METAPHOR, MEMORY AND MOVEMENT – CONCEPTUAL METAPHOR THEORY NOT SUPPORTED IN ASSOCIATION WITH RECOGNITION AND RECALL.

BY

JENNY WILLIAMS

A thesis submitted to the Victoria University of Wellington in fulfilment of the requirements for the degree of Master of Science in Psychology

Victoria University of Wellington

2021

Abstract

Conceptual Metaphor Theory (CMT) states that certain metaphor classes develop due to universal embodied human experiences and that these metaphor classes then influence cognitive processes and behaviour. In this paper, we investigated if movements made during learning affected memory. The hypothesis was that a movement that was congruent with the metaphor class *up is good/down is bad* would assist learning of valenced words and result in better recognition or recall, while incongruent movements would hinder learning and lead to poorer recognition and recall. Using an online survey, undergraduates were asked to learn nonsense words and their valenced meaning in English. They were then asked to make a movement – depending on the word's valence – that was either congruent, incongruent or neutral to the metaphor class *up is good/down is bad* before being tested on recognition or recall accuracy. An Analysis of Variance was carried out. Participants were less accurate in recognition when they made a movement that was incongruent with the metaphor class *up is good/down is bad* than when they made a congruent or neutral movement. When recalling valenced meanings, the movement made by participants did not have any impact on recall accuracy. We concluded CMT may not be as robust as previous research has indicated.

Keywords: conceptual metaphor theory, memory, movement, recognition, recall

Acknowledgements

I would like to thank Associate Professor John McDowall for his supervision and help in the process of completing this thesis. From pointing me in the direction of a fascinating area of study to his guidance on when to stop, John has offered unflappable support, sensible advice and good humour, making this a much more enjoyable and less stressful process than I could have imagined.

I would also like to thank all the students who generously gave their time to being part of the experiments carried out for this research.

Finally, I would like to thank my family. My mum for her lifelong belief in me, my partner Phil for his unstinting support and for being willing to step in all the times I couldn't and especially to Lila and Evie for being remarkably understanding and very patient – letting me keep writing when they would have rather I'd been playing football.

Table of Contents

List of Figures

Metaphor, memory and movement – Conceptual Metaphor Theory not supported in association with recognition and recall.

In an uncertain world where only change can be relied on, we constantly look for how to better understand our experiences and navigate our way through life. The Covid-19 pandemic has altered our world significantly. When trying to comprehend something so unfamiliar, metaphors can help us. In March and April 2020, New Zealanders experienced *lockdown* - when no one was actually locked within their houses, but millions of New Zealanders stayed at home for weeks, in order to *battle* COVID-19 19. During lockdown, we lived within *our bubble* – which is a metaphor for our household group – to prevent the virus's spread. We were not allowed to *break our bubble* – which referred to restricting physical and social contact. These actions had the intended goal of *flattening the curve* – meaning to slow and even halt the proliferation of the virus within our community. All of these metaphors help us to comprehend the confusing new world we inhabit and, by comprehending it, to manage better and hopefully change the course of the Covid-19 pandemic.

We can see from this example that metaphors assist with communication (Amin et al., 2015; Crawford, 2009; Lakoff & Johnson, 1980b). They can enhance clarity of meaning or enlarge understanding. They can make language richer and more vivid and so increase the information available to those listening or reading. They can make language more engaging and by making it more engaging, make it more relevant to the intended audience. Metaphors do this by linking two disparate concepts, allowing one to be better explained by associating it with another. This can be seen in the word's linguistic roots, metaphor comes from the Greek word *metapherin* which means transference (McGlone, 2007) – in that the meaning of

one concept is transferred to another. But metaphors do more than just enliven and facilitate communication. Metaphors are thought to be a vital part of the way we think and structure our knowledge, and for that reason they are a worthy target for study.

Within this thesis I will firstly outline the history of the study of metaphor before introducing and expanding on Conceptual Metaphor Theory (CMT) as described by Lakoff and Johnson (1980a, 1980b). I will go on to describe research findings that have developed our understanding of CMT and its impact on behaviour. I will then explain how CMT sits within the wider theory of embodied cognition and the experimental findings that illustrate this relationship. Following this, I will focus on Orientational Metaphors, their key role in much of the empirical work to date and why metaphors which refer to the vertical axis are particularly salient. Subsequently I will discuss the role of CMT in memory. I will also address the inconsistency and lack of replicability in this area of research and the subsequent relevance of the experimental work this thesis documents.

Metaphors - a brief history

Metaphors are pervasive in English. One study, analysing television scripts, showed metaphors were used once every 25 words (Graesser et al., 1989). Metaphors are also pervasive across language and culture (Aksan & Kantar, 2008; Kövecses, 2000; Yu, 1995), with research indicating that nearly all languages feature metaphors (Kövecses, 2005). Not only is the use of metaphor universal, but many cultures share similar associations between the concepts used in metaphors. The metaphor class of *anger as a container* has been observed in a number of languages—English ‘she was filled with anger’, Hungarian ‘he could not keep his anger inside’, Chinese ‘to have one’s body cavities filled with anger’ and Japanese ‘I contained my anger’ (Kövecses, 2000). As well, both English and Chinese use the metaphor class of *up is happy,* examples include ‘that boosted my spirits’ (English) and

‘they are in a high-mood’ (Chinese) (Yu, 1995). English and Turkish share the metaphor class *love is a journey* with both languages using phrases like ‘our relationship is not on track’ (Aksan & Kantar, 2008).

Given its key role in human language, metaphor has long been a focus of enquiry. The ancient Greeks studied metaphor. Aristotle identified comparison as the key function of metaphor. He explained that metaphor took two entities or ideas and identified things they had in common (as cited in McGlone, 2007). In the intervening centuries, scholarship focused on metaphor and its role in literature. However, in the twentieth century, theory about metaphor developed away from conceiving of it as a purely literary device. Black (1962) posited that metaphor was more about the interaction of the two concepts than simple comparison between two entities. Then in 1980, Lakoff and Johnson published two papers which placed metaphor at the centre of cognitive function (Lakoff & Johnson, 1980a, 1980b). They introduced the idea that metaphor was not just a “matter of words” but was instead a matter of “thought and action” (Lakoff & Johnson, 1980a, p. 453-454). They called their approach Conceptual Metaphor Theory (CMT).

Conceptual Metaphor Theory

Conceptual Metaphor Theory states that our understanding of concrete concepts is fundamental to our comprehension of more abstract ideas. That metaphors, which link our corporeal experiences within the physical world to more intangible notions, are key to the way we think (Lakoff & Johnson, 1980a). An illustration of the function of metaphor in CMT can be seen using the metaphor class of *love as a journey.* CMT suggests that emotions, like love, are felt by an individual but cannot be understood solely from that relatively ephemeral experience. Instead, these abstract concepts need to be mediated by a more concrete entity or experience, like that of a journey (Lakoff & Johnson, 1980b;

McGlone, 2007). The concrete concept—in this case the journey—is the source and is used as a comparison to better understand the abstract concept—love—called the target (Fetterman, et al., 2016; Lakoff & Johnson, 1980b). For this metaphor class, metaphorical phrases have developed such as 'our relationship is at a crossroads', 'we're going in different directions' and 'after a bumpy start, our marriage is back on track' to aid in understanding the complexity of human relationships.

But the theory extends further and claims that metaphors do not just help to understand abstract concepts but provide a framework for organising conceptual thought (Bowdle & Gentner, 2005; Crawford, 2014; Lakoff & Johnson, 1980b; Shutova et al., 2013). Metaphors do this by providing a mental template (Lakoff et al., 2001). The cognitive representation and organisation of the concrete concept like a journey provides cognitive representation and organisation for a more abstract concept like love (Lakoff & Johnson, 1980b; Shutova et al., 2013). By being able to 'map' the abstract concept onto the framework provided by the concrete concept, comprehension of the abstract concept is expanded and improved (Bowdle & Gentner, 2005; Shutova et al., 2013). The use of metaphor also removes the need for two mental representations for both source and target concepts. Instead, an understanding of the source domain provides a structure for understanding the target and cognitive economy is preserved (Shutova et al., 2013).

In their 1980 publication, Lakoff and Johnson describe three types of metaphor—structural, ontological and orientational (Lakoff & Johnson, 1980b). Structural metaphors are when one kind of experience is perceived in the context of another type of experience e.g., *understanding is seeing* (giving rise to metaphors like 'I see your point' and 'it looks different from my perspective') and *life is a journey* ('he's come to the end of the road' and 'she's taken a more spiritual path'). In a very similar fashion, ontological metaphors project the characteristics of one entity on to another. For example, *the mind is a container* ('my

brain is bursting after that lecture' and 'keep that in the back of your mind') and *vitality is a substance* ('I've run out steam' and 'she's brimming with energy'). The final kind of metaphor are orientational metaphors. Orientational metaphors take abstract concepts and orientate them within space. These metaphors often form pairs with one pole of the spatial dimension associated with one abstract concept and the other pole associated with the opposite abstract concept. An example of this is the metaphor class *up is happy/down is sad* giving rise to metaphors like 'my spirits soared' or 'his heart lifted' and at the opposite end of the spectrum 'she's down in the dumps' and 'I'm feeling very low'.

Lakoff and Johnson (1980) developed their theory beyond mere metaphor classification. They explain that within CMT metaphors have 'entitlements', meaning the attributions of the concrete concept can transfer to the abstract entity. An example they give is of the metaphor class *time is money.* As money has value, this attribute is also applicable to time, resulting in metaphors like 'I've invested a lot of time' or 'how did you spend your weekend?'. They temper this by also explaining that although some characteristics of the source concept are transferable, metaphorical definitions are 'partial, inconsistent and overlapping' (Lakoff & Johnson, 1980b, p. 198), meaning that no one metaphor can ever completely capture an abstract concept, so often an abstract target is defined by more than one concrete source. Ideas can be represented by a number of metaphor classes—'*ideas are people* 'the theory is her brainchild', *ideas are plants* 'that seeded an idea in his mind' and *ideas are money* 'they have a wealth of ideas'—in order to fully capture the complexity of the abstract concept. Lakoff et al. (1991) went on to develop a master metaphor list, which makes an attempt to capture many of the main conceptual metaphor classes in English. They produced 211 pages of metaphor classes – from *opportunities are open paths* to *morality is straightness*.

The Development of Conceptual Metaphor Theory

Conceptual Metaphor Theory prompted significant further discussion and study. Subsequent research posited that metaphors are 'cognitive phenomena' (Shutova et al., 2013, p. 1263), which do not just shape our thinking but also drive behaviour. Experimental work has shown that conceptual metaphors appear to affect evaluation, attention and perception (Bayer et al., 2012; Meier & Robinson, 2004; Meier et al., 2008; Meier et al., 2004; Open Science, 2015; Silvera et al., 2002; Yap et al., 2013). Meier et al. (2008) showed that participants were faster to evaluate both positively valenced words in a larger font and negatively valenced words in a smaller font, consistent with the metaphor class *big is good*, than they were if the words were of a size that was inconsistent with the metaphor class. In similar research examining attention and the metaphor class *up is good/down is bad,* university students were quicker to see a target within the top half of a screen when they had been primed by seeing a positive word first and to see a target in the bottom half of a screen when they had just seen a negative word, than they were when the target appeared within the portion of the screen that was inconsistent with *up is good/down is bad* (Meier & Robinson, 2004). Perception has also been shown to be impacted by Conceptual Metaphor Theory. Participants were first primed to feel powerful or powerless by writing about an instance when they were in a position of power or when someone was in a position of power over them. Then they were asked to estimate someone's height and weight from a picture. Those who were primed to feel powerful under-estimated the weight and height of the stranger, with the opposite result for those who were primed to feel powerless (Yap et al., 2013).

There is also evidence of these effects in more 'real world' situations (Jackson & Ervin, 1992; Parzuchowski et al., 2016; Seidel & Prinz, 2018; Slepian et al., 2012; Zhong et al., 2010). To test CMT with the metaphor class *bright is moral/dark is immoral*, participants were placed in either a well-lit room or one that was more dimly illuminated. Those in the

darker room were more likely to cheat at a maths game than their peers in the brighter room; associating ethical behaviour with the lighter environment and unethical behaviour with the darker room (Zhong et al., 2010). With reference to the metaphor class *big is good*, undergraduates were asked to judge two pieces of art—the only difference being size. The larger reproductions were preferred over the smaller works (Seidel & Prinz, 2018). Another study found that when participants were asked to remember an important secret, they estimated a hill was steeper than participants asked to recall a more inconsequential one—the theory being participants were more *weighed down* by the big secret (Slepian et al., 2012).

