior, might be not be true, or as important, to autistic individuals; indeed, the interpretation of politeness might be a 'typical' neurotypical concept as well.

5.4.2.2. Replication in a more ecologically valid setting

Our replication of Forgas (1999a) was conducted in a more ecologically valid, 'real-life' setting: a conversation. Instead of reading vignettes of hypothetical difficult situations, participants in the current study interacted in dyads by asking each other a question about a pre-set, personal topic. In this way, we were able to create, or mimic, a socially difficult situation in need of a request in a natural setting, enhancing the external validity of our findings. In our sample of neurotypical students, we did not replicate Forgas (1999a). Given that studying to which extent findings are generalizable is an important part of scientific research, we think that this non-replication enriches the field.

5.4.2.3. Contribution to the literature

Third, our study contributes to the sparse literature about politeness in autistic adults (e.g., Belek, 2018; Manett, 2020; Zalla et al., 2014), as well creating a dataset containing spontaneous conversational spoken language by autistic individuals (Yang et al., 2020).

5.4.3. Limitations

Next to strengths, our study also had some limitations.

5.4.3.1. Low to moderate interrater agreement

Interrater agreement was moderate or low, for Forgas' request characteristics, as well as the politeness strategies by B&L. However, it is worth emphasizing that other studies, including for example Imtiaz et al. (2018), found similar low interrater agreement between two pairs of raters (weighted κ = .36 and .48) on various politeness measures as well, including B&L's positive- and negative politeness. Given that our annotation scheme for politeness strategies was inspired by Imtiaz et al. (2018), this might partly explain the relatively low interrater agreement. However, in politeness research, it is found more often that individuals show high variability in what they consider to be (im)polite behavior or speech (e.g., Eelen, 2001; Haugh & Chang, 2019; Winans, 2020). While low interrater agreement might be inconvenient from an experimental and statistical point of view, we conjecture that it might also reflect true variability in what people consider to be (im)polite utterings or behavior.

5.4.3.2. Rating requests on politeness strategies in an experimental setting

We must also keep in mind that the original purpose of B&L politeness theory was to use it as an 'ethnographical tool' to analyze social relationships by studying the speaking practices of its members (Brown, 2015). Indeed, in a recent article, one of the original authors on B&L's politeness theory, Brown (2015), argues that it can be hard to code politeness in concrete situations, because 'it is not always possible to be certain what interlocutors' intentions are at a particular point in natural interaction'. Keeping this in mind, we conjecture that categorizing politeness strategies might be less useful in experimental settings to detect verbal politeness behavior.

5.4.3.3. Groups and settings are heterogenous

Although we aimed to keep the studies as similar as possible, there were clear differences between Study 1 and Study 2. First, our sample of NT participants was quite homogenous, in contrast to the more diverse group of autistic participants. As an illustration, all 166 NT participants were communication science students, with the vast majority being younger than 25 years old (91.5%, *n* = 152) with no (diagnosed) mental illness (97.6%, *n* = 162). In Study 2, although all 29 participants reported to have an ASD, they varied in many other aspects. For example, 35.5% of autistic individuals (*n* = 10) were younger than 25 years, 35.5% (*n* = 10) were between 25 and 34 years old, and 31% were 35 years or older (*n* = 9). In a similar vein, variations in occupation and educational level, among other things, were observed as well.

Second, whereas all NT students were tested in the (often quiet and familiar) university's laboratory, autistic participants were tested in various (occasionally noisy and unfamiliar) environments (see also *5.2.1.5. Procedure* for Study 1; *5.3.1.4. Procedure* for Study 2). These different settings might make it harder to compare the two studies. Nevertheless, we argue that prioritizing the convenience and comfort of our (especially autistic) participants was more important to us than higher similarity between Study 1 and Study 2.

Third, as individuals with autism are often harder to recruit than neurotypicals (see e.g., Haas et al., 2016), we overestimated how many autistic individuals we could recruit for participation in Study 2. As a result, the sample size of participants in Study 2 (*N* = 29) is much smaller than in Study 1 (*N* = 166), which makes it harder to (statistically) compare the two groups. Therefore, we explored the results of Study 2 in a more exploratory fashion. However, we must note that, compared to other studies on verbal language production in autistic individuals, our sample size of autistic participants is common (see, e.g., Sirota, 2004; Volden & Sorenson, 2009).

5.4.4. Conclusion

In this study, we aimed to replicate and extend in a more realistic setting Forgas' seminal finding that individuals in a negative affective state tend to be more polite. We studied this in both neurotypical and autistic individuals. Although we did not replicate the original findings by Forgas in our neurotypical sample, our findings show a limited effect of affect on (im)polite request formation in spontaneous, free conversation in autistic speakers.

Supplementary tables

Supplementary Table 1

Instructions autobiographical memory task

Dutch (original)	English (translation)
Positive condition	
Herinner je een specifieke situatie die je hebt meegemaakt in je sociale leven waarvan je heel blij werd. Probeer de situatie zo levendig mogelijk voor de geest te halen. Stel je voor hoe het was toen je deze gebeurtenis meemaakte. Probeer alle details van de situatie weer te ervaren; voel dezelfde gevoelens die je destijds voelde. Beschrijf de situatie die je je herinnert zo levendig mogelijk, inclusief alle belangrijke details. Je krijgt hiervoor 10 minuten, pas daarna kun je doorklikken naar de volgende pagina.	*Remember a specific situation that has occurred in your social life that has made you very happy. Imagine the situation as vividly as possible. Imagine what it was like when you experienced this event. Try to re-experience all the details of the situation; feel the same feelings you felt at the time. Describe the situation you remember as vividly as possible, including all the important details. You get 10 minutes for this task, only after this you can click through the next page.*
Neutral condition	
Herinner je je een gebruikelijke dag in je leven. Probeer je de dag zo levendig mogelijk voor de geest te halen: welke activiteiten doe je? Wat is het eerste wat je in de ochtend doet? Ga je naar school, werk, en hoe kom je daar? Wat zijn je avond activiteiten? Beschrijf de dag die je je herinnert zo levendig mogelijk, inclusief alle belangrijke details. Je krijgt hiervoor 10 minuten, pas daarna kun je doorklikken naar de volgende pagina.	*Remember a usual day in your life. Try to picture the day as vividly as possible: what activities do you do? What is the first thing you do in the morning? Do you go to school, work, and how do you get there? What are your evening activities? Describe the day you remember as vividly as possible, including all the important details. You get 10 minutes for this task, only after this you can click through the next page.*
Negative condition	
Herinner je een specifieke situatie die je hebt meegemaakt in je sociale leven waardoor je heel verdrietig werd. Probeer de situatie zo levendig mogelijk voor de geest te halen. Stel je voor hoe het was toen je deze gebeurtenis meemaakte. Probeer alle details van de situatie weer te ervaren; voel dezelfde gevoelens die je destijds voelde. Beschrijf de situatie die je je herinnert zo levendig mogelijk, inclusief alle belangrijke details. Je krijgt hiervoor 10 minuten, pas daarna kun je doorklikken naar de volgende pagina.	*Remember a specific situation that has occurred in your social life that has made you very sad. Imagine the situation as vividly as possible. Imagine what it was like when you experienced this event. Try to re-experience all the details of the situation; feel the same feelings you felt at the time. Describe the situation you remember as vividly as possible, including all the important details. You get 10 minutes for this task, only after this you can click through the next page.*

Supplementary Table 2

Conversation example on excessive alcohol use (English translations from Dutch originals)

Study 1	Study 2
S (man, 24 years old): So, [name L], if you do not want to talk about it, then, <uh> L (man, 24 years old): Yes, yes. S: You don't have to, of course, but <uh>, do you have experience with excessive use of alcohol? L: Uh, yeah I don't know, I I drink, I did have a period of time that I drank a lot but I think it's also kind of part of student life of something I think. S: <Uhuh>. L: Not like it was really problematic, <uh> and in my environment, actually also no [problematic use of alcohol], actually no-one. Actually yes, a friend of mine who <uh>, we are a bit afraid that he <uh>, is kind of going in the wrong direction. S: <Uhuh>. L: Quite some, quite some psychiatric problems and he, say, quickly reaches for the bottle. S: Yes. L: <uh> so that is the only experience I have with it and it is also a bit a bit close call you know. And otherwise in my environment not really but alright, no one who abused alcohol or something. S: Yes. L: And you, have you ever abused alcohol yourself? S: No yes I have abused alcohol but. L: But not not not then. S: More just [at] a party or something. L: Yes exactly but not not really structurally <unintelligible>. S: No I also don't drink [on a] weekly [basis] or something so. L: No exactly. S: And. L: And do you have someone in your environment? S: I cannot really imagine [someone] it is something that happens a lot but. L: Yes, I do share that idea that it happens a lot and that I have never really experienced it or something. S: No, but <unintelligible> it is a topic that you can discuss or. L: Yes. S: <unintelligible> L: Difficult because he he it is yes with all due respect really a very quiet guy <beep> and rather rigid and ignores advice rapidly you know. S: Yes.	S (woman, 23 years old): <uh> <uh>, have you ever suffered from <uh> using too much alcohol yourself or those around you? L (non-binary individual, 23 years old): <uh> a little bit, <unintelligible> I was quite an I didn't go out very often but when I went out I always ended up somewhere <unintelligible>. But- but a good friend of mine who <uh> we would eventually have almost send her to the <Al-> <uh> Alcoholic Anonymous or something like that because she was actually just always drinking. S: Oh yes. L: She was also going to do an <uh> education at a lower level so that she could go to the pub more often. S: Oh <unintelligible> yes. L: <laughter> S: Those are <uh>. L: Yeah <laughter> S: Yes questionable moments yes. L: Yes so <uh> yeah. S: <uh> myself <uh>, I don't, I once <uh> made a few wrong decisions when I had relatively little experience with alcohol. L: <laughter> S: With <uh> I am for example, I know from myself L: <uhuh> S: That I can't handle flavors like licorice and all. L: Oh yes. S: And then in ten minutes I took four different types of shots including 1 anise and 1 licorice. L: <laughter> S: Well, I really vomited and I did make a mistake there, but that is a one-off and not much more [often]. But also in my relationship and family and circles of friends I have little experience with it, yes. Which <beep> L: Positive <laughter>. S: Yes that is really nice actually.

Note. S = speaker, L = listener. <beep> indicates the stopwatch going off, indicating that their conversation time was up. Between [brackets]: clarification

Supplementary Table 3

Krippendorff's alphas for Forgas' request characteristics scales

	RaterRequestNT	RaterRequestASD
Request characteristic	α	α
Politeness	.345	.036
Directness	.352	.310
Friendliness	.310	.060
Elaboration	.333	.148
Hedging	.444	.202
Complexity	.249	.124
General	.339	.147

Note. For each scale, the measurement level was ordinal.

Supplementary Table 4

Means and standard deviations of elaboration (aggr.) and politeness (aggr.)

	Study 1: NT					Study 2: ASD				
		Elaboration[1] (aggr.)		Friendly[2] (aggr).			Elaboration[1] (aggr.)		Friendly[2] (aggr.)	
Condition	$N_{request}$	*M*	*SD*	*M*	*SD*	$N_{request}$	*M*	*SD*	*M*	*SD*
Positive	81	3.85	0.54	3.62	1.39	108	3.80	0.51	3.47	1.59
Neutral	72	3.87	0.42	3.33	1.38	116	3.82	0.62	3.40	1.48
Negative	69	3.87	0.58	3.55	1.33	112	3.79	0.61	3.61	1.28
Total	222	3.86	0.52	3.50	1.37	336	3.80	0.58	3.49	1.45

Note. $N_{request}$ represents the number of requests rated.
[1]Direct-indirect, elaborate-simple, hedging-not hedging, simple-complex
[2]Polite-impolite, friendly-unfriendly

Supplementary Table 5

Means and standard deviations per condition for Forgas' request characteristics scales

		Study 1: NT		Study 2: ASD	
Request characteristic	Condition	*M*	*SD*	*M*	*SD*
Polite-impolite	Positive	3.69	1.57	3.55	1.79
	Neutral	3.33	1.57	3.41	1.69
	Negative	3.62	1.50	3.77	1.56
	Total	3.55	1.55	3.58	1.68
Direct-indirect	Positive	3.27	1.99	2.21	1.36
	Neutral	2.94	1.58	2.88	1.81
	Negative	3.17	1.81	2.89	1.87
	Total	3.14	1.81	2.67	1.73
Friendly-unfriendly	Positive	3.54	1.43	3.38	1.58
	Neutral	3.33	1.35	3.38	1.43
	Negative	3.48	1.31	3.45	1.18
	Total	3.45	1.37	3.41	1.40
Elaborate – simple	Positive	4.52	1.86	5.38	1.53
	Neutral	4.97	1.63	5.05	1.53
	Negative	4.49	1.76	4.85	1.74
	Total	4.66	1.76	5.09	1.61
Hedging - not hedging	Positive	4.56	2.03	4.93	2.00
	Neutral	4.74	1.78	4.49	1.98
	Negative	4.29	2.04	4.19	2.03
	Total	4.53	1.95	4.53	2.02
Simple – complex	Positive	3.05	1.69	2.67	1.59
	Neutral	2.85	1.52	2.84	1.59
	Negative	3.48	1.64	3.22	1.84
	Total	3.12	1.63	2.92	1.69

Supplementary Table 6

Frequencies of B&L's politeness strategies for requests by neurotypicals, per rater

		Politeness strategies							
		Bald on record		Positive politeness		Negative politeness		Off-record	
Condition	*n* requests	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
Positive	38	12	11	7	8	10	14	9	5
Neutral	34	12	13	7	5	12	13	3	3
Negative	33	11	11	5	3	12	14	5	5
Total	105	35	35	19	16	34	41	17	13

Supplementary Table 7

Frequencies of B&L's politeness strategies for requests by autistic speakers, per rater

		Politeness strategies							
		Bald on record		Positive politeness		Negative politeness		Off-record	
Condition	*n* requests	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
Positive	27	15	15	4	3	6	7	2	2
Neutral	29	13	10	4	3	4	12	8	4
Negative	28	12	8	3	3	2	11	11	6
Total	84	40	33	11	9	12	30	21	12

APPENDIX

Additional information AQ-Short

The AQ-Short is divided into two scales: difficulties in social behavior, and preoccupation with numbers and patterns. Participants respond to each statement on a 4-point Likert scale, with 1 'definitely agree', 2 'slightly agree', 3 'slightly disagree' and 4 'definitely disagree'. Thirteen items were reversed, in which 'agree' with the statements represent autistic characteristics. Item scores are summed into a sum score, with the minimum of 28 indicating no autistic traits, and the maximum of 112 indicating full affirmation of all autistic traits. We settled for a cut-off score of 70, based on the 'more stringent cut-off score' proposed by Hoekstra et al. (2011). However, as can be read in the main text, this cut-off score was merely used as a check.

Missing Data

The AQ-Short is considered unreliable if, per participant, more than 3 items are missing (Hoekstra et al., 2001). Therefore, we excluded participants with four or more missing items. Regarding three or less missing items, the AQ-Short score can be corrected by calculating: total AQ-Short score + (mean item score x number of missing items) (Hoekstra et al., 2011).

Study 1

Given that the study conducted in phase 1 did not include the AQ-Short, a few months after their participation in the lab study, participants in the positive and negative condition were approached by via e-mail. They were reminded of their earlier participation in the study, and requested to, as part of the study, fill in one last questionnaire online. Almost 60% (n = 57 out of n = 96) filled in the AQ-Short; all participants received compensation (course credit, monetary donation to charity, or a candy bar).

In phase 2, AQ-Short was included in the experimental study.

6

General discussion

In this dissertation, we studied the relationship between affect and language production. In four experimental studies, we investigated the following three questions. First, whether verbalizing affective content, categorized by the dimensions of valence and arousal, can induce affective states. Second, whether affective states, as well as specific emotions, can influence semantic and pragmatic aspects of spoken language production in conversation, specifically, alignment (referential expressions, conceptual pact formation) and verbal politeness. Third, whether we could replicate the results of previous studies in the field of (affective) language production, using conceptual replications in more naturalistic domains. In the following section, we will first provide a summary of each chapter, answering the research questions formulated in the introduction of this dissertation, followed by a concise answer to the general research question (see Chapter 1, *1.6. Research questions and overview of studies*). Subsequently, in light of our conceptual replication studies, we discuss our experience with replication research, focusing on both the importance and the difficulties of replication. Then, we provide theoretical implications of our work, including the importance of using naturalistic settings in research, the consequences of the lack of consensus on the definition of 'affect', and potential benefits of integrating both affect and language in existing theories in affective science and psycholinguistics. This chapter ends with a general conclusion.

