# Cross-Sensory Correspondences: Cross-Activation of Connotative Feature Dimensions through the Felt Heaviness of Lifted Objects

Gabrielle Scallon, M.Sc.

A thesis submitted to Lancaster University

in partial fulfilment of the requirements for

the degree of Doctor of Philosophy

Lancaster University

2018

## Declaration

I declare that this thesis is my own work and has not been submitted in substantially the same form for the award of higher degree elsewhere.

Gabrielle Scallon

## Abstract

Cross-sensory correspondences are the systematic associations demonstrated to arise between various feature dimensions such that their relative extremes are aligned. It has been proposed that correspondences arise as a result of cross-talk between abstract, amodal connotations of a core set of feature dimensions (P. Walker, 2016). Although there is some evidence to suggest that a dimension denoting heaviness may be included among a set of aligned feature dimension, the evidence to demonstrate this is limited. The present work explores whether heaviness, as received through the lifting of weighted objects, may enter into this scheme of correspondences. In addition, the separate contributions of the heaviness and size of lifted objects to the cross-activation of other feature dimensions is also considered. The influences of size and heaviness were explored in light of the size-weight illusion; a phenomenon where the size of equally weighted objects alters their felt heaviness such that the smaller object is experienced to be heavier than the larger object. A series of rating scale tasks were conducted, examining whether heaviness can induce predicted correspondences with other feature dimensions. It was demonstrated that heavier objects were consistently aligned with *dark* and *low pitch*. Further confirmation for these mappings was sought through a series of speeded classification tasks. The heaviness-brightness mapping was demonstrated to influence response speed in a brightness classification task, where objects varying in heaviness were used as response keys. In both rating scale and speeded classification tasks, the heaviness-brightness correspondence continued to form the basis of cross-sensory interactions despite the potential for a size-brightness correspondence. The present work confirms that cross-activation between dimensions can be accessed through the manipulation of felt objects. What is more, support is provided for a framework of aligned feature dimensions, their conceptual nature and the inclusion of heaviness among this proposed network of dimensions.

**Acknowledgements**

I would like to thank the Economic and Social Research Council for funding my postgraduate training. I would also like to thank Lancaster University and the Department of Psychology for supporting me during my time as an undergraduate and postgraduate student.

I would particularly like to express my gratitude to Dr. Peter Walker, for his insights, support and patience; it's been a privilege, and a pleasure to work alongside you. I am also highly appreciative of the expertise and guidance I received from Prof. Brian Francis over the course of the project. I would like to acknowledge Dr. Laura Walker for allowing me to use the stimuli she developed, and whose work was, in many ways, a basis for understanding my own. I am grateful for the hard work of the technical and support staff in the department. Especially, I would like to thank Barrie Usherwood and Dave Gaskell for developing the objects and equipment used in the research and for always being willing to help when assistance was needed. A special thank you is extended to Michael Scallon, for reading through early written drafts of the thesis and for undertaking the momentous task of teaching me how to construct sentences. And especially, to all the volunteers who participated in the experiments, your essential contribution to the work is very much appreciated.

I would also like to acknowledge the support I have received from my fellow students, colleagues and friends. Those who saw me embark on this journey, and those who I have met along the way- you have made my time in Lancaster an extraordinary chapter full of many fond memories. And finally, to Mum, Dad, Catriona, Michael and Anthony, your encouragement, love and support has been invaluable. And especially to Simon, for your unwavering belief in me, I can't thank you enough. And finally, to Sophie, Jess, Heather Matthew, Aaron, and everyone from Lees Street AMDS, for your enthusiastic support and at times welcome distraction! I couldn't have done it without any of you- thank you all!

**Publication**

Walker, P., Scallon, G., & Francis, B., J. (2017). Cross-sensory correspondences: heaviness is dark and low pitch. Perception, 46 (7), 772-792

# List of Tables

# List of Figures

# Contents

# Chapter 1

## 1.1 Cross-Sensory Correspondences

Our perceptual experience is made up of information that is drawn together from various sensory channels. In some cases, the same information can be received by different channels, for example we can experience the shape or size of an object through touch and through vision. However in other cases, these channels capture aspects of our environment which are quite different: auditory pitch is a very different sensory feature to visual brightness. However, despite their obvious differences, these quite distinct sensory features seem to be more closely related than on first inspection. It has been demonstrated that people tend to align relative extremes of various feature dimensions in a particular way. Evidence of these systematic associations, termed cross-sensory correspondences, has emerged through several independent research areas including psychophysics, language, and study of the condition synaesthesia. It has now become an area of research in its own right, and is the topic of interest in the present thesis.