Conceptual Metaphor Theory and Embodied Cognition

Conceptual Metaphor Theory sits within the wider idea of embodied cognition. Embodied cognition states that the whole body is involved in cognitive processes (Barsalou, 2008; Crawford, 2014). It challenges the traditional view of our brains as processing units into which our bodies fed perceptual information in a one-way process (Amin et al., 2015). Embodied cognition maintains that influence and information flows in both directions and our mental processes are also mediated by our physical experiences within a wider social and cultural context (Barsalou, 2008). This mirrors CMT, in that the concrete influences the cognitive.

Research into embodied cognition shows that language is impacted by the corporeal world, with semantic meaning grounded in physical experiences. Neurological studies show sensory-motor experiences influences language processing (Hauk & Pulvermüller, 2004; Moody & Gennari, 2010; Rueschemeyer et al., 2010; Willems et al., 2009). A study by Hauk and Pulvermüller (2004) used an electroencephalogram (EEG) to record cortical activity while participants read verbs which related to leg (e.g. kick), arm (e.g. pick) and head (e.g. lick) movements. Reading the words resulted in cortical activations that were comparable to making the actual leg and arm movements. Research findings extend beyond verbs to nouns

as well. Using functional magnetic resonance imaging (fMRI), activation of the motor cortex was found when participants read words for manipulatable nouns—like hammer or cup—but not for words like clock and bookend (Rueschemeyer et al., 2010). Even more nuanced evidence of this embodied effect has been demonstrated. The amount of effort implied by a sentence appears to be reflected in the amount of cortex activated. Participants were asked to read sentences that suggested a range of physical effort while having an fMRI. Sentences like 'the man pushed the piano' activated more cortex than sentences implying less exertion— 'the man pushed the chair' and less cortex was activated again when no effort was implied—'the man forgot the piano' (Moody & Gennari, 2010). This kind of cerebral activation has also been shown to be body specific, with differences between left and right handers. Willems et al. (2009) asked participants to imagine performing an action while having an fMRI. When participants imagined complicated hand movements, right-handers activated their left cortex and left-handers activated their right cortex; the brain hemisphere related to their dominant side.

As well as empirical evidence for more generic embodied cognition, neuroimaging experiments also show that these types of effects extend to metaphorical language and so also to Conceptual Metaphor Theory (Citron & Goldberg, 2014; Desai et al., 2011; Lacey et al., 2012; Lai et al., 2019). In a comparison of cortical activation when reading about literal action ('the daughter grasped the flowers'), metaphorical action ('the public grasped the idea') and abstract action ('the public understood the idea'), fMRI studies showed that literal and metaphorical actions both activated the part of the brain associated with action planning but abstract sentences did not (Desai et al., 2011). An event related potential (ERP) study examined the timing of cortical activity for literal and metaphorical sentences. Individuals read literal sentences ('the bodyguard bent the rod') and metaphorical sentences ('the church bent the rules'). The ERP showed similarities between the timing of neural activity for both

sentence types, suggesting that both started out with sensorimotor simulation (Lai et al., 2019)

Activation of sensory cortex in relation to CMT has also been investigated. While having an fMRI, seven participants were asked to read 54 sentence pairs, with one sentence using a textural metaphor and the other sentence matched for meaning but using literal language ('she had a rough day', 'she had a bad day'). The metaphorical sentences activated texture-sensitive somatosensory cortex while the literal sentences did not, which is also supportive of CMT (Lacey et al., 2012). Gustatory metaphors have also been investigated using fMRI. Participants read sentences using metaphors relating to taste—'she looked at him sweetly'—that were matched for meaning with literal sentences—'she looked at him kindly'. The metaphorical sentences activated the parts of the brain associated with taste more than the literal sentences (Citron & Goldberg, 2014).

However, this is not a robust finding as neural imaging does not always detect sensory or motor cortical activation in association with metaphorical language (Aziz-Zadeh et al., 2006; Raposo et al., 2009). Participants were presented with a verb in isolation ('kick'), a verb in a literal sentence ('kick the ball') and a verb used metaphorically ('kick the bucket') while in an fMRI. Researchers found activation of the part of the brain associated with movement occurred for the first two stimuli but not the third (Raposo et al., 2009). Willems and Casasanto (2011) suggest that this is because 'kick the bucket' is a frozen idiom; that is an idiom that can only be written in one way to maintain its meaning. Conversely, conceptual metaphors are flexible and can be composed in a number of ways as their meaning is derived from their component words and not the full phrase (Kacinik, 2014). So, it is possible to 'grasp' an idea, but also it is possible to have a 'tenuous hold' on an idea, or to 'not be grabbed' by an idea. 'Kick the bucket' however, only holds true in that exact form—kicking a smaller container does not mean to merely be injured. It is theorised this

lack of flexibility reduces those links between abstract and concrete with a resulting lack of motor cortex activation (Willems & Casasanto, 2011).

Another fMRI study also failed to associate metaphorical language with sensory-motor cortex activity (Aziz-Zadeh et al., 2006). Individuals were asked to watch an action (grasping keys), read an action sentence ('grasp the keys') and read a metaphorical sentence ('grasp an idea'). Motor cortex activation was recorded for the observation and the literal sentence but not for the metaphorical use. The researchers in this study suggested the reason for this was activation of the motor cortex occurred when the metaphor was novel but, in time, this activation was reduced as the metaphor developed other representations not associated with the actual movement. It has been theorised these inconsistencies suggest the link between metaphorical language and sensory-motor activation not automatic but instead is dependent on context (Raposo et al., 2009; Willems & Casasanto, 2011).

Orientational Metaphors

Of the three metaphor types described by Lakoff and Johnson (1980b), orientational metaphors appear to be the most pervasive in language and cultural (Kövecses, 2000, 2005; Meier & Robinson, 2004; Meier et al., 2004; Yu, 1995). The key concept is the perception of space is based on universal bodily experiences and many of these experiences are associated with characteristic affective responses that nearly all humans experience. This has led to the development of links between particular spatial dimensions, emotion and valence across culture and language (Barsalou, 2008; Lakoff & Johnson, 1980b; Marmolejo-Ramos et al., 2013; Taylor et al., 2015).

In terms of this work, the physical world is divided into three dimensions – vertical, sagittal and horizontal (Crawford, 2009; Franklin & Tversky, 1990; Koch et al., 2011; Marmolejo-Ramos et al., 2017; Taylor et al., 2015; Tversky, 2011, 2014; Tversky & et al.,

1991). All three dimensions have metaphor classes associated with them. In relation to the vertical axis examples are, *up is happy/down is sad*—'his spirits soared' 'I feel down in the dumps'; *up is moral/down is immoral*—'it was all above board', 'her behaviour is underhand'. In terms of the sagittal axis examples include, *forward is positive/backward is negative* 'it's a forward-looking company' 'that was a backward step'. An example of a metaphor associated with the horizontal axis is *right is good/left is bad*—'he's my right-hand man', 'she's a bit left of centre'. These spatial associations even permeate our day-to-day activities – with premium brand liquor being placed on the top shelf, business class being at the front of the plane and Tinder using a right swipe to show attraction and a left swipe for rejections.

This review of the literature has found the sagittal plane is the least studied and has fewer conceptual metaphors associated with it than the other two dimensions. It is linked to metaphors that relating to agency—*forward is decisive/backwards is indecisive* (Koch et al., 2011) and to achievement—*forward is success/backwards is failure* (Robinson & Fetterman, 2015). It is also associated with temporal metaphors, with *forwards being future/backwards being past*, although the direction can be culturally dependent (Koch et al., 2011; Tversky & et al., 1991; Ulrich & Maienborn, 2010). For example, the Māori concept of the past is very different to that of Pākehā. While Pākehā 'leave the past behind them', the Māori view of the past is well explained by the whakatoukī (proverb) *Ka mua, ka muri*—walking backwards into the future—keeping the past very much in view (Reese et al., 2008). Cultural differences like these are reflected in behavioural experiments. In a cross-cultural study, both adults and children were asked to place stickers on paper to provide a graphic representation of time. For example, a participant would be given three stickers representing breakfast, lunch and dinner. They were then asked to place the stickers on a sheet of paper to represent the temporal relationship between the meals. The researchers found these representations were

influenced by written language direction. English speakers representing time, from past to present, in a left to right direction. Arabic speakers representing time in a right to left direction mirroring their dextrosinistral (right to left) language. Hebrew speakers had mixed results and this was attributed to the fact that most Hebrew speakers are at least bilingual and their second language is often sinistrodextral (left to right) (Tversky & et al., 1991).

The horizontal dimension appears to be more culturally uniform but often differs on an individual level (Casasanto, 2009; Casasanto & Chrysikou, 2011; Casasanto & Henetz, 2012; Ping et al., 2009). In many cultures *right is good/left is bad*, and this does not change with reading/writing direction (Tversky, 2011; Tversky & et al., 1991). This partiality is illustrated in both language and action. In English, we talk about 'being on the right hand of God' or 'having two left feet'. In French, droite refers to the right-hand side but also means an entitlement or privilege, and gauche (left) means distasteful or clumsy. In Islam, the right hand is reserved for eating and the left hand is used for unsanitary jobs. However, experimental research reveals a different picture at an individual level. When asked to draw images of liked and unliked animals to the left or right of a cartoon figure, right-handers placed liked animals on the right, while left-handers placed them on the left. This result was repeated when the participants were asked to respond verbally rather than by drawing. This same study also showed that right-handers also prefer products presented to their right side with left-handers showing a preference for products on the opposite side (Casasanto, 2009). This partiality is also evident in young children, with five-year-olds attributing more amiability and intelligence to animals presented on their dominant side (Casasanto & Henetz, 2012). This suggests that the link between horizontal space and valence may be body specific (Casasanto, 2009). Research by Ping et al. (2009) showed individuals preferred kitchen utensils they could reach more easily than those that were more difficult to grasp. The researchers posited that people have a preference for greater fluency and ability, and it is

assumed that this is what drives the individual preference for the dominant side (Ping et al., 2009). This supposition is supported by a study examining stroke patients. Right-handed patients who lost the use of their left hand continued to prefer the right-hand side but right-handers who lost the use of their right hand and were forced to use their left hand switched their preference to their left side (Casasanto & Chrysikou, 2011).

The final dimension of verticality has many metaphor classes associated with it. Some of these are *God is up/the devil is down, more is up/less is down, healthy is up/unhealthy is down, power is up/powerless is down* and *developed is up/undeveloped is down.* These can largely be summarised by the more generic valenced response of *up is good/down is bad* (Koch et al., 2011; Marmolejo-Ramos et al., 2013; Tversky, 2011, 2014). The universality of this metaphor class is mirrored in the cultural spread of this orientational metaphor, as *up is good/down is bad* appears to be a widespread association across the world (Marmolejo-Ramos et al., 2017; Marmolejo-Ramos et al., 2013; Tversky & et al., 1991; Yu, 1995). In the cross-cultural study mentioned before which investigated differences between English speakers, Arabic speakers and Hebrew speakers, Tversky et al. (1991) asked the adults and children to provide a graphic representation of preference. In one of the tasks, a participant would be given three stickers representing a loved television programme, a liked television programme and a less-liked television programme. They were then asked to place the stickers on paper to illustrate the spatial relationship between the programmes that indicated preference. All three cultures associated upwards with an increase in preference (Tversky & et al., 1991). In another study, researchers asked 2153 university students, who spoke 22 different languages, to use a Likert scale to rate the valence of the words ‘up’ and ‘down’. All language groups rated ‘up’ as a positive word and all but three (Cebuano-, Chinese- and Hebrew-speakers) rated ‘down’ as a negative word (Marmolejo-Ramos et al., 2013). The same researchers gave English and Japanese speakers a two-by-two grid and

asked them to place valenced words that related to the qualities of a job applicant within the grid. All participants place the positive words in the top half of the grid, but there was variation in terms of placing the valenced words in the left- or right-hand sides of the grid (Marmolejo-Ramos et al., 2013). A similar experiment asked individuals from six different linguistic backgrounds (English, German, Hindi, Japanese, Spanish and Vietnamese) to place the words joy, surprise and sadness within a grid. Regardless of language or cultural background, joy was placed in higher locations than either of the other two words. Sadness was placed in the lower parts of the grid, with surprise being placed in the middle (Marmolejo-Ramos et al., 2017).