6.1. Overview of the findings

6.1.1. Chapter 2: Can verbalizing affective pictures induce affective states?

In Chapter 2, we investigated whether verbalizing affective pictures could induce the corresponding affective states in speakers. We were inspired by Velten (1967; 1968), who found that reading out loud self-referring statements that gradually increase in positive or negative content induces the corresponding moods in the speaker. In this chapter, we investigated whether verbalizing affective images would have a similar effect. We conducted two experiments in which participants described (Study 1) or passively viewed (Study 2) affective (IAPS) pictures that gradually increased from neutral to positive or negative content, or remained neutral. We measured the self-reported affective state before and after participants were exposed to the pictures. We hypothesized that successful affect induction would occur in both studies, but would be stronger for speakers who verbally described the pictures compared to individuals who merely passively viewed the same pictures.

As predicted, our results showed that speakers felt more negative after describing negative pictures and that describing neutral pictures did not result in a change of affective state. However, contrary to our expectations, describing positive pictures did not significantly result in a positive affective state. We also found no differences in effectivity of affect induction between actively describing and passively viewing the arrays of pictures.

In addition to testing our hypotheses, we aimed to investigate the linguistic characteristics used by speakers that described affective pictures. To do this, picture

descriptions were audio recorded and transcribed. Using the word-counting software LIWC (Pennebaker et al., 2001), we explored whether the descriptions contained gradually more affectively laden language in the expected directions. The results confirmed that this was indeed the case: compared to the neutral pictures, positive and negative pictures were gradually described with more affective words, which were more positive and negative, respectively.

The results from the two studies in Chapter 2 provide support for the affect induction potency of viewing, and viewing and verbalizing affective pictures that gradually increase in affective content, but only for negative and not for positive pictures. Based on the results of these studies, we conclude that verbalizing (negative) affective pictures can indeed induce the corresponding (negative) affective states. Interestingly, verbalizing or passively viewing the negative and positive affective content were equally effective.

6.1.2. Chapter 3: Do emotions influence alignment between conversation partners, in specific, alignment in referential expressions?

In Chapter 2, we investigated whether viewing and verbalizing affective content results in the corresponding affective states. In the subsequent chapters, we investigated whether (induced) affective states can influence various linguistic aspects of spoken language production in conversations. In Chapter 3, we focused on the influence of specific emotions, namely, amusement and disgust, on alignment in the production of referential expressions. We extended and replicated the study by Goudbeek and Krahmer (2012), who found that while speakers generally prefer to use color to describe an object, they start to use other, dispreferred attributes such as orientation ('the chair seen from the front') when they are primed by a pre-recorded partner using these dispreferred attributes. In other words: speakers adjusted their use of referential expressions, aligning to their (hypothetical) conversation partner.

In Chapter 3, we replicated this study in a more naturalistic setting, namely, a face-to-face interaction in which dyads described pictures of furniture to each other. We hypothesized that, as in the study by Goudbeek and Krahmer (2012), participants would align with the referential expressions used by their conversation partner. Additionally, we expected that the induced emotional state of the speakers, amused or disgusted, would influence the conceptual alignment between dyads in these interactions, hypothesizing that disgusted speakers would align more with their conversation partners than amused speakers. Our hypothesis was based on earlier studies finding support that speakers in a negative mood, compared to a positive mood, tend to take a less egocentric perspective (e.g., Converse et al., 2008) and a more narrow attentional scope (e.g., Gasper & Clore, 2002). We presumed that the results of these studies might translate into negative, disgusted speakers, compared to positive, amused speakers, being more narrowly focused on their conversation partner, therefore displaying more referential alignment. Given that that speakers naturally prefer to use color in their descriptions (e.g., Pechmann, 1989; Sedivy, 2003), we focused on alignment with the dispreferred attribute size.

Supporting our first hypothesis, we replicated the results of Goudbeek and Krahmer (2012): participants indeed aligned to the referential expressions of their conversation partner by using the same type of property, even when this property was not preferred (i.e., size, instead of color). Presumably because of the more naturalistic nature of the task, many participants used overspecified descriptions (using both color and size; 'the large blue chair'), which was taken into account in the analysis. Even so, our results indicated that speakers primed with size attributes were more likely to use size, with or without considering the presence of overspecified descriptions (i.e., including color in the description).

Emotion induction was generally successful, especially for disgust. We found a limited effect of emotion on alignment between conversation partners, with disgusted individuals, compared to amused individuals, being (slightly) more likely to align to their conversation partner, in that they were more likely to use only size when their conversation partner did so, too. Based on these findings, we conclude that our second hypothesis was only partly supported.

In conclusion, we demonstrated that emotion can have an, albeit small, influence on alignment in referential expressions between conversation partners. These findings were in line with the findings of earlier studies, supporting the presumption that affective state can influence various cognitive processes important for language production, including attention and perspective taking.

6.1.3. Chapter 4: Do affective pictures, or affective states, influence the production of conceptual pacts in reference?

In Chapter 4, we studied another type of collaboration in conversation, namely, conceptual pact formation. We replicated and extended the study of Clark and Wilkes-Gibbs (1986), who found that individuals repeatedly describing abstract figures to each other gradually and efficiently create temporary agreements, or conceptual pacts, on how to refer to specific figures. As a result, over time, they need fewer words and speaking turns to describe the figures. We predicted that we would replicate their results, investigating whether dyads conversing about naturalistic, affectively laden pictures, varying in high or low (un)pleasant and (un)arousing content, also create successful conceptual pacts. Additionally, as in Chapter 2, we hypothesized that verbalizing the content of the affective pictures would induce the corresponding affective state in the speaker. In contrast to the individual-focused setting in Chapter 2, we now studied this within dyads, describing and matching affective pictures in a director-matcher task. Subsequently, we explored whether affective state of speakers influenced the conceptual pact formation within dyads, and whether the affect induction effect of the referential task might be stronger for directors, compared to matchers.

To study this, we used a paper-based director-matcher task, in which two participants were seated in front of each other, with an opaque screen in the middle of the table obscuring their ability to view the other person's activities. Participants were presented with identical sets of pictures, depicted in different, fixed orders, on six sheets (the director)

or as a stack of separate cards, accompanied by a sheet depicting empty boxes (the matcher). The director reported to the matcher which picture occupied which position in each particular sequence; the matcher put the pictures in the correct order, in the boxes on their sheet. This process was repeated until all six trials were completed, and therefore, each picture was described six times. To analyze the number of words used, and turns taken to describe the pictures, the experiment was audio recorded and transcribed. Before and after engaging in the director-matcher task, participants indicated their affective state by filling in two Self-Assessment Manikins (Bradley & Lang, 1994) measuring the dimensions pleasantness and arousal.

The results showed that dyads indeed formed conceptual pacts successfully: over time, directors used less words, and less speaking turns, to describe the pictures. Surprisingly, and contrary to our prediction, participants exposed to the four affective picture categories generally reported an increase in pleasantness and arousal, regardless of condition. Apparently, the task is engaging and pleasant, whether one describes pleasant content or not. While keeping this in mind, we also found differences between directors and matchers. Regardless of affective picture condition, directors, compared to matchers, generally reported substantially higher levels of arousal after viewing the highly (compared to low) arousing pictures, and (slightly) more pleasantness after engagement in the task.

Given that the affective pictures generally did not induce the corresponding affective states in the speakers, we instead examined the influence of the affective picture sets, based on their (un)pleasant and (un)arousing properties, on conceptual pact formation. Our results indicated that the strength of the decline in word use and turn taking was, albeit limited, influenced by the affective picture category. We found complex relationships between valence and arousal dimensions (of the pictures) and measures of alignment (decline in word use and turn taking) that warrant further investigation.

In conclusion, our results support the collaborative model: conceptual pact formation emerges when dyads describe naturalistic images, as well as abstract figures. Given that the affective content of the pictures did not induce the expected, corresponding affective states, but generally increased both the positive affect and arousal levels in participants, the influence of affect remains somewhat unclear in this particular setting.

6.1.4. Chapter 5: Do affective states influence the production of (im)polite language?

In Chapter 5, we replicated and extended the study by Forgas (1999a), who found that, in hypothetical difficult situations in need of a request, individuals in a negative affective state formulate more polite requests than those in a positive affective state. We replicated his study in a more naturalistic setting, by creating a conversational setting in which participants needed to formulate requests (about potentially sensitive topics) to gain information about their conversation partner. After an affect induction, participants were paired up to ask each other about their experiences with intimate, personal affairs such as bullying and infidelity. Conversations were audio recorded and transcribed for analysis. We

hypothesized that we would replicate Forgas's finding that speakers formulate more polite requests in a negative mood, compared to speakers in a positive mood. Politeness of the requests was assessed based on the same six scales indicating request characteristics used in the original study, including (im)politeness. Additionally, we explored whether there was a relationship between affective state of the speaker and the usage of politeness strategies (Brown & Levinson, 1987).

Furthermore, to explore whether there are differences between autistic and non-autistic (so-called 'neurotypical') speakers, the experiment was performed with both neurotypical students (Study 1) and a small sample of autistic individuals (Study 2). We investigated whether there was a relationship between affective state and politeness for autistic speakers as well. Comparing the results of these two studies, we were able to explore whether autistic individuals were more, less, or equally polite as their neurotypical peers.

Affect induction was generally successful in both studies. But, contrary to our expectation, we were unable to replicate Forgas's findings, as the degree of politeness expressed by neurotypicals in Study 1 seemed independent of their affective state. In Study 2, however, we did find an effect of affect for autistic speakers in a negative affective state, who formulated more elaborate, hedged, indirect and complex requests, compared to their peers in a positive affective state.

We found that, regardless of affective state, autistic speakers did not produce more or less polite requests, compared to neurotypical speakers. A more detailed analysis of different politeness strategies revealed no differences of affective state on politeness strategies used by either group of speakers.

Concluding, affective states influenced the language characteristics related to politeness in autistic individuals, but not in neurotypical students. We found an effect based on (some) request characteristics by Forgas, as described above, but not based on the ratings of politeness strategies. Based on the judgement by their peers, neurotypical and autistic speakers were about equally polite.

6.2. Answering our main research question

In Chapter 2 to 5, we investigated the relationship between affect and language production in four experimental studies, in order to address our main research question: 'To which degree do affective states influence (spoken) language production in an interactive setting?' Based on the findings of the preceding chapters, we conclude that affective state and language production do influence each other, sometimes to a small, and sometimes to a large degree. For example, as would intuitively be expected, the speakers who described affectively-laden pictures in Chapter 2 used more affective words in their descriptions, compared to non-affective pictures. When they described pictures with negative (but not positive) content, they reported to feel more negative as well.

However, the nature of the relationship between affect and language remains somewhat unclear, and its strength is variable. Sometimes our findings indicated that affect, and affective state, do influence language production, although this effect was subtle and not

so easy to detect. For instance, in Chapter 5, we found a small effect of affective state on politeness in autistic, but not neurotypical speakers. In Chapter 3, disgusted speakers, compared to amused speakers, tended to align somewhat more with their conversation partner by using the same referential expressions in their descriptions.

Some results of our studies were surprising, seemed to be counterintuitive, or even contradicted each other. For instance, after speakers verbalized negative affective pictures in Chapter 2, they, as we expected, reported to experience the corresponding negative affective state. In Chapter 4, however, speakers reported to feel more positive after they described negative pictures. As we discussed in Chapter 4, the latter finding might be due to the light-hearted and positive setting of the director-matcher task in this specific experiment. However, as suggested by D. Moulds (personal communication, 30 June 2021), another reason might be that dyads can so-called 'talk themselves out of their negative state'. Moulds found that, after a director-matcher task preceded by a sad emotion induction, interlocutors generally showed improvement in their self-reported affective state, when they were both allowed to verbally interact with each other (e.g., provide feedback), and share their negative affective state. In light of his findings, it is also conceivable that the negative speakers in Chapter 4 experienced something similar.

Another surprising result was reported in Chapter 5, where we found that affective state (slightly) influenced some of the request characteristics related to politeness of autistic, but not neurotypical speakers. Indeed, while affect induction was generally successful for both groups, we only partially replicated Forgas's findings in a very different population than he originally tested, while simultaneously failing to replicate his findings in his original target population, namely, (presumably) neurotypical college students. While we are aware that our sample of college students is undoubtedly different from the sample tested by Forgas (see also, e.g., Anderson et al., 2016), the naturalistic setting of our experiment is probably the biggest reason why we have not been able to replicate the original findings. For instance, the social interactions between dyads were likely influenced by a number of unintended, uncontrolled factors, making the results 'noisy', but arguably also more realistic.

6.3. Replication research

In the current dissertation, we performed conceptual replications of several previous studies. As we emphasized in the introduction, replication is an essential aspect of research (e.g., Makel et al., 2012; Van Berkel & Crandall, 2018). By replicating seminal studies on affect and language production, we were able to increase the validity and (mostly ecological) reliability of well-known phenomena as conceptual pact formation and referential alignment between conversation partners.

In general, we replicated the findings from original studies to a large degree. In Chapter 4, we found strong support for the collaborative model of Clark and Wilkes-Gibbs (1986), as the general pattern of decline in word use and turn taking over time was highly similar to the results presented in their original paper. In a similar vein, in Chapter

3, we replicated Goudbeek and Krahmer's (2012) findings that dyads tend to align their referential expressions when describing objects. Moreover, we replicated this study in a more naturalistic, ecological valid setting, contributing to the generalizability of conceptual alignment between individuals engaging in a conversation.
Chapter 2 similarly builds and extends existing research. In this chapter, we found partial support for the affect induction effect of reading out loud affective content, as designed by Velten (1967; 1968). In our study, however, participants reading out loud negative, but not positive affective content, reported to experience the corresponding affective state. Given that previous studies including a positive and negative affect induction have found similar results (see, e.g., Uhrig et al., 2016; Westermann et al., 1996), we conjecture that it generally might be more challenging to induce positive affect than negative affect.

In Chapter 5, we found some significant differences when aiming to replicate the original findings by Forgas (1999a) that speakers in a negative, compared to positive affective state tend to be more polite when formulating requests. As discussed above, regardless of the effectivity of the affect induction, we found no effect of affective state on politeness for neurotypical speakers in Study 1, and only a small effect for autistic speakers in Study 2, as indicated by some request characteristics related to politeness, as directness. One possible reason that we only found modest support for the replication of the original results by Forgas (1999a) might be that our experimental setting was more natural, and therefore rather different from the original procedure, hence no effect of affective state on politeness could be detected. Given that few experimental studies have examined verbal politeness in autistic individuals (see *5.1. Introduction*), let alone *emotional* autistic individuals, interpreting or explaining the effect for autistic individuals is challenging, but it does indicate the importance of including more diverse populations in both affective and psycholinguistic studies.

Taken together, we replicated the majority of the findings of the original studies, contributing to the efforts of many labs to increase the reliability of psychological (and psycholinguistic) research.

6.4. Theoretical implications

This dissertation focused on how affect and emotions influence language production. Based on the results of our studies, we draw the following theoretical implications for affective science, as well as psycholinguistics.

6.4.1. Ecological validity

One of the foci of this dissertation was on replicating seminal studies in psycholinguistics in a more naturalistic, social setting. In the experiments described in this dissertation, we let participants engage in real social interactions as conversing about a film clip (Chapter 3), sharing personal experiences with each other (Chapter 5), and describing and matching pictures in a two-player game, formatted as a table-top (Chapter 4) or computer game

(Chapter 3). These naturalistic, more ecologically valid paradigms are in contrast to more conventional, artificial experimental research, in which, for example, participants read vignettes describing hypothetical social interactions, and answer questions about their (equally) hypothetical behavior in these imagined situations. Given that the purpose of most scientific studies, including the studies presented in this dissertation, is to gain knowledge about the world in general, instead of, for example, specific laboratory situations (Kaplan, 1964; Rosenthal & Rosnow, 1984), we were happy that we could replicate the findings of important psycholinguistic phenomena as referential alignment (Chapter 3) and conceptual pact formation (Chapter 4) in more naturalistic settings. First, as described in Chapter 1, replicating these findings in more naturalistic settings contributes to the generalization of these phenomena to real-life situations. Second, taking a broader perspective, our findings support the sentiment that this type of research, namely, conversational settings in which dyads converse freely, is possible and feasible, and can give interesting and important new information on how individuals truly converse with each other.