### 1.1.1Cross-Sensory Correspondences in Language

In language, the interrelatedness of different types of sensory information can be observed in our use of metaphors such as the sharpness of a taste or the heaviness of a scent. In many different languages, the labels used to mean high or low pitch are words which also refer to other sensory feature dimensions. For example, in English the same verbal labels are used to mean high and low in spatial elevation as well as high and low pitch. Speakers of Farsi use thick (for low pitch) and thin (for high pitch); in Kpelle the word heavy (for low pitch) and light (for high pitch) are used; and speakers of Norwegian use dark (for low pitch) and bright (for high pitch) (Eitan and Timmers, 2010). Eitan and Timmers (2010) explored whether there is universal agreement about how the extremes of different feature dimensions are assigned to extremes of auditory pitch in various languages. To test this, Western participants (familiar with a pitch-spatial elevation mapping) were asked to assign extremes of 29 different dimensions, used in non-Western or historical languages to mean high and low pitch, to describe segments of music varying in pitch register. These labels included sensory feature dimensions such as light/heavy, big/small, and sharp/blunt, as well as more complex ideas such as alert/sleepy, happy/sad, and feminine/masculine. Substantial agreement was found in the way extremes were mapped onto high and low pitch. High pitch being rated by the majority of participants as, among other things, active (vs. passive), fast (vs. slow), light (vs. heavy), bright (vs. dark), small (vs. big), sharp (vs. blunt and heavy), alert (vs. sleepy), happy (vs. sad) and feminine (vs. masculine).

### 1.1.2 Cross-Sensory Correspondences in Synaesthesia

Systematic mappings between various senses have also been found in the perceptual experience of people with the condition synaesthesia. For people with synaesthesia, perceptual experience in one sensory channel is induced automatically as a result of input from another. For example, in hearing-visual synaesthesia, a person may see shapes and colours in their visual field in response to hearing sounds such as music or speech (Day, 2005). Other examples include experiencing tastes in response to hearing words (Ward and Simner, 2003) or experiencing various visual arrays in response to smells (Cytowic, 1993). Typically, the specific mapping of one particular sensory feature onto another is idiosyncratic in nature, varying from person to person. However, systematic patterns have been observed. For example, there is a tendency for higher pitch sounds to induce visual images that are brighter, smaller and pointier than low pitch sounds (Marks, 1974; 1975; 1978; Marks, Hammeal and Bornstein, 1987; Ward, Huckstep and Tsakanikos, 2006).

Interestingly, the same cross-sensory patterns found in synaesthesia are consistent with mappings made by people who do not have synaesthesia (Ward, Huckstep and Tsakanikos, 2006; Karwoski, Odbert and Osgood, 1942). Karwoski et al (1942) asked people with and without synaesthesia to draw what they experience visually (or what might be suggested) when listening to different pieces of clarinet music. Both groups exhibited consistencies in the visual images they produced. For example, a rising trill followed by a descending trill was represented with, among other things, lines that ascended and descended in visual space; a dark colour which became bright and then dark again; or a coil of rope that started off thick, became thinner and then thicker again.

### 1.1.3 Cross-Modal Matching

The systematic associations that people make between various feature dimensions have become a topic of interest in their own right. Early work focused predominantly on associations between audio-visual feature dimensions, such as how sounds varying in loudness or in pitch map onto visual stimuli varying in luminance, brightness[1] and hue with findings indicating that louder and lower pitch sounds are matched to stimuli that are dimmer and darker visual stimuli (Bond and Stevens, 1969; Marks, 1974; Root and Ross, 1965; Stevens and Marks, 1965; Wicker, 1968). The mapping between brightness and pitch is consistent with the alignment experienced by people with hearing-colour synaesthesia (Karwoski, et al., 1942) and has been demonstrated to arise with auditory stimuli varying in degree of complexity. For example, segments of classical music that varied in pitch register and pairs of tones ascending or descending in pitch (Collier and Hubbard, 2001; Hubbard, 1996; Karwoski, et al., 1942).

[1] “Brightness” has been used by different sources as a term to refer to both luminance i.e. the intensity of light being emitted (vs. dim), and achromatic colour (vs. dark in shade). In the present thesis, “brightness” will be used to refer to variation in achromatic colour and “luminance” is used for intensity of emitted light.

Other feature dimensions have also been observed to enter into systematic associations. Size has been demonstrated to correspond with both pitch and brightness such that people are consistent in matching *bigger* with *darker* and *lower in pitch* (P.Walker and Smith 1985; P. Walker and Walker, 2012; L. Walker, Walker and Francis, 2012). Similarly, adults expect pointier shapes to make a higher pitch sound (e.g. if they were struck by another object or if they were to come to life and make a sound) compared to more rounded shapes (P.Walker, et al., 2010; P. Walker, 2012). Other consistencies relating to pointiness have been observed in sound symbolism research where it has been demonstrated that people show consensus in assigning nonsense words to visual stimuli varying in characteristics such as size and pointiness. For example, Köhler (1929) demonstrated that when people are asked to label a pointy and a rounded shape with two nonsense words, people are more likely to assign the word “baluma” to a more rounded shape and “takete” to a pointed/angular shape, rather than vice versa. This finding has subsequently been replicated with the words “bouba” and “kiki” (Ramachandran and Hubbard, 2001; 2003) and named the bouba/kiki effect.

In a comprehensive exploration of the associations people make between feature dimensions, L.Walker, et al. (2012) asked participants to rate stimuli which contrasted in either size (perceived visually or through touch), pointiness, brightness and pitch on a set of twelve rating scales which captured the dimensions: bright-dark, small-big, fast-slow, light-heavy, high-low (in spatial elevation and in pitch), thin-thick, active-passive, sharp-blunt, shallow-deep, weak-strong and hard-soft. It was found that irrespective of which feature dimension was being rated, the same assignment of extremes to stimuli varying in brightness, pitch, size and pointiness emerged (see Figure 1.1).