It is results like this that lend support to the idea that the vertical axis is the most salient to CMT (Barsalou, 2008; Ekman et al., 1972; Franklin & Tversky, 1990; Koch et al., 2011; LaFrance & Mayo, 1978; Tversky, 2011, 2014). Embodied cognition and CMT state that our mental processes are structured by a connection between abstract concepts and our physical experiences. In terms of the vertical axis, the metaphor class *good is up/bad is down* is not arbitrary. We live in a world ruled by gravity so being upright implies a degree of health and strength (Tversky, 2011, 2014). As we grow older, stronger and more socially powerful we grow upwards (Barsalou, 2008; Tversky, 2011, 2014). Upward movements are associated with our physical experiences of positive emotions and negative emotions are linked to downward movement. These physical responses seem to be largely universal. Ekman et al. (1972) showed 18 photographs of individuals displaying one of six emotional expressions (happiness, surprise, sadness, fear, disgust and anger) to people from ten cultures. They found strong agreement in emotional identification across all cultures. Happiness, represented by smiling faces, was the most commonly agreed upon emotion across cultures. A review of the literature reinforced this finding, with a number of other studies finding similar results (LaFrance & Mayo, 1978). Smiling raises the corners of the mouth and the

cheeks. The association with upwards movement and positive affect can also be seen in physical posture. A study comparing the physical reactions to winning of blind, congenitally blind and sighted athletes found that all three groups responded to victory in a similar way with raised arms and faces tilted up, suggesting these upwardly directed actions are innate and not learned (Tracy & Matsumoto, 2008). This link between upwards bodily movement and positive affect is theorised to be a key mechanism which leads to the association between up and positive valence (Casasanto & de Bruin, 2019; Crawford, 2014; Marmolejo-Ramos et al., 2013; Tversky, 2011, 2014).

In contrast, the other two dimensions appear to have less saliency. Barsalou (2008) observed that our bodies are most asymmetrical in the vertical axis—which may increase its salience, then the sagittal axis, with the asymmetry within the horizontal, or left/right, axis being relatively minor. This asymmetry effect may also explain the relative flexibility in individual behaviour with regards to the linking of valence to left and right which is theorised to be the least salient of the dimensional axes. This idea of the dominance of the vertical axis is supported by an experiment by Franklin and Tversky (1990). Individuals were given a description of a scene to read and memorise, then they were asked to recall if objects within the scene were above, below, in front, behind, to the left or to the right of the narrator. Participants were fastest to identify objects' locations in the vertical plane (Franklin & Tversky, 1990).

Memory and Conceptual Metaphor Theory

This saliency of the vertical axis may account for *up is good/down is bad* being the most studied metaphor class within Conceptual Metaphor Theory, with subsequently more empirical evidence supporting it than any other orientational metaphor (Crawford, 2014). It is also the only metaphor class studied with regards to CMT and its effect on memory

(Crawford, 2014). While much of the other research has focused on online tasks like perception and evaluation, the studies with memory are important because they allow researchers to see if CMT is robust enough to impact an offline process where the individual is not actively engaged in the current environment and so embodiment may be less relevant.

As with online processes, CMT effects have been observed when investigating autobiographical memory (Casasanto & Dijkstra, 2010; Riskind, 1983). Students were asked to smile or frown and then recall happy or sad memories. The participants were faster to recall memories that were congruent with their facial expression. Comparable results were obtained when the students' facial expressions were manipulated into smiles and frowns by touch, allowing the emotionally loaded words of 'smile', 'happy expression', 'frown' and 'sad expression' to be avoided. A later experiment also examined the impact of movement on autobiographical memory, this time the movement investigated was emotionally-neutral (Casasanto & Dijkstra, 2010). This was to avoid any effect of encoding specificity—that being individuals are more likely to remember things if they experience the same conditions at retrieval as at encoding (Godden & Baddeley, 1975), in this case the physical sensation of smiling. Casasanto and Dijkstra (2010) had students move marbles in an upwards or downwards direction while recalling either positive or negative personal memories. Participants were faster to recall memories whose valence was consistent with the movement, that is happy memories when moving the marbles up and sad memories for moving marbles down. In an extension of this work, participants were given a neutral prompt (they were asked about what happened the day before) and the students who moved marbles up while remembering, recalled more positive memories, while students who moved marbles down recalled more negative memories.

These effects of CMT have also been found in more general memory tasks (Casasanto & de Bruin, 2019; Crawford et al., 2006; Meier et al., 2007; Palma et al., 2011). Undergraduates were shown 60 emotionally valenced pictures in different locations onscreen. Then the participants were shown the images again and asked to drag them back to the position where they had seen them initially. Positive images tended to be placed higher on the screen in comparison to their initial position than comparably placed negative images (Crawford et al., 2006). In an experiment with the same procedure but using images of God and the devil, students were more likely to remember pictures of God being higher than they were when initially viewed and images of the devil as being lower, supporting the metaphor class *holiness is up/evil is down* (Meier et al., 2007). Crawford (2014) argues these two results show the influence of metaphor occurs at retrieval, meaning it is a memory effect; rather than encoding, which would suggest a perceptual effect. If the impact of CMT led to a biased perception of the image's location, it would occur both times the image was seen—at encoding and retrieval—cancelling each other out and so no effect would be observed. However, she theorises an effect is obtained because there is additional bias at retrieval, which is an illustration of CMT's effect on memory (Crawford, 2014).

In another CMT experiment, this time using words as stimuli, individuals were more likely to remember attributes of a fictious job-seeker that were presented in a position onscreen that was consistent with the valence of the attribute, for example 'friendly' at the top of the screen or 'insensitive' at the bottom of the screen. The researchers ran a second experiment where they asked participants to place cards with attributes printed on them on either a high or low shelf. Participants better remembered the attributes which required a movement congruent with the valence of the characteristic, that is an upward movement to place a positive attribute on a high shelf or a downward movement to a lower shelf for a negative attribute (Palma et al., 2011). This technique was used again in a recent

experiment, where students were asked to learn the meaning associated with a nonsense word while placing a card—with the word and meaning on it—on a high or low shelf (Casasanto & de Bruin, 2019). As with previous findings, the undergraduates were more likely to recognise the learned meaning when they had made a movement which was consistent with the meaning's valence. It is posited that this reinforcement of valence with congruent action may be the mechanism which leads to improved memory (Crawford, 2014). A similar effect was seen in another experiment which paired word learning with head movements but without metaphor. Researchers found participants who nodded their heads were more likely to remember positive adjectives than negative ones and participants who shook their heads were more likely to remember negative adjectives (Förster & Strack, 1996b).

However, the results of experiments testing CMT and memory are not consistent (Crawford et al., 2014). In an experiment where participants were shown valenced words at either the top of the bottom of a screen, both recognition and recall revealed that words shown in an incongruent position when compared to their valence were better remembered. The researchers theorised that because the words were in an incongruent spatial location the resulting perceptual mismatch required more effort to remember them and this increased the depth of processing, leading to better remembering (Crawford et al., 2014). This is consistent with previous memory experiments which have shown that words that are more demanding to read, either due to an interfering mask (Mulligan, 1996) or because they are in a difficult-to-read font (Diemand-Yauman et al., 2011) are better remembered.

Inconsistency and Replicability

As well as inconsistency in results for the effect of Conceptual Metaphor Theory on memory, it is important to note that inconsistency, in the form of poor replicability, has been a feature of experimental research into CMT in general (Earp et al., 2014; Johnson et al., 2014; LeBel & Campbell, 2013; Lynott et al., 2014; Pashler et al., 2012). An example is

research where Brandt et al. (2014) were unsuccessful in replicating a study showing participants who recalled unethical events perceived a room as darker than those who remembered ethical situations (Banerjee et al., 2012). There was also a failure to replicate results in a study where participants who cleaned their hands after watching disgusting imagery found fictional immoral actions less wrong than participants who had not had the opportunity to 'purify' themselves (Schnall et al., 2008). There have also been failures to replicate the research by Williams and Bargh (2008), which demonstrated that holding a warm object – in this case, a warm cup – resulted in a warmer assessment of other people or carrying out more prosocial acts (Chabris et al., 2019; Lynott et al., 2014).

The inconsistency in CMT experimental results in memory, as well as more general failures to replicate, suggest CMT may not be as robust as the initial theorists proposed. Context and detail may play a significant part. Nonetheless, the breadth of research implies there is an interaction between metaphor and cognition. Only further rigorous study will allow this interaction to be fully elucidated and understood.

The Current Study

Given the importance of replication in this area of study, the purpose of this current study was to see if the effect of Conceptual Metaphor Theory on recognition was robust enough to be replicated. There was also an intention to see if this research could then be broadened to see if Conceptual Metaphor Theory would affect recall. We aimed to replicate the first experiment in the Casasanto and Bruin (2019) study—*Metaphors we learn by: Directed motor action improves word learning.* We hypothesised that if CMT impacted memory, then words with a meaning that was congruent with a movement made during the study phase—an upwards movement for a positive word and a downwards movement for a negative word—would be better remembered in the test phase than words where there was incongruence between valence and movement, that is a downward movement with positive

meaning and an upward movement with a negative word. We then extended the work of Casasanto and Bruin (2019) and repeated the experiment with recall. If the effect of CMT on memory is robust, then we would expect words that with a meaning that was congruent with the movement made would be better recalled than words with a meaning incongruent with the movement made.

The initial intention was to exactly replicate the Casasanto and Bruin (2019) study with in-person data collection using concrete stimuli (words on card) within a laboratory setting. However, the advent of Covid-19 meant the experimental methodology needed to be adapted to accommodate physical distancing, so an online protocol was designed. While this is not ideal in terms of exact replication, there are benefits in seeing if CMT effects on memory could also be observed in a virtual setting. Replication of results combined with replication in another medium would imply the impact of CMT on memory was robust.

Experiment One

Method

Experiment 1 was a replication of a study done by Casasanto and Bruin (2019). However, because of the restrictions due to Covid-19, data collection had to move from in-person to online. Participants were presented with nonsense words with positively and negatively valenced meanings in English. Participants were then asked to make a movement depending on the meaning's valence as they attempted to commit the word to memory. After studying the words, the participants were required to pick the correct meaning from the target and a foil.

Participants

Fifty-seven undergraduates participated in the experiment, however four participants were excluded as they did not follow the experimental protocol by overriding the survey

timing for the study and test phases or because they did not open the study and test components of the survey.

The survey was completed by 53 first-year students attending Victoria University of Wellington. The students were enrolled in a first-year psychology paper and were recruited via an online request. Participants received course credit for taking part in the study. Participation was voluntary and anonymous. Within the sample there were 33 women and 20 men. There was one participant in the under 18-year age group, there were 36 participants in the 18—19-year age group, 10 in the 20—21-year age group, one in the 22—24-year age group and five in the 25-year and older age group. There were three participants who identified as Māori, 34 who identified as Pākehā, four who identified as Pasifika, eight as Asian and four as being from another ethnicity. Ethical approval was granted for this study by the Victoria University of Wellington Human Ethics Committee.

Materials

In this study 16 pronounceable six-letter nonsense words were used, which were created for the study. Sixteen English words were selected – half were positively valenced and the remaining eight were negatively valenced, according to the Affective Norms for English Words (ANEW) word database (Bradley & Lang, 1999). Although matching was not required by the fully counterbalanced design, the positively and negatively valenced words were equally as extreme from a neutral midpoint, $F(1, 14) = 0.71$, $p = .415$. There was no significant difference between word frequency for the positive and negative groups of words, $F(1, 14) = 0.48$, $p = .502$. Also, average word length in letters was similar for the positive and negative groups of words, $F(1, 14) = 1$, $p = 1.00$; as was the average syllable length, $F(1, 14) = 1$, $p = 1.00$. The assignment of each nonsense word to a positive or negative English word was counterbalanced across participants (Appendix A).

Procedure

The students were recruited online. They were required to attend a scheduled online meeting using Zoom. At the beginning of the Zoom meeting, the researcher gave a basic introduction to the study. Participants were told the study was investigating if there was an association between movement, emotion and memory and they needed to learn the English meaning of 16 nonsense words. While they were learning each word, they were required to make a movement—dependent on the word's valence—to help in their learning. The movements were demonstrated. After the introduction, participants were asked to open an online survey. This survey was created within Qualtrics, an online computer survey programme. Participants were asked to read information about the study explaining the purpose of the study in more detail, what being a participant would entail and contact details for the researchers were made available. Participants could withdraw from the study up until data collection.

Once participants had read the information, they were asked if they were willing to continue. If they were, participants were requested to mute their device and minimise Zoom and open the survey full screen. For the remainder of the study, the participants were observed by the researcher to ensure they made the movements the experimental instructions required. The Zoom meeting was not recorded.