6.4.2. Construct validity and consensus in research

As mentioned in the introduction of this dissertation, there is no general consensus in affective science how affect, or emotion, should be defined. Izard (2010) found that this is echoed by researchers in affective science: when asking 37 scholars to give their definition of 'emotion', most individuals did not agree on a unitary definition, although they did express the need for this. After all, a lack of consensus on a scientific concept is not without costs. As Izard (2010) rightfully argues, when 'emotion' remains a fuzzy concept, it might result in inconsistencies and confusion, hindering progress in the scientific field of affect. For example, given the multitude of theories on affect, researchers can use their own operational definition. This lack of standardization results in researchers using their preferred combination of ways to induce, interpret and measure affective state, and report their findings in light of one, or multiple, affect theories. This might explain why, for example, the effectivity of affect induction procedures tends to vary (see, e.g., Joseph et al., 2020; Westermann et al., 1996).

In a similar vein, there are also multiple ways to measure and interpret politeness. In Chapter 5, raters often disagreed when indicating which politeness strategy was implemented for which request. Upon reading their rationale accompanied with their decision, this was either based on a difference in opinion, *or* a difference in the interpretation of the politeness strategy. Given that our annotation scheme was tested and discussed with both raters simultaneously, and updated for clarification, this was an unexpected and interesting finding. Based on this experience, we conjecture that, as affect, the (scientific, but also 'folk') definition and conceptualization of politeness might not be uniform either, which is endorsed by several scholars (see, e.g., Dimitrova-Galaczi, 2002; Qiu, 2015). As a result, the construct validity of (at least our own) politeness strategies might be low, meaning that we may not precisely have measured what we intended to measure.

While the need for standardization in affective research is acknowledged (e.g., Izard, 2010), to date, there are still not many commonly accepted standards in affective science. In fact, the same can be said about psychological research more in general. Eronen and Bringmann (2021) even state that psychology research is 'facing a theory crisis', in which scholars point out that theoretical underpinnings of psychology theories are often 'shaky' (p. 1), resulting in only a few good, robust theories in psychology. Eronen and Bringmann offer multiple reasons for this, including the following two causes. First, they argue that psychology researchers tend to formulate vague theories, which makes it hard to properly investigate, and, therefore, falsify them. Furthermore, they state, even when theories are considered to be faulty or inadequate, psychology researchers tend to use them anyway, resulting in an abundance of partially overlapping, inadequate theories on the same phenomenon. Second, they argue that scholars do not focus enough on the improvement and validation of psychological constructs, but, instead, keep introducing new constructs and terms for presumably existing constructs. Their argument was based on Hagger's (2014) commentary on Skinner (1996), who, as early as in the 1990's, already found more than 30 different constructs connected to 'perceived control'; this number is still growing (Hagger, 2014). In short, for more progress in the study of the influence of affect on language production both further standardization and theory development are called for.

6.4.3. The relationship between affect and language production

As we mentioned in the preceding chapters, research on the relationship between affect and language production is growing. We think that researchers in affective science should focus more on the influence of language in their experiments, and, conversely, researchers in (psycho)linguistics should focus more on the influence of affect in language production. For instance, we argue that the influence of affect could be integrated in speech production models, for example, the model by Levelt (1989). This model states that speakers first have to decide what they are going to say. To do this, they conceptualize their message into a 'preverbal message'. In the next phase, the speaker decides how they will transpose their preverbal message into a verbal message, i.e., how they are going to say it, resulting in an utterance plan. Finally, this resulting plan is phonologically encoded and articulated, resulting in spoken language. Previous research shows that these different stages can be influenced by external processes as perspective taking (Levelt, 1999), conversational factors (Pickering & Garrod, 2013), as well as affect (as this dissertation research has shown). The majority of studies on the influence of affect on language production focus on voice characteristics as pitch (e.g., Bachorowski, 1999; Goudbeek & Scherer, 2010; Scherer, 2003), providing support for the influence of affect on later stages in speech production, i.e., articulation. Now, based on the results in the current dissertation, as well as earlier work by other researchers (e.g., Forgas, 1999a; Kempe et al., 2013), we argue that affect likely influences the earlier speech production phases as well. For example, the affective state of (autistic) speakers in Chapter 5 might have influenced their verbal politeness in the earlier stages of language production, i.e., conceptualizing of the request

and the subsequent utterance plan. Although the specifics have not been fleshed out, we conjecture that affect likely influences all stages of language production. Perhaps affect should be integrated as an overall, contextual factor in speech production models, influencing all stages of spoken language production. It would be interesting to develop these ideas further into a full-fledged model of affect and language production, and test the predictions of such an integrated model.

6.5. Conclusion

We have devoted this dissertation to the study of affective language production, focusing on social interactions. We created conceptual replications of original studies, using experimental settings that were as naturalistic as possible. The findings of our studies indicated that emotion and affect, although sometimes subtly, influence language production. Affective state of speakers influenced their usage of affectively laden words (Chapter 2), adaptation to their conversation partner (Chapter 3 and Chapter 4) and verbal politeness when formulating requests (Chapter 5). Furthermore, using our conceptual replications, we replicated the results of original studies to a substantial degree, but not completely. Based on our results, we conclude that affect influences how we converse together, affecting both our social behaviors, such as mimicking our conversation partner, as well as choosing the appropriate words to deliver a message.

SUMMARY

In this dissertation, we studied the relationship between affect and language production. In four experimental studies, we investigated the following three questions. First, whether verbalizing affective content, categorized by the dimensions of valence and arousal, can induce affective states. Second, whether affective states, as well as specific emotions, can influence semantic and pragmatic aspects of spoken language production in conversation, specifically, alignment (referential expressions, conceptual pact formation) and verbal politeness. Third, whether we could replicate the results of previous studies in the field of (affective) language production, using conceptual replications in more naturalistic domains.

In the following section, we will provide a summary of each chapter, answering the research questions formulated in the introduction of this dissertation.

In **Chapter 2**, we investigated whether verbalizing affective pictures could induce the corresponding affective states in speakers. We were inspired by Velten (1967; 1968), who found that reading out loud self-referring statements that gradually increase in positive or negative content induces the corresponding moods in the speaker. In this chapter, we investigated whether verbalizing affective images would have a similar effect. We conducted two experiments in which participants described (Study 1) or passively viewed (Study 2) affective (IAPS) pictures that gradually increased from neutral to positive or negative content, or remained neutral. We measured the self-reported affective state before and after participants were exposed to the pictures. We hypothesized that successful affect induction would occur in both studies, but would be stronger for speakers who verbally described the pictures compared to individuals who merely passively viewed the same pictures.

As predicted, our results showed that speakers felt more negative after describing negative pictures and that describing neutral pictures did not result in a change of affective state. However, contrary to our expectations, describing positive pictures did not significantly result in a positive affective state. We also found no differences in effectivity of affect induction between actively describing and passively viewing the arrays of pictures.

In addition to testing our hypotheses, we aimed to investigate the linguistic characteristics used by speakers that described affective pictures. To do this, picture descriptions were audio recorded and transcribed. Using the word-counting software LIWC (Pennebaker et al., 2001), we explored whether the descriptions contained gradually more affectively laden language in the expected directions. The results confirmed that this was indeed the case: compared to the neutral pictures, positive and negative pictures were gradually described with more affective words, which were more positive and negative, respectively.

The results from the two studies in Chapter 2 provide support for the affect induction potency of viewing, and viewing and verbalizing affective pictures that gradually increase in affective content, but only for negative and not for positive pictures. Based on the results of these studies, we conclude that verbalizing (negative) affective pictures can indeed induce

the corresponding (negative) affective states. Interestingly, verbalizing or passively viewing the negative and positive affective content were equally effective.

In Chapter 2, we investigated whether viewing and verbalizing affective content results in the corresponding affective states. In the subsequent chapters, we investigated whether (induced) affective states can influence various linguistic aspects of spoken language production in conversations.

In **Chapter 3**, we focused on the influence of amusement and disgust on alignment in the production of referential expressions. We extended and replicated the study by Goudbeek and Krahmer (2012), who found that while speakers generally prefer to use color to describe an object, they start to use other, dispreferred attributes such as orientation ('the chair seen from the front') when they are primed by a pre-recorded partner using these dispreferred attributes. In other words: speakers adjusted their use of referential expressions, aligning to their (hypothetical) conversation partner.

In Chapter 3, we replicated this study in a more naturalistic setting, namely, a face-to-face interaction in which dyads described pictures of furniture to each other. We hypothesized that, as in the study by Goudbeek and Krahmer (2012), participants would align with the referential expressions used by their conversation partner. Additionally, we expected that the induced emotional state of the speakers, amused or disgusted, would influence the conceptual alignment between dyads in these interactions, hypothesizing that disgusted speakers would align more with their conversation partners than amused speakers. Our hypothesis was based on earlier studies finding support that speakers in a negative mood, compared to a positive mood, tend to take a less egocentric perspective (e.g., Converse et al., 2008) and a more narrow attentional scope (e.g., Gasper & Clore, 2002). We presumed that the results of these studies might translate into negative, disgusted speakers, compared to positive, amused speakers, being more narrowly focused on their conversation partner, therefore displaying more referential alignment. Given that that speakers naturally prefer to use color in their descriptions (e.g., Pechmann, 1989; Sedivy, 2003), we focused on alignment with the dispreferred attribute size.

Supporting our first hypothesis, we replicated the results of Goudbeek and Krahmer (2012): participants indeed aligned to the referential expressions of their conversation partner by using the same type of property, even when this property was not preferred (i.e., size, instead of color). Presumably because of the more naturalistic nature of the task, many participants used overspecified descriptions (using both color and size; 'the large blue chair'), which was taken into account in the analysis. Even so, our results indicated that speakers primed with size attributes were more likely to use size, with or without considering the presence of overspecified descriptions (i.e., including color in the description).

Emotion induction was generally successful, especially for disgust. We found a limited effect of emotion on alignment between conversation partners, with disgusted individuals, compared to amused individuals, being (slightly) more likely to align to their conversation partner, in that they were more likely to use only size when their conversation partner did

so, too. Based on these findings, we conclude that our second hypothesis was only partly supported.

In conclusion, we demonstrated that emotion can have an, albeit small, influence on alignment in referential expressions between conversation partners. These findings were in line with the findings of earlier studies, supporting the presumption that affective state can influence various cognitive processes important for language production, including attention and perspective taking.

In **Chapter 4**, we studied another type of collaboration in conversation, namely, conceptual pact formation. We replicated and extended the study of Clark and Wilkes-Gibbs (1986), who found that individuals repeatedly describing abstract figures to each other gradually and efficiently create temporary agreements, or conceptual pacts, on how to refer to specific figures. As a result, over time, they need fewer words and speaking turns to describe the figures. We predicted that we would replicate their results, investigating whether dyads conversing about naturalistic, affectively laden pictures, varying in high or low (un) pleasant and (un)arousing content, also create successful conceptual pacts. Additionally, we hypothesized that verbalizing the content of the affective pictures would induce the corresponding affective state in the speaker (i.e., describing pleasant content would induce a positive state). Subsequently, we explored whether affective state of speakers influenced the conceptual pact formation within dyads, and whether the affect induction effect of the referential task might be stronger for directors, compared to matchers.

To study this, dyads engaged in a director-matcher task, describing and matching affective pictures of the affective condition they were assigned to (e.g., jumping individuals, representing the positive valence/high arousal category). To analyze the number of words used, and turns taken to describe the pictures, the experiment was audio recorded and transcribed. Before and after the director-matcher task, participants indicated their affective state by filling in two Self-Assessment Manikins (Bradley & Lang, 1994) measuring pleasantness and arousal.

The results showed that dyads indeed formed conceptual pacts successfully: over time, directors used less words, and less speaking turns, to describe the pictures. After engaging in the director-matcher task, participants generally reported an increase in pleasantness and arousal, regardless of condition. Apparently, the task is engaging and pleasant, whether one describes pleasant content or not. While keeping this surprising finding in mind, we also found differences between directors and matchers. Regardless of affective picture condition, directors, compared to matchers, generally reported substantially higher levels of arousal after viewing the highly (compared to low) arousing pictures, and (slightly) more pleasantness after engagement in the task.

Given that the affective pictures generally did not induce the corresponding affective states in the speakers, we instead examined the influence of the affective picture sets, based on their (un)pleasant and (un)arousing properties, on conceptual pact formation. Our results indicated that the strength of the decline in word use and turn taking was, albeit limited, influenced by the affective picture category. We found complex relationships

between valence and arousal dimensions (of the pictures) and measures of alignment (decline in word use and turn taking) that warrant further investigation.

In conclusion, our results support the collaborative model: conceptual pact formation emerges when dyads describe naturalistic images, as well as abstract figures. Given that the affective content of the pictures did not induce the expected, corresponding affective states, but generally increased both the positive affect and arousal levels in participants, the influence of affect remains somewhat unclear in this particular setting.

In **Chapter 5**, we replicated and extended the study by Forgas (1999a), who found that, in hypothetical difficult situations in need of a request, individuals in a negative affective state formulate more polite requests than those in a positive affective state. We replicated his study in a more naturalistic setting, by creating a conversational setting in which participants needed to formulate requests (about potentially sensitive topics) to gain information about their conversation partner. After an affect induction, participants were paired up to ask each other about their experiences with intimate, personal affairs such as bullying and infidelity. Conversations were audio recorded and transcribed for analysis. We hypothesized that we would replicate Forgas's finding that speakers formulate more polite requests in a negative mood, compared to speakers in a positive mood. Politeness of the requests was assessed based on the same six scales indicating request characteristics used in the original study, including (im)politeness. Additionally, we explored whether there was a relationship between affective state of the speaker and the usage of politeness strategies (Brown & Levinson, 1987).

Furthermore, to explore whether there are differences between autistic and non-autistic (so-called 'neurotypical') speakers, the experiment was performed with both neurotypical students (Study 1) and a small sample of autistic individuals (Study 2). We investigated whether there was a relationship between affective state and politeness for autistic speakers as well. Comparing the results of these two studies, we were able to explore whether autistic individuals were more, less, or equally polite as their neurotypical peers.

Affect induction was generally successful in both studies. But, contrary to our expectation, we were unable to replicate Forgas's findings, as the degree of politeness expressed by neurotypicals in Study 1 seemed independent of their affective state. In Study 2, however, we did find an effect of affect for autistic speakers in a negative affective state, who formulated more elaborate, hedged, indirect and complex requests, compared to their peers in a positive affective state. We found that, regardless of affective state, autistic speakers did not produce more or less polite requests, compared to neurotypical speakers. A more detailed analysis of different politeness strategies revealed no differences of affective state on politeness strategies used by either group of speakers.

Concluding, affective states influenced the language characteristics related to politeness in autistic individuals, but not in neurotypical students. We found an effect based on (some) request characteristics by Forgas, as described above, but not based on the ratings of politeness strategies. Based on the judgement by their peers, neurotypical and autistic speakers were about equally polite.

We have devoted this dissertation to the study of affective language production, focusing on social interactions. The findings of our studies indicated that emotion and affect, although sometimes subtly, influence language production. Affective state of speakers influenced their usage of affectively laden words (Chapter 2), adaptation to their conversation partner (Chapter 3 and Chapter 4) and verbal politeness when formulating requests (Chapter 5).

Based on our results, we conclude that affect influences how we converse together, affecting both our social behaviors, such as mimicking our conversation partner, as well as choosing the appropriate words to deliver a message.