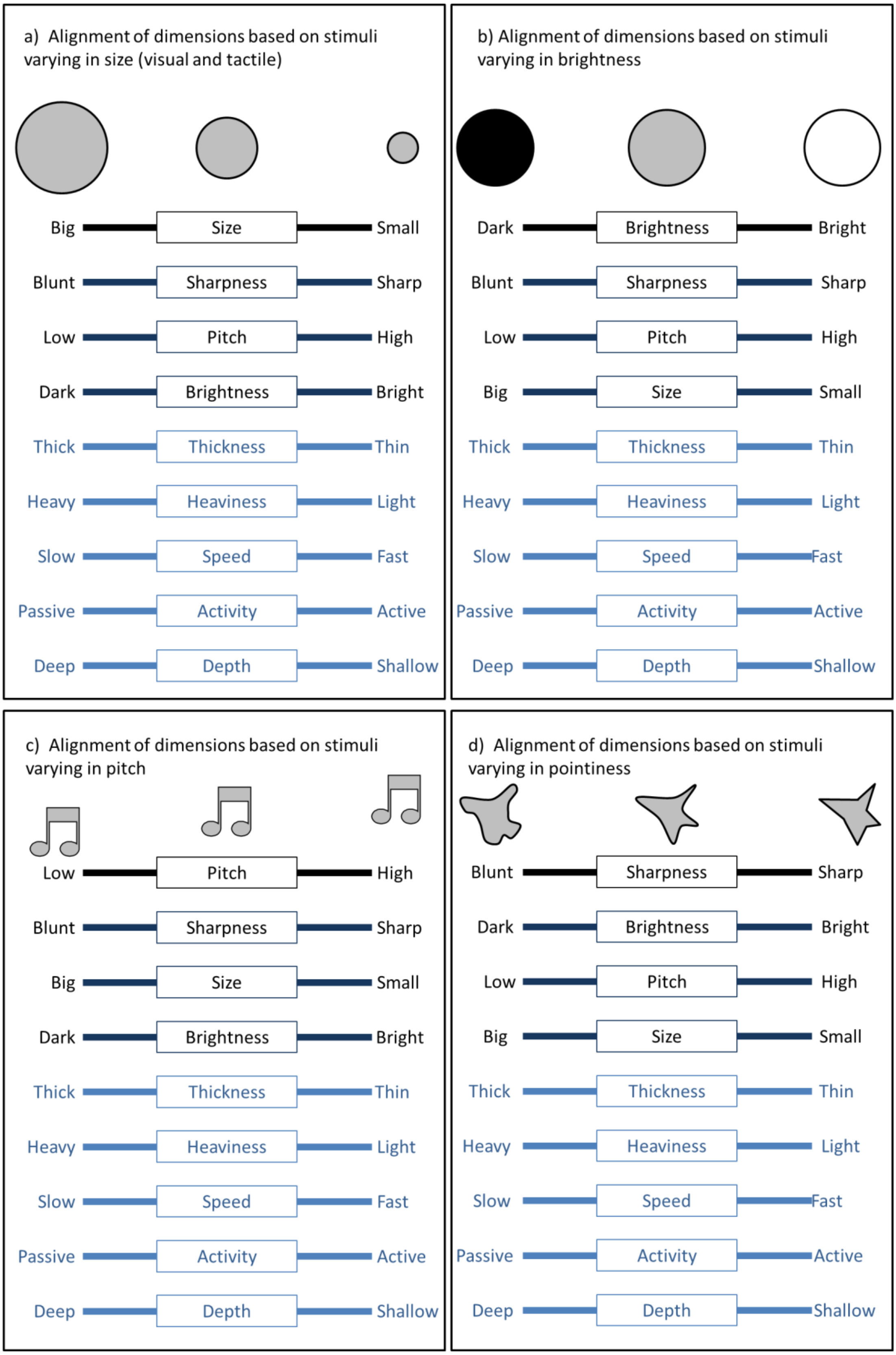

a) Alignment of dimensions based on stimuli varying in size (visual and tactile)
Big Size Small
Blunt Sharpness Sharp
Low Pitch High
Dark Brightness Bright
Thick Thickness Thin
Heavy Heaviness Light
Slow Speed Fast
Passive Activity Active
Deep Depth Shallow
b) Alignment of dimensions based on stimuli varying in brightness
Dark Brightness Bright
Blunt Sharpness Sharp
Low Pitch High
Big Size Small
Thick Thickness Thin
Heavy Heaviness Light
Slow Speed Fast
Passive Activity Active
Deep Depth Shallow
c) Alignment of dimensions based on stimuli varying in pitch
Low Pitch High
Blunt Sharpness Sharp
Big Size Small
Dark Brightness Bright
Thick Thickness Thin
Heavy Heaviness Light
Slow Speed Fast
Passive Activity Active
Deep Depth Shallow
d) Alignment of dimensions based on stimuli varying in pointiness
Blunt Sharpness Sharp
Dark Brightness Bright
Low Pitch High
Big Size Small
Thick Thickness Thin
Heavy Heaviness Light
Slow Speed Fast
Passive Activity Active
Deep Depth Shallow

*Figure 1.1.* The alignment of feature dimensions observed by L. Walker, Walker and Francis (2012) when presented with stimuli varying in a) size b) brightness c) pitch and d) pointiness. Dimensions in black were presented to participants as 3 stimuli, and also appeared as rating scale dimensions when other stimuli were being rated (dark blue). The light blue dimensions represent the other features asked about. The same systematic alignment of dimensions was observed irrespective of which dimension was being presented. Note the dimensions: hard-soft, weak-strong and high-low in spatial elevation were rated but have not been included as they did not correspond with all feature dimensions (in the case of spatial elevation, this did not correspond with other dimensions as it was controlled for in the stimuli presented).

