UNIVERSITY OF CALIFORNIA

SANTA CRUZ

LONELY OR JUST DISTANT?
THE ROLE OF INTERPRETATION IN THE EMOTIONAL IMPACT OF A METAPHORICAL GAME

A dissertation submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PSYCHOLOGY

by

Christopher R Karzmark

June 2020

The Thesis of Christopher R Karzmark
is approved:

_____
Professor Benjamin Storm, Chair

_____
Professor Nicolas Davidenko

_____
Professor Leila Takayama

_____
Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

Table of Contents

List of Tables and Figures

Abstract

Lonely or Just Distant?

The Role of Interpretation in the Emotional Impact of a Metaphorical Game

Christopher R Karzmark

When playing a metaphorical game, what is the role of the player's interpretation in the game's emotional impact? If players are consciously aware of the metaphor, do they experience a stronger effect of that metaphor? This dissertation extends part of the debate on deliberate metaphor (Gibbs, 2015b; Steen, 2008, 2017) into the empirical space of video games. Participants played a version of the game *Loneliness* (Magnuson, 2007). In the game, the player controls a square. When the player approaches other squares, the other squares move away as if rejecting the player. Since this game communicates primarily through the metaphor AFFECTION IS PROXIMITY, any change in the player's emotions can be attributed to this metaphor. In a series of experiments, participants played the game described above, a variation on that game in which the other squares approach the player's square, or *Tetris* as a baseline. Participants self-reported their emotional state before and after playing the game, and the change from before to after was the primary dependent variable. Experiment 1 established that *Loneliness* influences feelings of loneliness and acceptedness and piloted methods of measuring participants' perceptions of metaphor deliberateness. In Experiment 2, participants who played *Loneliness* were split into Aware and Unaware groups based on ratings of agreement with statements adapted from Gibbs (2015b). The Aware showed a significant change in both loneliness and

acceptedness, while the Unaware and baseline groups showed little or no change. These results are interpreted based on a set of hypotheses motivated by the paradox of metaphor (Steen, 2008, 2017), responses to the paradox (Gibbs, 2015), and implicit metaphor studies (Thibodeau & Boroditsky, 2011, 2013). The overall pattern of results is consistent deliberate metaphor theory, in that conscious awareness of metaphor in *Loneliness* seems to be critical to its emotional impact.

Acknowledgements

I would first like to thank my parents, Janet and Peter Karzmark for their unflinching support through what amounts to 24 straight years of school. Without the knowledge that you were always in my corner and the curiosity you instilled in me, there is no way in hell I would have come this far. Thanks, Mom, for always being ready to hear about my newest obsession, even back to naming Pokémon in order. Thanks, Dad, for your love of esoteric knowledge and even of tests, setting on the bedroom floor and doing your psych assessments for fun took the anxiety out of testing later in life. You've also been such a great example of life after the PhD, it's helped to know that there's a lot more to life than publications.

Thanks of course to my dissertation committee, Ben Storm, Nick Davidenko, and Leila Takayama. Leila, you always mange to bring in new literature to the discussion without making me feel silly for not already knowing it, and you've provided an excellent lifeline to the world outside academia when it was sorely needed. Nick, you have a remarkable ability to help condense a wide-ranging discussion into actionable steps, which has been invaluable in getting the interesting topics we throw around in meetings into design and writing decisions. Ben, thank you for helping keep me focused on ideas that were likely to produce significant, publishable findings, as clearly, I have a tendency to go in weird directions. More importantly, thank you so much for accepting me into your lab, and for allowing me the autonomy to do the work I love. You made a welcoming home for a lab orphan, and I sincerely do not know what I would have done if you hadn't taken me in.

Thanks also to my first advisor Ray Gibbs. Ray, you accepted a kid who wanted to do something with games and metaphor and took me so much further. You showed me so much more of the embodiment world, and after about 4 years I think I've finally got a handle on dynamical systems (mostly). I can't wait to drink Budweiser at the Rush Inn again and hopefully other beers when we're at another conference in who-knows-where. Even after you retired you've been a sounding board and a stabilizing presence in my life, and I know that will continue for many years and many miles to come.

Thanks also to the many lab mates I've had over the years. Pat, I've said this before but if feels more true every day, you're the sister I never had. You've been an inspiration throughout my entire grad school career, even if you sometimes set the bar too high – we never could convince you to half-ass anything. Julia, one of the bright sides of changing labs was having you as a lab mate and getting to know you better. I'm so glad that we got to finish this process together comrade.

Thanks to my fabulous cognitive cohort, Julia Sores, Jennifer Day, Peter Krause, and Alina Larson. Hanging out with you in first year is what made Santa Cruz feel like home, and I'll always remember our Community parties and the famous pirate ship. We went into this together, we all made it through, we have the best brunch game, and no matter where we all go in life, I know we'll be friends.

Thanks to my research assistants, there have been too many of you to name but I will highlight Karlee Sprowl, Kameela Afifi, Grayson Mullen, Evan Amsden, Elizabeth Trimble, Alex Supica, Tyler Mendoza, Hannah Lloyd, Eric Kellenberger,

Zixuan Zheng, and Auveen Maghsoodnia who all helped collect and analyze data for this dissertation. You were often the first audience for a new research idea, you helped me refine designs, we waded through papers together (some papers were more arduous than others), you listened to an occasional rant about academia, and of course you ran innumerable participants and coded vast piles of qualitative data. I hope to stay in contact with you and can't wait to see where your lives take you – and as always, don't hesitate to ask for help.

Thanks to my comrades from the picket line, who reminded me that this school has a beautifully supportive grad community despite an administration that keeps choosing to actively harm its students. The strike for COLA brought some of my most cherished memories of grad school, and I hope I can live up to the commitment to praxis I saw in so many of you. Every interaction I have had with the UCSC administration has been a hindrance to my graduate education, and the inclusion of a particularly odious administrator's name on the first page of this document is done under protest.

Thanks to the wonderful housemates I've had over these six years. In particular, thanks to Mary Harpin for reminding me of the world outside of academia, and always being down to watch a movie.

Thanks to all my fellow tabletop players: Anie, Allie, John, Jen, BJ, Andy, Lee, Brad, Julia, Chris, Emily, Allison, Ryan, Kelsey, Justin, Jordan, Roxy, Mike, Nicole, Sam, Chris, Ben, Allen, and Trevor. This hobby has given me a creative outlet for 6 years of grad school, and after hundreds of hours I still can't wait to keep

playing. We explored and crafted fantastical worlds together, infiltrated cities, fought demons, and some of you even went along with three sessions of my deepest nerd coming out in a fantasy constitutional convention.

Thanks to Jordan Magnuson for making the game *Loneliness* – a beautiful piece of interactive art that has neatly book-ended my graduate career. Thanks even more-so for helping with this dissertation by sending your alternate versions that formed the conditions in Experiment 1. Thanks also to James Portnow and the rest of the Extra Credits crew for bringing this game to my attention back in 2012 – I can trace my graduate research back to this video more clearly than other single factor in my life.

Finally, thanks to my girlfriend Anie Thompson. You've been a shoulder to cry on, an assurance that life after academia exists, a singing-and-dancing in the kitchen companion, a trivia competitor, a complementary nerd, and after three months quarantined together I still wouldn't pick anyone else. And hey, you play a pretty great angry psychic spider.

CHAPTER I

Introduction

Suppose you are playing a simple browser game called *Loneliness* (Magnuson, 2007). You use the arrow keys to control a black square, moving it around a two-dimensional space with a cloudy grey background. You eventually encounter other groups of black squares, and when you approach the other squares, they move away from you, and disappear after a few seconds. There is no way to lose the game, and the goal is left open to the player. The game ends when you travel a certain distance upward, which is somewhat conventional for other two-dimensional games. When you reach the end of the game, the screen fades to black and some lines of text appear, reading: "Children and adolescents in Korea are the least satisfied with their lives among 26 member countries of the OECD. Many report **loneliness** as a major factor. I taught those kids for a year. This notgame was made for them. And for anyone who has ever felt lonely."

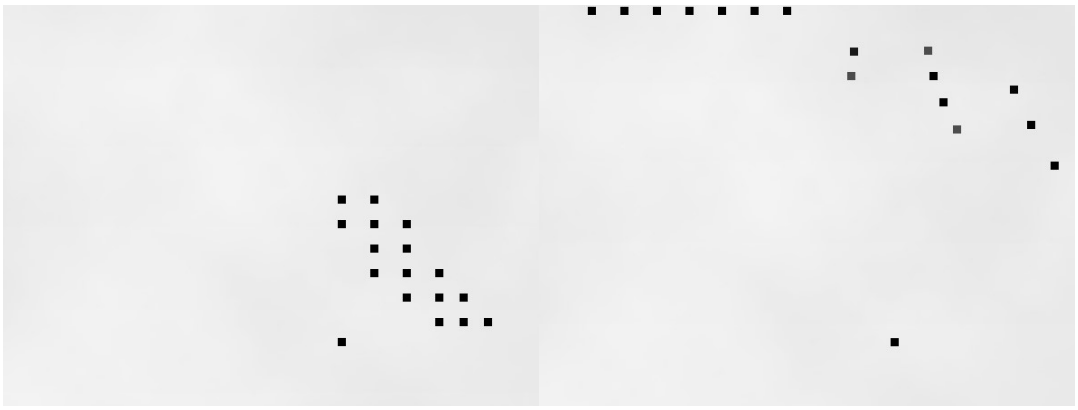

Figure 1: Sequential images of Loneliness. Image on the left shows the player's square approaching a group of other squares. Image on the right shows the other squares dispersing in response to the proximity of the player's square.

How does this game say something about loneliness? Most contemporary metaphor researchers would suggest that this game relies on the metaphor AFFECTION IS PROXIMITY to communicate a message about loneliness. We see this metaphor in phrases like "I'm close to my family," "We've grown distant lately," or "I feel like I just can't get close to people." Conceptual metaphor theory (CMT) holds that we process these phrases by making on-line conceptual mappings from the source domain of PROXIMITY to the target domain of AFFECTION (Gibbs, 2011, 2017; Lai & Curran, 2013). In the game, the other squares are rejecting the player by moving away from the player every time the player tries to get close to them. This seems to be the only way the game communicates; thus, any effect upon the player (i.e. feeling lonely) should come from Proximity-Affection mappings. This allows us to study which players are making these cross-domain mappings and allows us to start identifying what role explicit awareness of the metaphor plays in whether we make cross-domain mappings. CMT would hold that people perform these cross-domain mappings without being aware of the mappings or that they are processing anything metaphorical. Empirical CMT studies tend to support this account of implicit mappings, which will be explored in the Empirical Metaphor Studies section.

Deliberate metaphor theory (DMT), at least in its stronger form (Steen, 2008), would suggest that the player can only draw these cross-domain mappings (and thus experience an emotional effect) if they are explicitly aware of the metaphor[1] (the

[1] Strict DMT would also require that participants attribute the metaphor to a deliberate choice by the author, however awareness should be a pre-requisite for that, and the present studies are better suited to measuring awareness than deliberateness.

strong deliberate hypothesis). A weaker form of DMT, more reconcilable with mainline CMT, would suggest that any player may draw cross-domain mappings, but that awareness of the metaphor will predict stronger emotional impact of the game (the *weak deliberate hypothesis*). Conversely, the bulk of empirical work on CMT would suggest that any player might draw cross-domain mappings, and that conscious awareness does not moderate the emotional impact of the metaphor (the *non-deliberate hypothesis*). Finally, many implicit priming studies only find priming effects in participants who are unaware of the priming manipulation. If this game functions like such a prime, then we might expect emotional impact to be weaker or absent in participants who are aware of the metaphor (the *implicit hypothesis*).

In this dissertation, I will provide a brief review of the literature on conceptual metaphor theory (CMT), multimodal metaphor theory (MMT), and deliberate metaphor theory (DMT), and review the evidence for each hypothesis presented above. I will then present the findings of a first experiment demonstrating the impact of the game and piloting an interpretive questionnaire, as well as a second experiment addressing the above hypotheses.

The experiments investigate the role of interpretation in the impact of a metaphorical game. The first experiment demonstrates the emotional impact of the game. The second experiment tests how the emotional impact of the game interacts with the player's awareness of the metaphor AFFECTION IS PROXIMITY. These experiments are one of the first empirical tests of metaphor in a video game. This work provides a test of non-linguistic metaphor, which both CMT proponents

(Forceville et al., 2012) and critics (Murphy, 1996) have called for, and contributes a new form of evidence to the debate over deliberate metaphor.

CHAPTER II

Conceptual Metaphor Theory

Conceptual metaphor theory (CMT) began with the somewhat radical notion that metaphor was an everyday phenomenon (Lakoff & Johnson, 1980). In this theory, metaphor is part of everyday language, not just a poetic device used by poets or orators to make us look at a subject in a new light (for example, "**look** at a subject in a **new light**" implies the metaphor SEEING IS UNDERSTANDING). Not only is metaphor a part of everyday language, it shapes how we think. The standard CMT position is that processing any metaphor requires on-line conceptual mappings between a source domain and target domain. The target domain is what the metaphor is describing (e.g. understanding), often the more abstract domain, and the source domain is what we are using to describe the target (e.g. seeing).

CMT draws a distinction between conceptual and linguistic metaphors. A conceptual metaphor describes a set of metaphorical mappings in your mind, for example mappings from the domain of PROXIMITY to the domain of AFFECTION (i.e. thinking of caring for someone in terms of physical distance). A linguistic metaphor is a specific instantiation of metaphor in language, often an instance of a conceptual metaphor. For example, the phrase "I've been trying to get closer to my brother" is itself a linguistic metaphor, and an instantiation of the conceptual metaphor AFFECTION IS PROXIMITY.

CMT draws a distinction between conventional and novel metaphors. Conventional metaphors are those which a native speaker of a particular language has

encountered enough for the metaphor to be part of their cognitive architecture. Conventional metaphors are sprinkled throughout language, although people rarely notice them, and they tend to seem cliché if attention is brought to them. Novel metaphors are those which an individual generates anew, as in Shakespeare's "Juliet is the Sun." Novel metaphors are essentially the opposite of conventional metaphors.

Linguistic Evidence

Early work on CMT was largely focused on analyses of language, from introductory work (Lakoff & Johnson, 1980) to later work on the inherent emotional models in conventional expressions (Lakoff & Kövecses, 1987; chapter 2 of Gibbs, 2017 provides an overview). The basic premise of this work is that by systematically analyzing the language people use to talk about a particular subject, researchers can find the metaphors underlying how we think about those subjects.

Phrases like "You're *wasting* my time," "I don't *have* the time to *give* you," and "I've *invested* a lot of time in her" suggest that there is an underlying structure to common expressions (Lakoff & Johnson, 1980, p 8, italics in original). Some would see all these phrases as "dead metaphors," those which are used so often they are no longer processed metaphorically. CMT claims that these phrases, along with many others, emerge from an underlying TIME IS MONEY (or more broadly TIME IS A VALUED COMMODITY) conceptual metaphor which gives us a way to think about time. This shared conceptual metaphor frames time as something that can be wasted, something that can be given but only if the giver has enough of it, and something which can be invested, presumably with an expectation of a reward or return on investment. This

highlights how time is limited, and encourages us to use it efficiently, investing time when we think it will pay off and not wasting time where it would be inefficient or pointless. These linguistic analyses generally assume that these metaphors are structuring thought even if everyday people are unaware of them. This assumption is part of the motivation for the non-deliberate hypothesis.

These analyses of metaphorical language show how the way we talk about abstract concepts like time systematically uses words from other domains, from "investing time" to "moving an appointment forward" to "seeing past events." This line of research is still very productive, but some critics have emerged over the years. The most frequent criticism of these methods is that they rely too much on the introspection and intuition of the researchers (Vervaeke & Kennedy, 1996), are too susceptible to confirmation bias (McGlone, 2011), and that their reasoning is circular (Murphy, 1996). These concerns have merit but are too often used to dismiss all of CMT, as if these linguistic analyses were the only source of evidence for CMT.

Empirical Metaphor Studies

There are other additional forms of evidence for CMT, principally metaphorical framing and priming, metaphorical consistency effects, and multimodal metaphor analysis. Demonstrating that people show metaphor-consistent behavior in a variety of situations, sometimes with no linguistic presentation of the metaphor at all, illustrates that metaphorical thinking goes deeper than language.