REFERENCES

Achim, A. M., Achim, A., & Fossard, M. (2017). Knowledge likely held by other affects speakers' choices of referential expressions at different stages of discourse. *Language, Cognition and Neuroscience, 32*(1), 21-36. https://doi.org/10.1080/23273798.2016.1234059

Al-Duleimi, H. Y., Rashid, S. M., & Abdullah, A. N. (2016). A critical review of prominent theories of politeness. *Advances in Language and Literary Studies*, *7*(6), 262-270. http://dx.doi.org/10.7575/aiac.alls.v.7n.6p.262

Al-Hindawi, F. H., & Alkhazaali, M. A. R. (2016). A critique of politeness theories. *Theory and Practice in Language Studies*, *6*(8), 1537-1547. http://dx.doi.org/10.17507/tpls.0608.03

American Psychiatric Association (2013): Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Anderson, C. J., Bahník, Š., Barnett-Cowan, M., Bosco, F. A., & Chandler, J. Chartier, C. R., ... Zuni, K. (2016). Response to comment on estimating the reproducibility of psychological science. *Science*, *351*(6277), 1037-1039. https://doi.org/10.1126/science.aad9163

Arbuckle, T. Y., Nohara-LeClair, M., & Pushkar, D. (2000). Effect of off-target verbosity on communication efficiency in a referential communication task. *Psychology and Aging*, *15*(1), 65-77. https://doi.org/10.1037/0882-7974.15.1.65

Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics*, *43*(1), 361-374. https://doi.org/10.1016/j.pragma.2010.07.013

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Bachorowski, J.A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science, 8*(2), 53-57. https://doi.org/10.1111/1467-8721.00013

Bachorowski, J. A., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science*, *6*(4), 219-224. https://doi.org/10.1111/j.1467-9280.1995.tb00596.x

Baikie, K. A., & Wilhelm, K. (2005). Emotional and physical health benefits of expressive writing. *Advances in Psychiatric Treatment*, *11*(5), 338-46. https://doi.org/10.1192/apt.11.5.338

Balada, F., Blanch, A., & Aluja, A. (2014). Arousal and habituation effects (excitability) on startle responses to the International Affective Picture Systems (IAPS). *Journal of Psychophysiology*, *28*(4), 233-241. https://doi.org/10.1027/0269-8803/a000115

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5-17. https://doi.org/10.1023/A:1005653411471

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain.* Houghton Mifflin Harcourt.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *Lme4: Linear mixed-effects models using eigen and S4 (R package version 1).* ArXiv. https://arxiv.org/abs/1406.5823

Belek, B. (2018). Autism and the proficiency of social ineptitude: Probing the rules of 'appropriate' behavior. *Ethos*, *46*(2), 161-179. https://doi.org/10.1111/etho.12202

Bella, S., & Ogiermann, E. (2019). An intergenerational perspective on (im)politeness. *Journal of Politeness Research*, *15*(2), 163-193. https://doi.org/10.1515/pr-2017-0033

Betella, A., & Verschure, P. F. (2016). The affective slider: A digital self-assessment scale for the measurement of human emotions. *PloS ONE*, *11*(2), Article e0148037. https://doi.org/10.1371/journal.pone.0148037

Beukeboom, C. J., & Semin, G. R. (2006). How mood turns on language. *Journal of Experimental Social Psychology*, *42*(5), 553-566. https://doi.org/10.1016/j.jesp.2005.09.005

Bezuidenhout, A. (2013). Perspective taking in conversation: A defense of speaker non-egocentricity. *Journal of Pragmatics*, *48*(1), 4-16. https://doi.org/10.1016/j.pragma.2012.11.007

Binyamin-Suissa, L., Moyal, N., Naim, A., & Henik, A. (2019). Perspective taking and emotion: The case of disgust and sadness. *Consciousness and Cognition*, *74*, Article 102773. https://doi.org/10.1016/j.concog.2019.102773

Bird, G., & Cook, R. (2013). Mixed emotions: The contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry*, *3*(7), Article e285. https://doi.org/10.1038/tp.2013.61

Bohner, G., Crow, K., Erb, H. P., & Schwarz, N. (1992). Affect and persuasion: Mood effects on the processing of message content and context cues and on subsequent behaviour. *European Journal of Social Psychology*, *22*(6), 511-530. https://doi.org/10.1002/ejsp.2420220602

Bohn-Gettler, C. M., & Rapp, D. N. (2014). Emotions in reading and writing. In R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 437-457). Routledge.

Bölte, S., Feineis-Matthews, S., & Poustka, F. (2008). Brief report: Emotional processing in high-functioning autism—physiological reactivity and affective report. *Journal of Autism and Developmental Disorders*, *38*(4), 776-781. https://doi.org/10.1007/s10803-007-0443-8

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, *44*(2), 123-147. https://doi.org/10.1177/00238309010440020101

Boyes, M. E., Clarke, P. J. F., & Hasking, P. A. (2020). Relationships between dispositional and experimentally elicited emotional reactivity, intensity, and perseveration. *Personality and Individual Differences*, *152*, Article 109573. https://doi.org/10.1016/j.paid.2019.109573

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49-59. https://doi.org/10.1016/0005-7916(94)90063-9

Bradley, M. M., & Lang, P. J. (2017). International Affective Picture System. In V. Zeigler-Hill, & T. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences*. Springer. https://doi.org/10.1007/978-3-319-28099-8_42-1

Branigan, H. P., Catchpole, C. M., & Pickering, M. J. (2011). What makes dialogues easy to understand? *Language and Cognitive Processes*, *26*(10), 1667-1686. https://doi.org/10.1080/01690965.2010.524765

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*(2), 274-291. https://doi.org/10.1111/j.1756-8765.2009.01019.x

Broekens, J., & Brinkman, W. P. (2013). AffectButton: A method for reliable and valid affective self-report. *International Journal of Human-Computer Studies*, *71*(6), 641-667. https://doi.org/10.1016/j.ijhcs.2013.02.003

Brown, P. (2015). Politeness and language. In J. Wright (Eds.), *The International Encyclopedia of the Social and Behavioural Sciences* (2nd ed., pp. 326-330). Elsevier.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge University Press.

Burklund, L. J., Creswell, J. D., Irwin, M., & Lieberman, M. (2014). The common and distinct neural bases of affect labeling and reappraisal in healthy adults. *Frontiers in Psychology*, *5*, Article 221. https://doi.org/10.3389/fpsyg.2014.00221

Bynion, T. M., & Feldner, M.T. (2017). Self-Assessment Manikin. In V. Zeigler-Hill, & T. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences*. Springer. https://doi.org/10.1007/978-3-319-28099-8_77-1

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In M. Lewis, & J.M. Haviland-Jones (Eds.) *Handbook of Emotions* (2nd ed., pp. 173-191). Guilford Press.

Cai, R. Y., Richdale, A. L., Uljarević, M., Dissanayake, C., & Samson, A. C. (2018). Emotion regulation in autism spectrum disorder: Where we are and where we need to go. *Autism Research, 11*(7), 962-978. https://doi.org/10.1002/aur.1968

Capecelatro, M. R., Sacchet, M. D., Hitchcock, P. F., Miller, S. M., & Britton, W. B. (2013). Major depression duration reduces appetitive word use: An elaborated verbal recall of emotional photographs. *Journal of Psychiatric Research*, *47*(6), 809-815. https://doi.org/10.1016/j.jpsychires.2013.01.022

Carbary, K., & Tanenhaus, M. (2011). Conceptual pacts, syntactic priming, and referential form. In *Proceedings of the CogSci Workshop on the Production of Referring Expressions: Bridging the Gap Between Computational, Empirical and Theoretical Approaches to Reference (PRE-CogSci 2011).* (pp. 1-6). https://pre2011.uvt.nl/pdf/carbary-6pgs-wksp-2011.pdf

Carretero, L. M., Latorre, J. M., Fernández, D., Barry, T. J., & Ricarte, J. J. (2020). Effects of positive personal and non-personal autobiographical stimuli on emotional regulation in older adults. *Aging Clinical and Experimental Research*, *32*(1), 157-164. https://doi.org/10.1007/s40520-019-01147-0

Castro, N., & James, L. E. (2014). Differences between young and older adults' spoken language production in descriptions of negative versus neutral pictures. *Aging, Neuropsychology, and Cognition*, *21*(2), 222-238. https://doi.org/10.1080/13825585.2013.804902

Charash, M., McKay, D., & Dipaolo, N. (2006). Implicit attention bias for disgust. *Anxiety, Stress & Coping*, *19*(4), 353-364. https://doi.org/10.1080/10615800601055915

Chen, X., Fang, H., Lin, T. Y., Vedantam, R., Gupta, S., Dollár, P., & Zitnick, C. L. (2015). *Microsoft coco captions: Data collection and evaluation server.* arXiv. https://arxiv.org/abs/1504.00325

Clark, H. H. (1996). *Using language.* Cambridge University Press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1-39. https://doi.org/10.1016/0010-0277(86)90010-7

Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences*, *11*(9), 393-399. https://doi.org/10.1016/j.tics.2007.08.005

Cohen, J. E. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

Converse, B. A., Lin, S., Keysar, B., & Epley, N. (2008). In the mood to get over yourself: Mood affects Theory-of-Mind use. *Emotion, 8*(5), 725-730. https://doi.org/10.1037/a0013283

Cummins, C., Pellicano, E., & Crane, L. (2020). Autistic adults' views of their communication skills and needs. *International Journal of Language & Communication Disorders*, *55*(5), 678-689. https://doi.org/10.1111/1460-6984.12552

Dale, R., & Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, *19*(2), 233-263. https://doi.org/10.1016/0364-0213(95)90018-7

Darwin, C. R. (1872). *The expression of the emotions in man and animals*. John Murray.

DeBrabander, K. M., Morrison, K. E., Jones, D. R., Faso, D. J., Chmielewski, M., & Sasson, N. J. (2019). Do first impressions of autistic adults differ between autistic and nonautistic observers? *Autism in Adulthood*, *1*(4), 250-257. https://doi.org/10.1089/aut.2019.0018

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

De Looze, C., Scherer, S., Vaughan, B., & Campbell, N. (2014). Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*, *58*, 11-34. https://doi.org/10.1016/j.specom.2013.10.002

Derksen, B. J., Duff, M. C., Weldon, K., Zhang, J., Zamba, K. D., Tranel, D., & Denburg, N. L. (2015). Older adults catch up to younger adults on a learning and memory task that involves collaborative social interaction. *Memory*, *23*(4), 612-624. https://doi.org/10.1080/09658211.2014.915974

De Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in Cognitive Science*, *4*(2), 232-248. https://doi.org/10.1111/j.1756-8765.2012.01183.x

De Villiers, J., Fine, J., Ginsberg, G., Vaccarella, L., & Szatmari, P. (2007). Brief report: A scale for rating conversational impairment in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 37*(7), 1375-1380.
https://doi.org/10.1007/s10803-006-0264-1

Dhaka, S., & Kashyap, N. (2017). Explicit emotion regulation: Comparing emotion inducing stimuli. *Psychological Thought*, *10*(2), 303-314. https://doi.org/10.5964/psyct.v10i2.240

Dimitrova-Galaczi, E. (2002). Issues in the definition and conceptualization of politeness. *Studies in Applied Linguistics and TESOL*, *2*(1), 1-20. https://doi.org/10.7916/salt.v2i1.1650

Eelen, G. (2001). *A critique of politeness theory*. St. Jerome.

Ekkekakis, P. (2012). Affect, mood, and emotion. In G. Tenenbaum, R. C. Eklund, & A. Kamata (Eds.), *Measurement in sport and exercise psychology* (pp. 321–332). Human Kinetics.

Ekman, P., Freisen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology*, *39*(6), 1125-1134. https://doi.org/10.1037/h0077722

Engelen, T., & Mennella, R. (2020). *What is it like to be an emotion researcher?* PsyArXiv. https://doi.org/10.31234/osf.io/k34hp

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, *40*(6), 760–768. https://doi.org/10.1016/j.jesp.2004.02.002

Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science*, *16*(4), 779-788. https://doi.org/10.1177/1745691620970586

Fan, R., Varol, O., Varamesh, A., Barron, A., Van de Leemput, I. A., Scheffer, M., & Bollen, J. (2019). The minute-scale dynamics of online emotions reveal the effects of affect labeling. *Nature Human Behaviour*, *3*(1), 92-100.
https://doi.org/10.1038/s41562-018-0490-5

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

Feng, H. (2015). Politeness theory. In C. R. Berger, & M. E. Roloff (Eds.), *The International Encyclopedia of Interpersonal Communication* (1st ed., pp. 1-13). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118540190.wbeic164

Fernández-Aguilar, L., Navarro-Bravo, B., Ricarte, J., Ros, L., & Latorre, J. M. (2019). How effective are films in inducing positive and negative emotional states? A meta-analysis. *PloS ONE*, *14*(11), Article e0225040. https://doi.org/10.1371/journal.pone.0225040

Ferrer, R., Grenen, E., & Taber, J. (2015). Effectiveness of Internet-based affect induction procedures: A systematic review and meta-analysis. *Emotion*, *15*(6), 752-762. https://doi.org/10.1037/emo0000035

Fodor, J. A. (1985). Précis of the modularity of mind. *Behavioral and Brain Sciences*, *8*(1), 1-5. https://doi.org/10.1017/S0140525X0001921X

Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science*, *18*(12), 1050-1057. https://doi.org/10.1111/j.1467-9280.2007.02024.x

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, *117*(1), 39-66. https://doi.org/10.1037/0033-2909.117.1.39

Forgas, J. P. (1999a). On feeling good and being rude: Affective influences on language use and request formulations. *Journal of Personality and Social Psychology*, *76*(6), 928–939. https://doi.org/10.1037/0022-3514.76.6.928

Forgas, J. P. (1999b). Feeling and speaking: Mood effects on verbal communication strategies. *Personality and Social Psychology Bulletin*, *25*(7), 850-863. https://doi.org/10.1177/0146167299025007007

Forgas, J. P. (Ed.). (2012). *Affect in social thinking and behavior.* Psychology Press.

Forgas, J. P. (2013). Don't worry, be sad! On the cognitive, motivational, and interpersonal benefits of negative mood. *Current Directions in Psychological Science*, *22*(3), 225-232. https://doi.org/10.1177/0963721412474458

Forgas, J. P., & Matovic, D. (2020). Mood effects on humor production: Positive mood improves the verbal ability to be funny. *Journal of Language and Social Psychology*, 39*(5-6)*, 701-715. https://doi.org/10.1177/0261927X20917994

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998. https://doi.org/10.1126/science.1218633

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *The American Psychologist*, *56*(3), 218–226. https://doi.org/10.1037/0003-066X.56.3.218

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition and Emotion*, *19*(3), 313-332. https://doi.org/10.1080/02699930441000238

Gable, P. A., & Harmon-Jones, E. (2008). Approach-motivated positive affect reduces breadth of attention. *Psychological Science*, *19*(5), 476-482. https://doi.org/10.1111/j.1467-9280.2008.02112.x

Gable, P. A., & Harmon-Jones, E. (2010). The blues broaden, but the nasty narrows: Attentional consequences of negative affects low and high in motivational intensity. *Psychological Science*, *21*(2), 211–215. https://doi.org/10.1177/0956797609359622

Gable, P. A., & Harmon-Jones, E. (2016). Assessing the motivational dimensional model of emotion–cognition interaction: Comment on Domachowska, Heitmann, Deutsch, et al., (2016). *Journal of Experimental Social Psychology*, *67*, 57-59. https://doi.org/10.1016/j.jesp.2015.10.004

Gangamohan, P., Kadiri, S. R., & Yegnanarayana, B. (2016). Analysis of emotional speech—a review. In A. Esposito, & L. Jain (Eds.), *Toward robotic socially believable behaving systems (Vol. I)* (pp. 205-238). Springer.
https://doi.org/10.1007/978-3-319-31056-5_11

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 133–177). Academic Press.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11. https://doi.org/10.1016/j.tics.2003.10.016

Gasper, K. (2004). Do you see what I see? Affect and visual information processing. *Cognition and Emotion*, *18*(3), 405-421. https://doi.org/10.1080/02699930341000068

Gasper, K., & Clore, G. L. (2002). Attending to the big picture: Mood and global versus local processing of visual information. *Psychological Science*, *13*(1), 34-40. https://doi.org/10.1111/1467-9280.00406

Gatt, A., Krahmer, E., Van Deemter, K., & Van Gompel, R. P. G. (2017). Reference production as search: The impact of domain size on the production of distinguishing descriptions. *Cognitive Science*, *41*, 1457-1492. https://doi.org/10.1111/cogs.12375

Gatt, A., Van der Sluis, I., & Van Deemter, K. (2007). Evaluating algorithms for the generation of referring expressions using a balanced corpus. In *Proceedings of the Eleventh European Workshop on Natural Language Generation (ENLG 07)* (pp. 49-56). Association for Computational Linguistics.