Participants were randomly assigned to one of three groups within the study. The Good is Up (UiG) group who made a movement consistent with the metaphor class – *good is up/down is bad*. The movement this group made was to look at the ceiling when shown a positively valenced word and to look at the floor or their lap when shown a negatively valenced word. The second group was Good is Down (DiG). This group made a movement inconsistent with the metaphor class – *good is up/down is bad*. The movement this group

made was to look at the floor or their lap when shown a positively valenced word and at the ceiling when shown a negatively valenced word. The final group was the control group (CON) who were asked to blink twice when memorising the words, regardless of valence.

In the UiG and DiG groups the participants were given examples of idioms which supported the premise of the group. For the UiG condition the idioms were ‘my spirits soared’ and ‘down in the dumps’. For the DiG group the idioms were ‘she’s got her feet on the ground’ as a positive idiom and ‘he thinks he’s above the rest of us’ as a negative idiom.

For the study phases of this experiment, the participants were shown the words in two groups of eight. After attempting to learn each group of eight words, the participants were tested on those words. This meant the structure of the experiment was study phase, test phase, study phase and test phase.

In each study phase, the participants were shown eight nonsense words and their meaning in English, three times each, for six seconds each time. The timing of the presentation of the stimuli was controlled by the Qualtrics programme. Within each group of eight, the words with their meanings were in random order. The participants were asked to read the nonsense word, identify the valence of its meaning and then make the movement required for that valence as they attempted to memorise the word and its meaning.

Once they had studied the words, the participants were tested. Each nonsense word was presented for six seconds and the participants were asked to choose between the correct meaning and a foil. The foil was one of the other words in the study, so each English meaning was used as both a target and a foil during testing. For each participant, half the correct responses were presented on the right and the other half were presented on the left. The nonsense words with their target and foil were presented randomly.

The task took approximately twenty minutes to complete. Once the participants had completed the experiment, they were provided with a debriefing document which explained the purpose of the research.

Results

Results were analysed using a one-way between subjects Analysis of Variance (ANOVA) to compare the effect of movement in congruent, incongruent or control groups on recognition accuracy. Analysis showed there was a significant effect of movement for the three conditions, $F(2, 50) = 4.03$, $p = .024$, $\eta_p^2 = 0.139$. Post hoc comparisons using the Tukey test indicated that recognition accuracy was significantly lower in the incongruent movement condition ($M = 11.80$, $SD = 2.86$) than in the congruent movement condition ($M = 13.42$, $SD = 2.99$) and in the control condition ($M = 14.43$, $SD = 2.14$). The congruent movement condition did not significantly differ from the control condition.

The participants were less accurate in their recognition of the meaning of nonsense words when they made a movement that was incongruent with the metaphor class *up is good/down is bad* than when they made a congruent or neutral movement. There was no difference in recognition accuracy when participants either made a congruent or neutral movement. These findings are shown in Figure 1.

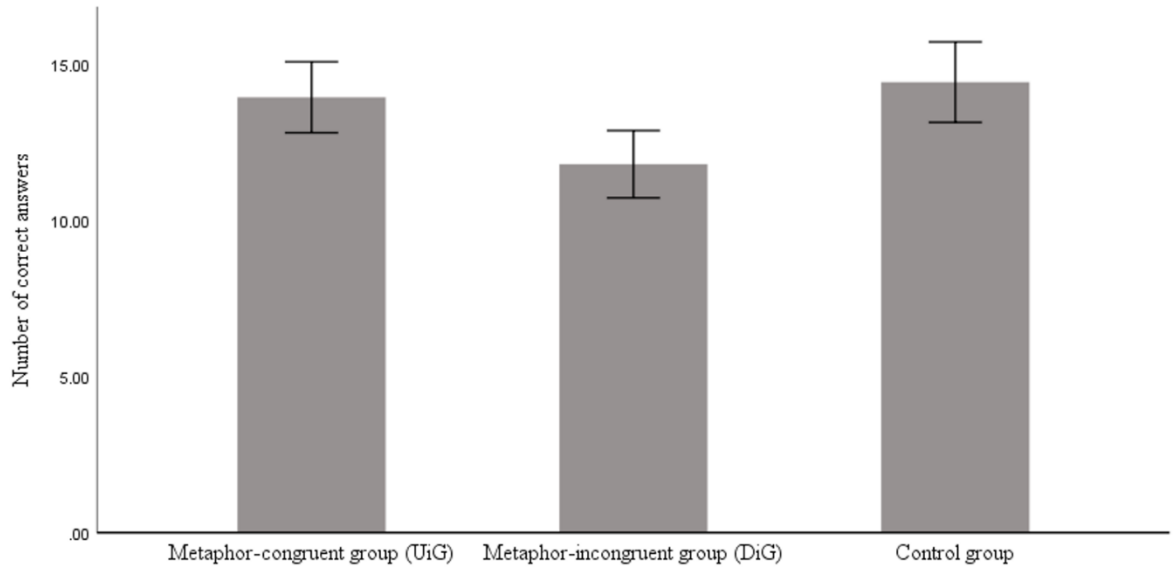

Fig. 1. Number of correctly recognised answers by experimental group

Results were analysed using a 3 (group: metaphor congruent movement, metaphor incongruent movement, control) x 2 (valence: positive, negative) mixed-design repeated measures Analysis of Variance (ANOVA). As the previous analysis showed, there was a main effect of group, $F(2, 50) = 15.11$, $p = .024$. There was no main effect for valence, $F(2, 50) = 0.76$, $p = .389$ and there was no interaction between valence and group, $F(2, 50) = .01$, $p = .992$.

The word valence—be it positive or negative—did not have a significant effect on recall accuracy for the English meaning of the nonsense words. There was no interaction between word valence and experimental group. The results are shown in Figure 2.

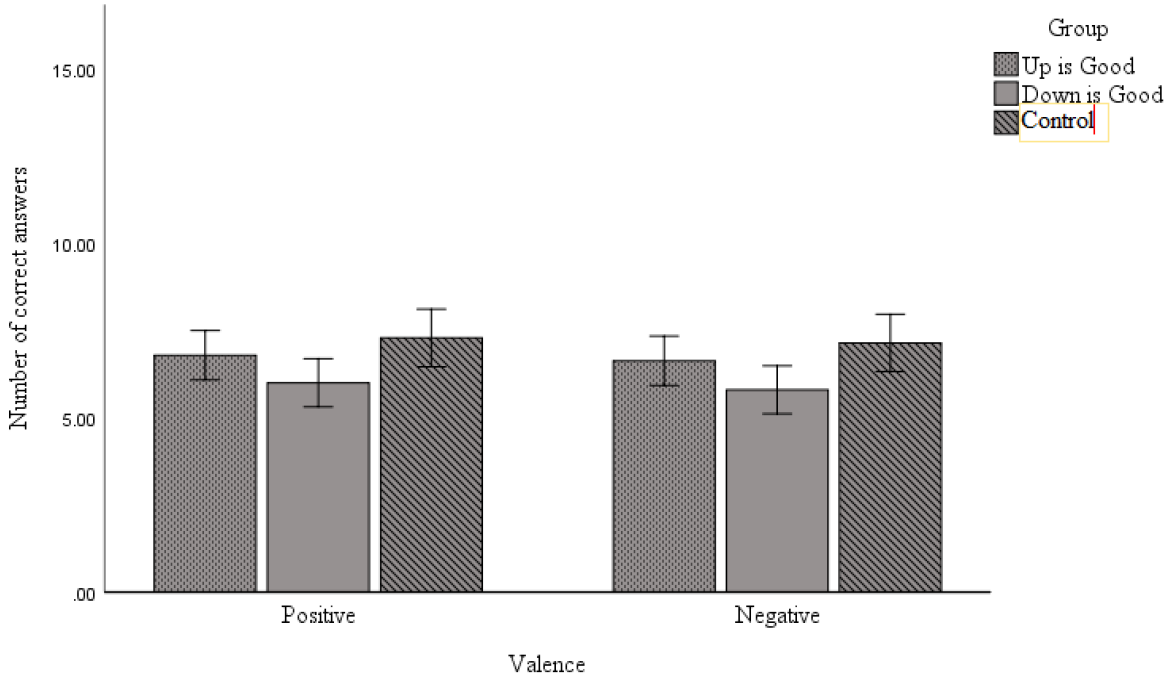

Fig. 2. Number of correctly recognised answers by experimental group and word valence

Discussion

Our first study attempted to replicate the work of Casasanto and Bruin (2019) with regards to their research into Conceptual Metaphor Theory (CMT), specifically the effect of movements, made in relation to the metaphor class *up is good/down is bad*, on recognition of nonsense word meaning. We found participants were better able to recognise nonsense word meanings when they had made a movement that was congruent with the metaphor class or neutral, than if they made a movement that was incongruent with the metaphor class. The study by Casasanto and Bruin (2019) showed movements that were congruent with the metaphor class resulted in the best rate of recognition, then neutral movements and finally movements that were incongruent to the metaphor class *up is good/down is bad* resultred in the lowest scores for recognition. While we replicated the difference between the incongruent group and the other two groups, we did not replicate any difference between the congruent group and the control group. However, it is important to consider the impact of

changes to experimental design that were a result of Covid-19 restrictions. These will be discussed in more detail in the general discussion.

Casasanto and Bruin (2019) did not examine the impact of valence on recognition, however in our study we found there was no impact of the valence of the meaning of the nonsense word on recognition.

Experiment Two

Method

Experiment 2 extended the study done by Casasanto and Bruin (2019). Instead of asking the participants to recognise the nonsense words' meanings, they were asked to recall the words' meanings.

As in Experiment 1, participants were presented with nonsense words with positively and negatively valenced meanings in English. They were then asked to make a movement, depending on the meaning's valence, as they attempted to commit the word to memory. After studying the words, the participants were shown the nonsense words again and were required to recall the nonsense words' meanings in English.

Participants

Fifty-one undergraduates participated in the experiment, however three participants were excluded who did not follow the experimental protocol by overriding the survey timing for both study and test phases or who had incomplete results due to internet connection problems or computer malfunction.

The survey was completed by 48 first-year students attending Victoria University of Wellington. The students were enrolled in a first-year psychology paper and were recruited via an online request. Participants received course credit for taking part in the study. Participation was voluntary and anonymous. Within the sample there were 29 women and 19

men. There were 41 participants in the 18—19-year age group, six in the 20—21-year age group and one in the 25-year and older age group. There were 42 participants who identified as Pākehā, two who identified as Pasifika, three as Asian and one as being from another ethnicity. Ethical approval was granted for this study by the Victoria University of Wellington Human Ethics Committee.

Materials

In this study the same 16 pronounceable six-letter nonsense words were used, as were used in Experiment 1.

Procedure

The procedure was identical to those in Experiment 1 with the exception that the participants were asked to recall the words rather than recognise them. This meant during testing the participants were shown the nonsense word and asked to type the meaning they recalled into the online survey.

Results

Results were analysed using a one-way between subjects Analysis of Variance (ANOVA) to compare the effect of movement in congruent, incongruent or control groups on recall accuracy. Analysis showed there was no significant effect of movement for the three conditions, $F(2, 45) = 0.13$, $p = .877$.

An examination of the results revealed a number of participants had made spelling or typing errors resulting in an answer that was marked incorrect by the computer programme, but it was clear they had correctly remembered the word. For example, ‘greif’ instead of ‘grief’, ‘excellenc’ instead of ‘excellence’. The analysis was repeated allowing these inaccuracies to be included. A full list of the included inaccuracies is recorded in Appendix

C. The results, including the misspelt words, were analysed using a one-way between subjects Analysis of Variance (ANOVA) to compare the effect of movement in congruent, incongruent or control groups on recall accuracy. Analysis showed there was no significant effect of movement for the three conditions, $F(2, 45) = 0.09$, $p = .918$. The movement made by participants—be it congruent, incongruent or neutral with reference to the metaphor class *up is good/down is bad*—did not have a significant effect on recall accuracy for the English meaning of the nonsense words. The results are shown in Figure 3.