In the metaphor consistency effect, people are slower to respond to a sentence based in a new metaphor if previous sentences have all been based in a different

metaphor. For example, after reading several sentences that treat time as a moving object (TIME-MOVING, e.g. "I can't wait for the weekend **to get here**"), you will be slower to respond to a new sentence treating time as a space that you move through (EGO-MOVING, e.g. "I can't wait to **get to the weekend**") (Gentner et al., 2002). Gibbs (2011) gives an overview of several similar reading time studies, all showing faster processing for sentences that follow the metaphors used previously. This suggests that exposure to linguistic metaphor can prime us to think in the relevant conceptual metaphor, at least insofar as we get more efficient at processing subsequent instances of a metaphor and suffer a cost from switching metaphors.

Metaphorical framing or priming can also affect decision-making in the target domain. For example, framing crime as either a beast or a virus affects what policies people endorse to solve a crime problem (Thibodeau & Boroditsky, 2011, 2013). Participants who received the virus frame were more likely to endorse social reform proposals like education reform and funding or economic interventions, which fit the viral metaphor as ways of "inoculating" the community against the virus of crime. Participants who received the beast metaphor frame were more likely to favor enforcement solutions such as larger police forces and longer prison sentences, which fit the beast metaphor as hunting and caging the beast. Participants were not aware of the metaphor manipulation. These studies have received some criticism, including a failure to replicate (Steen, Reijnierse, & Burgers, 2014). However, this failure to replicate can be explained by shifting societal views of terms such as "community policing," and when Thibodeau and Boroditsky (2015) ran new norming studies on

the terms used in the study, they found that these new findings (Steen et al., 2014) actually replicate their original findings. This series of studies is particularly critical in forming the non-deliberate hypothesis, since participants did not attribute their policy decisions to the metaphorical language and yet it influenced their decisions. This could also be construed as support for the implicit hypothesis, as the small number of participants who indicated the metaphorical language as driving their decision were removed from later analyses. Insofar as deliberateness of metaphors can be defined in terms of the receiver's awareness, these studies seem to show some metaphorical thinking without deliberate metaphor.

The TIME IS MOTION metaphor has been especially fruitful for empirical work because of two variations which are distinguished by direction. The first metaphor is called the EGO-MOVING metaphor, and in this case, time is conceptualized as a fixed space which you move through. The alternative is the TIME-MOVING metaphor, in which you are standing still as time moves around you, sometimes visualized as a river or a conveyor belt carrying events towards you (see Boroditsky, 2000; Gentner et al., 2002 for useful illustrations). Each has a different direction of motion, such that "forward" in the EGO-MOVING metaphor is away from you, in the same direction that you are walking, while "forward" in the TIME-MOVING metaphor is towards you, aligned with the movement of events towards you. This is generally tested by asking participants a question in which "forward" is ambiguous (called Ambiguous Time Questions or ATQs) and testing how the resolution of that ambiguity depends on recent experiences of motion. For example, people who have just arrived in an airport

are more likely to give an EGO-MOVING answer than those who have been sitting and waiting for someone's plane to come in (Boroditsky & Ramscar, 2002). The EGO-MOVING priming can come from simulated motion in addition to literal motion, as people who have just looked at a picture of a figure rolling in a chair are more likely to give an EGO-MOVING answer than those who see a picture of a figure pulling a chair towards them (Boroditsky & Ramscar, 2002). Even viewing very simple forms of motion, such as watching boxes move across a screen, can prime the EGO-MOVING metaphor relative to stationary boxes (Núñez et al., 2006). Many other studies have found variations on this effect (Boroditsky, 2000), some using different versions of the ATQ (Feist & Duffy, 2015). Duffy has also explored other factors to explain the variance in this paradigm, in particular the role of calendars and timekeeping devices (Duffy, 2014), personality (Duffy et al., 2014), and emotional state (Hauser et al., 2009). Together, these studies comprise one of the most established paradigms in this still young field, and broadly support the non-deliberate and implicit hypotheses in that the activation of these time metaphors does not seem to be deliberate, or at least is not consciously recognized as metaphorical by the participants.

Multimodal Metaphor

One of the central tenets of CMT is that metaphor is not just a linguistic phenomenon, but part of how we think. The empirical work I have discussed demonstrates that metaphor can be seen in behavior, yet many of those experiments represent at least one domain in a linguistic modality (e.g. the "Wednesday meeting" question after a behavioral prime). One of the most productive new directions in

metaphor research is multimodal metaphor, in which researchers analyze media for its metaphorical content. This work moves CMT beyond its linguistic roots with Lakoff and Johnson, showing how metaphor can arise in the drawings of a cartoon, the shot composition and music of a film, or even the structure of a videogame. Multimodal metaphor analyses generally come out of departments of communication, semiotics, or media studies, so it should come as no surprise that the field's primary method is analysis of media artifacts themselves rather than experimentation.

Multimodal metaphor started with discussions of pictorial metaphor (Forceville, 1994, 2002), and eventually broadened to include other modalities (Forceville, 2006). For Forceville, a metaphor is multimodal if each of its domains is presented in a different modality (2006), although most researchers now use "multimodal" to refer to any non-linguistic metaphor, which is how this dissertation will use the term. Many early analyses of multimodal metaphor focused on commercials, from explorations of how cross-domain mappings are shaped by the promotion of a product (Forceville, 2007) to investigating cultural differences (Lantolf & Bobrova, 2012). This approach has become common enough that some have even proposed procedures for identifying multimodal metaphor in commercials (Bobrova, 2015) and calls to begin empirically testing these metaphors (Forceville et al., 2012). Political cartoons are another productive medium, usually with the source domain presented in the pictorial mode and the target domain presented in the written mode. Bounegru and Forceville (2011) showed how cartoons about the 2008 financial crisis used different source domains to convey different commentary (e.g. a "natural

disaster" source domain implies that the crisis is a force of nature, no one is at fault). Analyses of gesture have also been very fruitful in exploring how abstract ideas are conveyed in physical space. Cienki and Müller (2008) analyzed the gestures of students discussing academic honesty, finding gestures consistent with HONESTY IS STRAIGHT even when the metaphor did not appear in a student's speech.

Multimodal metaphor researchers have devoted a great deal of work to film, music, dance, and other traditional art forms. However, more recently video games have been the subjects of several multimodal metaphor analyses. Kromhout and Forceville (2013) first made the jump to video games, comparing uses of the LIFE IS A JOURNEY metaphor in the games *Half-Life 2*, *Heavy Rain*, and *Grim Fandango*. Kromhout and Forceville specifically discussed the narrative of each game in terms of different instantiations of the LIFE IS A JOURNEY metaphor, concluding that there was a tradeoff between quality of story and quality of gameplay. In essence, giving the player too much gameplay freedom on their journey compromised authorial control over narratives about life. Stamenković and Jaćević's (2015) later multimodal metaphor paper helps demonstrate a less antagonistic view, exploring how the game *Braid* (Blow, 2008) uses complementary metaphorical gameplay and narrative to build the game's themes of obsession and control. *Braid* is structured like a traditional two-dimensional platformer, (e.g. *Super Mario*) with the added feature that pressing a certain button rewinds time, allowing the player to fix their mistakes and solve puzzles. The authors suggest that the game violates our expectations of TIME IS MOTION metaphors both to enhance the expression of its themes and to unsettle the

player. The authors address the gameplay/story tradeoff that Kromhout and Forceville (2013) proposed, suggesting that *Braid* uses gameplay to reinforce its narrative themes (i.e. the ability to rewind and fix your mistakes reinforces the obsessiveness of the protagonist). Therefore, they suggest, *Braid* shows that the gameplay/story tradeoff is not universal, and that at least in some cases these elements of a game can work together to create a cohesive experience. *Braid* accomplishes this in part with its use of metaphor, which may indicate that metaphor can help resolve this gameplay/story tradeoff in other cases.

While neither of these papers is specifically concerned with a player's metaphor awareness or metaphor deliberateness, they may still illustrate two sides of a part of the deliberate metaphor debate. The instances of LIFE IS A JOURNEY that Kromhout and Forceville (2013) discuss seem to be relatively implicit. There is no particular reason that the journeys in these games would strike a player as metaphorical, and thus we can loosely call this an instance of non-deliberate metaphor. Conversely, the subversion of TIME IS MOTION metaphors that Stamenković and Jaćević (2015) point out in *Braid* may fit Steen's definition of deliberate metaphor by calling attention to the source domain, thus making the player more aware of the metaphor, and such cases of metaphor subversion in games may benefit from DMT analysis.

By moving away from linguistic modalities, multimodal metaphor makes a strong case that metaphor exists not only in language, but in the mind. These analyses carry the same risk as the linguistic analyses of CMT: that metaphor experts see

metaphors in some artifact and assume that regular audiences are making those same cross-domain mappings. Of course, the body of empirical work supporting CMT covered above should allay these concerns, and the experiments in this paper may help bridge the gap between multimodal metaphor and the empirical tests of CMT that tend to focus on linguistic metaphor. Most MMT analyses make no particular assumptions about the audience's awareness of metaphors, and in this sense, they may align with the empirical CMT work in supporting the non-deliberate hypothesis. However, some multimodal metaphor could reasonably be called deliberate, as I have suggested above in the case of *Braid*, so there may be some support for the weak or strong deliberate hypotheses amongst these analyses.

CHAPTER III

Deliberate Metaphor Theory

The work discussed above (and most metaphor research since 1980) has been focused on conventional metaphors implicit in everyday language and the role of those metaphors in the mind. A newer branch of metaphor research, deliberate metaphor theory (DMT), shifts focus towards the rarer metaphors that are deliberately chosen with a specific communicative purpose (Steen, 2008, 2017). Although different definitions of deliberate metaphor have emerged over the years, they usually involve drawing the audience's attention towards the source domain as a salient part of the discourse (Steen, 2017). DMT's most controversial claim is that metaphors only involve on-line cross-domain mappings when they are deliberate, in this sense, metaphorical thinking is only provoked by deliberate metaphors; Steen calls this the "paradox of metaphor" (2008, 2017). This claim is motivated by previous work (Bowdle & Gentner, 2005; Glucksberg & Haught, 2006) suggesting that conventional metaphors might not be processed with cross-domain mappings, although Gibbs (2011b) suggests other interpretations of those findings and points out bodies of evidence for CMT that Steen has not accounted for. DMT claims that these metaphors that are not processed with cross-domain mappings must be non-deliberate, and are instead understood by lexical disambiguation (Steen, 2008, 2017). For example, when you read "now I see what you mean" instead of mapping between the seeing and understanding domains, your mental lexicon simply has a definition of "see" that means "understand." This claim is so central to deliberate metaphor that some

definitions become circular, claiming that a deliberate metaphor is any metaphor that involves cross-domain mappings, and cross-domain mappings are only evoked by deliberate metaphors (Ng & Koller, 2013; Steen, 2017). The claim that other metaphor isn't really processed metaphorically (i.e. with cross-domain mappings) has been met with stiff resistance (Gibbs, 2011a, 2015a, 2015b). However, a version of deliberate metaphor theory without this paradox of metaphor has been proposed that would be more compatible with CMT (Gibbs, 2015b), though it has not been more widely adopted.

The vast majority of empirical work on metaphor has shown effects of implicit, presumably non-deliberate metaphor (see Empirical Evidence section above). Steen (2017) has attempted to reconcile DMT with this evidence by calling some of the existing metaphor studies deliberate and publishing some competing empirical studies. For example, he suggests that the political metaphors used as stimuli in a paper (Read et al., 1990) published before the advent of DMT may have actually been deliberate. While this explains this particular paper, it does not necessarily account for the larger body of empirical CMT work. Steen, Reijnierse, and Brugers (2014) do provide a potential new explanation of the CRIME IS A VIRUS/BEAST findings, but as mentioned in Empirical Metaphor Studies, their replication is marred by assuming stable meanings of the political phrases involved. Thibodeau and Boroditsky (2015) have accounted for their results by running another norming study that found the connotations of "neighborhood watch programs" had shifted from 2008-2011 during their initial data collection to 2014. A later test found

limited evidence that these effects increase with more extended versions of the metaphors (Reijnierse et al., 2015). If we accept extendedness as a signal of deliberate metaphor, this finding lends some support to the weak deliberate hypothesis, in that more deliberate metaphors have slightly stronger effects on participants. In a more direct test of DMT, participants did not perceive metaphorical phrases like "We really have come a long way since the wedding" as more deliberate (Gibbs, 2015b) if they were given the linguistic marks that Steen (2008) has suggested indicate deliberateness (e.g. "one might say..."). Although not explicitly framed as using deliberate metaphor, Hendricks and Boroditsky's (2017) use of explicit training in a novel metaphor could be construed as deliberate metaphor, although their use of action compatibility effects is still in the tradition of implicit metaphor.

When exactly a metaphor counts as deliberate is still an unsettled matter, given that it relies heavily on the intentions of a speaker/author in the moment of production. Steen (2010) suggests that textual markers such as simile, breaches of register, or wordplay may indicate deliberateness, and Krennmayr (2011) has suggested an expanded list of textual signals. On this basis DMT scholars have performed corpus analyses (Charteris-Black, 2012; Nacey, 2013) assuming that any metaphorical language that appears alongside these textual markers must be deliberate. These corpus analyses have revealed some surprising relationships. For example, Nacey (2013) found that these deliberate markers were more likely to appear alongside conventional metaphors than novel metaphors, despite the traditional association between novel metaphor and deliberate metaphor. However,

Gibbs (2015b) has called into question this assumption that markers signal deliberateness, given that readers did not increase their ratings of deliberateness in the presence of textual markers. Charteris-Black (2012) has even suggested using the term purposeful metaphor, instead of deliberate metaphor, in part to relax the assumption of the author's intentions. The vast majority of work on DMT has assumed deliberateness based on features of the text, but thus far those textual features have not aligned with empirical tests of participants' sense of deliberateness (Gibbs, 2015b). Since assuming deliberateness on the part of the author based on features of the text has been problematic, this dissertation will focus more on the audience's interpretations and their awareness of the metaphor. If deliberate metaphor draws attention to the metaphor (as Steen, 2017 suggests), then it seems reasonable to use awareness as a proxy for deliberate metaphor. Therefore this dissertation's deliberate hypotheses refer to participants' awareness of the metaphor – if some participants are aware of the metaphor, it stands to reason that the metaphor is at least "more deliberate" for those participants.

The present study also deviates from previous work on DMT in its use of non-linguistic metaphor. Steen has defined DMT exclusively in terms of textual and verbal metaphor. However, he has not explicitly rejected the possibility of metaphors in other modalities being deliberate, and other researchers have just begun to explore this possibility. Ng and Koller (2013) analyzed the visual metaphors in the branding of Singapore Management University. They concluded that the university combined visual and textual instances of the metaphor UNIVERSITY IS BODY, and that this

metaphor use was deliberate. Thus, the present study is not the first to apply DMT to multimodal metaphor, but the approach is still rare.

There may be an additional benefit to this approach to DMT. Consider that our definition of deliberate metaphor relies upon the producer attempting to bring the source domain to their audience's attention. In previous empirical tests of DMT (Gibbs, 2015b), the text is presented to participants as a conversation between two people, Mark (the producer) and Larry (the audience). In this conversation, the participants are not the audience. Perhaps the participants did not respond to the expected cues of deliberateness because their role is that of an overhearer rather than Mark's audience.

Games do not suffer from this problem. The audience for a game is whoever happens to play it. The participants have no reason to believe the maker of a game intended its message for someone else; if you play the game, the creator is communicating directly to you. Thus, using a game as our metaphorical stimulus may sidestep this potential problem with prior empirical tests of DMT.

The game used in this study also has the advantage of communicating non-linguistically, thus the lexical-disambiguation account which has been applied to some CMT studies should not apply in this case. Players would have to see the distance between squares, think of the word "close" and then make the connection to the affection sense of "close" to perform lexical disambiguation – this seems less likely than participants more directly seeing the distance between squares as representing physical distance. Giving participants a verbal distractor task while they

play the game might more conclusively rule out lexical disambiguation, but for the present study this seemed too likely to affect interpretations.