### 1.1.4 Cross-Sensory Correspondences and Information Processing

The same pattern of correspondences has been demonstrated to influence performance in tasks where explicit matching between feature dimensions is not required.  For example, the pitch of a sound can influence people's ability to find a visual target varying in brightness in accordance with a pitch-brightness correspondence (Klapetek, Ngo and Spence, 2012). In addition, the Implicit Association Task (IAT) has been used to confirm correspondences including between pitch and size, and associations from sound symbolism research including between pointy and rounded stimuli with words *takete* and *maluma* (Parise and Spence, 2012).The task most widely used to explore whether correspondences can influence information processing is the speeded-classification task. In this task, participants are required to make speeded judgements about stimuli varying along one feature dimension, such as whether visual stimuli are "bright" or "dark", while receiving information about another feature varying orthogonally/independently with the first one that is incidental to the task demands (i.e. does not require attending to). Although this second feature is, strictly speaking, irrelevant to the classification decision, it has been demonstrated that performance (speed and accuracy of responses) can be influenced by whether or not the two features presented happen to correspond with one another in accordance with the alignments demonstrated in conscious rating tasks.

Bernstein and Edelman (1971) used this paradigm to explore if the pitch of incidentally presented sounds interferes with the ability to categorise stimuli varying in vertical location. The demonstration that responses are faster when position in space is congruent with auditory pitch (a stimulus at a high position in space is presented with a high pitch sound and a stimulus at a low position in space is presented with a low pitch sound), as opposed to incongruent, has subsequently been replicated several times (Ben-Artzi and Marks, 1995; Chiou and Rich, 2012; Evans and Treisman, 2010; Lidji, Kolinsky, Lochy and Morais, 2007; Patching and Quinlan, 2002; Rusconi, Kwan, Giordano, Umilta and Butterworth, 2006). The speeded-classification task has also been used to demonstrate the same alignment of pitch with other feature dimensions including size (Evans and Treisman, 2010; Gallace and Spence, 2006), brightness (Melara, 1989; Martino and Marks, 1999; Marks, 1987) and pointiness (Marks, 1987).

#### *1.1.4.1 Levels of Processing*

The involuntary nature of congruity effects in the speeded classification task may suggest that they arise at lower stages of perceptual processing. For example, presenting a tone of a certain pitch may alter the threshold at which a simultaneously presented light appears brighter or darker, subsequently influencing the speed with which it can be classified. One attempt to isolate low level interactions in a speeded-classification task comes from Evans and Treisman (2010). They demonstrated that a congruency effect arises between pitch and spatial location, and pitch and spatial frequency (a grating pattern containing more or less lines) even when both of the feature dimensions were task irrelevant (the task required participants to identify the timbre of a sound or orientation of the grating pattern). In these cases, the size of the effect was demonstrated to be similar in the indirect task as in a direct task where one of the target feature dimensions was being classified. According to Evans and Treisman (2010) this suggests that higher cognitive levels of processing, such as attention, are not required for congruity effects to arise.

Chiou and Rich (2012) argue that the speeded-classification task cannot isolate the stage in processing that cross-sensory correspondences may influence performance. This is because the task involves response selection, and it is possible that feature congruity has an influence at this level, even in an indirect task. Chiou and Rich (2012) used a speeded-detection task, to determine if the congruity effect between vertical location and pitch arise at a point earlier than response selection- specifically, influencing attentional mechanisms. The speeded-detection task required participants to make the same response whenever a visual stimulus appears; removing any response selection/discriminatory elements. They found that presenting sounds varying in pitch induced shifts in attention to higher or lower position in space and as such argued that the association between these feature dimensions interferes with performance at an attentional level. The demonstration that this cueing effect is influenced by the context defining relative pitch values and that it can influenced by top-down control, places the congruity effect at late stages of 'voluntary attention orienting'.

Although it is difficult to isolate lower level processes in tasks such as the speeded-classification task; in some cases, congruity effects can only be explained by higher level processes. For example, Gallace and Spence (2006) demonstrated that the use of the spoken words "high" and "low" interfered with the classification of visual size, such that responses were faster when the word *high* was presented with smaller visual stimuli and *low* was presented with bigger visual stimuli, in the same way as occurs when the sounds themselves are used. P. Walker and Smith (1984; 1985) demonstrated that the size of response keys and/or the pitch of a sound influenced participants' ability to classify a range of words describing extremes of feature dimensions, such as dull vs bright and heavy vs light. The findings suggest that the congruity effect occurs after the linguistic processing of the features, and must be interacting at a conceptual/semantic level, not at a lower sensory level since no physical property was being presented (see also, Martino and Marks, 1999).