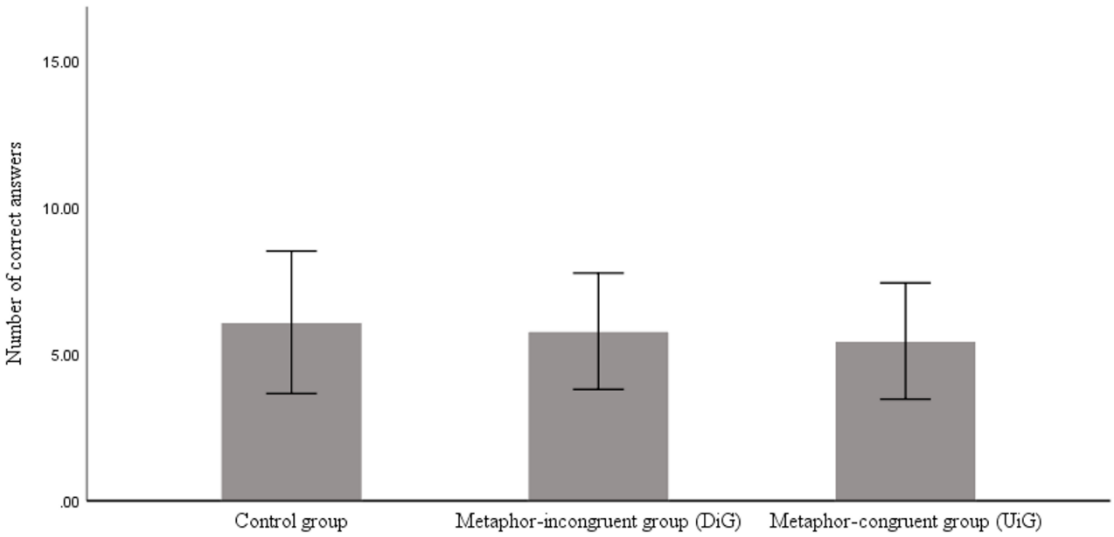

Fig. 3. Number of correctly recalled answers by experimental group with lenient scoring

Results were analysed using a 3 (group: metaphor congruent movement, metaphor incongruent movement, control) x 2 (valence: positive, negative) mixed-design repeated measures Analysis of Variance (ANOVA). As the previous analysis showed, there was no main effect of group, $F(2, 45) = 0.01$, $p = .918$. However, there was a main effect for valence, $F(2, 45) = 5.55$, $p = .023$, $\eta_p^2 = .023$, such that positively valenced meanings (M =

3.15, *SD* = 2.32) were better recalled than negatively valenced meanings (*M* = 2.58, *SD* = 2.07). There was no interaction between valence and group, $F(2, 45) < .01$, $p = .999$.

The participants were more accurate in their recall of positively valenced meanings of nonsense words than negatively valenced meanings of nonsense words. The results are shown in Figure 4.

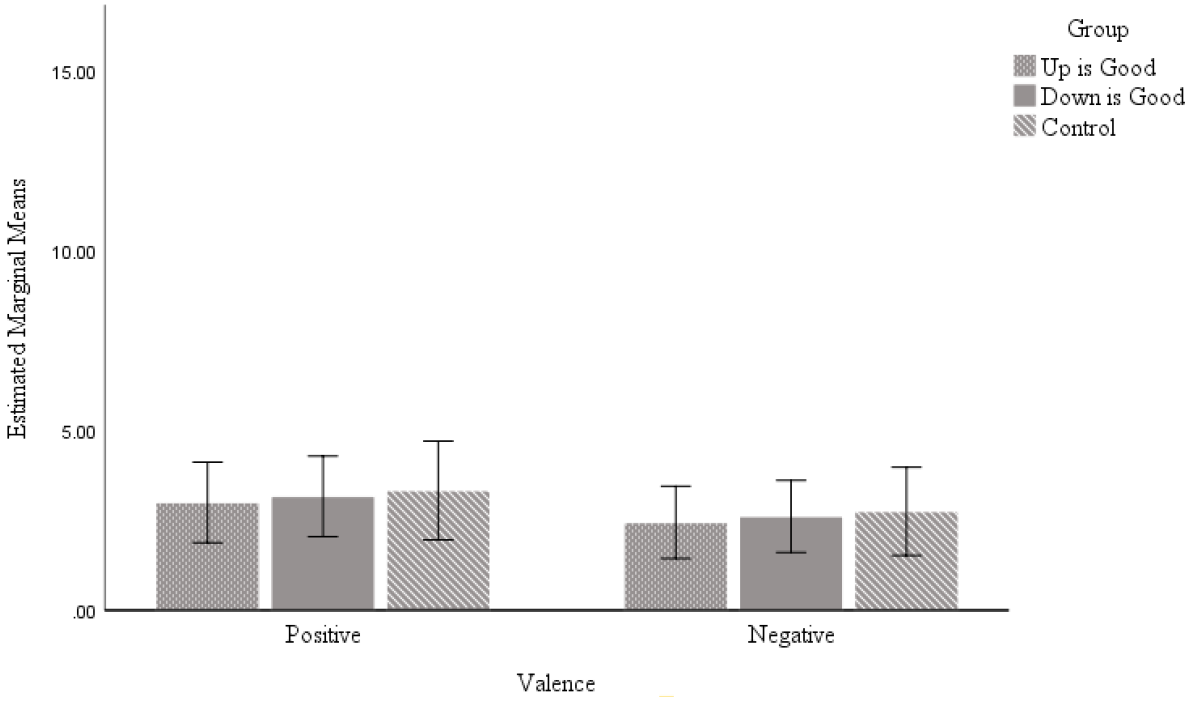

Fig. 4. Number of correctly recalled answers by experimental group and word valence

Discussion

Our second study extended the work of Casasanto and Bruin (2019) by testing the effect of movements, made in relation to the metaphor class *up is good/down is bad,* on recall rather than recognition of nonsense word meaning. Analysis indicated that while there was an effect of valence with participants recalling more positive words than negative words, participants' recall ability for the meanings of nonsense words was not affected by any

movements made – be they congruent, incongruent or neutral to the metaphor class *up is good/down is bad.* These results were inconsistent with those predicted by CMT.

General Discussion

Conceptual Metaphor Theory (CMT) is centred on the premise that some universal human experiences influence our cognitions and that conceptual metaphors reflect this phenomenon. The effect of this impact is observed when the metaphors are 'activated' by certain experimental tasks resulting in differences in measures like reaction times (Gibbs, 2011; Lakoff & Johnson, 1980a; Tolaas, 1991) and accuracy (Casasanto & de Bruin, 2019; Meier et al., 2004).

There has been a great deal of experimental research which supports CMT (Casasanto, 2009; Meier et al., 2008; Meier et al., 2004; Peetz & Soliman, 2016; Seidel & Prinz, 2018; Silvera et al., 2002) and, in the last decade, there has also been increased research endeavouring to replicate the work already done to test the robustness of the theory (Brandt et al., 2014; Chabris et al., 2019; Earp et al., 2014; E. LeBel & Wilbur, 2014; Meier et al., 2015). In our first study we sought to replicate an experiment demonstrating how CMT was reflected in relation to movement and the recognition of words (Casasanto & de Bruin, 2019). Our hypothesis was that we would see better recognition of the valenced meanings of nonsense words when a movement was made during learning which was congruent with the metaphor *up is good/down is bad* than when a neutral movement was made. We also hypothesised that we would see better recognition of the valenced meanings of nonsense words when a movement was made during learning which was neutral when compared to a movement that was incongruent to the metaphor class, as was seen in the study we were replicating (Casasanto & de Bruin, 2019). We went on to extend the study to see if we could observe this effect with the more cognitively challenging task of recall. In this case, our hypothesis was we would see the best results for recall of valenced meanings when a

movement during learning was congruent with the metaphor *up is good/down is bad.* We also predicted we would observe the poorest results for recall when the movement made was incongruent with the metaphor class. Following on from that, we hypothesised recall scores for words learnt when a neutral movement was made would sit between these two results.

In our replication of Casasanto and Bruin's (2019) study, we found there was some support for our hypothesis. Participants were better able to recognise valenced words when they had made a movement that was congruent with the metaphor class *up is good/down is bad* or when they made a neutral movement. They were less able to recognise nonsense word meanings if they made a movement that was incongruent with the metaphor class. These results mean there was some validation of CMT.

In our second study, we sought to extend the research by investigating if movements made during learning, that was either congruent, incongruent or neutral in relation to the metaphor class *up is good/down is bad,* had any effect on recalling the valenced meaning of nonsense words. Our hypothesis was not supported as there was no impact on recall accuracy in any of the movement groups. Our results relating to recall do not support CMT.

As with much of everyday life and with other experimental studies, Covid-19 had a marked impact on this research. The physical distancing required during New Zealand's Level Four lockdown meant that in-person data collection was no longer possible. This change then went on to affect our replication of Casasanto and Bruin's (2019) research, with notable differences in the method. The central change was a move to online data collection, which then required subsequent changes to many elements of the method.

The original experiment (Casasanto & de Bruin, 2019) asked participants to read a nonsense word on a card, turn the card over to read the nonsense word's meaning, attempt to memorise the word and its meaning and then place the card on a shelf –either high or low— once the study period was over. The shelf the card was placed on depended on the meaning's

valence. Participants did this three times for each of 16 nonsense words. Once the study phase was completed, the participants were tested on their ability to recognise the words' meanings by choosing between the target and a foil on a laptop.

With the change to online data collection, there were several modifications required during the experiment's study phase. Firstly, lockdown requirements meant all participants were confined to their place of residence and could not come into a laboratory for the experiment. This required all participants to carry out the experiment in their own home or university hostel accommodation. The most notable impact was the inability to provide individual access to uniform shelving. Consideration was given to delivering shelving to participants to assemble themselves but, as well as the obvious complexities of this approach (including would participants have a suitable table to assemble the shelving on and would they be able to assemble the shelving to the configuration required), it was also not ethically appropriate to deliver shelving to participants given the risk, albeit small, of contamination with the Covid-19 virus. Without the shelving, the up/down movement had to be changed to one that could be carried out independently of any equipment.

Another impact of the study moving online, was the variety of devices participants could use to carry out the experiment. Given the experiment was to be carried out at participants' homes or residence, devices used for the experiment would be decided by the participant, dependent on what they had available. This meant we had to adapt the experiment to allow for the use of desktop computers, laptops, tablets and mobiles phones. If participants used laptops, tablets or mobile phones they would not necessarily be seated at a desk, or even seated at all. If they took part on a tablet or mobile phone, we had to allow for the fact they might hold the device in one hand. These issues placed further restrictions on the possible movements. Finally, the up/down movements had to be observable on Zoom

while the participant was still looking at the device screen, as this was where the stimuli were to be presented.

The up/down movements chosen that met all these criteria were head and eye movements. For the valenced up/down movements, participants were invited to either look up at the ceiling or to look down at the floor or their lap (to allow for standing or sitting). A double blink (double so it could easily be observed virtually) was selected for the neutral control movement.

The change to the new movements could well have had an effect on our results (Förster & Strack, 1996a). The head movement –looking up to the ceiling or looking down to the floor or the participant's lap—is analogous to a head nod which has been shown to affect memory for valenced words. In an experiment examining what effect head movement might have on learning, individuals were given headphones and told they were going to assess the headphones' performance by moving their heads while listening to word lists. There were three conditions – some participants made a vertical head movement like nodding; some participants made a horizonal movement like head shaking and a final group were told to move their head in a circular motion. The participants were played positively and negatively valenced words through the headphones. They were then tested on their recognition of the words they heard. Participants who made the vertical movement, similar to a head nod, were more likely to remember positively valenced words than negatively valenced words, unlike the other two groups for whom there was no difference in recognition for the differently valenced words. (Förster & Strack, 1996a).

This effect may have also impacted the double eye blink. Observation of participants during the experiments revealed that some individuals nodded their head slightly when they blinked their eyes. Although our results showed that participants did recall positive meanings for the nonsense words better than negative meanings, there was no similar effect observed in

the recognition experiment. While these effects may explain some of the reasons why our results did not replicate those of Casasanto and Bruin (2019), they do not fully explain the pattern of results we did obtain, as we would expect any differences in word valence remembered to be seen across the congruent, incongruent and control groups in both the recognition and recall experiments, and we did not observe this. However, there may be interactions with other experimental conditions that differed from the original research.