CHAPTER IV

The Present Studies

The following experiments investigate the role of the player's interpretation of a metaphorical game in how the game impacts their emotional state. More specifically, do players have to be consciously aware of the metaphor for the game to have an effect on them? Results of two experiments are reported. Each experiment uses a pre-scale and post-scale of self-reported feelings of loneliness and acceptedness as a primary dependent variable. Experiment 1 demonstrates the emotional impact of the Avoid and Approach versions of the game and tests qualitative and quantitative questions about interpretation. In Experiment 2 participants play either the Avoid version of the game or play *Tetris* as a baseline. Those in the Avoid condition were split into two groups based on their interpretations of the game, treating metaphor awareness as a quasi-experimental variable. If then examine the role of metaphor awareness in emotional impact of the game relative to the baseline condition.

Review of Hypotheses

Strong Deliberate Hypothesis

Participants will only process the game through the metaphor AFFECTION IS PROXIMITY if they are explicitly aware of the metaphor. This hypothesis predicts that the game will only have a detectable effect on the emotions of participants who are aware of the metaphor, taking awareness as a pre-requisite to deliberateness. This hypothesis is based on Steen's paradox of metaphor (2008, 2017), the claim that non-

deliberate metaphors are not processed by cross-domain mappings. In other words, a linguistic metaphor might be present in language but not processed metaphorically. Steen made this claim based on research suggesting that some metaphorical statements are processed by pointing out membership in a superordinate category (Bowdle & Gentner, 2005; Gentner & Bowdle, 2001; Glucksberg & Haught, 2006). For example, "My lawyer is a shark" may be processed as both lawyers and sharks sharing a predatory nature, rather than drawing mappings between lawyers and sharks. Steen has also explained the large body of evidence for traditional CMT by calling some metaphors deliberate, and suggesting that many metaphorical terms (e.g. using "close" to mean "affectionate/intimate") are understood by lexical disambiguation rather than cross-domain mappings. This account is one of the reasons that this study will use games rather than text as its metaphorical stimulus, as the distance between squares in the game cannot be resolved with lexical disambiguation.

Weak Deliberate Hypothesis

Participants will experience stronger metaphoricity (Kövecses, 2015; Müller, 2009) if they are aware of the game's metaphor. Rather than the all-or-nothing nature of the strong deliberate hypothesis, this weaker version simply predicts that the game will have a stronger effect on participants depending on their interpretation, perhaps even gradually scaling up with higher degrees of awareness. This hypothesis is based on the recent shift in DMT away from an absolute paradox of metaphor stance and towards an exploration of what the notion of deliberateness can add to the study of

metaphor. Steen's recent writing (2017) softens or revises some previous claims, shifting emphasis towards deliberate metaphor arising where someone has a communicative purpose, contested metaphors (e.g. what is the best way to characterize cancer), and how deliberateness can arise from humor. Gibbs' recent work has also shifted, focusing more on what deliberate metaphor can add to the larger CMT literature, although the paradox of metaphor is still rejected (2015a). If this hypothesis is supported, then there is some role for player's explicit awareness of metaphor in the impact of games, and thus a role for deliberate metaphor theory in the study of multimodal metaphor.

Non-Deliberate Hypothesis

Participants do not need to be consciously aware of the game's metaphor to process it metaphorically. This hypothesis predicts that the game's effect on emotion will be the same regardless of awareness. This is based on the many empirical metaphor studies which find evidence of metaphorical processing without explicit metaphorical awareness of the participants, particularly Thibodeau and Boroditsky's work on crime metaphors (2011, 2013). In nine experiments reported, less than 5 percent of participants showed an awareness of the metaphorical prime, yet the effect of covert metaphors seems to be reliable. The preponderance of empirical metaphor studies show effects of non-deliberate metaphor, but given Steen's (2008, 2011, 2017) lexical disambiguation account, it is worth testing non-linguistic metaphor which lexical disambiguation cannot account for. If participants experience an

emotional impact of the game's metaphor without conscious awareness, this will be an instance of metaphorical processing that DMT cannot account for.

Implicit Hypothesis

Participants will experience stronger metaphoricity if they do not perceive the game's metaphor deliberately. Many priming studies (metaphorical priming and other forms of priming) assume that participants who are aware of the priming manipulation may experience a reduced or absent priming effect. Though this isn't often tested directly, it is implied by the removal of any participants aware of the manipulation in studies such as the crime metaphor experiments mentioned previously (Thibodeau & Boroditsky, 2011, 2013). However, this hypothesis relies on an assumption that *Loneliness* acts like a metaphorical prime, which may not be the case. Recent meta-analyses also suggest that, at least in the case of primes that align with a participant's goals, awareness can even increase priming effects (Bargh, 2016).

The Game *Loneliness*

The games used in this study are variations on *Loneliness*, the game described at the beginning of this paper (Magnuson, 2007). The original game uses many cues to signal isolation, chiefly the other squares moving away when you approach them, but also somber minor-key music and a background that slowly grows darker. The original game also displays the title "Loneliness" at the beginning of the game and basic instructions, and the text about teens in South Korea at the end of the game. For use in this study, I obtained two variations of the game from Jordan Magnuson, and made a few minor adjustments. The variations used for this study have removed the

music and background color transition so that the behavior of squares in the game is the only emotional cue. The title of the game has been removed, although basic instructions to use arrow keys to control have been retained. The text about South Korean teens has also been removed and replaced with an instruction to click a button at the bottom of the screen leading to a questionnaire.

Without the title, music, and gradient, the game has only the motion of the squares as a means of communication. In order to be emotionally impacted by the game, the player must make the connection between the other squares moving away from or towards their square – otherwise they have no reason to see the events of the game as indicating rejection or acceptance, and no reason to feel differently after playing. This is the basis for using emotion change as a measure of cross-domain mappings; if the game impacts a participant's emotions, that participant must have made some mapping from physical distance to emotional distance.

The only difference between the two variations of the game is in how the other squares react when the player's square approaches them. In the Avoid condition, the other squares behave as described in the original game, moving away as soon as the player's square approaches and slowly fading from view after a few seconds. This is reversed in the Approach condition, in which the other squares move towards the player's square when it gets close enough. Most disappear after a few seconds, but a small number (4-5 over the course of the game) will continue to follow the player's square, as if it has made a few enduring friendships. The Avoid condition is expected to increase ratings of loneliness and decrease ratings of acceptedness,

while the Approach condition is expected to decrease ratings of loneliness and increase ratings of acceptedness. How this interacts with the player's interpretations of the game and awareness of the metaphor is the focus of Experiment 2.

Related Stimuli

The reader may notice this game bears some resemblance to the video stimuli famously used by Heider and Simmel (1944) to demonstrate the ability to read a narrative into the movement of shapes on a screen. In this video, a large triangle appears to bully a smaller triangle while a small circle hides. This game is similar in the use of simple shapes to depict social interaction. However, there are several ways that *Loneliness* deviates from their stimuli, allowing the present studies to build on their work. Loneliness is interactive as opposed to a video for participants to passively watch. This allows the player to explore a system and see how it responds to them, rather than watching a predetermined narrative. Heider and Simmel's stimuli also have different shapes of different sizes which occupy distinct roles in the narrative, whereas Loneliness uses only squares of the same size and the only distinct role is the square controlled by the player.

Williams' (1997) Cyberball game, used in studies on ostracism, also bears some resemblance to *Loneliness* – not in visual properties like the Heider and Simmel video, but in the way the game is used to evoke a feeling of isolation in participants. In the Cyberball game, characters on a screen pass a ball back and forth, and how often the ball is passed to a certain character (usually the character associated with the participant) is manipulated by experimenters. Cyberball does not require metaphorical

mappings to understand, it is a direct representation of a playground game and can evoke the feeling of being excluded by other players. In contrast, *Loneliness* requires both that the player see the squares as people and that they map the movement of the squares to social distance. Whether the player must project themselves into the square they control is less certain, and the Cyberball literature offers one possibility: Wesselmann, Bagg, and Williams (2009) found that participants' negative affect when they had been told to take the perspective of an ostracized player was similar to when they simply observed a player being ostracized. If this effect extends to the present studies, then participants taking the perspective of the square they control may not be necessary to being emotionally impacted.

CHAPTER V

Experiment 1: Emotional Impact of Avoid and Approach Variants of *Loneliness*

Experiment 1 was performed to test the emotional impact of the Avoid and Approach versions and to test various ways of eliciting interpretations of the game and measuring metaphor awareness. The Avoid version was predicted to increase feelings of loneliness and decrease acceptedness, while the Approach version was expected to have the opposite effect. The Avoid version had the predicted effect, both increasing loneliness and decreasing acceptedness. The Approach version had a less clear effect, with a small decrease in loneliness but no change in acceptedness, although it had the expected effect relative to the Avoid version. Based on these findings, Experiments 2 used the Avoid version as an experimental manipulation compared to a baseline. Participants played the Approach version at the end of Experiment 2 to mitigate the emotional impact of the Avoid version.

The interpretive questionnaire was initially all qualitative questions. However, after several iterations there were no questions that could reliably show metaphor awareness interpretation without the question being overly leading. Because of this, a set of agreement ratings of statements adapted from Gibbs (2015b) was added to the questionnaire. These ratings are key to the measurement of awareness in Experiment 2, but since they were introduced late in Experiment 1, they are not part of the primary analysis in this experiment.

Participants

Participants were undergraduate students from the University of California, Santa Cruz. All participants were native English speakers. Participants were given one hour of course credit for participating. 81 participants were run in this experiment; 3 participants were removed from analysis due to incomplete responses, 12 were removed because they had additional interactions with the research assistant or were already acquainted with the research assistant which might compromise the sense of isolation during the experiment, and 6 participants were excluded because they seemed agitated or ill when they arrived or seemed to rush through the experiment[2]. Of the 60 participants included in analysis, 36 identified as female, 22 identified as male, and 1 identified as genderqueer and 1 declined to report gender. Participants' average age was 20.1 years old (SD = 1.71 years). All participants were native English speakers. Participants were randomly assigned to the Avoid and Approach conditions, after excluding the participants as described above the Avoid condition was left with 28 participants and the Approach condition with 32 participants.

Design

This experiment used a two-level between subjects manipulated design (game version: Approach vs Avoid); emotional change was measured as the dependent variable. Participants were randomly assigned to play either the Approach or Avoid

[2] Analyses were also performed with the full 78 participants (only excluding incomplete responses), and a similar pattern of results was observed – see appendix A

condition. Emotional change was measured by the difference between pre-scale and post-scales of self-reports of loneliness and acceptedness.

Materials

Emotion scale. An emotion scale was administered both before and after playing the game. The scale asks for the participants' current level of ten emotions. "How lonely do you feel right now?" and "How accepted do you feel right now?" are the questions of interest, the other eight are treated as distractors. Under each question is a horizontal line measuring 140 mm, with a label at each end (e.g. "Not lonely at all" on the left and "Extremely lonely" on the right). Participants are asked to indicate how they currently feel by marking a vertical mark through the horizontal line. After the participant left, the experimenter measured the distance in millimeters between the left end of the scale and the mark (i.e. 0 mm of loneliness is a mark at the "Not lonely at all" end of the scale). The continuous scale was used instead of a Likert-type scale to avoid the participants anchoring to their pre-scale responses, given the short time span between the pre and post-scales (i.e. the game takes only 3-5 minutes to play). "Lonely" and "accepted" were used instead of "ostracized" because they seemed less conspicuous in a list of emotions, and thus less likely to alert participants to the nature of the study before playing the game. Two randomized orders of questions were generated for counterbalancing. Each participant saw one order for the pre-scale and one for the post-scale; this was counterbalanced across participants. These scales were administered on paper.

Game versions. General features of the game are described earlier in the *Loneliness* section. In this experiment, two versions of the game are used, the Avoid version and the Approach version. In the Avoid version, other squares move away from the player's square when they get close, much like the original game. In the Approach version, other squares come closer when the player's square gets close to them, and some squares continue to follow the player's square throughout the game. Neither game version uses the musical or darkening gradient cues from the original game, and the only text displayed to participants is instructional.

Both game versions have text above them explaining the controls of the game and a note that "This game is about people." This note was added because of 16 pilot participants: 5 had no idea what the game was about, 10 thought it was about something other than people (e.g. "pattern recognition") and only 1 thought that the game was about people. Leaving the game fully open to interpretation would be preferable, but even participants with an implicit metaphorical interpretation should be able to identify the topic of the metaphor when prompted. Since 15 of 16 participants' interpretations were irrelevant to even the non-deliberate hypothesis, "This game is about people" was added. Even with this identification of topic, it should take some interpretative work from the participants to discern the details of the metaphor (e.g. squares moving away is rejection).

Interpretive questionnaire. Participants were asked a series of questions about their interpretation of the game and what they thought the creator was trying to communicate. These questions were broken into five sections in the Google Form so

that participants could not change their answers to early open-ended questions based on more leading later questions. These questions will be described here in the order they were given to participants, although the participants only saw numbered section headings (e.g. Part 3) rather than the headings used here. Some of the questions below were added during data collection, in those cases the number of participants who saw those questions are noted. Because this part of Experiment 1 changed during data collection, it will be treated as a pilot of interpretation questions and not used in analysis of emotional change.

Initial impressions. In this section, participants were asked: "What do you think the game was about? What made you think that?" and to "Describe your experiences playing the game." While these questions were too broad to reveal metaphor awareness or deliberate metaphor interpretations, some participants did demonstrate a level of confusion that likely precludes awareness or deliberateness. Since later questions may lead participants towards interpretations they did not think of while playing the game, these initial questions can help find participants whose initial impressions differ substantially from their later responses.

Interpretive statements. This section was added for the final 9 participants in Experiment 1 to add a quantitative measure of perceived deliberateness and metaphor awareness. In this section, participants were asked to rate their level of agreement with nine statements about the game. These statements are adapted from the seven statements used in Gibbs (2015a) to assess participants sense of whether a textual metaphor was being used deliberately. In the original Gibbs paper, participants rated

each statement on a 7-point scale from Strongly Disagree to Strongly Agree, and these ratings were the dependent variable of primary interest. In this study, participants rated each statement on a 10-point scale from Strongly Disagree to Strongly Agree, and these ratings will be used in Experiment 2 as a quasi-experimental variable grouping variable. The scale was changed from 7-points to 10-points to ensure that each response clearly agrees (6-10) or disagrees (1-5) with the statement. While some of these statements were used in Gibbs (2015a) to measure participants' sense of deliberateness, in this dissertation they will be treated as measuring metaphor awareness. This is based on a factor analysis performed on Experiment 2 data that suggests that the statements are measuring a single construct, and on feedback from the dissertation committee that the final versions of the statements used in Experiment 2 seemed to address metaphor awareness rather than deliberateness. Some statements have a different version to match what happens in each game condition. These are reported in the format: (Approach version/Avoid version).

1) *The game implied that the player's square is being (accepted/rejected) by the other squares*. This statement should capture whether participants understood the most basic implication of the game. While this acceptance/rejection is entailed by the AFFECTION IS PROXIMITY metaphor, this question does not use language specific to that metaphor.

2) *The game implied that the player's square finds it (easy/hard) to get close to people*. The phrasing of this statement fits the AFFECTION IS PROXIMITY metaphor.

High agreement ratings on this statement will indicate a participants' awareness of the game's metaphor.

3) *The game implied that the player's square finds it (easy/hard) to make connections with people*. The phrasing of this statement fits the AFFECTION IS CONNECTEDNESS metaphor. This metaphor has the same target domain, but a different source domain. This was included as a distractor, so that not all metaphorical statements reflected the canonical metaphor.

4) *The game implied that the other squares are being (warm/cold) towards the player's square*. The phrasing of this statement fits the AFFECTION IS WARMTH metaphor, but it otherwise serves the same purpose of Statement 3.

5) *The game was just about squares moving around*. High levels of agreement with this statement should be mutually exclusive with metaphor awareness. This statement was reverse-coded when included in the Metaphor Awareness scale in Experiment 2.

6) *The creator of the game was exactly sure what they wanted to say about people*. Statements 6-9 will be initially intended to assess perceived deliberateness. However, in Experiment 2 these were re-written to remove reference to the creator based on work on source orientation (Sundar & Nass, 2000). Based on this change and a factor analysis mentioned previously and explained further in the results of Experiment 2, these were included as measures of metaphor awareness.

7) *The creator consciously wanted you (the player) to think hard about the meaning of the game*.