### 1.1.5 Perceptual Judgement Tasks

As well as influencing the speed in which classification decisions can be made, cross-sensory correspondences have been demonstrated to influence people's perceptual judgements (Maeda, Kanai and Shimojo, 2004; Parise and Spence, 2009; Bien, Oever, Goebel and Sack, 2012). Maeda, et al. (2004) demonstrated that visual grating patterns moving ambiguously were more likely to be judged as ascending (or descending) when a simultaneously presented sound ascended (or descended) in pitch. This was only the case when the stimuli onsets were within 100ms of each other (see also Miller, Werner and Wapner, 1958). Interestingly, unlike with speeded-classification tasks, the presentation of the spoken word 'up' or 'down' did not have the same effect on motion judgements (Maeda, et al., 2004). Similarly, hand movements gesturing either upwards or downwards in direction have been demonstrated to bias how high or low pitch an observer will perceive a sung note. Notes that were accompanied with a hand gesture moving upwards were more likely to be considered as higher and notes accompanied with a hand gesture moving downwards were more likely to be considered as lower in pitch compared to tones of equivalent pitch presented beforehand (Connell, Cai and Holler, 2012).

In addition, cross-sensory congruence can influence judgements about amodal aspects of multi-sensory stimuli. For example, Bien, et al. (2012) found that when a circle is presented visually to the left or right of a central point, a participant's judgement of where a simultaneously presented sound is coming from, is biased by whether the relative size of the circle is consistent with the relative pitch of the sound (in accordance with the size-pitch mapping). Participants were more likely to judge both features as coming from the same position when they were congruent (a high pitch sound being presented with a smaller circle and a low pitch sound being presented with a larger circle) compared to when they were incongruent. Also, when presented with audio-visual stimuli that were discrepant in temporal onset or spatial location, participants' ability to discern whether auditory and visual information were separate in temporal onset or spatial location was more difficult when they were congruent versus incongruent in accordance with a pitch-size and with a pitch-shape correspondences (Parise and Spence, 2009).

#### *1.1.5.1 Cross-Sensory Correspondences as Cues for Sensory Integration*

In light of evidence to suggest that correspondences can influence our perceptual experience, Parise and Spence (2009) argue that this indicates that cross-sensory correspondences have a role in sensory-integration such that features that are congruent, in accordance with the alignment of correspondences, are more likely to be bound together. Parise (2015) suggests that a meaningful way of understanding cross-sensory correspondences is within the context of sensory cue integration. According to this account, perception is an "inference problem" (Helmholtz, 1909) in which available sensory information is combined with prior knowledge to arrive at a final estimate of our environment. In the case of cross-sensory correspondences, the congruence between extremes of various feature dimensions is argued to be a form of prior knowledge which has a subsequent influence of the final perceptual experience (Spence, 2011; Ernst, 2006).

#### *1.1.5.2 A Bayesian Approach*

A Bayesian approach has been used as a way to understand cross-sensory correspondences in terms of cue integration (Ernst, 2006; Parise and Spence, 2009). Within this approach, cross-sensory correspondences are represented as coupling priors: probability distributions which reflect the likelihood of a value along one feature dimension according to the value on another dimension. It is proposed that perception is a process of integrating estimates from these priors with estimates from the sensory input being received in a statistically optimal way. The stronger the coupling of two features (the less variability in the shared probability distribution) the more likely these features will be bound.

The findings described in Section 1.1.5 demonstrate how the integration of sensory information tends to shift perceptual judgement towards the relative extremes of the sensory feature values presented. However, in some cases the judgements shift away from the alignment of the presented extremes. For example, it has been observed that feature dimensions including size and brightness can influence the felt heaviness of lifted objects in the opposite direction to the features are thought to align in a phenomena known as weight-illusions. In the size weight illusion, for example, the larger of two equally weighted objects is felt to be lighter (contrary to expectation) than a smaller object (Murray, Ellis, Bandomir and Ross, 1999). Similarly, in the brightness weight illusion, the darker of two, otherwise identical, objects is expected to be heavier but felt to be lighter than the brighter object when lifted (P. Walker, Francis and Walker, 2010). In weight illusions the judged heaviness of objects is the opposite of what may have been anticipated by the feature dimension that is acting as a cue to heaviness (in these cases size and brightness). On the one hand, this does indicate that prior expectations about the relationships between feature dimensions, such as size and brightness with heaviness, do interact with sensory input to influence the resulting perceptual experience. However, since the final perceptual experience of heaviness is a contrast effect, the size-weight illusion (and by extension other weight illusion) have been termed “anti-Bayesian” (Ernst, 2009) and is somewhat problematic for explaining cross-sensory correspondences within a unified Bayesian framework.

## 1.2 Origins of Cross-Sensory Correspondences

### 1.2.1 Co-Occurrences in our Environment

The suggestion that correspondences have a role in sensory integration, including the use of the Bayesian approach, relies on an assumption that the observed alignment of feature dimensions reflects a relationship present in our external environment. If this were not the case, the integration of sensory input with prior knowledge including cross-sensory correspondences would offer no advantage in estimating aspects of our environment. Parise (2016) proposes that the relationships between different feature dimensions can be described along a continuum, pairs of features can be considered anywhere between redundant (where two sources provide the same information, for example haptic size and visual size) or completely uncorrelated (where neither feature can predict anything about the other, for example hue and hardness). Cross-sensory correspondences, therefore, lie somewhere in between, where one feature dimension does go some way to inform or generate a prediction about another feature, and therefore can be considered semi-redundant feature dimensions.