Postural differences between participants when they completed the experiment may have also influenced results because body position has been shown to affect memory (Michalak et al., 2014; Peper et al., 2017; Riskind, 1983; Schulman & Shontz, 1971). Participants in this research were observed in a range of positions while they completed the experiment – some lay on their fronts or their backs, some sat on upright chairs at a desk or table, while others assumed a more relaxed posture on a sofa or armchair with their device on their knee or in their hand. In an experiment examining mood and memory, undergraduates were asked to assume a 'happy' posture and expression (described as sitting upright and smiling) or a 'sad' posture and expression (described as sitting slumped and frowning). The students who sat erect were more efficient at remembering positively valenced autobiographical memories and the students who slumped were more efficient at remembering negatively valenced autobiographical memories (Riskind, 1983). Similar results were obtained in other research. Students were asked to recall positive and negative memories in both an erect seated position and a slumped seated position. The participants found it easier to remember positive memories in an upright position and negative memories in a slouched position (Peper et al., 2017). More specifically in research concerning memory of word lists, psychiatric inpatients who were suffering from major depressive disorder, were shown two lists of words – one positively valenced and one negatively valenced. Those who were slumped showed better recall of negative words, than those who were seated with an

upright posture (Michalak et al., 2014). However, research findings are mixed, with another experiment finding posture did not affect memory efficiency for words (Schulman & Shontz, 1971). Participants were given one of four postures – standing, sitting erect, sitting bent and lying supine. Then a list of words was played to them and they were asked to verbally recall as many of the words as they could. There was no difference in the number of words remembered across the groups but posture did influence which words were remembered with words relating to standing (e.g. energetic, decisive) and sitting erect (e.g. attend, perceptive) being best remembered by those who assumed those postures (Schulman & Shontz, 1971). Similar to the evidence about head movement and memory, body posture could have played a part in our results. In particular it could have affected the valence of the words each individual might recall. However, again it would be presumed these effects would have had a comparable impact across all groups and across both experiments. So, while they may have had an effect, they cannot fully account for the configuration of our results.

Also of interest, is the impact posture could have on motivation for the experimental tasks (Riskind & Gotay, 1982). In research examining emotion and posture, individuals who were placed in a slumped position were less likely to persist with a frustrating task than those who were in an erect posture (Riskind & Gotay, 1982). In our study, those individuals who did not sit upright may have had poorer motivation, resulting in poorer application to the memory task and so poorer results. The effect of this may well be more marked in the more cognitively difficult recall task. This possible lack of motivation to learning the meanings may have obscured any effect of congruent or incongruent movement.

The differences in results between our recognition study and our recall study could be due to recall being a more difficult task than recognition (Haist et al., 1992; Hollingworth, 1913; Myers, 1914). In some of the earliest studies in memory research, Hollingworth asked five individuals to read 50 adjectives, and for each adjective to speak out loud its opposite.

They were then asked to write down as many of the read adjectives as they could recall. Immediately after that, they were asked to recognise the 50 adjectives from 100. For all five participants the number of words recalled was lower than those recognised, with three out of the five recognising all 50 words and the other two recognising 49. The average number of words recalled was 28.8. The work was extended with the participants asked to look at words, pictures, geometric shapes and nonsense syllables. Across all types of stimuli, recognition was better than recall (Hollingworth, 1913). In an experiment a year later, 687 school children were given a spelling test of 20 words, they were then asked to recall and recognise the words either a day or a week later. Again, recognition was much more successful than recall (Myers, 1914). In more recent research, 12 amnesic participants were compared to 19 non-amnesic controls. Both groups were shown word lists and then asked to recall or recognise the words for periods ranging from 15 seconds to eight weeks. For both groups, across all times, recognition was better than recall (Haist et al., 1992). For our research, the impact of the congruent or incongruent movements may have been overshadowed by the effect of the increased cognitive effort required for recall versus recognition.

Failure to replicate the earlier significant statistical difference could, of course, be due simply to chance. As the probability value of $p = .05$ was used, then would then be expected that for five per cent of experimental replications, a non-significant result would be obtained even if the hypothesis was valid.

There has been some successful reproduction of experiments which support CMT. Meier et al. (2015) replicated their previous work investigating the metaphors *white is good*/*black is bad* (Meier et al., 2004). In this study, they had a large sample size with 980 participants, more five times the original sample size. As with the earlier research, the results upheld CMT. Results showed that the congruent pairings of positive words in a white font

and negative words in a black font were recognised more quickly than the incongruent pairings of positive words in a black font and negative words in a white font. However, further searching of the literature failed to reveal any other successful replications of CMT experiments.

Replication is considered to be key to robust research. This is because reproducibility is deemed to be the strongest test for any scientific study (Koole & Lakens, 2012; Pashler & Wagenmakers, 2012; Tackett et al., 2019). This has been debated with particular reference to psychological studies (Koole & Lakens, 2012; Pashler & Wagenmakers, 2012; Tackett et al., 2019). In a now famous international collaboration, which replicated 100 psychological experiments, a mere 36 per cent of the repeated research resulted in significant outcomes (Open Science, 2015). Studies of this kind have led to what is called the 'replication crisis' in psychology. Publication bias been suggested as a main cause for this crisis (Koole & Lakens, 2012; Simmons et al., 2016; Tackett et al., 2019; Wiggins & Christopherson, 2019). Publication bias is outlined as scientific journals' disinclination to feature replications of research or to publish studies which support their null hypothesis (Simmons et al., 2016; Tackett et al., 2019; Wiggins & Christopherson, 2019). The proposed reason for this bias is that original and narratively compelling work (Shrout & Rodgers, 2018) holds more appeal to the general media and so leads to wider publicity.

Limitations

There were several limitations in this research, most due to the changes imposed to experimental design due to the move from in-person data collection to online data collection because of Covid-19 lockdown requirements.

In the original experiment, the environment was completely within the control of the researchers. With online data collection this control was lost. Attempts were made to

manage noise and distraction as much as possible. The email participants received when they signed up for the experiment specified that they should be in a quiet place with no one else in the room. If it was obvious the participants were in an inappropriate environment, they were asked to shift somewhere more suitable, but it was not possible to police this stringently, as the Zoom format only allowed for a view of the space behind the participant, not the entire room. Also, if the participants used headphones it was not possible to hear any extraneous noise like speech, music, or environmental noise.

Research has demonstrated that a noisy environment can impair memory (Banbury & Berry, 1998; Boman et al., 2005; Kantner, 2009; LeCompte, 1994; Shield & Dockrell, 2003). In an experiment, 50 participants were shown a list of 12 letters in one of three environments – quiet, white noise and irrelevant speech. They were then asked to recall as many letters as they could. The individuals in the irrelevant speech condition had reduced recall when compared to performance in quiet spaces and with white noise (LeCompte, 1994). The impact of irrelevant speech on recognition was also examined (LeCompte, 1994). In this instance, 52 undergraduates were shown onscreen lists of words and then asked to decide if a probe word had been present in the list. Of the two conditions, participants in the white noise condition were more likely to correctly recognise a previously seen word than those in the irrelevant speech condition (LeCompte, 1994). Listening to music with lyrics has also been shown to reduce scores in serial recall tests of lists featuring semantically linked words when compared with individuals who listened to silence or instrumental music (Kantner, 2009). Office noise can impair memory in relation to a prose recall task and mental arithmetic (Banbury & Berry, 1998). Even noise external to the room, like road traffic noise has been shown to reduce cued recall and recognition, as well as impact children's learning. In a study which compared silence with irrelevant speech and road traffic noise, cued recall and recognition in a text memory task were negatively affected in both noise conditions when

compared to silence (Boman et al., 2005). A review by Shield and Dockrell (2003) showed high-noise classrooms resulted in poor academic performance. Our inability to control the environment of our participants could have resulted in poorer learning and therefore lower performance. This would be seen across both experiments and all groups.

Another change with online data collection was that the researcher was unable to check that the participants had correctly identified the valence of the meanings. In the original experiment (Casasanto & de Bruin, 2019) the researchers checked the participants had placed the word/meaning cards on the correct shelf as determined by the meaning's valence. The online survey design meant this was not possible in the current study. However, in Casasanto and de Bruin's (2019) work none of the participants mis-identified valence so this was deemed a low risk for this research.

Another limitation was sample size. A larger sample size had been planned but changes due to the Covid-19 pandemic led to lower participant numbers. Participants for this study were drawn from undergraduates who were required to take part in experiments for course credit. In order to qualify for the credits, only a small number of studies was required to be completed. This experiment was notably more taxing than the other studies available to participants. All the other studies were online surveys that undergraduates could complete in their own time. This study required participants to complete the study at a specified time. Also, the study asked those individuals taking part to be observed by Zoom and many students were unwilling to do this. Finally, the study required students to perform an active memory task, where nearly all of the remaining studies were asking students to give information or opinions. These additional conditions meant a much lower uptake of the experiment, with only 60 per cent of available experimental slots being filled. A bonus of .25 credit was offered as an inducement to increase uptake but had little effect.

Future directions

Given the changes due to Covid-19 impacts, an exact replication of Casasanto and Bruin's (2019) experiment would be an important future study. As would the extension of their research examining recall. Testing recall would test not only the transferability of Conceptual Metaphor Theory into another facet of memory but would also address any concerns about possible ceiling effects in their study (Casasanto & de Bruin, 2019).

It would be valuable to extend the research to examine the impact of movement on memory using different metaphor classes which reflect different dimensional axes. The experiment could be repeated in the horizontal plane with the metaphor class *right is good/left is bad.* This plane has been found to be less salient than the vertical plane (Franklin & Tversky, 1990), so it would be useful to explore if the learning effect was replicated in this dimension. Researching this axis would also allow the exploration of any impact of handedness on CMT and memory. Previous research has shown while there is cultural uniformity in relation to the horizontal dimension and valence, individuals often associate positive valence with the side of their dominant hand (Casasanto, 2009; Casasanto & Chrysikou, 2011; Casasanto & Henetz, 2012; Ping et al., 2009).

Research into the horizontal axis would also allow experimentation into the effect of CMT on a metaphor class unrelated to valence. Written language direction has been found to influence metaphorical representations of time in terms of the horizontal dimension. English speakers represent time, from past to present to future, in a left to right direction, while Arabic speakers representing time in a right to left direction mirroring the dextrosinistral nature of their written language (Tversky & et al., 1991). Given the inconsistency of some experimental results into CMT which largely focus on the vertical axis, it would be useful to see if an effect can be observed in this instance.

Extending this work by examining the recognition and recall of images rather than words for stimuli would be beneficial. This would permit the testing of younger or non-literate participants. This would allow research to see if this effect is robust enough to be replicated in other participant groups and investigate when CMT can be first observed in children.

Conclusion

The idea that universal physical experiences become associated with more abstract concepts and this association then goes on to affect our behaviour and cognitive processes is the central construct in Conceptual Metaphor Theory (CMT). While an appealing and engaging theory with considerable published evidence, the lack of consistency in CMT research results implies these effects may not as robust as initially thought. Further research to elucidate more information about where and when these effects are reliably observed and continued attempts at replication of previous studies remains a priority in this area.

References

Aksan, Y., & Kantar, D. (2008). No Wellness Feels Better than This Sickness: Love Metaphors from a Cross-Cultural Perspective. *Metaphor and Symbol, 23*(4), 262-291. doi:10.1080/10926480802426795

Amin, T., Fredrik, J., & Haglund, J. (2015). Conceptual Metaphor and Embodied Cognition in Science Learning: Introduction to special issue. *International Journal of Science Education, 37*. doi:10.1080/09500693.2015.1025245

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current biology, 16*(18), 1818-1823. https://search.proquest.com/docview/68871749?accountid=14782

Banbury, S., & Berry, D. C. (1998). Disruption of office-related tasks by speech and office noise. *British Journal of Psychology, 89*(3), 499-517.

Banerjee, P., Chatterjee, P., & Sinha, J. (2012). Is it light or dark? Recalling moral behavior changes perception of brightness. *Psychological Science, 23*(4), 407-409. doi:http://dx.doi.org/10.1177/0956797611432497

Barsalou, L. W. (2008). Grounded cognition. *Annual review of psychology, 59*, 617-645. https://search.proquest.com/docview/70158572?accountid=14782

Bayer, M., Sommer, W., & Schacht, A. (2012). Font size matters-emotion and attention in cortical responses to written words. *PLoS One, 7*(5), 1. doi:http://dx.doi.org/10.1371/journal.pone.0036042

Black, M. (1962). *Models and Metaphors.* Cornell University Press.

Boman, E., Enmarker, I., & Hygge, S. (2005). Strength of noise effects on memory as a function of noise source and age. *Noise & Health, 7*(27), 11-26. doi:http://dx.doi.org/10.4103/1463-1741.31636

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review, 112*(1), 193-216. https://search.proquest.com/docview/67341919?accountid=14782

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings*. (Tech. Report C-1). Gainesville: University of Florida, Center for Research in Psychophysiology.