8) *The game was intended to compare the space between squares to social closeness*.

9) *The game was creative or poetic*.

Interpreting game elements. This section contained four questions: "Did you see the squares as representing anything? If so, what?", "Why do you think the other squares moved?", "Did you see the distance between the squares as representing anything? If so, what?", and "Did you see yourself as one of the squares?" These questions assess what participants believe is going on in the game, and how they see their role. Whether they see themselves as one of the squares may be of interest, as this may be necessary for the game to influence participants' emotions. These questions may be biased by the interpretive statements above but given the possibility of biases in either direction the ratings of statements were prioritized.

Creator's intent. This section contained two questions: "What do you think was the creator's intent in making this game?" and "What was it about the game that made you think that was the creator's intent?" Before the addition of the Interpretive Statement section, these questions were the primary measure of perceived deliberateness. Because many participants gave short responses such as "The purpose was to make me feel sad", these questions were insufficient to elicit deliberate metaphor interpretations that may have existed. These questions were still included after the addition of the Interpretive Statement section in order to assess whether participants are interpreting the game in ways that the statements above do not capture.

Reflecting on change. This section contained two questions: “How is your current interpretation of the game different from when you first played it?” and “Has your emotional response to the game changed? If so, how?” With these questions we will be able to assess the impact that the questionnaire has on participants and determine whether participants' responses on the questionnaire may differ from their initial reaction to the game.

General game experience. After the questions about the game they just played, participants answered two questions on a nine-point scale: “How experienced are you at playing video games?” with 1 being “not experienced at all” and 9 being “very experienced”. “How frequently do you play video games?” with 1 being “never” and 9 being “very frequently”.

Demographics. Participants then filled out a demographic questionnaire, covering age, gender identity, major and year in school, ethnicity and whether they are a native English speaker.

Procedure

Participants were first given a consent form. If they decided to sign it, they were taken into a small experimental room and given a copy of the 10-item emotion scale (on paper). When they completed this scale, the experimenter set up the game, either the Avoid or Approach version depending on condition (instructions for the game are presented on screen). A copy of the 10-item emotion scale, items in a different order than they saw previously, was laid face down next to the keyboard. At the end of the game, text appeared instructing participants to click the “next” button

below the game window which took them to a Google Form. The Google Form directed participants to turn over the emotion scale and fill it out, then click to the next section of the form. The form then proceeded as described in Materials above: participants gave their initial impressions of the game, they rated their level of agreement with statements about the game, interpreted some basic elements of the game, described the intent of the game's creator, reflected on whether their interpretation of the game has changed, and finally answer some basic questions about demographics and experience with video games. When the participant finished with this Google Form, their part in this study is complete, although they were asked to participate in a second, unrelated study, time permitting.

After the participant left, the experimenter processed the emotion scales. The experimenter measured the distance from the participant's response to the left end of the scale in millimeters. This was recorded for the feelings of "lonely" and "accepted" on both the pre-scale and post-scale.

Results

The Avoid version of the game was predicted to increase feelings of loneliness and decrease feelings of acceptedness. The Approach version of the game was predicted to decrease feelings of loneliness and increase acceptedness. All emotional ratings were measured as the distance in millimeters (mm) between the leftmost point on a line and the mark a participant made on that line to indicate their current emotional state. The line was 140 mm long, so for the question "How lonely do you feel right now?" a 0 mm response would indicate being not lonely at all, a 140

mm response would indicate being extremely lonely, and a 70 mm response would indicate being moderately lonely, as this is the middle of the scale. Descriptive statistics are presented in Table 1 below.

Group	Emotion	Pre-Scale	Post-Scale	Mean Difference	Cohen's *d*	d_z
Avoid	Lonely	35(27)	47(32)	12(24)	.40	.50
Approach	Lonely	40(31)	31(24)	-9(21)	.31	.41
Avoid	Accepted	88(30)	73(29)	-15(33)	.50	.46
Approach	Accepted	83(32)	82(29)	-1(17)	.02	.05

Table 1: Mean (standard deviation) of Loneliness and Acceptedness ratings for Avoid and Approach conditions. All means and standard deviations are measured in millimeters. Cohen's d is calculated with pooled standard deviations while d_z is calculated with the standard deviations of difference scores. Cohen's d is provided as the more standard measure for comparison to other studies, d_z is provided as a more sensitive measure for within-subjects comparisons.

On the pre-scale of emotional state, participants in the Avoid condition reported slightly lower levels of loneliness than participants in the Approach condition. Since the pre-scale is administered before the manipulation is introduced, this was concerning; however the difference is not significant according to a Welch independent samples t-test ($t(58.0 = -.635$, $p = .528$) and would be small relative to standard deviation ($d = .16$). After playing their respective games, the Avoid condition showed a moderate increase in loneliness ($t(27) = -2.655$, $p = .013$) and the Approach condition showed a decrease of slightly smaller magnitude in loneliness ($t(31) = 2.302$, $p = .028$), both significant within subjects differences according to paired samples t-tests. A one-way ANOVA comparing the change in loneliness across conditions found a significant effect of condition on loneliness change such that the increase in the Avoid condition was greater than the decrease in the Approach

condition ($F(1,58) = 12.511$, $p = .001$, $\eta^2 = .18$). These changes in loneliness match the predictions: playing Avoid increased loneliness, and playing Approach decreased loneliness, with a significant difference in emotional impact between the two games.

As with the loneliness measure, a small ($d = .16$) non-significant difference between conditions appeared on the pre-scale, with the Avoid condition reporting slightly higher levels of acceptedness than the Approach condition ($t(57.8 = .625$, $p = .535$). After playing their respective games, the Avoid condition showed a moderate, significant decrease in acceptedness ($t(27 = 2.445$, $p = .021$) while the Approach condition showed a negligible, non-significant decrease in acceptedness ($t(31) = .225$, $p = .801$). The decrease in acceptedness in the Avoid condition was significantly greater than the decrease in the Approach ($F(1,58) = 4.809$, $p = .032$, $\eta^2 = .08$). The decrease in acceptedness for the Avoid condition fits the prediction, as does the difference between conditions. However, the lack of change within the Approach condition was counter to predictions.

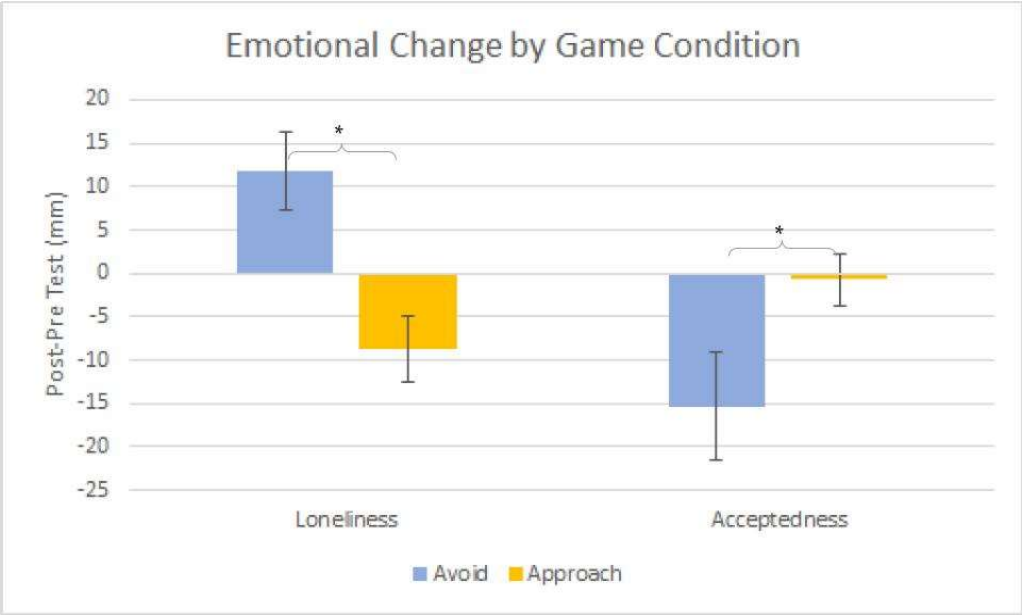

Figure 2: Difference between post and pre-scale of Loneliness and Acceptedness. Color of bars corresponds to condition. Bars above the x-axis are an increase in emotion from the pre-scale to the post-scale, bars below the x-axis are a decrease. Error bars are 1 standard error. Asterisks indicate significance at $p < .05$ level.

As discussed in the Interpretive Questionnaire section above, open-ended qualitative questions could not reliably elicit awareness of a specific metaphor or deliberate metaphor interpretations. Many responses were too short to reveal deliberate metaphor interpretations that participants may have had, despite the inclusion of wording such as "What was the creator's intent in making this game? What made you think that?" In addition, because deliberate metaphor definitions rely on a particular combination of what the creator did and why they did it, I was unable to devise a question that could elicit such an interpretation without being too leading. Holistic qualitative coding for any amount of metaphorical interpretation is ongoing and will be reported in future work.

For the above reasons, levels of agreement with statements about the game were introduced as part of the interpretive questionnaire for the last ten participants

(one of whom was removed from the analysis due to incomplete responses). Further explanations of these statements can be found in the Interpretive Statements section within Materials above. Below are preliminary means and standard deviations of each question. Each question was answered on a scale from 1 (strongly disagree) to 10 (strongly agree), so responses above and below 5.5 correspond to agreement and disagreement, respectively.

While the samples are too small to run inferential statistics (particularly for Statements 2 and 3, which were added to the questionnaire for the last 5 participants), some basic patterns do seem to emerge. The Avoid condition seems to have reasonably high agreement with the early metaphor awareness statements, low agreement with the non-metaphorical "just squares" statement and a mix of agreement on the deliberateness statements. In keeping with the previous finding that the Approach version is hard to interpret, agreement is low for basic metaphorical statements about the game (e.g. mean of 5.6 for "The game implied that the player's square is being rejected/accepted by the other squares.").

		Approach			Avoid		
#	Statement	Mean	SD	n	Mean	SD	n
1	The game implied that the player's square is **being accepted/rejected** by the other squares.	5.60	3.36	5	9.00	2.00	4
2	The game implied that the player's square finds it easy/difficult to **get close** to people.	5.00	4.24	2	6.67	3.06	3
3	The game implied that the player's square finds it easy/difficult to **make connections** with people.	4.00	2.83	2	6.67	3.51	3
4	The game implied that the other squares are **being warm/cold** towards the player's square.	5.80	3.11	5	4.00	4.08	4
5	The game was just about squares moving around.	4.80	3.35	5	3.00	2.83	4
6	The creator of the game was exactly sure what they wanted to say about people.	4.80	2.28	5	6.25	2.63	4
7	The creator consciously wanted you, the player, to think hard about the meaning of the game.	7.80	1.92	5	5.25	2.63	4
8	The game was intended to compare the space between squares to social closeness.	5.80	3.11	5	8.25	2.36	4
9	The game was creative or poetic.	4.20	3.56	5	3.50	2.65	4

Table 2: Mean and standard deviations of agreement ratings. Number of responses is also included as n. Bolding was added for clarity but was not present in the original questionnaire. Statements 1-4 have 2 versions to match what happens in each game condition, these are reported in the format: (Approach version/Avoid version).

Discussion

The Avoid version seems to have the predicted impact, both increasing loneliness and decreasing acceptedness. The Approach version has a less clear

impact, decreasing loneliness but with no reliable effect on acceptedness. The Approach version was harder to interpret – e.g. being chased was a common interpretation instead of being accepted. This greater diversity of interpretations may be why the effect was less clear in the Approach condition. Negativity bias might also explain some of the lack of effect in the Approach group – perhaps being rejected by the other squares was more salient than being accepted by them.

Experiment 2 used the Avoid version as the experimental condition, compared to Tetris as a baseline game. Participants who are assigned to the Avoid condition will play the Approach version after completing the rest of the experiment to wash out some of the negative emotions from the Avoid version. While the Approach version was hard to interpret in isolation, after playing the Avoid version it should be clearer in the context of having played the Avoid version. Due to time constraints, the participants were not asked to interpret the Approach version in Experiment 2, so this increased clarity is conjecture.

CHAPTER VI

Experiment 2: Role of Metaphor Awareness in Emotional Impact of *Loneliness*

Experiment 2 will explore the role of player's interpretations of a metaphorical game in how it affects their emotional state. Does the player have to be aware of the AFFECTION IS PROXIMITY metaphor to be affected by the Avoid version of the game? The strong deliberate hypothesis would predict that the game will only affect the emotions of players who demonstrate an explicit awareness of this metaphor. The weak deliberate hypothesis predicts that explicit awareness of the metaphor will strengthen the emotional impact, which may still be present in other participants. The non-deliberate hypothesis predicts that explicit awareness will not change the emotional impact of the game. The implicit hypothesis predicts that explicit awareness will weaken the impact.

Participants

Participants were undergraduates from the University of California, Santa Cruz. All participants were native English speakers. Participants were given course credit for participating. Participants were randomly assigned to play the Avoid version of the game or a baseline, with ⅔ assigned to Avoid and ⅓ assigned to Baseline to account for the Avoid version being further subdivided based on interpretation.

A total of 167 students participated in this experiment. However, 4 participants were removed from analysis due to incomplete responses, 13 participants were excluded from analyses due to additional interactions with the research assistant

which may have compromised the sense of isolation during the experiment. An additional 7 participants were excluded due to seeming ill or agitated when they arrived or seeming to rush through the experiment. After these exclusions, 146 participants were included in the final analyses: 49 in the baseline condition and 97 in the avoid condition[3]. Random assignment to conditions was conducted at a 2:1 ratio since the Avoid condition would be further split based on metaphor awareness. The average age of the 146 analyzed participants was 19.6 years (SD=1.96), and 100 were female, 45 were male, and 1 was non-binary. All participants were native English speakers.

Design

This experiment used a three group between-subjects design with emotional change as the dependent variable. The primary comparison will be Avoid-Aware vs Avoid-Unaware vs Baseline, where Baseline vs Avoid is randomly assigned and Aware vs Unaware is split within the Avoid condition based on the Metaphor Awareness scale. Emotional change was measured by the difference between pre-scale and post-scale of self-reported loneliness and acceptedness.

Materials

Emotion scale. Experiment 2 used the same emotion scale as Experiment 1.

Games. Participants in the Avoid condition played the Avoid version of *Loneliness* game, as described in Experiment 1. Participants in the Baseline condition

[3] Analyses were also performed with the full 167 participants (only excluding incomplete responses), and a similar pattern of results was observed – see appendix A.

played a version of *Tetris* coded in Javascript called blockrain.js (Aerolab, n.d.). This will be referred to as the Tetris condition from here on.

Metaphor Awareness scale. Metaphor awareness was measured for each participant as the average of statements 1, 2, 6-9 and a reverse code of statement 5. Participants were analyzed in the Aware group if they were equal to or above the median on this scale, and as Unaware if they were below the median. Secondary analyses were performed using the scale as a continuous variable and grouping based on a midpoint (3.5) split of the scale. Further detail can be found in the subsection Grouping Based on Awareness within Results below.

Conditions. Participants were randomly assigned to play the Avoid version of the game or play Tetris for 4 minutes (the average duration of the Avoid version of the game).

Interpretive questionnaire. After the post-scale of emotion, participants were asked a series of questions about their interpretation of the game and what they thought the creator was trying to communicate. This questionnaire was only administered to participants in the Avoid condition, participants in the Tetris conditions moved directly from the post-scale of emotion to General Game Experience and Demographics. These questions were broken into seven sections in the Google Form so that participants could not change their answers to early open-ended questions based on more leading later questions. These sections will be described here in the order they were given to participants, although the participants

only saw numbered section headings (e.g. Part 3) rather than the headings used here. This questionnaire was piloted in Experiment 1.

Initial impressions. In this section participants were asked: "What do you think the game was about? What made you think that?" and to "Describe your experiences playing the game." This will establish a general interpretation of the game without leading participants towards a particular interpretation. This can be used to confirm that participants' initial impressions do not differ substantially from their responses to later, more directed questions.