Evidence of this is drawn from examples of natural co-occurrences of various feature dimensions. For example, the pitch-size alignment can be observed in the animal kingdom, where the size of an animal can be determined based on the pitch of its vocalisation (Bee, Peril and Owen, 2000; Harrington, 1987). In addition, according to laws of resonance, larger objects make lower pitch sounds. It has been demonstrated that we are sensitive to this association since we are able to estimate the size of a falling object based on acoustic qualities of the impact sound such as pitch (Carello, Anderson and Kunkler-peck, 1998; Grassi, 2005). In addition, the mapping of pitch and position in space has also been demonstrated to be present in our environment. Parise, Knorre and Ernst (2013) took recordings from two microphones positioned at an upper and lower position on a person's head as they moved freely across many different types of settings (rural, urban, indoors, outdoors). A trend for high frequency sounds to originate from higher positions in space was observed. What is more, the filtering properties of the outer ear results in high pitch sounds being received at the top of the ear, and lower pitch sounds at the bottom. This indicates that we are sensitive to quite subtle statistical relationships between feature dimensions in our environment, and that these relationships are consciously available to us when considering the expected heaviness of objects.

Despite the occurrence of cross-sensory correspondences in our environment being a compelling explanation for their origin; some have warned against relying on the "just-so" type explanations of cross-sensory correspondences (Dolsheid et al, 2014). Firstly, because often one can find examples of experiences where the relationship between feature dimensions does not reflect the internalised correspondence between feature dimensions. For example, small animals that squeak such mice are associated with lower spatial height (which is contrary to the high pitch - high position in space association) and thunder is low pitch but is associated with a higher position in space. What is more, not all demonstrated correspondences have been observed as co-occurrences in our environment. In some cases it is quite unclear where such co-occurrences may arise. For example, the correspondences that brightness enters into with dimensions such as pitch and size. Although that does not necessarily mean they aren't present; more work exploring the correlations between dimensions in our environment is necessary. Often, a number of different potential sources of cross-sensory correspondences are acknowledged (Spence, 2011; Smith and Sera, 1992; Marks, 1978), including the role of language and underlying structures of sensory systems.

### 1.2.2 Language

As mentioned in Section 1.1.1, the same verbal labels are often used to refer to different feature dimensions in different languages. For example, the terms “high” and “low” used in English, and many other western languages, are used to describe extremes of pitch and position in space. It has been argued that exposure to these common labels may result in the emergence of some correspondences that have been observed (Spence, 2011). However, work with very young infants indicates that they are sensitive to some of the same mappings (Lewkowicz and Turkewitz, 1980; Mondloch and Maurer, 2004; Wagner, Winner, Chicchetti and Gardner, 1981; P.Walker et al., 2010). For example, children aged between 30-36 months selected a darker and/or larger bouncing ball as being responsible for making a low pitch sound and a smaller and/or brighter coloured ball as being responsible for making a high pitch sound (Mondloch and Maurer, 2004). Infants as young as 3- to 4-months old have been demonstrated to look longer at images changing in vertical location or changing from a pointy to a rounded shape when the accompanying sound changed in pitch in congruence with the visual changes (high pitch when in a high position in space or a pointier shape and low pitch when in a low position in space and more rounded) (P. Walker, Bremner, Spring, Mattock, Slater and Johnson, 2010). This suggests that the origins of cross-sensory correspondences may be non-verbal.

Even if correspondences precede language, this does not rule out the potential for language to subsequently influence the development of correspondences. Smith and Sera (1992) demonstrated that changes to children's representational organisation of feature dimensions coincided with the acquisition of words relating to different feature dimensions. For example, although infants at the age of 2 demonstrate a reliable size-brightness correspondence; the way in which children were found to assign extremes of size and brightness became much less systematic at the same time that children began to understand the words "dark" and "light". Dolsheid, Shayan, Majid and Casasanto (2013) asked participants to reproduce the pitch of a sound being heard, while accompanied by lines on a screen varying in height or thickness (features used as labels for pitch in Dutch and Farsi respectively). It was shown that the pitch of reproductions was influenced by the visual feature that was used to describe pitch in the participant's language. Dutch speakers reproduced tones in a higher pitch when presented with a line that was at a higher spatial position compared to the same tone when presented with a line at a lower spatial position. Similarly, Farsi speakers reproduced tones that were higher in pitch when accompanied by a thinner line compared to the same tone when accompanied with a thicker line. Interestingly, although the other dimension for each group of speakers (thickness for Dutch speakers and height for Farsi speakers) did not spontaneously generate the same effect on tone reproduction, Dutch speakers could be influenced by the thickness of lines accompanying tones after being trained to use thickness metaphors to describe sounds, but not when they were trained in the reverse mapping of thickness onto pitch. It was subsequently demonstrated that 4 month old Dutch infants were sensitive to both height-pitch and thickness-pitch mappings (Dolsheid, Hunnius, Cassanto and Majid, 2014). These findings suggest that although infants may be sensitive to a wide range of cross modal mappings preverbally, that language does have a role in strengthening certain cross-sensory mappings if they are reinforced by being present in the individual's language.