Gelati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*(1), 35-51. https://doi.org/10.1016/j.jml.2009.09.002

Gerrards-Hesse, A., Spies, K., & Hesse, F. W. (1994). Experimental inductions of emotional states and their effectiveness: A review. *British Journal of Psychology*, *85*(1), 55-78. https://doi.org/10.1111/j.2044-8295.1994.tb02508.x

Gilam, G., & Hendler, T. (2016). With love, from me to you: Embedding social interactions in affective neuroscience. *Neuroscience & Biobehavioral Reviews*, *68*, 590-601. https://doi.org/10.1016/j.neubiorev.2016.06.027

Gillioz, C., Fontaine, J. R., Soriano, C., & Scherer, K. R. (2016). Mapping emotion terms into affective space. *Swiss Journal of Psychology*, *75*(3), 141–148. https://doi.org/10.1024/1421-0185/a000180

Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior.* Anchor Books.

Goldrick, M., Ferreira, V., & Miozzo, M. (2014). *The Oxford Handbook of Language Production.* Oxford University Press.

Göritz, A. S., & Moser, K. (2003). Mood and flexibility in categorization: A conceptual replication. *Perceptual and Motor Skills*, *97*(1), 107-119. https://doi.org/10.2466/pms.2003.97.1.107

Göritz, A. S., & Moser, K. (2006). Web-based mood induction. *Cognition and Emotion*, *20*(6), 887-896. https://doi.org/10.1080/02699930500405386

Gotham, K., Unruh, K., & Lord, C. (2015). Depression and its measurement in verbal adolescents and adults with autism spectrum disorder. *Autism, 19*(4), 491-504. https://doi.org/10.1177/1362361314536625

Goudbeek, M., & Krahmer, E. (2012). Alignment in interactive reference production: Content planning, modifier ordering and referential overspecification. *Topics in Cognitive Science*, *4*(2), 269-289. https://doi.org/10.1111/j.1756-8765.2012.01186.x

Goudbeek, M., & Scherer, K. (2010). Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. *Journal of the Acoustical Society of America*, *128*(3), 1322-1336. https://doi.org/10.1121/1.3466853

Griffin, Z. M., & Ferreira, V. S. (2006). Properties of spoken language production. In M. Traxler, & M. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed.) (pp. 21-59). Academic Press.

Grossman, R. B., Mertens, J., & Zane, E. (2019). Perceptions of self and other: Social judgments and gaze patterns to videos of adolescents with and without autism spectrum disorder. *Autism*, *23*(4), 846-857. https://doi.org/10.1177/1362361318788071

Haas, K., Costley, D., Falkmer, M., Richdale, A., Sofronoff, K., & Falkmer, T. (2016). Factors influencing the research participation of adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *46*(5), 1793-1805. https://doi.org/10.1007/s10803-016-2708-6

Heasman, B., & Gillespie, A. (2019). Neurodivergent intersubjectivity: Distinctive features of how autistic people create shared understanding. *Autism*, *23*(4), 910-921. https://doi.org/10.1177/1362361317708287

Hagger, M. S. (2014). Avoiding the 'déjà-variable' phenomenon: Social psychology needs more guides to constructs. *Frontiers in Psychology, 5*, Article 52.
https://doi.org/10.3389/fpsyg.2014.00052

Harlé, K. M., & Sanfey, A. G. (2010). Effects of approach and withdrawal motivation on interactive economic decisions. *Cognition and Emotion*, *24*(8), 1456-1465. https://doi.org/10.1080/02699930903510220

Hatfield, E., Cacioppo, J. L., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Sciences*, *2*(3), 96-100.
https://doi.org/10.1111/1467-8721.ep10770953

Haugh, M., & Chang, W. L. M. (2019). 'The apology seemed (in) sincere': variability in perceptions of (im) politeness. *Journal of Pragmatics, 142*, 207-222. https://doi.org/10.1016/j.pragma.2018.11.022

Hayashi, E. C., Posada, J. E. G., Maike, V. R., & Baranauskas, M. C. C. (2016). Exploring new formats of the Self-Assessment Manikin in the design with children. *Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems* (Article 27, pp. 1-10). https://doi.org/10.1145/3033701.3033728

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*(1), 77-89. https://doi.org/10.1080/19312450709336664

Heasman, B., & Gillespie, A. (2019). Neurodivergent intersubjectivity: Distinctive features of how autistic people create shared understanding. *Autism, 23*(4), 910-921. https://doi.org/10.1177/1362361318785172

Hoekstra, R. A., Vinkhuyzen, A. A., Wheelwright, S., Bartels, M., Boomsma, D. I., Baron-Cohen, S., Posthuma, D., & Van Der Sluis, S. (2011). The construction and validation of an abridged version of the autism-spectrum quotient (AQ-Short). *Journal of Autism and Developmental Disorders*, *41*(5), 589-596.
https://doi.org/10.1007/s10803-010-1073-0

Hofvander, B., Delorme, R., Chaste, P., Nydén, A., Wentz, E., Ståhlberg, O., Herbrecht, E., Stopin, A., Anckarsäter, H., Gillberg, C., Råstam, M., & Leboyer, M. (2009). Psychiatric and psychosocial problems in adults with normal-intelligence autism spectrum disorders. *BMC Psychiatry*, *9*, Article 35.
https://doi.org/10.1186/1471-244X-9-35

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91-117. https://doi.org/10.1016/0010-0277(96)81418-1

Hot, P., & Sequeira, H. (2013). Time course of brain activation elicited by basic emotions. *NeuroReport*, *24*(16), 898-902. https://doi.org/10.1097/WNR.0000000000000016

Hsiao, E. Y., Schwartz, M. F., Schnur, T. T., & Dell, G. S. (2009). Temporal characteristics of semantic perseverations induced by blocked-cyclic picture naming. *Brain and Language*, *108*(3), 133-144. https://doi.org/10.1016/j.bandl.2008.11.003

Hupet, M., Seron, X., & Chantraine, Y. (1991). The effects of the codability and discriminability of the referents on the collaborative referring procedure. *British Journal of Psychology*, *82*(4), 449-462. https://doi.org/10.1111/j.2044-8295.1991.tb02412.x

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216-225. https://ojs.aaai.org/index.php/ICWSM/article/view/14550

Imtiaz, N., Middleton, J., Girouard, P., & Murphy-Hill, E. (2018). Sentiment and politeness analysis tools on developer discussions are unreliable, but so are people. *Proceedings of the 3rd International Workshop on Emotion Awareness in Software Engineering (SEmotion '18)* (pp. 55-61). https://doi.org/10.1145/3194932.3194938

Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review*, *2*(4), 363-370. https://doi.org/10.1177/1754073910374661

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jallais, C., & Gilet, A. L. (2010). Inducing changes in arousal and valence: Comparison of two mood induction procedures. *Behavior Research Methods*, *42*(1), 318-325. https://doi.org/10.3758/BRM.42.1.318

James, W. (1890). *The principles of psychology (Vol. 1)*. Henry Holt and Company. http://archive.org/details/theprinciplesofp01jameuoft

Jennings, P. D., McGinnis, D., Lovejoy, S., & Stirling, J. (2000). Valence and arousal ratings for Velten mood induction statements. *Motivation and Emotion*, *24*(4), 285-297. https://doi.org/10.1023/A:1010745016868

Joseph, D. L., Chan, M. Y., Heintzelman, S. J., Tay, L., Diener, E., & Scotney, V. S. (2020). The manipulation of affect: A meta-analysis of affect induction procedures. *Psychological Bulletin*, *146*(4), 355-375. https://doi.org/10.1037/bul0000224

Kamiloğlu, R. G., Fischer, A. H., & Sauter, D. A. (2020). Good vibrations: A review of vocal expressions of positive emotions. *Psychonomic Bulletin & Review*, *27*(2), 237-265. https://doi.org/10.3758/s13423-019-01701-x

Kanske, P., & Kotz, S. A. (2010). Leipzig affective norms for German: A reliability study. *Behavior Research Methods*, *42*(4), 987-991. https://doi.org/10.3758/BRM.42.4.987

Kaplan, A. (1964). *The conduct of inquiry: Methodology for behavioral science*. Chandler Publishing Company.

Kempe, V., Rookes, M., & Swarbrigg, L. (2013). Speaker emotion can affect ambiguity production. *Language and Cognitive Processes*, *28*(10), 1579-1590. https://doi.org/10.1080/01690965.2012.755555

Kenealy, P. M. (1986). The Velten mood induction procedure: A methodological review. *Motivation and Emotion*, *10*(4), 315-335. https://doi.org/10.1007/BF00992107

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism*, *20*(4), 442-462. https://doi.org/10.1177/1362361315588200

Koch, A. S., Forgas, J. P., & Matovic, D. (2013). Can negative mood improve your conversation? Affective influences on conforming to Grice's communication norms. *European Journal of Social Psychology*, *43*(5), 326-334. https://doi.org/10.1002/ejsp.1950

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, *43*(13), 3231-3250. https://doi.org/10.1016/j.pragma.2011.06.008

Koolen, R., Goudbeek, M., & Krahmer, E. (2013). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*, *37*(2), 395-411. https://doi.org/10.1111/cogs.12019

Krahmer, E. J., Van Dorst, J., & Ummelen, N. (2004). Mood, persuasion and information presentation: The influence of mood on the effectiveness of persuasive digital documents. *Information Design Journal*, *12*(3), 219-232. https://doi.org/10.1075/idjdd.12.3.10kra

Krauss, R. M., & Fussell, S. R. (1991). Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition*, *9*(1), 2-24. https://doi.org/10.1521/soco.1991.9.1.2

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*, 113-114. https://doi.org/10.3758/BF03342817

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Sage.

Kuppens, P., Tuerlinckx, F., Yik, M., Koval, P., Coosemans, J., Zeng, K. J., & Russell, J. A. (2017). The relation between valence and arousal in subjective experience varies with personality and culture. *Journal of Personality*, *85*(4), 530-542. https://doi.org/10.1111/jopy.12258

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82(13)*, 1-26. https://doi.org/10.18637/jss.v082.i13

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, 39-58. https://www2.unifesp.br/dpsicobio/adap/instructions.pdf

Laukka, P., Juslin, P., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition and Emotion*, *19*(5), 633-653. https://doi.org/10.1080/02699930441000445

Lee, S., Yildirim, S., Kazemzadeh, A., & Narayanan, S. (2005). An articulatory study of emotional speech production. *Ninth European Conference on Speech Communication and Technology* (pp. 497-500). https://www.researchgate.net/publication/221480126_An_articulatory_study_of_emotional_speech_production

Leech, G. (2005). Politeness: is there an East-West divide? *Journal of Foreign Languages*, *6*, 1-30.

Lench, H. C., Flores, S. A., & Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: A meta-analysis of experimental emotion elicitations. *Psychological Bulletin*, *137*(5), 834-855. https://doi.org/10.1037/a0024244

Levelt, W. J. M. (1989). *Speaking: from intention to articulation.* MIT Press.

Levelt, W. J. M. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown, & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83–122). Oxford University Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words: Affect labeling disrupts amygdala activity in response to affective stimuli. *Psychological Science*, *18*(5), 421-428. https://doi.org/10.1111/j.1467-9280.2007.01916.x

Limonero, J. T., Fernández-Castro, J., Soler-Oritja, J., & Álvarez-Moleiro, M. (2015). Emotional intelligence and recovering from induced negative emotional state. *Frontiers in Psychology*, *6*, Article 816. https://doi.org/10.3389/fpsyg.2015.00816

Lindquist, K. A. (2017). The role of language in emotion: Existing evidence and future directions. *Current Opinion in Psychology, 17,* 135-139. https://doi.org/10.1016/j.copsyc.2017.07.006

Mackie, D. M., & Worth, L. T. (1989). Processing deficits and the mediation of positive affect in persuasion. *Journal of Personality and Social Psychology*, *57*(1), 27-40. https://doi.org/10.1037/0022-3514.57.1.27

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*(6), 537-542. https://doi.org/10.1177/1745691612460688

Manett, J. (2020). *Somewhere in the middle: Understandings of friendship and approaches to social engagement among postsecondary students with autism spectrum disorder* (Doctoral dissertation, University of Toronto]. TSpace. https://hdl.handle.net/1807/101360

Marcusson-Clavertz, D., Kjell, O. N., Persson, S. D., & Cardeña, E. (2019). Online validation of combined mood induction procedures. *PloS ONE*, *14*(6), Article e0217848. https://doi.org/10.1371/journal.pone.0217848

Mathersul, D., McDonald, S., and Rushby, J. A. (2013). Understanding advanced theory of mind and empathy in high-functioning adults with autism spectrum disorder. *Journal of Clinical and Experimental Neuropsychology*, *35*(6), 655-668. https://doi.org/10.1080/13803395.2013.809700

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the International Affective Picture System. *Behavior Research Methods*, *37*(4), 626-630. https://doi.org/10.3758/BF03192732

Mills, S. (2003). *Gender and politeness* (Vol. 17). Cambridge University Press.

Mirandola, C., & Toffalini, E. (2016). Arousal - but not valence - reduces false memories at retrieval. *PloS ONE*, *11*(3). Article e0148716. https://doi.org/10.1371/journal.pone.0148716

Morrison, K. E., DeBrabander, K. M., Jones, D. R., Faso, D. J., Ackerman, R. A., & Sasson, N. J. (2020). Outcomes of real-world social interaction for autistic adults paired with autistic compared to typically developing partners. *Autism*, *24*(5), 1067-1080. https://doi.org/10.1177/1362361319892701

Morse, C. R., & Afifi, W. A. (2015). I'm not rude, I'm just moody: The impact of affect on politeness in initial interactions. *Communication Research, 42*(1), 87-106. https://doi.org/10.1177/0093650212465432

Munezero, M., Montero, C. S., Sutinen, E., & Pajunen, J. (2014). Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text. *IEEE Transactions on Affective Computing*, *5*(2), 101-111. https://doi.org/10.1109/TAFFC.2014.2317187

Nummenmaa, L., Glerean, E., Viinikainen, M., Jaaskelainen, I.P., Hari, R., & Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(24), 9599-9604. https://doi.org/10.1073/pnas.1206095109

Ortner, C. N. (2015). Divergent effects of reappraisal and labeling internal affective feelings on subjective emotional experience. *Motivation and Emotion*, *39*(4), 563-570. https://doi.org/10.1007/s11031-015-9473-2

Out, C., Goudbeek, M., & Krahmer, E. (2020a). Gradual positive and negative affect induction: The effect of verbalizing affective content. *PloS ONE*, *15*(5), Article e0233592. https://doi.org/10.1371/journal.pone.0233592

Out, C., Goudbeek, M., & Krahmer, E. (2020b). Do Speaker's emotions influence their language production? Studying the influence of disgust and amusement on alignment in interactive reference. *Language Sciences*, *78*, Article 101255. https://doi.org/10.1016/j.langsci.2019.101255

Pająk, K., & Trzebiński, J. (2014). Escaping the world: Linguistic indicators of suicide attempts in poets. *Journal of Loss and Trauma*, *19*(5), 389-402. https://doi.org/10.1080/15325024.2013.794663

Paul, R., Orlovski, S. M., Marcinko, H. C., & Volkmar, F. (2009). Conversational behaviors in youth with high-functioning ASD and Asperger syndrome. *Journal of Autism and Developmental Disorders, 39*(1), 115-125. https://doi.org/10.1007/s10803-008-0607-1

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, *27*, 89–110. https://doi.org/10.1515/ling.1989.27.1.89

Pennebaker, J. W. (1993). Putting stress into words: Health, linguistic, and therapeutic implications. *Behaviour Research and Therapy*, *31*(6), 539-548. https://doi.org/10.1016/0005-7967(93)90105-4

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science*, *8*(3), 162-166.
https://doi.org/10.1111/j.1467-9280.1997.tb00403.x

Pennebaker, J. W., & Beall, S. K. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology, 95*(3), 274–281. https://doi.org/10.1037/0021-843X.95.3.274

Pennebaker, J. W., & Chung, C. K. (2007). *Expressive writing, emotional upheavals, and health*. In H. S. Friedman, & R. C. Silver (Eds.), *Foundations of health psychology* (pp. 263–284). Oxford University Press.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001*. Lawrence Erlbaum Associates.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*(1), 547-577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-190. https://doi.org/10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329-347. https://doi.org/10.1017/S0140525X12001495

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(3), 715-734. https://doi.org/10.1017/S0954579405050340

Qiu, L. (2015). A survey of conceptualizations of politeness. *International Journal of English Linguistics*, *5*(6), 115-121. http://dx.doi.org/10.5539/ijel.v5n6p115

Quené, H., & Van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, *43*(1-2), 103-121. https://doi.org/10.1016/j.specom.2004.02.004

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/index.html

Ritz, T., & Thöns, M. (2006). Affective modulation of swallowing rates: Unpleasantness or arousal? *Journal of Psychosomatic Research*, *61*(6), 829-833. https://doi.org/10.1016/j.jpsychores.2006.05.008

Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behaviour research: methods and data analysis*. McGraw-Hill.

Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology, 37*(3), 322–336.
https://doi.org/10.1037/0022-3514.37.3.322

Rottenberg, J. (2005). Mood and emotion in major depression. *Current Directions in Psychological Science*, *14*(3), 167-170. https://doi.org/10.1111/j.0963-7214.2005.00354.x

Rottenberg, J., Ray, R. D., & Gross, J. J. (2007). Emotion elicitation using films. In J. A. Coan, & J. J. B. Allen (Eds.), *The handbook of emotion elicitation and assessment* (pp. 9-28). Oxford University Press.

Rowe, G., Hirsh, J. B., & Anderson, A. K. (2007). Positive affect increases the breadth of attentional selection. *Proceedings of the National Academy of Sciences*, *104*(1), 383-388. https://doi.org/10.1073/pnas.0605198104

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161-1178. https://doi.org/10.1037/h0077714

Sasson, N. J., Faso, D. J., Nugent, J., Lovell, S., Kennedy, D. P., & Grossman, R. B. (2017). Neurotypical peers are less willing to interact with those with autism based on thin slice judgments. *Scientific Reports*, *7*, Article 40700. https://doi.org/10.1038/srep40700

Schaeffer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion*, *24*(7), 1153 – 1172. https://doi.org/10.1080/02699930903274322

Schaeffer, A., & Philippot, P. (2005). Selective effects of emotion on the phenomenal characteristics of autobiographical memories. *Memory*, *13*(2), 148-160. https://doi.org/10.1080/09658210344000648

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1-2), 227-256.
https://doi.org/10.1016/S0167-6393(02)00084-5

Scherer, K. R. (2013). The nature and dynamics of relevance and valence appraisals: Theoretical advances and recent evidence. *Emotion Review, 5*(2), 150-162. https://doi.org/10.1177/1754073912468166

Scherer, K. R. (2019). Acoustic patterning of emotion vocalizations. In S. Frühholz & P. Belin (Eds.) *The Oxford handbook of voice perception.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.4

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13*(2), 90–100. https://doi.org/10.1037/a0015108

Schneider, K. P., & Placencia, M. E. (2017). (Im) politeness and regional variation. In J. Culpeper, M. Haugh, & D.Z. Kádár (Eds.), *Palgrave Handbook of Linguistic (Im)politeness* (pp. 539-570). Palgrave Macmillan.
https://doi.org/10.1057/978-1-137-37508-7_21

Schwarz, N. (1990). Feelings as information: Informational and motivational functions of affective states. In E. T. Higgins, & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 527-561). The Guilford Press.

Schwarz, N., & Clore, G. L. (1988). How do I feel about it? Informative functions of affective states. In K. Fiedler, & J. Forgas (Eds.), *Affect, cognition, and social behavior* (pp. 44-62). Hogrefe International.

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, *32*(1), 3-23. https://doi.org/10.1023/A:1021928914454

Seidel, E. M., Habel, U., Kirschner, M., Gur, R.C., & Berntl, B. (2010). The impact of facial emotional expressions on behavioral tendencies in women and men. *Journal of Experimental Psychology: Human Perception and Performance, 36*(2)*,* 500-507. https://doi.org/10.1037/a0018169

Siegman, A. W., & Boyle, S. (1993). Voices of fear and anxiety and sadness and depression: The effects of speech rate and loudness on fear and anxiety and sadness and depression. *Journal of Abnormal Psychology*, *102*(3), 430. https://doi.org/10.1037/0021-843X.102.3.430

Sirota, K. G. (2004). Positive politeness as discourse process: Politeness practices of high-functioning children with autism and Asperger syndrome. *Discourse Studies*, *6*(2), 229-251. https://doi.org/10.1177/1461445604041769

Shevchenko, Y., & Bröder, A. (2019). Noncompliance with online mood manipulations using film clips: How to detect and control for it. *Heliyon*, *5*(4), Article e01438. https://doi.org/10.1016/j.heliyon.2019.e01438

Skinner, E. A. (1996). A guide to constructs of control. *Journal of Personality and Social Psychology*, *71*(3), 549–570. https://doi.org/10.1037/0022-3514.71.3.549

Smith, J. C., Bradley, M. M., & Lang, P. J. (2005). State anxiety and affective physiology: Effects of sustained exposure to affective pictures. *Biological Psychology*, *69*(3), 247-260. https://doi.org/10.1016/j.biopsycho.2004.09.001

Soares, A. P., Macedo, J., Oliveira, H. M., Lages, A., Hernández-Cabrera, J., & Pinheiro, A. P. (2019). Self-reference is a fast-acting automatic mechanism on emotional word processing: Evidence from a masked priming affective categorisation task. *Journal of Cognitive Psychology*, *31*(3), 317-325. https://doi.org/10.1080/20445911.2019.1599003

Sobin, C., & Alpert, M. (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research*, *28*(4), 347-365. https://doi.org/10.1023/A:1023237014909

Soliday, E., Garofalo, J. P., & Rogers, D. (2004). Expressive writing intervention for adolescents' somatic symptoms and mood. *Journal of Clinical Child & Adolescent Psychology*, *33*(4), 792-801. https://doi.org/10.1207/s15374424jccp3304_14

Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics 9: Pragmatics* (pp. 315-332). Academic Press. https://doi.org/10.1163/9789004368873_013

Sterponi, L. (2004). Construction of rules, accountability and moral identity by high-functioning children with autism. *Discourse Studies*, *6*(2), 207-228. https://doi.org/10.1177/1461445604041768

Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, *63*(4), 517-522. https://doi.org/10.1097/00006842-200107000-00001

Thomas, A., & Bambara, L. M. (2020). Using peer-mediation to enhance conversation and reduce inappropriate communication acts in adolescents with autism. *Education and Training in Autism and Developmental Disabilities*, *55*(2), 185-200.

Todd, A. R., Forstmann, M., Burgmer, P., Brooks, A. W., & Galinsky, A. D. (2015). Anxious and egocentric: How specific emotions influence perspective taking. *Journal of Experimental Psychology: General*, *144*(2), 374-391. https://doi.org/10.1037/xge0000048

Tong, E. M. W. (2015). Differentiation of 13 positive emotions by appraisals. *Cognition and Emotion*, *29*(3), 484-503. https://doi.org/10.1080/02699931.2014.922056

Torre, J. B., & Lieberman, M. D. (2018). Putting feelings into words: Affect labeling as implicit emotion regulation. *Emotion Review*, *10*(2), 116-124. https://doi.org/10.1177/1754073917742706

Uhrig, M. K., Trautmann, N., Baumgärtner, U., Treede, R. D., Henrich, F., Hiller, W., & Marschall, S. (2016). Emotion elicitation: A comparison of pictures and films. *Frontiers in Psychology, 7*, Article 180. https://doi.org/10.3389/fpsyg.2016.00180

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences*, *113*(23), 6454-6459. https://doi.org/10.1073/pnas.1521897113

Van Berkel, L., & Crandall, C. (2018). Conceptual replication of the automaticity of hierarchy: Multiple methods to test a hypothesis. In *SAGE Research Methods Cases.* https://www.doi.org/10.4135/9781526438720

Van Berkum, J. J., De Goede, D., Van Alphen, P., Mulder, E., & Kerstholt, J. H. (2013). How robust is the language architecture? The case of mood. *Frontiers in Psychology*, *4*, Article 505. https://doi.org/10.3389/fpsyg.2013.00505

Van Deemter, K., Gatt, A., Van Gompel, R. P.G., & Krahmer, E. (2012). Toward a computational psycholinguistics of reference production. *Topics in Cognitive Science*, *4*(2), 166–183. https://doi.org/10.1111/j.1756-8765.2012.01187.x

Van Gompel, R. P., van Deemter, K., Gatt, A., Snoeren, R., & Krahmer, E. J. (2019). Conceptualization in reference production: Probabilistic modeling and experimental testing. *Psychological Review, 126*(3), 345–373. https://doi.org/10.1037/rev0000138

Van Raalte, J. L., Vincent, A., & Brewer, B. W. (2016). Self-talk: Review and sport-specific model. *Psychology of Sport and Exercise*, *22*, 139-148. https://doi.org/10.1016/j.psychsport.2015.08.004

Van der Zwaag, M. D., Janssen, J. H., Nass, C., Westerink, J. H., Chowdhury, S., & De Waard, D. (2013). Using music to change mood while driving. *Ergonomics, 56*(10), 1504-1514. https://doi.org/10.1080/00140139.2013.825013

Velikonja, T., Fett, A. K., & Velthorst, E. (2019). Patterns of nonsocial and social cognitive functioning in adults with autism spectrum disorder: a systematic review and meta-analysis. *JAMA Psychiatry*, *76*(2), 135-151. https://doi.org/10.1001/jamapsychiatry.2018.3645

Velten Jr, EC. (1967). *The induction of elation and depression through the reading of structured sets of mood-statements* (No. 6713045) [Doctoral dissertation, University of Southern California]. ProQuest Dissertations and Theses global.

Velten Jr, E. (1968). A laboratory task for induction of mood states. *Behaviour Research and Therapy, 6*(4), 473-482. https://doi.org/10.1016/0005-7967(68)90028-4

Vice. (2012, 14 December). うさぎピョンピョン大会 - *Cute Bunny Jumping Competition!* [Video]. YouTube. https://youtu.be/3UJh7x-rH8BQ

Viethen, J., & Dale, R. (2010). Speaker-dependent variation in content selection for referring expression generation. *Proceedings of the Australasian Language Technology Association Workshop 2010* (pp. 81-89).

https://www.aclweb.org/anthology/U10-1013.pdf

Viethen, J., Van Vessem, T., Goudbeek, M., & Krahmer, E. (2017). Color in reference production: The role of color similarity and color codability. *Cognitive Science, 41*(S6), 1493-1514. https://doi.org/10.1111/cogs.12387

Vigliocco, G., & Hartsuiker, R.J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin*, *128*(3), 442-472. https://doi.org/10.1037//0033-2909.128.3.442

Volden, J., & Sorenson, A. (2009). Bossy and nice requests: Varying language register in speakers with autism spectrum disorder (ASD). *Journal of Communication Disorders*, *42*(1), 58-73. https://doi.org/10.1016/j.jcomdis.2008.08.003

Vonk, P., Goudbeek, M., & Krahmer, E. (2021). The influence of affect on the production of referring expressions. *Language, Cognition and Neuroscience*, in press. https://doi.org/10.1080/23273798.2021.1955139.

Westermann, R., Spies, K., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: A meta-analysis. *European Journal of Social Psychology, 26*(4), 557-580.

https://doi.org/10.1002/(SICI)1099-0992(199607)26:4<557::AID-EJSP769>3.0.CO;2-4

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, *31*(2), 183-194.

https://doi.org/10.1016/0749-596X(92)90010-U

Wilting, J., Krahmer, E., & Swerts, M. (2006). Real vs. acted emotional speech. *Ninth International Conference on Spoken Language Processing* (pp. 805-808). https://www.researchgate.net/publication/221486724_Real_vs_acted_emotional_speech

Winans, M. (2020). Email requests: Politeness evaluations by instructors from diverse language backgrounds. *Language Learning & Technology*, *24*(2), 104-118. http://hdl.handle.net/10125/44728

Yang, C., Prud'hommeaux, E., Silverman, L. B., & Canfield, A. (2020). Toward characterizing the language of adults with autism in collaborative discourse. In D. Kokkinakis, K. Lundholm Fors, C. Themistocleous, M. Antonsson, & M. Eckerström (Eds.), *Language Resources and Evaluation Conference (LREC 2020)* (pp. 54-59).

Yang, M. L., Yang, C. C., & Chiou, W. B. (2010). When guilt leads to other orientation and shame leads to egocentric self-focus: Effects of differential priming of negative affects on perspective taking. *Social Behavior and Personality: An International Journal*, *38*(5), 605-614. https://doi.org/10.2224/sbp.2010.38.5.605

Yip, J. A., & Schweitzer, M. E. (2019). Losing your temper and your perspective: Anger reduces perspective-taking. *Organizational Behavior and Human Decision Processes*, *150*, 28-45. https://doi.org/10.1016/j.obhdp.2018.07.003

Yoon, S. O., & Brown-Schmidt, S. (2019). Contextual integration in multiparty audience design. *Cognitive Science*, *43*(12), Article e12807. https://doi.org/10.1111/cogs.12807

Young, M. J., Tiedens, L. Z., Jung, H., & Tsai, M. H. (2011). Mad enough to see the other side: Anger and the search for disconfirming information. *Cognition and Emotion*, *25*(1), 10-21. https://doi.org/10.1080/02699930903534105

Zalla, T., Amsellem, F., Chaste, P., Ervas, F., Leboyer, M., & Champagne-Lavau, M. (2014). Individuals with autism spectrum disorders do not use social stereotypes in irony comprehension. *PloS one*, *9*(4), Article e95568. https://doi.org/10.1371/journal.pone.0095568

Zijlstra, H., Van Middendorp, H., Van Meerveld, T., & Geenen, R. (2005). Validiteit van de Nederlandse versie van de Linguistic Inquiry and Word Count (LIWC*). Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden, 60*(3), 50-58. https://doi.org/10.1007/BF03062342

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*. Article e120.
https://doi.org/10.1017/S0140525X17001972

Dankwoord

DANKWOORD

Zoals velen van jullie weten, wist ik vanaf mijn tweede studiejaar (Psychologie, Leiden) dat ik een PhD wilde gaan doen. Toen ik eind 2015 in Tilburg een sollicitatie gesprek had voor een PhD positie en ik uitgeteld in de trein zat, kreeg ik al snel een e-mail dat ik deze wens eindelijk in vervulling mocht laten gaan. Na al die jaren is het dan zover: eindelijk heb ik mijn titel en mag ik zelfstandig de wetenschap bedrijven… al hoop ik dat vooral samen met mensen te blijven doen, want dat is toch veel leuker! Het was een lange weg om daadwerkelijk te promoveren en daarom wil ik graag een aantal mensen bedanken.

Ten eerste mijn ALP 'team': Emiel Krahmer, Martijn Goudbeek en (dr.!) Nadine Braun. Bedankt voor alle (veel te lange) gezellige meetings, waarbij Emiel altijd nog even koffie moest halen, Martijn nog even iets moest bespreken met Emiel (waarbij ik nooit wist waar het nou over ging), Nadine ons volstopte met al dan niet zelfgemaakte zoetigheden ('no, take it!') en ik op een gegeven moment ergens wel een emotie over had, die ik best uitgebreid wilde toelichten (#noshame). Jullie staan niet voor niets op de achterkant van dit boekje, want echt uitdrukken wat jullie voor mij en mijn wetenschappelijke leven hebben betekend, valt lastig in woorden te vatten. Maar ik zal het proberen.
Beste Emiel, bedankt dat je een psycholoog uit Leiden aannam om dit project te gaan doen (en nog meer bedankt dat je alvast een beurs had binnen gesleept). Als taalkundige ben je ongelofelijk goed met woorden, waardoor a) mijn artikelen nu ook elegante zinnen bevatten en b) ik je soms heb moeten lastigvallen met mijn kritische vragen… die je dan weer heel wetenschappelijk genuanceerd kon beantwoorden. Bedankt voor het delen van je brede kennis en enthousiasme, en de gesprekken over zowel de nieuwste rap albums als het wel en wee van de academische wereld. Je vriendelijkheid, geduld en oprechte interesse in anderen blijft niet onopgemerkt.
Beste Martijn, jouw sarcastische humor maakte elke meeting beter… bonuspunten als het wordt uitgeoefend op externen die een onschuldig praatje kwamen houden. Bedankt voor de lange sessies op de gang of in onze kantoren, waarbij we makkelijk een uur konden doorzagen over statistische analyses en Good Scientific Conduct. Op dat soort momenten voelde ik mij een heel gelukkige wetenschapper: in het gezelschap van een soortgenoot, lekker op alle slakken zout leggen om mijn werk zo goed mogelijk te doen. Bedankt hiervoor!
Dear Nadine, thank you for being my office mate and ALP-buddy. It took a while for us to get used to each other, but at some point, it 'clicked' (for me at least) and voila, harmony in D401. Thanks for patiently listening to my rants (about food, weightlifting, friends and cute people) and our discussions about, well, anything that came to mind, from family shizzle and broken bones to life in academia ('I'm so confused!'). I will never forget your passionate driving, and I will also never hug you (unless you sign an informed consent letter first, of course).