Interpretive statements. In this section, participants were asked to rate their level of agreement with nine statements about the game. The development and piloting of these statements is described in the Materials and Results sections of Experiment 1. What follows are changes introduced in Experiment 2. Since the Approach version is no longer a separate condition, the Approach versions of the statements were removed. Statements about what the creator of the game is trying to communicate were changed to what the game is communicating based on previous research on source orientation (Sundar & Nass, 2000). Agreement ratings were shifted from a 10 point scale to a 6 point scale (1=Strongly Disagree, 6=Strongly Agree), such that each level is more meaningful and to make it more obvious to participants that the scale has no midpoint. The statements were also reorganized: Statement 5 ("just squares") was put in its own section before the other statements so that participants could rate it without the suggestions of the other statements. Statements 5-9 were moved ahead of 1-4 so that the more general statements were

made before ratings of specific metaphorical phrases. Below are the statements with new phrasing and order, but with numbers held constant from Experiment 1 to allow easier comparison across experiments.

5) *The game was just about squares moving around*.

6) *The game has a specific message about people*.

7) *The game is intended to make you think about its meaning*.

9) *The game was creative or poetic*.

8) *The game was intended to compare the space between squares to social closeness*.

3) *The game implied that the player's square finds it difficult to make connections with people*.

1) *The game implied that the player's square is being rejected by the other squares*.

2) *The game implied that the player's square finds it difficult to get close to people*.

4) *The game implied that the other squares are being cold towards the player's square*.

Interpreting game elements. This section is unchanged from Experiment 1.

Creator's intent. This section is unchanged from Experiment 1.

Metaphorical communication. This is a new section containing two questions: "Did you think there was something metaphorical about the game? If so, what?" and "Do you think the creator deliberately used that metaphor to communicate

something to you? What were they trying to communicate?" These questions were added to more directly prompt participants to explain the potential metaphor and deliberateness of the game. Even if participants do not discuss metaphor in earlier sections, the ability to describe the metaphorical content of the game when prompted still reveals some metaphorical understanding of the game.

Communication with movement of squares. This is a new section containing two questions: "What do you think the creator was trying to communicate with the movement of squares? If you don't think the creator was trying to communicate, please respond "NA"" and "Did the squares moving in response to you seem like a deliberate choice by the creator?" (rated on a scale from 1-Not at all deliberate to 7-Definitely a deliberate choice). These questions were added, like those in Metaphorical Communication above, to prompt participants to explain the metaphorical content in the game. These questions are even more directed, specifically focusing attention on the way the squares move as a means of communication.

Reflecting on change. This section is unchanged from Experiment 1.

General game experience. This section is unchanged from Experiment 1.

Demographics. This section is unchanged from Experiment 1.

Procedure

Participants were first given a consent form. If they decided to sign it, they were taken into a small experimental room and given a copy of the 10-item emotion scale (on paper). Once they completed this scale, the experimenter set up the game,

either the Avoid version or Tetris depending on condition (instructions for the game were presented on screen). A copy of the 10-item emotion scale, items in a different order than they saw previously, was laid face down next to the keyboard. At the end of the game, text appeared instructing participants to click the "next" button below the game window which took them to a Google Form. The Google Form directed participants to turn over the emotion scale and fill it out, then click to the next section of the form. In the Avoid condition, the form then proceeds as described in Materials above: participants gave their initial impressions of the game, they rated their level of agreement with statements about the game, interpreted some basic elements of the game, describe the intent of the game's creator, reflect on whether their interpretation of the game had changed, and finally answered some basic questions about demographics and experience with video games. In the Tetris condition, the form only included the questions about demographics and experience with video games. Participants in the Avoid condition will then play the Approach version of the game to wash away lingering emotional influence. At this point, the study is complete, although they some were asked to participate in a second, unrelated study, time permitting.

After the participant left, the experimenter processed the emotion scales. The experimenter measured the distance from the participant's response to the left end of the scale in millimeters. This was recorded for the feelings of "lonely" and "accepted" on both the pre-scale and post-scale.

Results

The Avoid version of the game was predicted to increase feelings of loneliness and decrease feelings of acceptedness, as in Experiment 1. The control condition, Tetris, was not expected to impact emotions. Within the Avoid version, each hypothesis made a different prediction for how metaphor awareness would interact with emotional impact: 1) strong deliberate hypothesis: metaphor awareness is necessary for emotional impact – without awareness, impact will be similar to the baseline; 2) weak deliberate hypothesis: metaphor awareness will increase emotional impact; 3) non-deliberate hypothesis: metaphor awareness will not affect emotional impact; 4) implicit hypothesis: metaphor awareness will decrease emotional impact. Emotions were self-reported as marks on a 140 mm long line, further explanation available in the Results section of Experiment 1.

Emotional impact of Avoid vs Tetris (not separated by awareness). On the pre-scale of emotional state, participants in the Avoid condition reported low levels of loneliness similar to those in the Tetris condition. On the post-scale, participants in the Avoid condition reported higher levels of loneliness, while participants in the Tetris condition reported slightly lower levels of loneliness. Paired sample t-tests showed that the increase in loneliness in the Avoid condition was significant ($t(96) = -3.29$, $p = .001$), while the decrease in loneliness in the Tetris condition was not significant ($t(49) = 1.768$, $p = .083$). A one-way ANOVA comparing the change in loneliness between the Avoid and Tetris conditions revealed a significant interaction

such that the increase in loneliness in the Avoid condition was greater than the decrease in the Tetris condition, $F(1,144) = 10.379$, $p = .002$, $\eta^2 = .07$.

Group	Emotion	Pre-Scale	Post-Scale	Mean Difference	Cohen's d	d_z
Tetris	Lonely	43(32)	38(31)	-4(18)	.14	.25
Avoid	Lonely	42(31)	50(32)	8(24)	.25	.33
Tetris	Accepted	89(27)	78(28)	-11(25)	.39	.44
Avoid	Accepted	85(29)	68(27)	-18(28)	.63	.63

Table 3: Mean (standard deviation) of Loneliness and Acceptedness ratings for Avoid and Tetris groups. All means and standard deviations are measured in millimeters. Cohen's d is calculated with pooled standard deviations while d_z is calculated with the standard deviations of difference scores. Cohen's d is provided as the more standard measure for comparison to other studies, d_z is provided as a more sensitive measure for within-subjects comparisons.

Levels of acceptedness were also similar between groups on the pre-scale. On the post-scale, feelings of acceptedness decreased for both conditions; both decreases were significant (Avoid: $t(96) = 6.227$, $p < .001$, Tetris: $t(48) = 3.069$, $p = .0035$). However, perhaps because the effect went in the same direction in both conditions, a one-way ANOVA comparing the decreases in acceptedness across conditions did not show a significant interaction ($F(1,144) = 2.180$, $p = .142$ $\eta^2 = .01$).

These findings fit the prediction for the Avoid condition and replicates the overall pattern of results in Experiment 1. The Tetris condition was not as emotionally neutral as expected, with its significant decrease in feelings of acceptedness. This may be simply an effect of being alone in the testing room while playing the game - which demonstrates the value of having this baseline condition for comparison rather than assuming that playing the Avoid game is the only factor impacting participants' emotions.

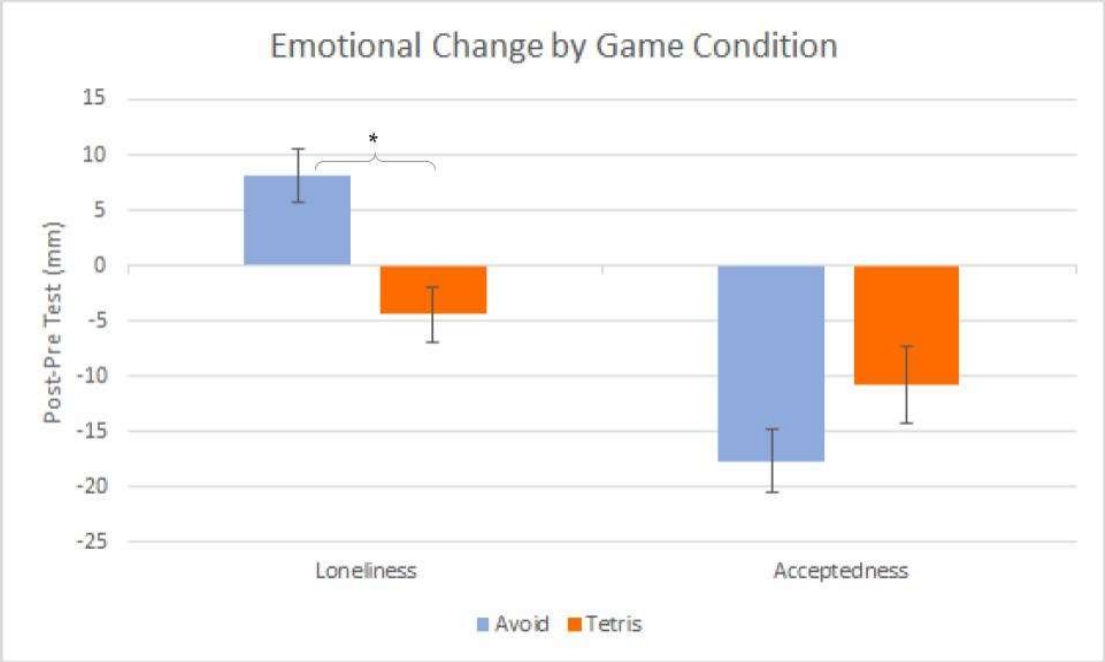

Figure 3: Difference between post and pre-scale of Loneliness and Acceptedness. Color of bars corresponds to condition. Bars above the x-axis are an increase in emotion from the pre-scale to the post-scale, bars below the x-axis are a decrease. Error bars are standard errors. Asterisks indicate significance at p < .05 level.

Agreement ratings and scales. This portion of the questionnaire was only given to participants in the Avoid condition since the questions and statements do not apply to the Tetris game, therefore the following ratings came from only the 97 participants in the Avoid condition. Ratings of agreement were made on a scale from 1 (Strongly Disagree) to 6, 1 (Strongly Agree), such that ratings above and below 3.5 represent agreement and disagreement, respectively. Medians, means, standard deviations and standard errors of the mean rating for each statement are presented in Table 4 below. Agreement ratings were generally high, all means are above the midpoint besides statement 5 (“just squares”) which was reverse coded in analyses.

As described in the Materials section above, these agreement ratings were averaged together to compose the Metaphor Awareness scale. The Metaphor

Awareness scale (α = .87) is a mean of statements 6-9 which generally address whether the game has special meaning, but also includes statements 1 and 2 which address the specific metaphor at work in the game (AFFECTION IS PROXIMITY) and a reverse code of statement 5, which is antithetical the game being metaphorical. The only statements excluded from the Metaphor Awareness scale were statements 3 and 4, which ask about other metaphors for affection and were used as distractor items. A principal components analysis was run to examine whether participants were treating the items in the Metaphor Awareness scale as representing more than one underlying construct (i.e. awareness and deliberateness), however, only the first factor had an eigenvalue above 1, so participants seem to be treating the seven items in the scale as tapping into a single construct. A Metaphor Specificity scale was also considered to assess whether participants were sensitive to the distinction between the game's metaphor (AFFECTION IS PROXIMITY) and other metaphors for affection (AFFECTION IS CONNECTION in statement 3 and AFFECTION IS WARMTH in statement 4). However, agreement ratings for all four metaphorical statements were high enough to make comparisons and median splits impractical (median of 5 for statements 2-4, median of 6 for statement 1). In addition, ratings of all four metaphorical statements (α = .92) seem to be very closely related, perhaps even more so than just statements 1 and 2 (α = .90). Therefore, Metaphor Specificity will not be used as a grouping variable in later analyses.

#	Statement	Median	Mean	SD	SE
5	The game was just about squares moving around.	2	2.72	1.43	0.14
6	The game has a specific message about people.	4	4.19	1.43	0.14
7	The game is intended to make you think about its meaning.	5	4.63	1.29	0.13
9	The game was creative or poetic.	4	3.51	1.35	0.14
8	The game was intended to compare the space between squares to social closeness.	5	4.32	1.56	0.16
3	The game implied that the player's square finds it difficult to **make connections** with people.	5	4.71	1.46	0.15
1	The game implied that the player's square is **being rejected** by the other squares.	6	4.99	1.44	0.15
2	The game implied that the player's square finds it difficult to **get close** to people.	5	4.82	1.40	0.14
4	The game implied that the other squares are **being cold** towards the player's square.	5	4.43	1.55	0.16
	Metaphor Awareness (mean of statements 6-9, 1, 2, and reverse coded statement 5	4.57	4.39	1.05	0.11

Table 4: Median, mean, standard deviation, and standard error of the mean of agreement ratings. Statistics for composite scales are also included at the bottom of the table. Bolding was added for clarity but was not present in the original questionnaire. Statement numbers were held constant from Experiment 1, but their order was changed, hence the statements not being presented in numerical order (more information in Materials section).

Metaphor Awareness groups. Awareness grouping within the Avoid condition was based primarily on a median split of the Metaphor Awareness scale: the 50 participants at or above the median (4.57) were grouped as Aware and the 47 participants below the median were grouped as Unaware. Splitting at the midpoint of the scale (3.5) was also used as a secondary grouping method, but because agreement

ratings were generally high this left very small groups below the midpoint. 17 of 97 participants were below the midpoint of the Metaphor Awareness scale. Such an imbalance in group size limits statistical power, so the midpoint split method was treated as secondary. The grouping method still has value in that the midpoint reflects the line between "Agree" and "Disagree" on the agreement ratings, so these groups correspond to overall agreement and disagreement on the items making up the Metaphor Awareness scale. These groups will be called Aware-Mid and Unaware-Mid, to distinguish them from the primary Aware and Unaware groups which used median split.

Emotional impact by median split of Metaphor Awareness scale. This section will compare the change in loneliness and acceptedness from the pre-scale to post-scale between the Tetris baseline condition and the Aware and Unaware groups of the Avoid condition, the construction of which is described in the Metaphor Awareness Groups section above.

Group	Emotion	Pre-Scale	Post-Scale	Mean Difference	Cohen's *d*	d_z
Tetris	Lonely	43(32)	38(31)	-4(18)	.14	.25
Aware	Lonely	41(33)	55(35)	14(23)	.42	.61
Unaware	Lonely	43(30)	44(29)	2(23)	.05	.06
Tetris	Accepted	89(27)	78(28)	-11(25)	.39	.44
Aware	Accepted	88(31)	64(29)	-23(35)	.78	.66
Unaware	Accepted	83(27)	71(26)	-12(16)	.45	.76

Table 5: Mean (standard deviation) of Loneliness and Acceptedness ratings for Aware, Unaware, and Tetris groups. All means and standard deviations are measured in millimeters. Cohen's d is calculated with pooled standard deviations while d_z is calculated with the standard deviations of difference scores.

Cohen's d is provided as the more standard measure for comparison to other studies, d_z is provided as a more sensitive measure for within-subjects comparisons.

On the pre-scale of emotional state, participants in the Aware group reported low levels of loneliness similar to those in the Unaware group and the Tetris condition. On the post-scale after playing their respective games, participants in the Aware group reported significantly higher levels of loneliness ($t(49) = -4.319$, $p < .001$), participants in the Unaware group reported a non-significant increase in levels of loneliness ($t(46) = -.431$, $p = .669$) while participants in the Tetris condition reported a non-significant change of loneliness ($t(48) = 1.768$, $p = .083$). A one-way ANOVA comparing the difference scores in each group found a significant effect of group on the change from pre to post-scale $F(2,143) = 9.702$, $p < .001$, $\eta^2 = .12$. Fisher-protected pairwise Welch t-tests comparing mean differences between conditions show that the increase in Loneliness in the Aware group was significantly different from both the decrease in the Tetris condition ($t(91.1) = -4.506$, $p < .001$) and the small increase in the Unaware group ($t(94.6) = 2.694$, $p = .008$). However, the difference between the Unaware group and the Tetris condition was not significant ($t(85.5) = -1.397$, $p = .166$).