### 1.2.3 Underlying Structure of the Sensory System

It has been suggested that the underlying organisation of the sensory system may also form some basis for the emergence of cross-sensory correspondences (Marks, 1978; Spence 2011; Smith and Sera, 1992). For example, although primary sensory cortices in adults tend to be specialised to a specific sensory channel, there is evidence to suggest that early in infancy these brain regions are much less specialised. And that many more connections between different sensory regions exist, which are thought to typically be pruned or disinhibited during development (Maurer & Maurer, 1988; Spector and Maurer, 2008). Given this, the occurrence of synaesthesia, and also demonstrations of cross-sensory correspondences in adulthood are argued to be an aftereffect of these early connections (Rouw & Scholte, 2007; see Spector and Maurer, 2008). Evidence of pre-verbal infants showing sensitivity to several cross-sensory correspondences is in keeping with the idea that these may be supported by early structural connections between different sensory areas. And what is more, work by Dolsheid et al. (2014, 2013) described above provides evidence of a potential pruning/inhibition process, such that the persistence of some connections in adulthood may be dependent on which connections are supported by exposure (either in language or in our environment).

Other ways that neural structures could give rise to correspondences between different types of sensory information are through the use of shared systems to code amodal sensory features such as intensity, irrespective of the channel in which it is perceived (Spector and Maurer, 2009; Spence, 2011). For example Walsh (2003) demonstrated that particular regions are activated to respond to intensity, irrespective of the channel in which sensory input is provided. The three factors above are not mutually exclusive and are each likely to have some role in determining the emergence of cross-sensory correspondences in adults (e.g. Smith and Sera, 1992; Marks 1978; Martino and Marks, 1999).

### 1.2.4 Convergence of Different Sources

Spence (2011) suggests that different classes of correspondences emerge as a result of these different origins: statistical, linguistic and structural. Statistical correspondences emerge from learned co-occurrences in our environment and can be understood in terms of Bayesian Integration Theory (see Section 1.1.5.1). Linguistically mediated correspondences are defined as being features sharing common linguistic terms including vertical position in space and pitch which share the verbal labels ‘high’ and ‘low’ in English. And structural correspondences are defined as those which are based on either neural structures or are underpinned by a common amodal dimension for example magnitude based correspondences. It is argued that each of these classes of correspondences would possess different properties, for example statistical correspondences being more likely to be universal, and linguistically mediated correspondences having a later developmental onset and interfering at higher processing levels. Spence (2011) also argues that pairs of feature dimensions which fall outside of these three categories would be unlikely to enter into cross-sensory correspondences.

However, this system for categorising correspondences does not successfully account for all known correspondences, in two ways. Some correspondences such as that between pitch and spatial elevation may fall into more than one category. For example, there is evidence to support the idea that the correspondence between pitch and vertical location has a statistical basis as well sharing the same verbal label in English (Parise, et al., 2013; see Section 1.2.1). In addition, it is unclear which category other correspondences would fall under, for example, the pitch-brightness correspondences and the size-brightness correspondence (Spence, 2011; P. Walker and Walker, 2012).

P. Walker and Walker (2012) propose that correspondences with a semantic basis would form an additional potential class of correspondences. This is in keeping with the Semantic Coding Hypothesis (Martino and Marks, 1999) which attempts to reconcile the different potential sources of correspondences suggesting that they converge on a "shared, abstract, semantic representation" that is available to both linguistic and perceptual systems. Whether correspondences at this level form a separate category of correspondences or whether they are the result of convergence from correspondences emerging from sources remains difficult to determine. Nonetheless, it is correspondences emerging at this level which are of particular interest for the present thesis.

### 1.3 Alignment of Feature Dimensions at a Level of Connotative Meaning

P. Walker and colleagues provide a framework for thinking about correspondences arising at a semantic level based on work by Karwoski et al. (1942) who proposed that the demonstrated correspondence between feature dimensions "*appears to be the parallel alignment of two gradients in such a way that the appropriate extremes are related" (p.217).* According to this framework, cross-talk arises between abstract, amodal connotations of various feature dimensions. Support for this comes from evidence that demonstrates that cross-sensory correspondences rely on the relative positioning of stimuli on these dimensions and engage in cross-activation that is transitive and bi-directional(P. Walker, Walker and Francis, 2015; P. Walker, 2016).