Brandt, M. J., Ijzerman, H., & Blanken, I. (2014). Does recalling moral behavior change the perception of brightness? A replication and meta-analysis of Banerjee, Chatterjee, and Sinha (2012). *Social Psychology, 45*(3), 246-252. doi:http://dx.doi.org/10.1027/1864-9335/a000191

Casasanto, D. (2009). Embodiment of Abstract Concepts: Good and Bad in Right- and Left-Handers. *Journal of Experimental Psychology: General, 138*(3), 351-367. doi:http://dx.doi.org/10.1037/a0015854

Casasanto, D., & Chrysikou, E. G. (2011). When left is "right": motor fluency shapes abstract concepts. *Psychological Science, 22*(4), 419-422. doi:http://dx.doi.org/10.1177/0956797611401755

Casasanto, D., & de Bruin, A. (2019). Metaphors we learn by: Directed motor action improves word learning. *Cognition, 182,* 177-183. doi:10.1016/j.cognition.2018.09.015

Casasanto, D., & Dijkstra, K. (2010). Motor action and emotional memory. *Cognition, 115*(1), 179-185. doi:https://doi.org/10.1016/j.cognition.2009.11.002

Casasanto, D., & Henetz, T. (2012). Handedness shapes children's abstract concepts. *Cognitive Science, 36*(2), 359-372. doi:http://dx.doi.org/10.1111/j.1551-6709.2011.01199.x

Chabris, C. F., Heck, P. R., Mandart, J., Benjamin, D. J., & Simons, D. J. (2019). No evidence that experiencing physical warmth promotes interpersonal warmth: Two

failures to replicate Williams and Bargh (2008). *Social Psychology*, *50*(2), 127-132. doi:http://dx.doi.org/10.1027/1864-9335/a000361

Citron, F. M. M., & Goldberg, A. E. (2014). Metaphorical sentences are more emotionally engaging than their literal counterparts. *Journal of cognitive neuroscience*, *26*(11), 2585-2595. doi:http://dx.doi.org/10.1162/jocn_a_00654

Crawford, L. E. (2009). Conceptual metaphors of affect. *Emotion Review*, *1*(2), 129-139. doi:http://dx.doi.org/10.1177/1754073908100438

Crawford, L. E. (2014). The role of conceptual metaphor in memory. In M. Landau, M. D. Robinson, & B. P. Meier (Eds.), *The power of metaphor: Examining its influence on social life* (pp. 65-83). American Psychological Association.

Crawford, L. E., Cohn, S. M., & Kim, A. B. (2014). "Good Is Up" Is Not Always Better: A Memory Advantage for Words in Metaphor-Incompatible Locations. *PLoS One*, *9*(9). doi:http://dx.doi.org/10.1371/journal.pone.0108269

Crawford, L. E., Margolies, S. M., Drake, J. T., & Murphy, M. E. (2006). Affect biases memory of location: Evidence for the spatial representation of affect. *Cognition and Emotion*, *20*(8), 1153-1169. doi:http://dx.doi.org/10.1080/02699930500347794

Desai, R. H., Binder, J. R., Conant, L. L., Mano, Q. R., & Seidenberg, M. S. (2011). The neural career of sensory-motor metaphors. *Journal of Cognitive Neuroscience*, *23*(9), 2376-2386. doi:http://dx.doi.org/10.1162/jocn.2010.21596

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition*, *118*(1), 111-115. doi:http://dx.doi.org/10.1016/j.cognition.2010.09.012

Earp, B. D., Everett, J. A. C., Madva, E. N., & Hamlin, J. K. (2014). Out, Damned Spot: Can the "Macbeth Effect" Be Replicated? *Basic and Applied Social Psychology*, *36*(1), 91. https://search.proquest.com/docview/1504639850?accountid=14782

Ekman, P,. Friesen, W,. O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., and Tomita, M. (1972). Universals and cultural differences in facial expressions of emotions. In J. K. Cole (Ed.). *Nebraska Symposium on Motivation* (pp. 712-717). University of Nebraska Press.

Fetterman, A. K., Bair, J. L., Werth, M., Landkammer, F., & Robinson, M. D. (2016). The scope and consequences of metaphoric thinking: Using individual differences in metaphor usage to understand how metaphor functions. *Journal of Personality and Social Psychology, 110*(3), 458-476. https://search.proquest.com/docview/1811844745?accountid=14782

Förster, J., & Strack, F. (1996). Influence of overt head movements on memory for valenced words: A case of conceptual-motor compatibility. *Journal of Personality and Social Psychology, 71*(3), 421-430. doi:http://dx.doi.org/10.1037/0022-3514.71.3.421

Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology; General, 119*(1), 63-76. https://search.proquest.com/docview/1297408438?accountid=14782

Gibbs, R. W., Jr. (2011). Evaluating Conceptual Metaphor Theory. *Discourse Processes, 48*(8), 529-562. doi:http://dx.doi.org/10.1080/0163853X.2011.606103

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, 66(3), 325-331. https://search.proquest.com/docview/64006152?accountid=14782

Graesser, A. C., Mio, J. S., & Millis, K. K. (1989). Metaphors in persuasive communication. In D. Meutsch & R. Viehoff (Eds.), *Comprehension of literary discourse: Results and problems of interdisciplinary approaches* (pp. 131-154). Walter de Gruyter.

Haist, F., Shimamura, A. P., & Squire, L. R. (1992). On the relationship between recall and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(4), 691-702. doi:http://dx.doi.org/10.1037/0278-7393.18.4.691

Hauk, O., & Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping, 21*(3), 191-201. https://search.proquest.com/docview/80139952?accountid=14782

Hollingworth, H. L. (1913). Characteristic differences between recall and recognition. *American Journal of Psychology, 24*, 532-544. https://search.proquest.com/docview/1289676648?accountid=14782

Jackson, L. A., & Ervin, K. S. (1992). Height stereotypes of women and men: The liabilities of shortness for both sexes. *Journal of Social Psychology, 132*(4), 433-455. https://search.proquest.com/docview/1290719595?accountid=14782

Johnson, D. J., Cheung, F., & Donnellan, M. B. (2014). Does cleanliness influence moral judgments? A direct replication of Schnall, Benton, and Harvey (2008). *Social Psychology, 45*(3), 209-215. doi:http://dx.doi.org/10.1027/1864-9335/a000186

Kacinik, N. A. (2014). Sticking your neck out and burying the hatchet: what idioms reveal about embodied simulation. *Frontiers in Human Neuroscience, 8*, 689. doi:http://dx.doi.org/10.3389/fnhum.2014.00689

Kantner, J. (2009). Studying with music: Is the irrelevant speech effect relevant? In M. R. Kelley (Ed.), *Applied Memory* (pp. 19-40). Nova Science Publishers.

Koch, S. C., Glawe, S., & Holt, D. V. (2011). Up and down, front and back: Movement and meaning in the vertical and sagittal axes. *Social Psychology, 42*(3), 214-224. doi:http://dx.doi.org/10.1027/1864-9335/a000065

Koole, S. L., & Lakens, D. (2012). Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science, 7*(6), 608-614. doi:http://dx.doi.org/10.1177/1745691612462586

Kövecses, Z. (2005). *Metaphor in culture : universality and variation*. Cambridge University Press.

Lacey, S., Stilla, R., & Sathian, K. (2012). Metaphorically feeling: Comprehending textural metaphors activates somatosensory cortex. *Brain and Language, 120*(3), 416-421. doi:http://dx.doi.org/10.1016/j.bandl.2011.12.016

LaFrance, M., & Mayo, C. (1978). Cultural aspects of nonverbal communication. *International Journal of Intercultural Relations, 2*(1), 71-89. doi:https://doi.org/10.1016/0147-1767(78)90029-9

Lai, V. T., Howerton, O., & Desai, R. H. (2019). Concrete processing of action metaphors: Evidence from ERP. *Brain research, 1714*, 202-209. doi:http://dx.doi.org/10.1016/j.brainres.2019.03.005

Lakoff, G., Espenson, J., & Goldberg, A. (1989). *Master Metaphor List.* (2nd ed.). University of California.

Lakoff, G., Johnson, M., & Neisser, J. U. (2001). Philosophy in the flesh. *Language, 77*(1), 166-168. https://search.proquest.com/docview/39045503?accountid=14782

Lakoff, G., & Johnson, M. (1980a). Conceptual metaphor in everyday language. *The Journal of Philosophy, 77*(8), 453-489. https://search.proquest.com/docview/1290532230?accountid=14782

Lakoff, G., & Johnson, M. (1980b). The metaphorical structure of the human conceptual system. *Cognitive Science, 4*(2), 195-208. doi:http://dx.doi.org/10.1207/s15516709cog0402_4

LeBel, E., & Wilbur, C. (2014). Big secrets do not necessarily cause hills to appear steeper. *Psychonomic Bulletin & Review, 21*(3), 696-700. doi:10.3758/s13423-013-0549-2

LeBel, E. P., & Campbell, L. (2013). Heightened sensitivity to temperature cues in individuals with high anxious attachment: Real or elusive phenomenon? *Psychological Science, 24*(10), 2128-2130. doi:http://dx.doi.org/10.1177/0956797613486983

LeCompte, D. C. (1994). Extending the irrelevant speech effect beyond serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1396-1408. doi:http://dx.doi.org/10.1037/0278-7393.20.6.1396

Lynott, D., Corker, K. S., Wortman, J., Connell, L., Donnellan, M. B., Lucas, R. E., & O'Brien, K. (2014). Replication of "Experiencing physical warmth promotes interpersonal warmth" by Williams and Bargh (2008). *Social Psychology, 45*(3), 216-222. doi:http://dx.doi.org/10.1027/1864-9335/a000187

Marmolejo-Ramos, F., Correa, J. C., Sakarkar, G., Ngo, G., Ruiz-Fernández, S., Butcher, N., & Yamada, Y. (2017). Placing joy, surprise and sadness in space: A cross-linguistic study. *Psychological Research, 81*(4), 750-763. doi:http://dx.doi.org/10.1007/s00426-016-0787-9

Marmolejo-Ramos, F., María Rosa, E., Yamada, Y., Hamm, N. F., & Noguchi, K. (2013). Appraisal of space words and allocation of emotion words in bodily space. *PLoS One, 8*(12), e81688. doi:http://dx.doi.org/10.1371/journal.pone.0081688

McGlone, M. S. (2007). What is the explanatory value of a conceptual metaphor? *Language & Communication, 27*(2), 109-126. doi:http://dx.doi.org/10.1016/j.langcom.2006.02.016

Meier, B. P., Fetterman, A. K., & Robinson, M. D. (2015). Black and white as valence cues: A large-scale replication effort of Meier, Robinson, and Clore (2004). *Social Psychology*, *46*(3), 174-178. doi:http://dx.doi.org/10.1027/1864-9335/a000236

Meier, B. P., Hauser, D. J., Robinson, M. D., Friesen, C. K., & Schjeldahl, K. (2007). What's "up" with God? Vertical space as a representation of the divine. *Journal of Personality and Social Psychology*, *93*(5), 699-710. doi:http://dx.doi.org/10.1037/0022-3514.93.5.699

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up. *Psychological Science*, *15*(4), 243-247. https://search.proquest.com/docview/195582726?accountid=14782

Meier, B. P., Robinson, M. D., & Caven, A. J. (2008). Why a Big Mac is a good Mac: Associations between affect and size. *Basic and Applied Social Psychology*, *30*(1), 46-55. doi:http://dx.doi.org/10.1080/01973530701866516

Meier, B. P., Robinson, M. D., & Clore, G. L. (2004). Why good guys wear white: Automatic inferences about stimulus valence based on brightness. *Psychological Science*, *15*(2), 82-87. doi:10.1111/j.0963-7214.2004.01502002.x

Mischnat, J., & Teismann, T. (2014). Sitting posture makes a difference: Embodiment effects on depressive memory bias. *Clinical Psychology & Psychotherapy*, *21*(6), 519-524. https://search.proquest.com/docview/1697761502?accountid=14782

Moody, C. L., & Gennari, S. P. (2010). Effects of implied physical effort in sensory-motor and pre-frontal cortex during language comprehension. *NeuroImage*, *49*(1), 782-793. doi:http://dx.doi.org/10.1016/j.neuroimage.2009.07.065

Mulligan, N. W. (1996). The effects of perceptual interference at encoding on implicit memory, explicit memory, and memory for source. *Journal of Experimental Psychology*, *22*(5), 1067-1087. https://search.proquest.com/docview/214371547?accountid=14782

Myers, G. C. (1914). A comparative study of recognition and recall. *Psychological Review, 21*(6), 442-456. doi:http://dx.doi.org/10.1037/h0071703

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science (New York, N.Y.), 349*(6251), e4716. doi:http://dx.doi.org/10.1126/science.aac4716

Palma, T. A., Garrido, M. V., & Semin, G. R. (2011). Grounding person memory in space: Does spatial anchoring of behaviors improve recall? *European Journal of Social Psychology, 41*(3), 275-280. doi:http://dx.doi.org/10.1002/ejsp.795

Parzuchowski, M., Bocian, K., & Gygax, P. (2016). Sizing up objects: The effect of diminutive forms on positive mood, value, and size judgments. *Frontiers in Psychology, 7*, 197-209. https://search.proquest.com/docview/1861220830?accountid=14782

Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PLoS One, 7*(8), e42510. doi:http://dx.doi.org/10.1371/journal.pone.0042510

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*(6), 528-530. doi:http://dx.doi.org/10.1177/1745691612465253

Peetz, J., & Soliman, M. (2016). Big money: The effect of money size on value perceptions and saving motivation. *Perception, 45*(6), 631-641. doi:http://dx.doi.org/10.1177/0301006616629033

Peper, E. P., Lin, I. M., Harvey, R. P., & Perez, J. (2017). How posture affects memory recall and mood. *Biofeedback, 45*(2), 36-41. doi:http://dx.doi.org/10.5298/1081-5937-45.2.01

Ping, R. M., Dhillon, S., & Beilock, S. L. (2009). Reach for what you like: The body's role in shaping preferences. *Emotion Review, 1*(2), 140-150. doi:http://dx.doi.org/10.1177/1754073908100439

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia, 47*(2), 388-396. doi:http://dx.doi.org/10.1016/j.neuropsychologia.2008.09.017

Reese, E., Hayne, H., & MacDonald, S. (2008). Looking back to the future: Māori and Pakeha mother–child birth stories. *Child Development, 79*(1), 114-125.