On the pre-scale, feelings of acceptedness were very similar in the Aware group and Tetris condition, the Unaware group was slightly lower. This small difference on the pre-scale does not seem to be significant ($F(2,143) = 0.636$, $p = .531$, $\eta^2 = .01$), but nevertheless is addressed in the Discussion section. After playing their respective games, the post-scale showed a large decrease in acceptedness in the Aware group ($t(49) = 4.663$, $p < .001$), and smaller decreases in the Unaware group

($t(46) = 5.186$, $p < .001$) and Tetris condition ($t(48) = 3.069$, $p < .001$), all significant changes from pre to post-scale. A one-way ANOVA comparing the difference scores in each group found a significant effect of group on the change from pre to post-scale $F(2,143) = 3.365$, $p = .037$, $\eta^2 = .05$. As with the loneliness measure, post-hoc t-tests showed that the change in acceptance for the Aware group was significantly larger than either the Unaware group ($t(68.4) = -2.088$, $p = .040$) or the Tetris condition ($t(87.5) = 2.052$, $p = .043$). The decrease in the Unaware group was not significantly different from the decrease in the Tetris condition ($t(81.9) = 0.253$, $p = .801$).

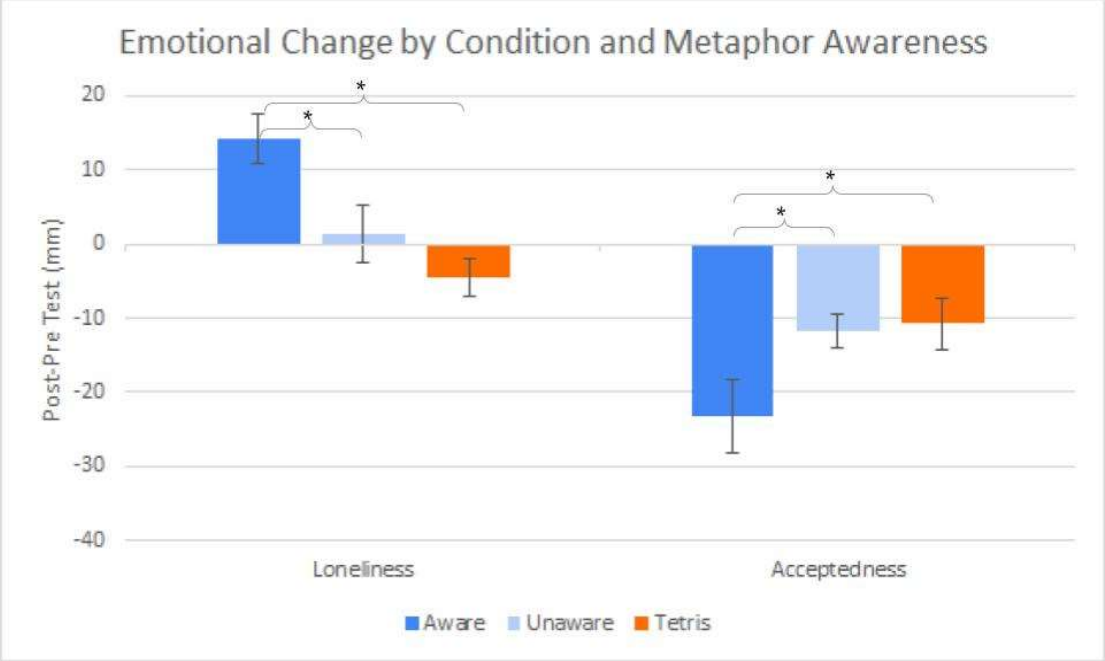

Figure 4: Difference between post and pre-scale of Loneliness and Acceptedness. Color of bars corresponds to condition. Bars above the x-axis are an increase in emotion from the pre-scale to the post-scale, bars below the x-axis are a decrease. Error bars are standard errors. Asterisks indicate significance at p < .05 level.

Emotional impact by midpoint split of Metaphor Awareness scale. This section will compare the change in loneliness and acceptedness from the pre-scale to post-scale between the Tetris baseline condition and the Aware-Mid and Unaware-

Mid groups of the Avoid condition, the construction of which is described in the Metaphor Awareness Groups section above. It should be noted that hypothesis tests will be underpowered in this section due to the small power of the Unaware-Mid group (n = 17), hence why these analyses are considered secondary.

Group	Emotion	Pre-Scale	Post-Scale	Mean Difference	Cohen's *d*	d_z
Tetris	Lonely	43(32)	38(31)	-4(18)	.14	.25
Aware-Mid	Lonely	41(32)	52(33)	11(25)	.34	.45
Unaware-Mid	Lonely	47(28)	41(29)	-6(16)	.21	.39
Tetris	Accepted	89(27)	78(28)	-11(25)	.39	.44
Aware-Mid	Accepted	85(30)	66(28)	-19(30)	.65	.63
Unaware-Mid	Accepted	87(24)	76(19)	-12(12)	.53	.95

Table 6: Mean (standard deviation) of Loneliness and Acceptedness ratings for Aware-Mid, Unaware-Mid, and Tetris groups. All means and standard deviations are measured in millimeters. Cohen's d is calculated with pooled standard deviations while d_z is calculated with the standard deviations of difference scores. Cohen's d is provided as the more standard measure for comparison to other studies, d_z is provided as a more sensitive measure for within-subjects comparisons.

On the pre-scale the Unaware-Mid group reported slightly higher levels of loneliness than other groups, but this difference was not significant $F(2,143) = .296$, $p = .744$, $\eta^2 < .01$. In paired sample t-tests from the pre to post-scale, the Aware-Mid group became significantly more lonely ($t(79) = -4.039$, $p < .001$), whereas the other groups both had small, non-significant decreases in loneliness (Unaware-Mid: $t(16) = 1.591$, $p = .131$, Tetris: $t(48) = 1.768$, $p = .083$). A one-way ANOVA comparing the difference scores in each group found a significant effect of group on the change from pre to post-scale $F(2,143) = 9.951$, $p < .001$, $\eta^2 = .12$. Fisher-protected pairwise Welch t-tests of these difference scores show that the increase in loneliness in the

Aware-Mid group is significantly different from the decreases in both the Unaware-Mid ($t(34.4) = 3.635$, $p = .001$) and Tetris groups ($t(123.7) = -4.172$, $p < .001$). The decreases in the Unaware-Mid and Tetris groups did not differ significantly ($t(30.6) = .369$, $p = .715$). These findings with the midpoint split method mirror the findings with the median split: The Aware group increased in loneliness, while the Unaware and Tetris groups both decreased in loneliness to similar degrees.

The acceptedness ratings of each group on the pre-scale are all similar. Paired-samples t-tests comparing pre and post-scale ratings show that participants in all three groups felt significant less accepted after playing their respective games, although the effect appears slightly bigger in the Aware-Mid group $t(79) = 5.631$, $p < .001$ than in the Unaware-Mid $t(16) = 3.932$, $p = .001$ and the Tetris group $t(48) = 3.069$, $p = .004$. However, a one-way ANOVA comparing the difference scores across groups did not find a significant difference between groups $F(2,143) = 1.644$, $p = .197$, $\eta^2 = .02$. This differs from the findings with the median-split grouping but given the similarity of the trends in acceptedness ratings (compare Figures 4 and 5) the overall patterns of results are comparable.

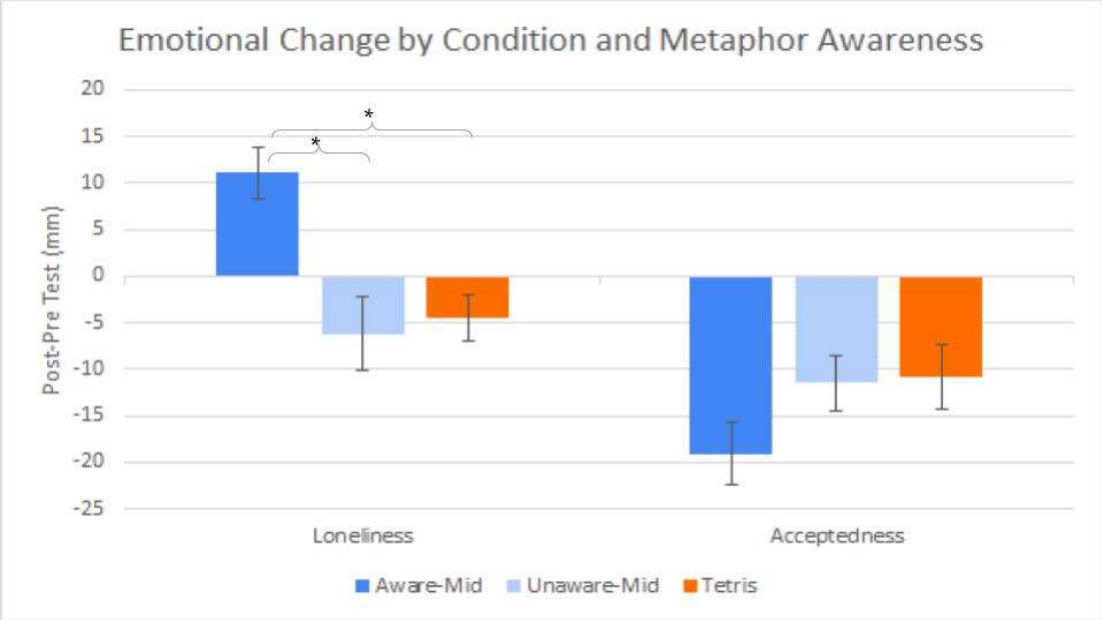

Figure 5: Difference between post and pre-scale of Loneliness and Acceptedness. Color of bars corresponds to condition. Bars above the x-axis are an increase in emotion from the pre-scale to the post-scale, bars below the x-axis are a decrease. Error bars are standard errors. Asterisks indicate significance at p < .05 level.

Emotional impact of metaphor awareness as a continuous variable. This section will treat the Metaphor Awareness scale as a continuous variable, rather than splitting into Aware and Unaware groups. This provides another approach to test the relationship between metaphor awareness and emotion change. Treating awareness as continuous is also more in keeping with the notion of metaphoricity as a scalable degree of metaphorical thinking, rather than all-or-nothing activation, as suggested in the description of the weak deliberate hypothesis. It is worth noting that this will exclude the Tetris group, who were not tested for metaphor awareness.

An examination of the scatterplot in Figure 6 suggests that participants with higher metaphor awareness tend to become lonelier after playing the Avoid version of the game. Participants are also clustered at the right side of the plot which illustrates

the overall high levels of metaphor awareness, which was noted in why the Unaware-Mid group was relatively small. Results of a Pearson correlation indicated a significant positive relationship between metaphor awareness and change in loneliness ($r(95) = .31$, $p = .002$) such that higher awareness of the metaphor predicted becoming lonelier after playing the Avoid game. A visual inspection also suggests that this relationship is roughly linear, or at least that there is not a strict cutoff between those who are or are not aware, nor does there appear to be a plateau of emotional impact once a certain level of awareness is reached. Rather, emotional impact seems to scale gradually with metaphor awareness, which is consistent with the weak deliberate hypothesis.

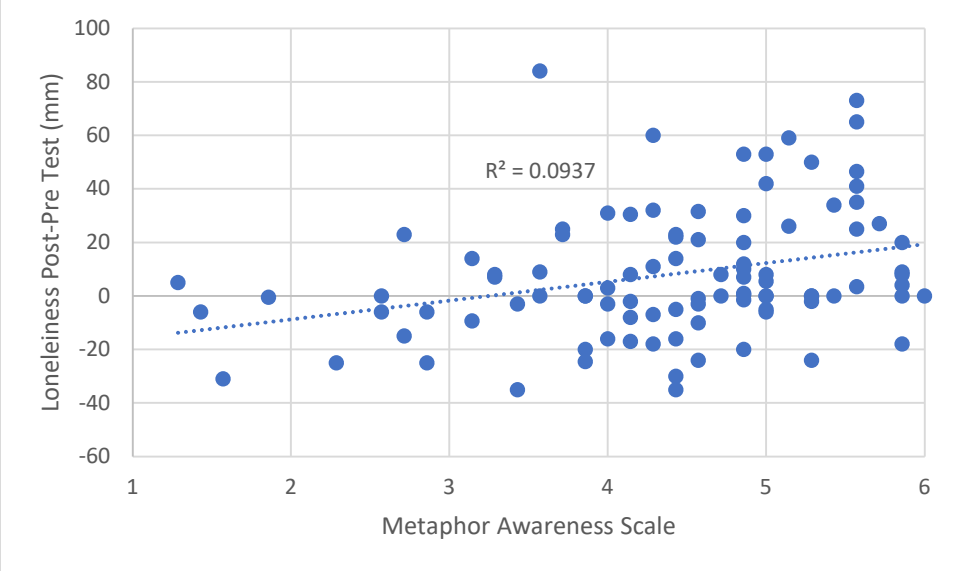

Figure 6: Scatterplot of change in loneliness and metaphor awareness scale. Positive numbers of the y-axis indicate becoming lonelier after playing the Avoid game, negative numbers indicate becoming less lonely.

An examination of Figure 7 shows a less clear relationship between acceptedness and metaphor awareness. There seems to be a very weak negative

relationship, with higher awareness predicting slightly lower acceptedness. This relationship was not significant according to a Pearson correlation ($r(95) = -.15$, $p = .133$). Acceptedness having a less clear relationship with awareness than loneliness did is not entirely surprising. Acceptedness showed decreases in Aware, Unaware, and Tetris groups, regardless of split method, as opposed to the crossover of effects on loneliness. Loneliness was also sensitive to differences between groups using both median and midpoint splits, whereas Acceptedness only showed differences between groups with the median split.

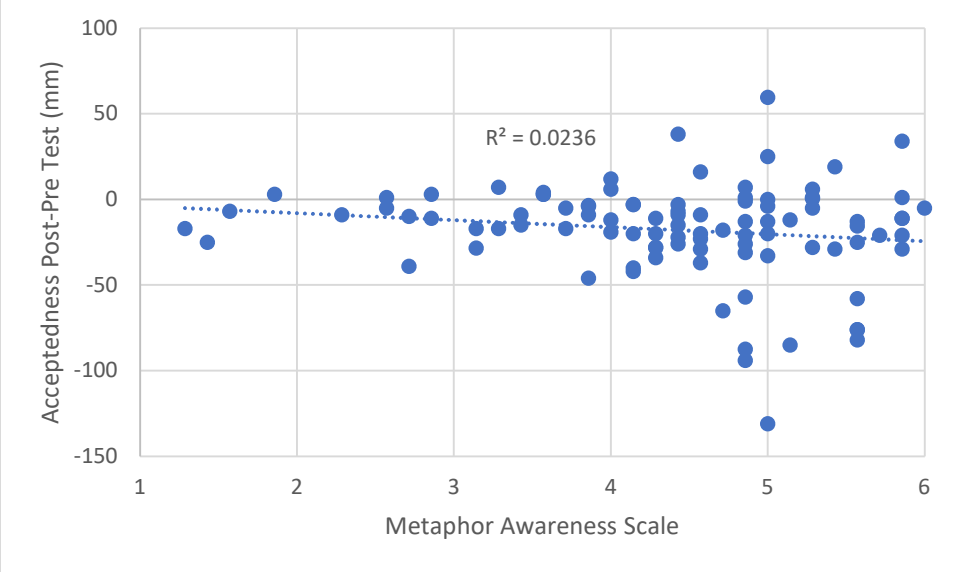

Figure 7: Scatterplot of change in acceptedness and metaphor awareness scale. Positive numbers of the y-axis indicate feeling more accepted after playing the Avoid game, negative numbers indicate feeling less accepted.

Game experience across metaphor awareness groups. This section will examine the relationship between overall experience playing games, frequency of game playing and metaphor awareness, as this may partly explain why some participants were more or less aware of the metaphor in *Loneliness*. On a scale of 1

(Not at all experienced) to 9 (Very experienced), participants were asked how experienced they were at playing video games. In the median-split groups, participants who were in the Aware group (M = 5.04, SD = 2.67) had slightly higher reported levels of experience than the Unaware group (M = 4.04, SD = 2.61), a marginally significant difference according to a Welch test ($t(94.9) = 1.859$, $p = .066$). The midpoint split groups showed a similar pattern, with the Aware-Mid group (M = 4.79, SD = 2.76) having higher experience levels than the Unaware-Mid group (M = 3.59, SD = 2.32), not quite a significant difference ($t(26.2) = 1.837$, $p = .078$). Finally, treating awareness as continuous reveals a marginally significant positive correlation between metaphor awareness and game experience ($r(95) = .192$, $p = .059$).