Ten tweede wil ik mijn commissie bedanken: Vera Kempe, Jos van Berkum, Mariët Theune, Mirjam Broersma en Fons Maes. Bedankt dat jullie de moeite en tijd hebben genomen om

dit proefschrift te lezen en van commentaar te voorzien, zodat ik het net wat beter kon maken. Jos, ik ken je niet, maar je pittige feedback maakte mij heel nieuwsgierig naar jou als mede kritische wetenschapper. Fons, voor mij zou je altijd de 'nestor' blijven van DCC. Ik hoop dat je tijdens de promotie mij hebt kunnen verblijden met je uitgebreide verhalen (blijf vooral doorgaan…).

Alle proefpersonen die deelnamen aan mijn experiment, hartelijk dank. Zonder jullie was dit niet mogelijk geweest, cliché maar waar. In het bijzonder wil ik de mensen bedanken die mijn experiment in Chapter 5 hebben waargemaakt. Wat begon als een wild idee, is uitgemond in een bijzondere ervaring. Door jullie is mijn passie voor onderzoek naar de 'algemene populatie' (niet-studenten) en neurodiverse mensen alleen maar groter geworden. Dank aan alle mensen bij AutiRoze, en in het bijzonder Iris Westhoff en Ivan Henczyk, voor jullie positieve houding en vriendelijkheid. Jullie doen goed werk en ik blijf mensen in mijn omgeving wijzen op jullie organisatie! Als het artikel is gepubliceerd kom ik nog een keer langs om het te presenteren (of eerder, want de resultaten kunnen jullie al lezen in dit boekje…). Ook wil ik graag bedanken: Jeanette van Rees (Universiteit Utrecht), PAS Nederland (en natuurlijk Diederik Weve), het Autsider forum en iedereen die mij een zetje heeft gegeven bij het rekruteren van proefpersonen en mij meer leerde over de autistische gemeenschap.

Ook wil ik alle collega's van DCC bedanken, in het bijzonder Lauraine en Yan. Tijdens corona hielden wij de vierde verdieping warm! Het was heel fijn om steeds toch nog twee collega's te kunnen zien, ondanks alle social distancing. Lauraine, bedankt voor je tomeloze positieve houding, verhalen over je wandelavonturen en je behulpzaamheid. Je bent altijd vriendelijk, geen wonder dat iedereen zo dol op jou is! Yan, thank you for being you. I appreciate all the talks we had about cultural differences, emotions, academia and expectations from the outside world (about life, being a woman, etc.). Jacqueline, ik mocht jou meteen. Bedankt voor alle informatie over het lab, de onderonsjes op de gang, je directheid, het delen van je zelfinzicht en je wijsheid over het leven. Ik hoop dat je lekker blijft doorgaan met programmeren en IT'en! Ik ga je missen.
Ook wil ik mijn andere (oud) collega's bedanken. Angelique, jij bent mooi, slim en vriendelijk… some have it all. Bedankt dat je mij ooit in vertrouwen nam en mij hielp toen ik koffie over mezelf heen gooide bij de Beekse Bergen. Ik kom je een keer opzoeken hoor! Nadine Bol, never change. Bedankt voor de mooie uitgaansverhalen, je chocolade en levendigheid. Jos: blijven genieten in Hong Kong, en blijven rennen hè! Jan de Wit, het was heel gezellig met jou bij het schrijfkamp. Kom vooral een keer een biertje doen en verder praten over dates! Saar en Ruben, bedankt voor alles wat jullie hebben georganiseerd en mij doen inzien dat wetenschap echt het leven van mensen kan verbeteren. Jullie doen mooi werk!
Verder: Alex, Alwin (one of us!), Peter Broeder (aka de bonusvader van al zijn studenten), Jan Engelen, Thiago, Neil, Chrissy (the christian Canadian… thanks for the dinnerparties and drinks!), Marije, Rein, Emmelyn, Debby, Hendrik, Renske, Myrthe, Frans, Marlies, Yulie, Linwei (cool and calm), Naomi, Ruud, Nynke, Christine (ik zal altijd blijven letten op slechte

slogans), Liesje (uitstekende danser met pretogen), Janneke, Maria (politiek!), Emiel v. M., Annemarie (jij kleed je echt goed), Tess (ja, jij dus ook… maar goed, dat heb ik jullie al vaker verteld), Leonoor, David (de beste opmerkingen), Monique (het kan BETER), Diana, Joost, Karin, Door, Marc, Kim (keep the party goin'), Mariek, Connie, Nadine van der Waal (lief, en jij kan goed dansen!), Carel, Per (en je hond!), Sander, Lane, Marieke (Haarlem!), Chris ('yeah'), Marjolijn en iedereen die ik ben vergeten.

Dan ook nog de mensen die mij hebben geholpen met mijn beursaanvraag voor ZonMw naar de maatschappelijke gevolgen van de corona crisis, toen we allemaal in lockdown zaten. Het is niet gelukt, maar het was op één van de hoogtepunten in mijn wetenschappelijke leven, tot nu toe! Ilja, bedankt dat je op mijn onhandige mailtjes heb gereageerd. Ik vond het best wel spannend, want ik had je al hoog zitten voordat we begonnen… ik bedoel, sociale exclusie, dat is mijn favoriete wetenschappelijke onderwerp… Bedankt voor de samenwerking, je soms vaderlijke houding en wie weet kunnen we in de toekomst nog eens onderzoek doen! Tineke, bedankt voor je enthousiasme en het delen van je kennis. Ik wist helemaal niets over kwalitatief onderzoek en nu… iets! Juliette, bedankt dat je insprong op het moment dat het echt hard nodig was. Zonder blikken of blozen zorgde jij voor een net financieel plan. Wauw. Arne Mellaard, wat gaaf dat ik het artikel van GGD Zuid Holland Zuid tegen kwam en bij jou uit kwam. Bedankt voor het vertrouwen in het project en eh… ben je al gepromoveerd?

Verder wil ik nog mijn nieuwe collega's bedanken bij Developmental Psychology. Stefan, bedankt dat je mij met open armen ontving toen ik langskwam om 'even te praten over de vacature' en mij de kans hebt gegeven om les te geven bij DP. Annet en Bart, de office mates van S216, bedankt voor alle gesprekken en wederzijdse steun als we weer eens proberen uit te vogelen hoe alles zit.

Yorim, bedankt dat je mijn paranimf en vriend bent. Het is altijd fijn om bij je te logeren en te kletsen over van alles en nog wat (toetsenborden, vrouwen, werk, geluidsoverlast…). Ik waardeer je directheid (al is het soms even slikken), creativiteit, je honderdeen projecten (die je dan ook echt doet!) en stevige knuffels. Nu ik ben gepromoveerd: laten we weer wat vaker bellen, en ik wil echt nog een goa trance feest meemaken!

Dan ook nog mijn andere vrienden. Jarno, we leerden elkaar kennen in de zomer van 2018, in de sportschool. Ik geniet van alle sportmomenten met jou – of we nou aan het knallen zijn, of een beetje aan het ouwehoeren en af en toe een dumbell pakken (al zal ik daar natuurlijk altijd wat van zeggen) – en ook alle niet-sportmomenten. Ik hoop nog vaak een bak koffie bij je te mogen drinken. Soms baal ik ervan dat ik je niet méér zie – je hebt altijd zoveel leuke ideeën! – maar dan hoor ik jou weer 'geen probleem!'. Het is een voorrecht je te kennen en ik hoop dat je je lekker blijft ontwikkelen (net als die triceps van jou).
Annemiek, het was zo fijn om met jou te kunnen praten over onze PhD: zowel de kickass 'powervrouw' momenten als de struggles. Ik waardeer je smaak (zowel qua eten als

esthetiek), de leuke quality momenten samen (bioscoop, etentjes, COCKTAILS en crea-middagen) en jouw doorzettingsvermogen. Heel veel plezier en geluk in je nieuwe huisje, met Fons (Fons Fons Fons, I'm looking for a good time!), je ontwikkeling als psychotherapeut en in de toekomst wellicht een extra mensje om je heen. En zo niet, dat je dan mag blijven genieten van je leuke nichtje en al het andere grut! Ik zou altijd aan jou denken bij witte chocolade en cocktails. En eh… ja, *dat* feestje moeten we nog doen!
Lieve, mooie Mandy: vanaf de eerste keer dat jij op de bank zat bij mij en Luc, was daar die beroemde klik: alsof we elkaar al een tijd kenden. We waren al snel onder de indruk van jouw ouderwetse hoffelijkheid: vaak een presentje mee, netjes op tijd, respect voor onze bedtijd… Wat volgde was een mooie vriendschap. Het is altijd leuk als jij langskomt: een kop koffie of thee, uitgebreide verhalen (van de dagelijkse beslommeringen tot je diepste zielenroerselen), en je soms heel duidelijk mening (waarbij je geïnteresseerd luistert naar mensen die het niet met je eens zijn). Je zelfontwikkeling is heel mooi om te zien. Je bent een sterke vrouw en voor jou houd ik graag een alcoholvrij biertje koud.
Beste, lieve Inge, bedankt dat jij het initiatief nam om in de corona crisis te beginnen met onze wekelijkse sportmomenten. Ik had nooit gedacht dat we het vandaag de dag nog steeds trouw doen! Je respect voor 'onze' donderdagavond is heel mooi en ik blijf heel blij dat je steeds weer leuke sportfilmpjes uitzoekt (en het dan niet erg vind als ik liever een andere doe). Bedankt voor je verhalen en dat je je kwetsbaar durft op te stellen, het blijft (als flapuit) heel leerzaam om met jou te praten! Je bent altijd welkom in Tilburg. En eh… gefeliciteerd met je nieuwe baan!

Alle leuke mensen uit Leiden, bedankt voor de borrels, thema feestjes en bbq's en al het leuks wat nog gaat komen: Berend, Sietske, Sander (de beste zachte en stevige knuffels), Kim, Anna, Martin (je bent weer in Nederland!), Rosina, Jolanda (lindy hop!), Mariska, Sandra, Tom en de rest. Hef het glas: 'CULTUUR!'
Verder nog een shout out naar Liselot (jij bent lief), Wouter en Esmee (party animals en huismussen in één, toch mooi), Chantal (met je personal trainer) en alle bezoekers van mijn borrels.

Mam, bedankt voor je steun bij alles wat ik doe en onze wekelijkse belsessies. Ik ben nu echt gepromoveerd! Bedankt dat je altijd vertrouwen hebt in mijn academische skills en mij respecteert in de keuzes die ik maak (ook als ik niet weet wat ik wil of moet doen). Je bent zelf enorm gegroeid en daar ben ik ook heel trots op.
Henk, bedankt voor je rust, respect en je humor. Ik bewonder je doorzettingsvermogen met het fietsen. Door jou zie ik dat je je leven lang kan sporten, dat inspireert.
Nicky, mijn grote broer, over groei gesproken. Ik kijk er altijd naar uit om je te zien en je te spreken. Niet alleen omdat we best op elkaar kunnen lijken, maar ook je rust, geduld en oplettendheid zijn heel prettig. Mooi om te zien dat je steeds meer je weg lijkt te vinden.
Carien, dondersteen, zorg goed voor jezelf en voor m'n broer hè? Ik hoop dat je nog veel mag blijven naaien en je het werk vind wat bij je past. Je kan goed koken, ik hoop nog vaak te mogen genieten van je gerechten en je getroetel met de beestjes.

Aileen en Danny, fijn dat we elkaar af en toe zien! Aileen, echt vet wat je hebt bereikt in je carrière! Ik vind het knap dat je jezelf zo in de spotlight kan zetten en je je bedrijf vanaf de grond hebt kunnen opbouwen tot wat het nu is. Zorg goed voor elkaar (en Suit!) en hopelijk tot Kerst, of een ander moment.

'Kom ik ook in het dankwoord dan?'. Ja Luc, jij ook. Lieve Luc, draakje, van een feestje in een Rotterdamse kelder (dat klinkt als een film hè) tot samenwonen in Tilburg Noord, wie had dat gedacht? Bedankt dat je er altijd voor mij bent geweest tijdens mijn promotie. Toen we elkaar leerden kennen was ik net begonnen aan mijn PhD en net in Tilburg. Jij reisde zonder moeite elke keer op en neer van Bennekom naar Tilburg, om mij te zien, vaak op vijf uur slaap. Wat hebben we ondertussen veel meegemaakt samen. Verschillende banen, de vakantie naar Nice, een vouwwagen, een enkelbreuk (bedankt dat je gewoon in het vliegtuig stapte, schat), longontsteking, een fantastische housewarming, Bastiaan modestad in Bavel, Holten, hobby'en, mountainbiken (en gaan!)… Ik ben blij om te zien hoe je steeds vaker weer je passies uitoefent (kamperen, honkballen…) en steeds meer leuke mensen om je heen verzameld. Ik ben trots op jou! Bedankt voor je geduld met mij tijdens de PhD en het respect als ik een zaterdagochtend 'nog even een paar uurtjes' wilde werken (…of ook nog een middag, shit). Bedankt voor alle ruimte die je mij geeft als ik even alleen wil zijn (om te rusten, te sporten of iets creatiefs te koken) of juist een weekendje de hort op wil….en dan het respect als ik toch nog even tegen je aan wil kruipen. Ik geniet van ons geklier, de gedeelde passie voor feestjes organiseren, jouw blije puppy momenten en kijk nog steeds vol bewondering naar jouw respectvolle kijk op alles en iedereen. Op naar een mooie toekomst samen, waarin we zelf kunnen invullen hoe traditioneel of modern onze levens eruit gaan zien, en met wie, wat en waar dat ook mag zijn.

LIST OF PUBLICATIONS

Journal publications

Out, C., Goudbeek, M., & Krahmer, E. (2021). *Is there a role for affective state in conceptual pact formation?* [Manuscript submitted for publication]. Department of Communication and Cognition, Tilburg University.

Out, C., Goudbeek, M., & Krahmer, E. (2021). An intimate conversation: The influence of affective state on (im)polite request formulations of neurotypical and autistic speakers in dyads. [Manuscript under revision]. Department of Communication and Cognition, Tilburg University.

Out, C., Goudbeek, M., & Krahmer, E. (2020). Gradual positive and negative affect induction: The effect of verbalizing affective content. *PloS ONE*, *15*(5), Article e0233592. https://doi.org/10.1371/journal.pone.0233592

Out, C., Goudbeek, M. G., & Krahmer, E. J. (2020). Do Speaker's emotions influence their language production? Studying the influence of disgust and amusement on alignment in interactive reference. *Language Sciences*, 78, Article 101255. https://doi.org/10.1016/j.langsci.2019.101255

Papers in conference proceedings

Out, C., Goudbeek, M. B., & Krahmer, E.J. (2017). Do speaker's emotions influence their language production?: Studying the influence of disgust and amusement on alignment in interactive reference. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 2846-2851). Cognitive Science Society. https://cogsci.mindmodeling.org/2017/cogsci17_proceedings.pdf

(Abstracts of) conference presentations

Out, C., Goudbeek, M. B., & Krahmer, E. J. (2020, postponed). Does describing emotional images in a conversation induce affect? [Conference presentation]. *8th Consortium of European Research on Emotion (CERE)*.

Out, C., Goudbeek, M. B., & Krahmer, E.J. (2019, 13-15 September). The influence of mood on politeness in autistic and neurotypical individuals [Conference presentation abstract]. *12th Autism-Europe International Congress*. Nice, France. https://www.autismeurope.org/wp-content/uploads/2019/10/ABSTRACT_BOOK_AE12th-Congress-2019.pdf

Goudbeek, M., Krahmer, E., Out, C., & Braun, N. (2019, 16-18 September). Producing affective language: experimental and corpus based approaches [Conference session]. *Emotions 2019, the 7th International Conference on Emotions, Well-being, and Health*, Tilburg, the Netherlands.

Out, C., Goudbeek, M. B., & Krahmer, E. J. (2019, 16-18 September). The influence of emotional state on politeness: Comparing neurotypical and ASD speakers [Conference presentation]. *Emotions 2019, the 7th International Conference on Emotions, Well-Being, and Health*, Tilburg, the Netherlands.