Riskind, J. H. (1983). Nonverbal expressions and the accessibility of life experience memories: A congruence hypothesis. *Social Cognition, 2*(1), 62-86. doi:http://dx.doi.org/10.1521/soco.1983.2.1.62

Riskind, J. H., & Gotay, C. C. (1982). Physical posture: Could it have regulatory or feedback effects on motivation and emotion? *Motivation and Emotion, 6*(3), 273-298. doi:http://dx.doi.org/10.1007/BF00992249

Robinson, M. D., & Fetterman, A. K. (2015). The embodiment of success and failure as forward versus backward movements. *PLoS One, 10*(2), e0117285. doi:http://dx.doi.org/10.1371/journal.pone.0117285

Rueschemeyer, S. A., van Rooij, D., Lindemann, O., Willems, R. M., & Bekkering, H. (2010). The Function of Words: Distinct neural correlates for words denoting differently manipulable objects. *Journal of Cognitive Neuroscience, 22*(8), 1844-1851. doi:http://dx.doi.org/10.1162/jocn.2009.21310

Schnall, S., Benton, J., & Harvey, S. (2008). With a clean conscience: cleanliness reduces the severity of moral judgments. *Psychological Science, 19*(12), 1219-1222. https://search.proquest.com/docview/37171897?accountid=14782

Schulman, D., & Shontz, F. C. (1971). Body posture and thinking. *Perceptual and motor skills, 32*(1), 27-33. https://search.proquest.com/docview/80933424?accountid=14782

Seidel, A., & Prinz, J. (2018). Great works: A reciprocal relationship between spatial magnitudes and aesthetic judgment. *Psychology of Aesthetics, Creativity, and the Arts, 12*(1), 2-10. doi:http://dx.doi.org/10.1037/aca0000100

Shield, B. M., & Dockrell, J. E. (2003). The effects of noise on children at school: A review. *Building Acoustics, 10*(2), 97-116. doi:10.1260/135101003768965960

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the Replication Crisis. *Annual Review of Psychology, 69*, 487-510. doi:http://dx.doi.org/10.1146/annurev-psych-122216-011845

Shutova, E., Devereux, B. J., & Korhonen, A. (2013). Conceptual metaphor theory meets the data: A corpus-based human annotation study. *Language Resources and Evaluation, 47*(4), 1261-1284. doi:http://dx.doi.org/10.1007/s10579-013-9238-z

Silvera, D. H., Josephs, R. A., & Giesler, R. B. (2002). Bigger is better: The influence of physical size on aesthetic preference judgments. *Journal of Behavioral Decision Making, 15*(3), 189. https://search.proquest.com/docview/214687007?accountid=14782

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2016). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359-1366.

Slepian, M. L., Masicampo, E. J., Toosi, N. R., & Ambady, N. (2012). The physical burdens of secrecy. *Journal of Experimental Psychology, 141*(4), 619-624. doi:http://dx.doi.org/10.1037/a0027598

Taylor, J. E. T., Lam, T. K., Chasteen, A. L., & Pratt, J. (2015). Bow your head in shame, or hold your head up with pride: Semantic processing of self-esteem concepts orients

attention vertically. *PLoS One, 10*(9) e0137704. doi:http://dx.doi.org/10.1371/journal.pone.0137704

Tackett, J. L., Brandes, C. M., King, K. M., & Markon, K. E. (2019). Psychology's Replication Crisis and Clinical Psychological Science. *Annual Review of Clinical Psychology, 15,* 579-604. http://dx.doi.org.helicon.vuw.ac.nz/10.1146/annurev-clinpsy-050718-095710

Tolaas, J. (1991). Notes on the origin of some spatialization metaphors. *Metaphor & Symbolic Activity, 6*(3), 203-218. doi:http://dx.doi.org/10.1207/s15327868ms0603_4

Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *Proceedings of the National Academy of Sciences of the United States of America, 105*(33), 11655-11660. doi:http://dx.doi.org/10.1073/pnas.0802686105

Tversky, B. (2011). Spatial thought, social thought. *Spatial Dimensions of Social Thought, 18*, 17-38.

Tversky, B. (2014). Visualizing thought. In W. Huang (Ed.), *Handbook of Human Centric Visualization* (pp. 3-40). Springer.

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology, 23*(4), 515-557. https://search.proquest.com/docview/62919010?accountid=14782

Ulrich, R., & Maienborn, C. (2010). Left-right coding of past and future in language: the mental timeline during sentence processing. *Cognition, 117*(2), 126-138. doi:http://dx.doi.org/10.1016/j.cognition.2010.08.001

Wiggins, B. J., & Chrisopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology, 39*(4), 202-217. DOI:10.1037/teo0000137

Willems, R. M., & Casasanto, D. (2011). Flexibility in embodied language understanding. *Frontiers in Psychology, 2*, 116. doi:http://dx.doi.org/10.3389/fpsyg.2011.00116

Willems, R. M., Toni, I., Hagoort, P., & Casasanto, D. (2009). Body-specific motor imagery of hand actions: Neural evidence from right- and left-handers. *Frontiers in Human Neuroscience, 3*, 39. doi:http://dx.doi.org/10.3389/neuro.09.039.2009

Yap, A. J., Mason, M. F., & Ames, D. R. (2013). The powerful size others down: The link between power and estimates of others' size. *Journal of Experimental Social Psychology*, *49*(3), 591-594. https://search.proquest.com/docview/1353281542?accountid=14782

Yu, N. (1995). Metaphorical expressions of anger and happiness in English and Chinese. *Metaphor and Symbolic Activity*, *10*(2), 59-92. https://search.proquest.com/docview/85602897?accountid=14782

Zhong, C.-B., Bohns, V. K., & Gino, F. (2010). Good lamps are the best police: darkness increases dishonesty and self-interested behavior. *Psychological Science*, *21*(3), 311-314. doi:http://dx.doi.org/10.1177/0956797609360754

Appendix A

The word list for Experiments 1 and 2. Both lists were presented to each participant, in two blocks of eight words and meanings. Half the participants saw each nonsense word with its positively valenced meaning and half with the negatively valenced meaning.

NONSENSE WORDS	**Positive English meaning (with valence score)**	**Negative English meaning (with valence score)**
List A		
BANTED	comedy (8.37)	hurt (1.90)
YINTAC	romantic (8.32)	disaster (1.73)
HIPLER	paradise (8.72)	murderer (1.53)
SLACRE	joke (8.10)	misery (1.93)
NITTEM	comedy (8.37)	murderer (1.53)
DROWAT	paradise (8.72)	hurt (1.90)
JASING	romantic (8.32)	misery (1.93)
KREDSI	joke (8.10)	disaster (1.73)
List B		
FUTVEN	affection (8.39)	depression (1.85)
GENALT	kiss (8.26)	depression (1.85)
MEWLUR	excellence (8.38)	rejected (1.50)
PODILS	miracle (8.60)	rejected (1.50)
REFFIP	affection (8.39)	funeral (1.39)
CROMIL	miracle (8.60)	grief (1.69)
TONFEY	excellence (8.38)	grief (1.69)
ZEEKON	kiss (8.26)	funeral (1.39)

Appendix B

INSTRUCTIONS EXPERIMENT RECALL UP IS GOOD

Scientists recently discovered a new method to learn words in a foreign language, based on Motion and Emotion. In this experiment you are going to use this method to make it easier to learn words in a language that is very foreign -- in fact, it's only spoken on the planet Fribbalia! You will see each Fribbalian word and the English translation below it. You have to learn the English translations for each Fribbalian word.

The Motion and Emotion Method is based on the following scientific discovery. Positive words are linked to upward movements and negative words are linked to downward movements. This is what often happens in language with metaphors - 'my spirits soared' is a positive event, but 'down in the dumps' is negative. In this way, positive feelings are linked to upward movements and negative feelings to downward movements.

To use the Motion and Emotion Method of word learning you will be shown each Fribbalian word one at a time with its meaning in English. Each time you see a word, you will decide on the emotional content of the word -- whether the meaning of the word is positive or negative. If you think the English meaning is positive, look up to the ceiling. If you think the meaning is negative, look down at your lap or the floor. You must make this movement; it is vital to the experiment. Think about the emotional content of the word as you try to memorise the meaning of the Fribbalian word.

Then you will be tested. You will be shown the word in Fribbalian and asked to write down the correct English meaning.

You are going to learn 16 words in total, in 2 groups of 8 words. You will see each word 3 times. Once you have been tested on the second group of words, the experiment is over.

INSTRUCTIONS RECALL GOOD IS DOWN

Scientists recently discovered a new method to learn words in a foreign language, based on Motion and Emotion. In this experiment you are going to use this method to make it easier to learn words in a language that is very foreign -- in fact, it's only spoken on the planet Fribbalia! You will see each Fribbalian word and the English translation below it. You have to learn the English translations for each Fribbalian word.

The Motion and Emotion Method is based on the following scientific discovery. Positive words are linked to downward movements and negative words are linked to upward movements. This is what often happens in language with metaphors. 'she's got her feet on the ground' is positive, but 'he thinks he's above the rest of us' is negative. In this way positive feelings are linked to downward movements and negative feelings to upward movements.

To use the Motion and Emotion Method of word learning you will be shown each Fribbalian word one at a time with its meaning in English. Each time you see a word, you will decide on the emotional content of the word -- whether the meaning of the word is positive or negative. If you think the English meaning is positive, look down at your lap or the floor. If you think the meaning is negative, look up to the ceiling. You must make this movement; it

is vital to the experiment. Think about the emotional content of the word as you try to memorise the meaning of the Fribbalian word.

Then you will be tested. You will be shown the word in Fribbalian and asked to write down the correct English meaning.

You are going to learn 16 words in total, in 2 groups of 8 words. You will see each word 3 times. Once you have been tested on the second group of words, the experiment is over.

INSTRUCTIONS RECALL CONTROL

Scientists recently discovered a new method to learn words in a foreign language, based on emotion. In this experiment you are going to use this method to make it easier to learn words in a language that is very foreign -- in fact, it's only spoken on the planet Fribbalia! You will see each Fribbalian word and the English translation below it. You have to learn the English translations for each Fribbalian word.

The Emotion Method is based on the following scientific discovery. Words are remembered better if you think about their emotional content – whether the meaning of the word is positive or negative. By thinking about these feelings, you will remember the words better.

To use the Emotion Method of word learning you will be shown each Fribbalian word one at a time with its meaning in English. Each time you see a word, you will decide on the emotional content of the word -- whether the meaning of the word is positive or negative. Keep thinking about the emotional content as you try to memorise the meaning of

each Fribbalian word. As you memorise the word, blink your eyes twice. You must make this movement; it is vital to the experiment.

Then you will be tested. You will be shown the word in Fribbalian and asked to write down the correct English meaning.

You are going to learn 16 words in total, in 2 groups of 8 words. You will see each word 3 times. Once you have been tested on the second group of words, the experiment is over.

Appendix C

Misspelled words in recall experiment which were accepted as correct in the lenient condition:

mircale

mircle

miricle

greif

excelence

excellen

disarste

disaste

depressio

deppression

miser

romatic

romantuic

paradis

murdere