Participants were also asked how frequently they play video games, rated from 1 (Never) to 9 (Very frequently). In the median-split groups, participants who were in the Aware (M = 4.00, SD = 2.55) reported more frequent game playing than the Unaware group (M = 3.11, SD = 2.324), also not quite a significant difference ($t(94.9) = 1.806$, $p = .074$). Using the midpoint-split, the Aware-Mid group (M = 3.78, SD = 2.52) reported significantly more frequent game playing than the Unaware-Mid group (M = 2.59, SD = 2.00, $t(28.0) = 2.114$, $p = .044$). Finally, treating awareness as continuous reveals a significant positive correlation between metaphor awareness and game play frequency ($r(95) = .215$, $p = .035$).

These findings suggest that people who play games more often and see themselves as more experienced are more likely to perceive deliberateness in

Loneliness. However, these questions were asked towards the end of the questionnaire, after the statement ratings and interpretations that formed the deliberateness scale. Therefore, it is possible that answers were biased by some order effect. For example, a participant who found themselves interpreting the game metaphorically, when asked about their overall experience with video games, may be more likely to see themself as a savvy gamer than someone who had trouble interpreting the game. Such an order effect is possible, but it seems more likely that participants who are already more familiar with video games are more likely to pick up on the metaphorical nature of the game.

UC Santa Cruz has several game design related majors, so we might also expect participants who have experience making games would be more likely to be aware of any messages behind the game. However, since the participants were recruited from psychology classes, only two participants reported a game-related major, both in the "Games and Playable Media" major. These two did have high scores on the Metaphor Awareness scale (5.14 and 5.29 on a scale from 1 to 6) but given that there are only 2 participants among 147 in experiment 2, further analyses are impractical.

Discussion

The results of Experiment 2 replicate and extend the findings of Experiment 1. The Avoid condition increased loneliness and decreased acceptedness as in Experiment 1, in this case with Tetris as a baseline comparison. The baseline had a greater impact on emotions than expected, with a significant decrease in acceptedness

which roughly matched the decrease in the Unaware group. Unless Tetris itself is more emotionally impactful than expected, this demonstrates the value of having a baseline condition to compare to, which was lacking in Experiment 1. If we assume that emotional change in the Tetris condition is due to being isolated in the testing room for several minutes, then changes in the other groups must go beyond this simple baseline effect to be attributed to the Avoid game.

The Aware groups (both median and midpoint split) clearly meet this standard, with an increase in loneliness and decrease in acceptedness both significant relative to the Tetris baseline. The Unaware groups did not seem to show such a distinction. Their decrease in acceptedness is significant as a within-subjects comparison but is almost indistinguishable from the decrease in the baseline, so this could be due to being isolated in the testing room. The Unaware groups show even less activity on the loneliness measure, with a negligible change within-subjects and no-significant difference compared to the baseline. This lack of emotional impact in the median-split Unaware group is surprising considering the way that the awareness groups were determined; the Metaphor Awareness scale was relatively right-skewed, with a median of 4.57 and mean of 4.39 on a scale from 1 to 6 (well above the 3.5 midpoint), so the median split grouping method placed some participants in the Unaware group who were above the midpoint of the scale and likely were aware of the metaphor to some extent. One would expect this inclusion of possible metaphor aware participants within the Unaware group to make this group more likely to show an impact like the Aware participants. Instead, despite the median split method

favoring an effect in the Unaware group, the emotional impact in this group largely mirrored the baseline, and the midpoint split method obtained similar results.

The pattern of results from both the midpoint and median splits are consistent with the strong deliberate hypothesis. The Aware group experienced an increase in loneliness and decrease in acceptedness as predicted, and both of these effects were greater for this group than either the Unaware or Tetris groups. The Unaware group's emotional impacts could not be distinguished from the impacts in the baseline Tetris group. This aligns with the strong deliberate predictions that only the Aware would show an emotional impact, while the impact in the Unaware group is indistinguishable from the baseline. However, this should be taken with a grain of salt, as the gradual increase in loneliness associated with awareness found when metaphor awareness was treated as continuous is more consistent with the weak deliberate hypothesis. Rather than needing to reach a certain level of awareness to be emotionally impacted, the scatterplot in Figure 6 suggests that the change in loneliness scales gradually with awareness.

The design of Experiment 2 does of course have some limitations. Awareness is operationalized relatively; the Aware and Unaware groups are based on median and midpoint splits of a scale developed alongside these experiments and is not intended as an absolute benchmark of which participants are aware of the metaphor. This experiment treats awareness as quasi-experimental and is not designed to determine what drives a participant towards explicit metaphor awareness. However, analysis of self-reported levels of experience with games suggests that more experienced game

players may be more likely to be explicitly aware of the metaphor. Furthermore, the small non-significant differences in acceptedness on the pre-scale for the Aware and Unaware groups may play some role in forming interpretations; for example, participants who start the experiment feeling more socially comfortable may be more willing to look for a deeper meaning in the game.

There is a possibility of reverse causality, wherein participants who were more emotionally affected by playing the Avoid game may interpret it more deeply to justify the change in their emotions. Given the small changes in emotion relative to the length of the scale (10 to 20 mm change on a 140 mm long line) and the lack of numbers along the scale to anchor responses to, it seems unlikely that participants are aware of the change from the pre to post-scale. However, it is still possible that participants are aware of their emotional experience while playing the game, and thus become aware of the metaphor. Whether the feeling of loneliness comes before awareness or vice versa is not the focus of this paper. In both cases I would claim that participants must subconsciously activate a mapping from movement of squares to rejection before either becoming more lonely or consciously aware of the metaphor.

It is also possible that some of the participants who scored low on the Metaphor Awareness scale and showed negligible emotional impact simply did not pay attention to the game or skipped ahead without finishing the game. While the instructions to progress through the study were only presented at the end of the game, it was still possible to click the “next” button before the game finished. This disengagement account could at least partially explain the lack of effects in the

Unaware groups. The removal of participants who seemed to rush through the experiment should mitigate this effect, but it is still possible that this contributed to the results. To address this possibility, further studies may need to record some activity within the game or otherwise check if participants were paying attention.

It is possible that participants' interpretations change as they go through the questionnaire, and thus do not entirely reflect their interpretation as they played the game. We took several measures to mitigate this possibility: (a) the statements that form the Metaphor Awareness scale appear early in the questionnaire, so the primary measure of metaphor awareness is not influenced by the rest of the questionnaire, (b) ongoing analyses of qualitative responses are factoring in the order in which questions were asked and making note of participants whose metaphor awareness changes over the course of the questionnaire, (c) at the end of the questionnaire the participants were asked how their interpretation has changed. In addition, even if participants' interpretations are shifting over the course of the questionnaire, both the pre-scale and post-scale of emotional state are administered before any interpretation questions, so the difference in emotional impact between the Aware and Unaware groups suggests that some difference between those groups existed before they were asked any interpretive questions. All of these control procedures aside, participants' ability to reflect on the game and explain metaphorical elements after the fact may be interesting on its own, particularly given that not all participants produced such a metaphorical interpretation even by the end of the questionnaire.

The inclusion of the text “This game is about people” above the game likely inflated metaphor awareness, contributing to the high scores on the Awareness scale. However, despite this inflation of awareness, the Awareness scores were still spread across the scale and the Unaware groups did not reliably show an emotional impact, which suggests the text hint was not so extreme as to cue all participants in on the metaphor.

CHAPTER VII

General Discussion

These results are consistent with something between the strong and weak deliberate hypotheses: when participants were split into groups in Experiment 2, the only reliable emotional impacts of the Loneliness game were observed in those aware of the game's metaphor, and when awareness was analyzed as continuous we observed a gradual increase in emotional impact with more metaphor awareness. This is surprising, as the bulk of empirical metaphor studies support the non-deliberate hypothesis, showing all sorts of implicit effects of metaphor. However, these previous studies have demonstrated that explicit awareness of metaphor is not necessary for metaphorical thinking (Thibodeau & Boroditsky, 2011, 2013), but not that awareness has no role in metaphorical thinking. The present experiments do not invalidate previous implicit findings, rather they suggest that there may be an understudied territory of empirical research on explicit awareness of metaphor – territory which a combination of deliberate metaphor theory and multimodal metaphor theory may be helpful in exploring.

An abstract game like Loneliness, what Möring (2015) might call an "artgame", may be an instance of metaphor that prompts conscious reflection, and in this case that reflection may be necessary to be emotionally impacted by the game. Particularly given the association between awareness and higher levels of game experience, perhaps the Unaware participants were less able or less inclined to reflect on a deeper meaning within a game. Participants who play games more often may be

more likely to find the minimalism of *Loneliness* striking compared to commercial games, and this may help prompt reflection. This is somewhat analogous to coming across minimalist art in a museum after walking past many baroque pieces – a painting of a square on a white background must have some sort of meaning, otherwise what would be the point? To continue the analogy, museumgoers with more experience looking at art, especially modern art, may be more able and willing to analyze such a minimalist piece. Similarly, the more experienced participants in Experiment 2 may have been more able and willing to analyze this strange game full of moving squares.

This support of the strong and weak deliberate hypotheses does not necessarily suggest that explicit, conscious metaphor is the only way for games to evoke cross-domain mappings (and therefore does not specifically support the paradox of metaphor claim of DMT). Rather, it demonstrates that awareness of metaphor has a role in the impact of an abstract, minimalist game like Loneliness, and perhaps other games or art which bring attention toward their metaphorical elements. This is consistent with deliberate metaphor theory's emphasis on explicit metaphors which draw the audience's attention. While strict DMT would also require the participants to attribute the metaphor to a deliberate communicative strategy by the game's creator, which these experiments could not establish, awareness of the metaphor should be a necessary condition for that fully deliberate interpretation. While future studies may take this further step of demonstrating fully deliberate

interpretations, the centrality of explicit metaphor awareness to DMT allows us to interpret these findings in the context of DMT.

This role for conscious interpretation of metaphor fits into the story that Gibbs (2015a) proposed, that deliberate, conscious metaphor can complement existing CMT work on political and medical metaphor that are used with communicative purposes (Harland, 2012; Lakoff, 2004) rather than being directly opposed to CMT. This fits a larger trend in CMT of studying different levels of activation of cross-domain mappings (Kövecses, 2015; Müller, 2009), in which explicit metaphor awareness may be just a high level of metaphoricity but not the only way of activating cross-domain mappings.

If this is the case, then abstract and minimalist games are a new context in which researchers can study such high levels of metaphoricity, with more control than researchers have over most other multimodal metaphor. Minimalist games can have their key metaphorical elements reprogramed with no other alterations to create experimental conditions (as in the Avoid vs Approach versions used in Experiment 1), whereas it is much harder to re-film a movie scene or redraw a cartoon in a way that only changes the metaphorical element of interest. There may yet be other games which evoke cross-domain mappings without conscious awareness, although these effects may be subtler and harder to test empirically - see Anderson, Karzmark, and Wardrip-Fruin (2019) for an example of the difficulty in testing games with more subtle messages.

Beyond the role of awareness, these experiments also support the basic claim of CMT, which spurred the development of MMT, that metaphorical thought goes beyond language to other modalities of experience. Given that: 1) participants produced metaphorical interpretations of the game[4], 2) that most participants agreed with metaphorical statements about the game, and 3) the game had a metaphorically consistent emotional impact on players, it seems reasonable to say that participants are processing the game as metaphorical even if it is not linguistic. The importance of testing CMT's predictions outside of linguistic instances of metaphor has been pointed out by proponents (Forceville, 2012) and critics (Murphy, 1996) of CMT alike.

These experiments expand upon the methodologies used in Williams' ostracism research (1997) and Heider and Simmel's animation (1944). The differences between these stimuli and *Loneliness* are explored in the Related Stimuli section, here I will focus on how the present studies add to the findings of these previous works. The agents in Williams' Cyberball game are shaped like people, whereas *Loneliness* employs black squares to represent people – so the finding of increased feelings of loneliness after being "rejected" by squares on a screen suggests that our ostracism detection system (Wesselmann et al., 2009) does not require human figures, but can even operate through metaphorical representations of people. Heider and Simmel's experiments focused on the interpretation of the animation and did not analyze the emotional impact of watching it. This dissertation's analysis of the

[4] According to preliminary analysis of qualitative responses

relationship between interpretation (metaphor awareness) and emotional impact extends Heider and Simmel's work, showing that not only can participants interpret minimalist stimuli, this interpretation has a direct impact on their emotional reaction to the stimulus. The interactive nature of *Loneliness* relative to a pre-determined film clip also expands upon Heider and Simmel's methodology, showing that participants can interpret deeper meaning into a system's reaction to their input, rather than simply reading meaning into a carefully planned sequence of animations.

Further investigation of the factors and individual differences in conscious interpretations of metaphor is warranted, such as the differences in prior experience with the medium or small differences in emotional state on the pre-scale that Experiment 2 found. I will also reiterate the call from Forceville and colleagues (2012) for more empirical testing of multimodal metaphor, and add that participants' metaphor awareness may be a useful lens in analyzing results or even grouping participants as in Experiment 2. Future studies may also seek to manipulate the way that a stimulus draws attention to metaphor such as varying titles or introductory text, which would more directly capture the DMT construct of deliberateness. Of course, further testing of multimodal metaphor should also test a broader swath of games, as other genres and perspectives may lend themselves to different metaphorical and emotional experiences. For example, metaphorical themes built up over dozens of hours of a role-playing game could be more nuanced than what Loneliness can offer in four minutes, and a game played in first-person perspective might make it easier to step into a character's shoes. Further study of the relationship between gaming

experience and metaphor awareness could also explore experience in specific game genres and styles, which these experiments did not include. Finally, experimental tests of MMT should be performed with other mediums such as film or political cartoons, though I have explained above how it may be harder to exercise experimental control when the stimuli cannot simply be reprogrammed.

CHAPTER VIII

Conclusions

These experiments expand upon a limited field of empirical study of MMT and DMT. In particular, these experiments extend the testing of DMT and CMT into multimodal metaphor, which researchers Forceville and colleagues (2012) have pointed out as a branch lacking empirical work. By testing DMT with a non-linguistic presentation of metaphor, we avoid the lexical disambiguation account that Steen has used to dismiss previous empirical evidence for CMT. In addition, using a game as a stimulus instead of a conversation between two other people avoids the potential problem of the participant not being the intended audience of metaphorical communication (Gibbs, 2015b). The critical role of metaphor awareness in the emotional impact of the game also draws an interesting contrast with the typical methods of testing CMT which emphasize implicit metaphor (Thibodeau & Boroditsky, 2011).

This dissertation also contributes a new methodological approach to the use of multimodal metaphor in studying metaphor theories. Experiment 1 demonstrates that differently programed variants of a metaphorical game can provide experimental conditions which have reliably different effects on participants. Experiment 1's Approach version also provides a cautionary note that variants of an existing game may not always be metaphorically coherent (e.g. the chasing ghosts/pigeons interpretations). Experiment 2 demonstrates that the emotional impact of an instance of multimodal metaphor depends on conscious awareness of metaphor. The effect of

awareness in this case suggests that as other researchers begin to empirically test multimodal metaphor theory it may be useful to focus on instances of metaphor which draw the audience's attention.

These experiments depart from the way games are often studied in psychology, either as improving spatial skills like mental rotation (De Lisi & Wolford, 2010; Feng et al., 2007) or in relation to aggression (Sherry, 2001). Instead, these experiments explore the role of metaphorical interpretation in how games affect their players. These findings suggest that explicitly interpreting a game as metaphorical can deepen your emotional relationship with the work. Furthermore, this adds empirical weight to the fundamental claim of CMT that metaphor pervades everyday human experience, not just language.

Appendix A: Analyses with Excluded Participants

In Experiment 1, 81 students participated in the experiment, but 3 were removed for incomplete responses while another 18 were removed for other reasons as described in the Participants section of Experiment 1, leaving 60 participants for the analyses in the Results section of Experiment 1. Included below are the primary analyses from the Results re-run including the 18 who were removed for the main results for 78 total participants.

With this inclusive set of 78 participants, the Avoid group increases in loneliness from the pre-scale (M = 40mm, SD = 31mm) to the post-scale (M = 45mm, SD = 32mm), although this within-subjects increase is not significant $t(38) = -1.056$, $p = .298$. The Approach group decreases from the pre-scale (M = 39mm, SD = 32mm) to the post-scale (M = 32mm, SD = 27mm), this decrease was significant $t(38) = 2.167$, $p = .037$. A one-way ANOVA comparing the change in loneliness across conditions found a significant effect of condition such that the increase in loneliness in the Avoid condition was greater than the decrease in the Approach condition ($F(1,76) = 4.107$, $p = .046$, $\eta^2 = .05$). The decrease in loneliness in the Approach condition and the difference between conditions mirrors the findings in the original tests from the Results section, although in this new test the decrease in the Avoid condition was not significant, unlike the original tests.