Out, C., Goudbeek, M., & Krahmer, E. J. (2018, 4-5 April). Gradual emotion induction with a visual Velten method using images [Conference presentation abstract]. *7th Consortium of European Research on Emotion (CERE),* Glasgow, Scotland (UK).
http://www.cere-emotionconferences.org/wp-content/uploads/2016/01/CERE-Full-Program_ABSTRACTS_April3.pdf
Out, C., Goudbeek, M., & Krahmer, E.J. (2017, 21-23 June). Alignment in naturalistic dialogue: Language production in interactive reference production [Conference presentation]. *7th biannual Experimental Pragmatics Conference*, Cologne, Germany.
https://xprag2017.uni-koeln.de/sites/xprag2017/user_upload/OutGoudbeekKrahmer.pdf

TICC PHD SERIES

1. Pashiera Barkhuysen. *Audiovisual prosody in interaction*. Promotores: M. G. J. Swerts, E. J. Krahmer. Tilburg, 3 October 2008.
2. Ben Torben-Nielsen. *Dendritic morphology: Function shapes structure*. Promotores: H. J. van den Herik, E. O. Postma. Co-promotor: K. P. Tuyls. Tilburg, 3 December 2008.
3. Hans Stol. *A framework for evidence-based policy making using IT*. Promotor: H. J. van den Herik. Tilburg, 21 January 2009.
4. Jeroen Geertzen. *Dialogue act recognition and prediction*. Promotor: H. Bunt. Co-promotor: J. M. B. Terken. Tilburg, 11 February 2009.
5. Sander Canisius. *Structured prediction for natural language processing*. Promotores: A. P. J. van den Bosch, W. Daelemans. Tilburg, 13 February 2009.
6. Fritz Reul. *New architectures in computer chess*. Promotor: H. J. van den Herik. Co-promotor: J. W. H. M. Uiterwijk. Tilburg, 17 June 2009.
7. Laurens van der Maaten. *Feature extraction from visual data*. Promotores: E. O. Postma, H. J. van den Herik. Co-promotor: A. G. Lange. Tilburg, 23 June 2009 (cum laude).
8. Stephan Raaijmakers. *Multinomial language learning*. Promotores: W. Daelemans, A. P. J. van den Bosch. Tilburg, 1 December 2009.
9. Igor Berezhnoy. *Digital analysis of paintings*. Promotores: E. O. Postma, H. J. van den Herik. Tilburg, 7 December 2009.
10. Toine Bogers. *Recommender systems for social bookmarking*. Promotor: A. P. J. van den Bosch. Tilburg, 8 December 2009.
11. Sander Bakkes. *Rapid adaptation of video game AI*. Promotor: H. J. van den Herik. Co-promotor: P. Spronck. Tilburg, 3 March 2010.
12. Maria Mos. *Complex lexical items*. Promotor: A. P. J. van den Bosch. Co-promotores: A. Vermeer, A. Backus. Tilburg, 12 May 2010 (in collaboration with the Department of Language and Culture Studies).
13. Marieke van Erp. *Accessing natural history. Discoveries in data cleaning, structuring, and retrieval*. Promotor: A. P. J. van den Bosch. Co-promotor: P. K. Lendvai. Tilburg, 30 June 2010.
14. Edwin Commandeur. *Implicit causality and implicit consequentiality in language comprehension*. Promotores: L.G. M. Noordman, W. Vonk. Co-promotor: R. Cozijn. Tilburg, 30 June 2010.
15. Bart Bogaert. *Cloud content contention*. Promotores: H. J. van den Herik, E. O. Postma. Tilburg, 30 March 2011.
16. Xiaoyu Mao. *Airport under control*. Promotores: H. J. van den Herik, E. O. Postma. Co-promotores: N. Roos, A. Salden. Tilburg, 25 May 2011.
17. Olga Petukhova. *Multidimensional dialogue modelling*. Promotor: H. Bunt. Tilburg, 1 September 2011.
18. Lisette Mol. *Language in the hands*. Promotores: E. J. Krahmer, A. A. Maes, M. G. J. Swerts. Tilburg, 7 November 2011 (cum laude).

19. Herman Stehouwer. *Statistical language models for alternative sequence selection*. Promotores: A. P. J. van den Bosch, H. J. van den Herik. Co-promotor: M. M. van Zaanen. Tilburg, 7 December 2011.
20. Terry Kakeeto-Aelen. *Relationship marketing for SMEs in Uganda*. Promotores: J. Chr. van Dalen, H. J. van den Herik. Co-promotor: B. A. van de Walle. Tilburg, 1 February 2012.
21. Suleman Shahid. *Fun & face: Exploring non-verbal expressions of emotion during playful interactions*. Promotores: E. J. Krahmer, M. G. J. Swerts. Tilburg, 25 May 2012.
22. Thijs Vis. *Intelligence, politie en veiligheidsdienst: Verenigbare grootheden?* Promotores: T. A. de Roos, H. J. van den Herik, A. C. M. Spapens. Tilburg, 6 June 2012 (in collaboration with the Tilburg School of Law).
23. Nancy Pascall. *Engendering technology empowering women*. Promotores: H. J. van den Herik, M. Diocaretz. Tilburg, 19 November 2012.
24. Agus Gunawan. *Information access for SMEs in Indonesia*. Promotor: H. J. van den Herik. Co-promotores: M. Wahdan, B. A. Van de Walle. Tilburg, 19 December 2012.
25. Giel van Lankveld. *Quantifying individual player differences*. Promotores: H. J. van den Herik, A. R. Arntz. Co-promotor: P. Spronck. Tilburg, 27 February 2013.
26. Sander Wubben. *Text-to-text generation using monolingual machine translation*. Promotores: E. J. Krahmer, A. P. J. van den Bosch, H. Bunt. Tilburg, 5 June 2013.
27. Jeroen Janssens. *Outlier selection and one-class classification*. Promotores: E. O. Postma, H. J. van den Herik. Tilburg, 11 June 2013.
28. Martijn Balsters. *Expression and perception of emotions: The case of depression, sadness and fear*. Promotores: E. J. Krahmer, M. G. J. Swerts, A. J. J. M. Vingerhoets. Tilburg, 25 June 2013.
29. Lisanne van Weelden. *Metaphor in good shape*. Promotor: A. A. Maes. Co-promotor: J. Schilperoord. Tilburg, 28 June 2013.
30. Ruud Koolen. *'Need I say more?' On overspecification in definite reference*. Promotores: E. J. Krahmer, M. G. J. Swerts. Tilburg, 20 September 2013.
31. J. Douglas Mastin. *Exploring infant engagement, language socialization and vocabulary development: A study of rural and urban communities in Mozambique*. Promotor: A. A. Maes. Co-promotor: P. A. Vogt. Tilburg, 11 October 2013.
32. Philip C. Jackson. Jr. *Toward human-level artificial intelligence – Representation and computation of meaning in natural language*. Promotores: H. C. Bunt, W. P. M. Daelemans. Tilburg, 22 April 2014.
33. Jorrig Vogels. *Referential choices in language production: The role of accessibility*. Promotores: A. A. Maes, E. J. Krahmer. Tilburg, 23 April 2014 (cum laude).
34. Peter de Kock. *Anticipating criminal behaviour*. Promotores: H. J. van den Herik, J. C. Scholtes. Co-promotor: P. Spronck. Tilburg, 10 September 2014.
35. Constantijn Kaland. *Prosodic marking of semantic contrasts: Do speakers adapt to addressees?* Promotores: M. G. J. Swerts, E. J. Krahmer. Tilburg, 1 October 2014.
36. Jasmina Marić. *Web communities, immigration and social capital*. Promotor: H. J. van den Herik. Co-promotores: R. Cozijn, M. Spotti. Tilburg, 18 November 2014.

37. Pauline Meesters. *Intelligent blauw.* Promotores: H. J. van den Herik, T. A. de Roos. Tilburg, 1 December 2014.
38. Mandy Visser. *Better use your head: How people learn to signal emotions in social contexts*. Promotores: M. G. J. Swerts, E. J. Krahmer. Tilburg, 10 June 2015.
39. Sterling Hutchinson. *How symbolic and embodied representations work in concert*. Promotores: M. M. Louwerse, E. O. Postma. Tilburg, 30 June 2015.
40. Marieke Hoetjes. *Talking hands. Reference in speech, gesture and sign.* Promotores: E. J. Krahmer, M. G. J. Swerts. Tilburg, 7 October 2015.
41. Elisabeth Lubinga. *Stop HIV/AIDS. Start talking? The effects of rhetorical figures in health messages on conversations among South African adolescents*. Promotores: A. A. Maes, C. J. M. Jansen. Tilburg, 16 October 2015.
42. Janet Bagorogoza. *Knowledge management and high performance: The Uganda financial institutions models for HPO*. Promotores: H. J. van den Herik, B. van der Walle. Tilburg, 24 November 2015.
43. Hans Westerbeek. *Visual realism: Exploring effects on memory, language production, comprehension, and preference*. Promotores: A. A. Maes, M. G. J. Swerts. Co-promotor: M. A. A. van Amelsvoort. Tilburg, 10 February 2016.
44. Matje van de Camp. *A link to the past: Constructing historical social networks from unstructured data*. Promotores: A. P. J. van den Bosch, E. O. Postma. Tilburg, 2 March 2016.
45. Annemarie Quispel. *Data for all: How professionals and non-professionals in design use and evaluate information visualizations*. Promotor: A. A. Maes. Co-promotor: J. Schilperoord. Tilburg, 15 June 2016.
46. Rick Tillman. *Language matters: The influence of language and language use on cognition.* Promotores: M. M. Louwerse, E. O. Postma. Tilburg, 30 June 2016.
47. Ruud Mattheij. *The eyes have it.* Promoteres: E. O. Postma, H. J. Van den Herik, and P. H. M. Spronck. Tilburg, 5 October 2016.
48. Marten Pijl. *Tracking of human motion over time.* Promotores: E. H. L. Aarts, M. M. Louwerse. Co-promotor: J. H. M. Korst. Tilburg, 14 December 2016.
49. Yevgen Matusevych. *Learning constructions from bilingual exposure: Computational studies of argument structure acquisition.* Promotor: A. M. Backus. Co-promotor: A. Alishahi. Tilburg, 19 December 2016.
50. Karin van Nispen. *What can people with aphasia communicate with their hands? A study of representation techniques in pantomime and co-speech gesture.* Promotor: E. J. Krahmer. Co-promotor: M. van de Sandt-Koenderman. Tilburg, 19 December 2016.
51. Adriana Baltaretu. *Speaking of landmarks: How visual information influences reference in spatial domains*. Promotores: A. A. Maes, E. J. Krahmer. Tilburg, 22 December 2016.
52. Mohamed Abbadi. *Casanova 2, a domain specific language for general game development*. Promotores: A. A. Maes, P. H. M. Spronck and A. Cortesi. Co-promotor: G. Maggiore. Tilburg, 10 March 2017.
53. Shoshannah Tekofsky. *You are who you play you are: Modelling player traits from video game behavior.* Promotores: E. O. Postma and P. H. M. Spronck. Tilburg, 19 June 2017.

54. Adel Alhuraibi. *From IT-business strategic alignment to performance: A moderated mediation model of social innovation, and enterprise governance of IT.* Promotores: H. J. van den Herik, Prof. dr. B. A. Van de Walle. Co-promotor: Dr. S. Ankolekar. Tilburg, 26 September 2017.
55. Wilma Latuny. *The power of facial expressions*. Promotores: E. O. Postma, H. J. van den Herik. Tilburg, 29 September 2017.
56. Sylvia Huwaë. *Different cultures, different selves? Suppression of emotions and reactions to transgressions across cultures*. Promotores: E. J. Krahmer, J. Schaafsma. Tilburg, 11 October, 2017.
57. Mariana Serras Pereira. *A multimodal approach to children's deceptive behavior*. Promotor: M. Swerts. Co-promotor: S. Shahid. Tilburg, 10 January, 2018.
58. Emmelyn Croes. *Meeting face-to-face online: The effects of video-mediated communication on relationship formation*. Promotores: E. J. Krahmer, M. Antheunis. Co-promotor: A. P. Schouten. Tilburg, 28 March 2018.
59. Lieke van Maastricht. *Second language prosody: Intonation and rhythm in production and perception*. Promotores: E.J. Krahmer, M.G.J. Swerts. Tilburg, 9 May 2018.
60. Nanne van Noord. *Learning visual representations of style.* Promotores: E. O. Postma, M. Louwerse. Tilburg, 16 May 2018.
61. Ingrid Masson Carro. *Handmade: On the cognitive origins of gestural representations.* Promotor: E. J. Krahmer. Co-promotor: M. B. Goudbeek. Tilburg, 25 June 2018.
62. Bart Joosten. *Detecting social signals with spatiotemporal Gabor filters.* Promotores: E. J. Krahmer, E. O. Postma. Tilburg, 29 June 2018.
63. Yan Gu. *Chinese hands of time: The effects of language and culture on temporal gestures and spatio-temporal reasoning.* Promotor: M. G. J. Swerts. Co-promotores: M. W. Hoetjes, R. Cozijn. Tilburg, 5 June 2018.
64. Thiago Castro Ferreira. *Advances in natural language generation: Generating varied outputs from semantic inputs*. Promotor: E. J. Krahmer. Co-promotor: S. Wubben. Tilburg, 19 September 2018.
65. Yu Gu. *Automatic emotion recognition from Mandarin speech.* Promotores: E. O. Postma, H. J. van den Herik, H. X. Lin. Tilburg, 28 November 2018.
66. Francesco Di Giacomo. *Metacasanova: A high-performance meta-compiler for domain-specific languages*. Promotores: P. H. M. Spronck, A. Cortesi and E. O. Postma. Tilburg, 19 November 2018.
67. Ákos Kádár. *Learning visually grounded and multilingual representations*. Promotores: E. O. Postma, A. Alishahi. Co-promotor: G. A. Chrupala. Tilburg, 13 November 2019.
68. Phoebe Mui. *The many faces of smiling: Social and cultural factors in the display and perception of smiles.* Promotor: M. G. J. Swerts. Co-promotor: M. B. Goudbeek. Tilburg, 18 December 2019.
69. Véronique Verhagen. *Illuminating variation: Individual differences in entrenchment of multi-word units*. Promotor: A. M. Backus. Co-promotores: M. B. J. Mos, J. Schilperoord. Tilburg, 10 January 2020 (cum laude).

70. Debby Damen. *Taking perspective in communication: Exploring what it takes to change perspectives*. Promotor: E. J. Krahmer. Co-promotores: M. A. A. Van Amelsvoort, P. J. Van der Wijst. Tilburg, 15 May 2020.
71. Alain Hong. *Women in the lead: Gender, leadership emergence, and negotiation behavior from a social role perspective*. Promotor: J. Schaafsma. Co-promotor: P. J. van der Wijst. Tilburg, 3 June 2020.
72. Chrissy Cook. *Everything you never wanted to know about trolls: An interdisciplinary exploration of the who's, what's and why's of trolling in online games*. Promotores: J. Schaafsma, M. L. Antheunis. Tilburg, 22 January 2021.
73. Nadine Braun. *Affective words and the company they keep: Investigating the interplay of emotion and language.* Promotor: E. J. Krahmer. Co-promotor: M. B. Goudbeek. Tilburg, 29 March 2021.
74. Yueqiao Han. *Chinese tones: Can you listen with your eyes? The influence of visual information on auditory perception of Chinese tones*. Promotor: M. G. J. Swerts. Co-promotores: M. B. J. Mos, M. B. Goudbeek. Tilburg, 18 June 2021.
75. Tess van der Zanden. *Language use and impression formation: The effects of linguistic cues in online dating profiles*. Promotor: E. J. Krahmer. Co-promotores: M. B. J. Mos, A. P. Schouten. Tilburg, 22 October 2021.
76. Janneke van der Loo. *Mastering the art of academic writing: Comparing the effectiveness of observational learning and learning by doing.* Promotor: E. J. Krahmer. Co-promotor: M. B. J. Mos. Tilburg, 1 December 2021.
77. Charlotte Out. *Does emotion shape language? Studies on the influence of affective state on interactive language production.* Promotor: E. J. Krahmer. Co-promotor: M. B. Goudbeek. Tilburg, 16 December 2021.