The Avoid condition in this inclusive data decreased in acceptedness ratings from the pre-scale (M = 88mm, SD = 32mm) to the post-scale (M = 64mm, SD =

34mm), a significant within-subjects decrease $t(38) = 3.548$, $p = .001$. The Approach condition decreased slightly from the pre-scale (M = 83mm, SD = 30mm) to the post-scale (M = 80mm, SD = 29mm), a non-significant decrease $t(38) = .939$, $p = .354$. A one-way ANOVA comparing the change in acceptedness across conditions found a significant effect of condition such that the decrease in acceptedness in the Avoid condition was greater than the increase in the Approach condition ($F(1,76) = 8.714$, $p = .004$, $\eta^2 = .10$). These differences in acceptedness mirror those found in the original tests, with a decrease in the Avoid condition, a negligible increase in the approach condition, and a significant effect of condition upon these changes.

In Experiment 2, 167 students participated in the experiment, but 4 were removed for incomplete responses while another 20 participants were removed for reasons described in the Participants section of Experiment 2, leaving 143 participants for the analyses in the Results section of Experiment 1. Included below are the primary analyses from the Results re-run including the 20 who were removed for the main results for 163 total participants.

With the full set of 163 participants, the median of the Metaphor Awareness scale is 4.57, as it was in the original analyses. Splitting the Avoid condition based on this median, the Aware group increased in loneliness from 39mm (SD = 32mm) on the pre-scale to 54mm (SD = 31mm) on the post scale; this within subjects increase was significant $t(47) = -4.395$, $p < .001$. The Unaware group showed negligible change in loneliness from the pre-scale (M = 45mm, SD = 29) to the post-scale (M = 45mm, SD = 29mm) ($t(60) = -.265$, $p = .793$). The Tetris group showed a small

decrease in loneliness from the pre-scale (M = 42mm, SD = 32) to the post-scale (M = 36mm, SD = 30), this decrease was significant ($t(54) = 2.266$, $p = .028$). A one-way ANOVA comparing the difference scores in loneliness across the three groups found a difference between groups ($F(2,160) = 12.128$, $p < .001$, $\eta^2 = .13$). Fisher-protected post-hoc Welsh t-tests found the increase in loneliness in the Aware group was larger than the smaller changes in the Unaware ($t(98.5) = 3.193$, $p = .002$) and the Tetris groups ($t(86.5) = 4.898$, $p < .001$). The negligible change in loneliness in the Unaware group was not significantly different from the decrease in the Tetris group ($t(111.6) = 1.658$, $p = .100$). These significance tests all came out with the same findings as in the original tests in the Results section.

In a median split with this larger group, the Aware group showed a significant decrease in feelings of acceptedness from the pre-scale (M = 88mm, SD = 31mm) to the post-scale (M = 65mm, SD = 30mm), a significant within-subjects decrease ($t(47) = 4.537$, $p < .001$). The Unaware group also decreased in acceptedness from the pre-scale (M = 81mm, SD = 28mm) to post-scale (M = 68mm, SD = 27mm), a smaller though still significant decrease ($t(60) = 5.590$, $p < .001$). The Tetris group also decreased in acceptedness from the pre-scale (M = 90mm, SD = 27mm) to the post-scale (M = 80mm, SD = 27mm), again a small but significant decrease ($t(53) = 2.848$, $p = .006$). A one-way ANOVA comparing the difference scores in acceptedness across the three groups found a difference between groups ($F(2,160) = 3.993$, $p = .020$, $\eta^2 = .05$). Fisher-protected post-hoc Welsh t-tests found the decrease in acceptedness in the Aware group was marginally greater than the decrease in the

Unaware group ($t(64.6) = -1.923$, $p = .059$), and significantly greater than the decrease in the Tetris group ($t(80.0) = -2.299$, $p = .024$). The decreases in the Unaware and Tetris groups were not significantly different ($t(95.6) = -.814$, $p = .418$). All of these significant differences are the same as those found in the original tests in the Results section, except that the comparison between the decreases in acceptedness in the Aware and Unaware groups was significant in the original results and was marginally significant in this test, likely because of increased noise from some of the participants who were removed for the original tests.

Using the Metaphor Awareness scale as a continuous variable, the significant positive correlation between change in loneliness and awareness comes out in this more inclusive sample $r(107) = .33$, $p = .001$. Just like the correlation found in the original results, as metaphor awareness increases, participants tend to get more lonely after playing the game. There is a weaker, non-significant negative correlation between change in acceptedness and awareness in this inclusive sample $r(107) = -.15$, $p = .126$. This is also similar to the weak, nonsignificant negative correlation between change in acceptedness and awareness in the original results.

The similarity of results displayed in this appendix to the results in the body of the paper illustrate that the exclusion of participants who has additional interactions with the experimenter or who seemed to hurry through the experiment did not noticeably impact the results.

References

Aerolab. (n.d.). *Blockrain.js—A Tetris game in HTML5 + Javascript*. Blockrain.Js. Retrieved May 19, 2020, from http://aerolab.github.io/blockrain.js

Anderson, B. R., Karzmark, C. R., & Wardrip-Fruin, N. (2019). The psychological reality of procedural rhetoric. *Proceedings of the 14th International Conference on the Foundations of Digital Games*, 1–9. https://doi.org/10.1145/3337722.3337751

Bargh, J. A. (2016). Awareness of the prime versus awareness of its influence: Implications for the real-world scope of unconscious higher mental processes. *Current Opinion in Psychology*, *12*, 49–52. https://doi.org/10.1016/j.copsyc.2016.05.006

Blow, J. (2008). *Braid*. Number None Inc.

Bobrova, L. (2015). A procedure for identifying potential multimodal metaphors in TV commercials. *Multimodal Communication*, *4*(2), 113–131.

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, *75*(1), 1–28.

Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, *13*(2), 185–189.

Bounegru, L., & Forceville, C. (2011). Metaphors in editorial cartoons representing the global financial crisis. *Visual Communication*, *10*(2), 209–229.

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, *112*(1), 193.

Charteris-Black, J. (2012). Forensic deliberations on 'purposeful metaphor.' *Metaphor and the Social World*, *2*(1), 1–21. https://doi.org/10.1075/msw.2.1.01cha

Cienki, A., & Müller, C. (2008). Metaphor, gesture, and thought. *The Cambridge Handbook of Metaphor and Thought*, 483–501.

De Lisi, R., & Wolford, J. L. (2010). Improving children's mental rotation accuracy with computer game playing. *The Journal of Genetic Psychology*, *163*(3). https://doi.org/10.1080/00221320209598683

Duffy, S. E. (2014). The role of cultural artifacts in the interpretation of metaphorical expressions about time. *Metaphor and Symbol*, *29*(2), 94–112.

Duffy, S. E., Feist, M. I., & McCarthy, S. (2014). Moving through time: The role of personality in three real-life contexts. *Cognitive Science*, *38*(8), 1662–1674.

Feist, M. I., & Duffy, S. E. (2015). Moving beyond 'Next Wednesday': The interplay of lexical semantics and constructional meaning in an ambiguous metaphoric statement. *Cognitive Linguistics*, *26*(4), 633–656.

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, *18*(10), 850–855. https://doi.org/10.1111/j.1467-9280.2007.01990.x

Forceville, C. (1994). Pictorial metaphor in advertisements. *Metaphor and Symbol*, *9*(1), 1–29.

Forceville, C. (2002). The identification of target and source in pictorial metaphors. *Journal of Pragmatics*, *34*(1), 1–14.

Forceville, C. (2006). Non-verbal and multimodal metaphor in a cognitivist framework: Agendas for research. *Applications of Cognitive Linguistics*, *1*, 379–402.

Forceville, C. (2007). Multimodal metaphor in ten Dutch TV commercials. *Public Journal of Semiotics*, *1*(1), 15–34.

Forceville, C. (2008). Pictorial and multimodal metaphor in commercials. In E. F. McQuarrie, & B. J. Phillips (Eds.), Go figure! New directions in advertising rhetoric (pp. 178-204). Armonk, NY: M.E. Sharpe.

Gentner, D., & Bowdle, B. F. (2001). Convention, form, and figurative language processing. *Metaphor and Symbol*, *16*(3–4), 223–247.

Gentner, D., Imai, M., & Boroditsky, L. (2002). As time goes by: Evidence for two systems in processing space→ time metaphors. *Language and Cognitive Processes*, *17*(5), 537–565.

Gibbs, R. W. (2011a). Are 'deliberate' metaphors really deliberate?: A question of human consciousness and action. *Metaphor and the Social World*, *1*(1), 26–52.

Gibbs, R. W. (2011b). Evaluating conceptual metaphor theory. *Discourse Processes*, *48*(8), 529–562.

Gibbs, R. W. (2015a). Does deliberate metaphor theory have a future? *Journal of Pragmatics*, *90*, 73–76. https://doi.org/10.1016/j.pragma.2015.03.016

Gibbs, R. W. (2015b). Do pragmatic signals affect conventional metaphor understanding? A failed test of deliberate metaphor theory. *Journal of Pragmatics*, *90*, 77–87. https://doi.org/10.1016/j.pragma.2015.05.021

Gibbs, R. W. (2017). *Metaphor Wars*. Cambridge University Press.

Glucksberg, S., & Haught, C. (2006). On the relation between metaphor and simile: When comparison fails. *Mind & Language*, *21*(3), 360–378.

Harland, P. (2012). *Trust Me, I'm the Patient: Clean Language, Metaphor, and the New Psychology of Change*. Wayfinder Press.

Hauser, D. J., Carter, M. S., & Meier, B. P. (2009). Mellow Monday and furious Friday: The approach-related link between anger and time representation. *Cognition and Emotion*, *23*(6), 1166–1180.

Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *The American Journal of Psychology*, *57*(2), 243–259. JSTOR.

https://doi.org/10.2307/1416950

Hendricks, R. K., & Boroditsky, L. (2017). New space–time metaphors foster new nonlinguistic representations. *Topics in Cognitive Science*, *9*(3), 800–818.

Kövecses, Z. (2015). *Where Metaphors Come from: Reconsidering Context in Metaphor*. Oxford University Press. https://books.google.com/books?hl=en&lr=&id=Ny3RBQAAQBAJ&oi=fnd&pg=PP1&ots=ykYR89wvHH&sig=r0p9rW3CzWlL_ANFw9ZRvNAsfhU#v=onepage&q&f=false

Krennmayr, T. (2011). *Metaphor in newspapers*. Netherlands Graduate School of Linguistics.

Kromhout, R., & Forceville, C. (2013). LIFE IS A JOURNEY: Source–path–goal structure in the videogames "Half-Life 2", "Heavy Rain", and "Grim Fandango." *Metaphor and the Social World*, *3*(1), 100–116.

Lai, V. T., & Curran, T. (2013). ERP evidence for conceptual mappings and comparison processes during the comprehension of conventional and novel metaphors. *Brain and Language*, *127*(3), 484–496.

Lakoff, G. (2004). *Don't Think of an Elephant! Know Your Values and Frame the Debate: The Essential Guide for Progressives*. Chelsea Green Publishing.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago press.

Lakoff, G., & Kövecses, Z. (1987). The cognitive model of anger inherent in American English. *Cultural Models in Language and Thought*, 195–221.

Lantolf, J. P., & Bobrova, L. (2012). Happiness is drinking beer: A cross-cultural analysis of multimodal metaphors in American and Ukrainian commercials. *International Journal of Applied Linguistics*, *22*(1), 42–66.

Magnuson, J. (2007). *Loneliness* [Java-based game]. https://www.necessarygames.com/my-

games/loneliness/flash

McGlone, M. S. (2011). Hyperbole, homunculi, and hindsight bias: An alternative evaluation of conceptual metaphor theory. *Discourse Processes*, *48*(8), 563–574.

Möring, S. (2015). What Is a Metaphoric Artgame? A Critical Analysis of Metaphor in the Artgame Discourse and in Artgames. *Embodied Metaphors in Film, Television, and Video Games: Cognitive Approaches*, *76*, 269–285.

Müller, C. (2009). *Metaphors dead and alive, sleeping and waking: A dynamic view*. University of Chicago Press, Chicago.

Murphy, G. L. (1996). On metaphoric representation. *Cognition*, *60*(2), 173–204.

Nacey, S. (2013). *Metaphors in Learner English* (Vol. 2). John Benjamins Publishing.

Ng, C. J. W., & Koller, V. (2013). Aware conventional metaphor in images: The case of corporate branding discourse. *Metaphor and Symbol*, *28*(3), 131–147. https://doi.org/10.1080/10926488.2013.797807

Núñez, R. E., Motz, B. A., & Teuscher, U. (2006). Time after time: The psychological reality of the ego-and time-reference-point distinction in metaphorical construals of time. *Metaphor and Symbol*, *21*(3), 133–146.

Read, S. J., Cesa, I. L., Jones, D. K., & Collins, N. L. (1990). When is the federal budget like a baby? Metaphor in political rhetoric. *Metaphor and Symbol*, *5*(3), 125–149.

Reijnierse, W. G., Burgers, C., Krennmayr, T., & Steen, G. J. (2015). How viruses and beasts affect our opinions (or not): The role of extendedness in metaphorical framing. *Metaphor and the Social World*, *5*(2), 245–263. https://doi.org/10.1075/msw.5.2.04rei

Sherry, J. (2001). The effects of violent video games on aggression: A meta-analysis. *Human Communication Research*, *27*(3), 409–431. https://doi.org/10.1111/j.1468-

2958.2001.tb00787.x

Stamenković, D., & Jaćević, M. (2015). Time, Space, and Motion in Braid: A Cognitive Semantic Approach to a Video Game. *Games and Culture*, *10*(2), 178–203.

Steen, G. J. (2008). The Paradox of Metaphor: Why We Need a Three-Dimensional Model of Metaphor. *Metaphor and Symbol*, *23*(4), 213–241. https://doi.org/10.1080/10926480802426753

Steen, G. J. (2010). When is metaphor deliberate. *Selected Papers from the Stockholm 2008 Metaphor Festival*, 43–63.

Steen, G. J. (2011). From three dimensions to five steps: The value of deliberate metaphor. *Metaphork.De*, 83–110.

Steen, G. J. (2017). Deliberate Metaphor Theory: Basic assumptions, main tenets, urgent issues. *Intercultural Pragmatics*, *14*(1), 1–24. https://doi.org/10.1515/ip-2017-0001

Steen, G. J., Reijnierse, W. G., & Burgers, C. (2014). When do natural language metaphors influence reasoning? A follow-up study to Thibodeau and Boroditsky (2013). *PloS One*, *9*(12), e113536.

Sundar, S. S., & Nass, C. (2000). Source Orientation in Human-Computer Interaction: Programmer, Networker, or Independent Social Actor. *Communication Research*, *27*(6), 683–703. https://doi.org/10.1177/009365000027006001

Thibodeau, P. H., & Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. *PloS One*, *6*(2), e16782.

Thibodeau, P. H., & Boroditsky, L. (2013). Natural language metaphors covertly influence reasoning. *PloS One*, *8*(1), e52961.

Thibodeau, P. H., & Boroditsky, L. (2015). Measuring effects of metaphor in a dynamic opinion landscape. *PloS One*, *10*(7), e0133939.

Vervaeke, J., & Kennedy, J. M. (1996). Metaphors in language and thought: Falsification and multiple meanings. *Metaphor and Symbol*, *11*(4), 273–284.

Wesselmann, E. D., Bagg, D., & Williams, K. D. (2009). "I Feel Your Pain": The effects of observing ostracism on the ostracism detection system. *Journal of Experimental Social Psychology*, *45*(6), 1308–1311. https://doi.org/10.1016/j.jesp.2009.08.003

Williams, K. D. (1997). Social Ostracism. In R. M. Kowalski (Ed.), *Aversive Interpersonal Behaviors* (pp. 133–170). Springer US. https://doi.org/10.1007/978-1-4757-9354-3_7