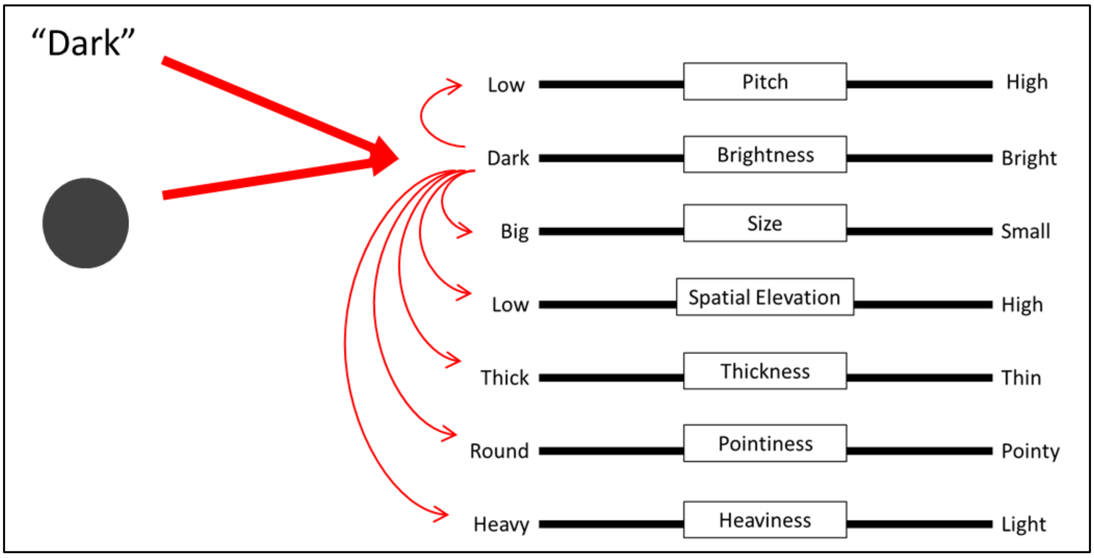



*Figure 1.2.* A depiction of the proposed cross-talk between aligned feature dimensions induced by activation of an amodal representation of *dark*

### 1.3.1 Relativity

If correspondences arise at a level which is amodal and conceptual in nature, it assumes that it is not the specific value along a feature dimension that an object possesses which results in cross-activation of other feature dimensions. Instead, it is the relative identity of that value which induces connotations of being at one or other pole of the particular feature dimension in question. For example, it is not the specific measurements of an object which gives it an identity as big or small, but the context within which its size has meaning as being either bigger or smaller and would subsequently give rise to cross-talk at this level.

As mentioned in Section 1.1.4.1, tasks where words are used as opposed to values of the physical property itself indicate that the absolute feature values are not necessary to induce correspondences as opposed to their relative identity (Gallace and Spence, 2006; Martino and Marks, 1999; P. Walker and Smith, 1984; 1985). Additionally, several studies have demonstrated that the relative context in which one value on a feature dimension is presented can change the nature of the association that same absolute value induces with other sensory features (Chiou and Rich, 2012; Gallace and Spence, 2006; Marks, 1987; L. Walker and Walker, 2016). Chiou & Rich (2012) demonstrated that when the same frequency of a sound was contextually the higher or lower tone in a speeded detection task, it differentially biased attention toward higher or lower position in space.

Similarly, L. Walker and Walker (2016) varied the relative context of response keys and brightness stimuli in a size-brightness speeded-classification task where participants were asked to respond to circles presented on a screen as being brighter or darker than the grey background. Responses were made by pressing one of two response keys which varied in size. Previously, this task induced a congruity effect such that responses were faster when the larger object was being used to respond to darker stimuli and the smaller object was being used to respond to brighter stimuli compared in the other way around (P. Walker and Walker, 2012). The relative brightness of visual stimuli were altered by changing the grey background upon which the stimuli appear such that two intermediate brightness values would be contextually “darker” in one condition and “brighter” in the other. A congruity effect between brightness and size was found to alter based on the relative context of the intermediate brightness value. Response times were faster when the larger object was used to respond to the intermediate value was it was contextually “darker” and when the smaller object was used to respond to the same intermediate brightness value when it was contextually “brighter”. In a second experiment, a middle sized ball (5cm diameter) was paired with either a smaller (2.5cm diameter) or larger (7.5cm diameter) ball to act as response keys in the task. It was found that the same sized object induced congruity effects with both darker classifications and brighter classifications depending on the object with which it was paired.

The demonstration of relative versus absolute mappings between values on different feature dimensions has implications for whether cross-sensory correspondences arise at a sensory level or at a higher cognitive level (Marks, 1987; Chiou and Rich, 2012; Gallace and Spence, 2006). If a relative context is required for a value of one feature dimension to correspond with another, it suggests that the association is happening after higher level interpretive processes, what L. Walker and Walker (2016) call a “post-categorical level”. In contrast, correspondences reflecting absolute mappings of dimension values are likely to arise during lower level perceptual processes and therefore should emerge irrespective of context (for example Guzman-Martinez, Ortega, Grabowecky, Mossbridge, and Suzuki, 2012; Lunghi and Alais, 2013;).

### 1.3.2 Transitivity

In logic, the principle of transitivity means that when there is a relationship between two elements and one of these elements shares the same type of relationship with a third element then it necessarily follows that the other of the initial elements also shares the same relationship with the third element (e.g. if A=B and B=C then A=C). Evidence shows that correspondences between different sensory features demonstrate transitivity. The pattern of relating between any pair of sensory feature dimensions does not contradict correspondences each share with any third feature dimension. For example, the existence of a size-brightness correspondence (P. Walker and Walker, 2012) where dark aligns with big and bright aligns with small is an illustration of this. This size-brightness correspondence was predicted given the previously demonstrated correspondences between pitch and size (where low aligns with big and high aligns with small), and pitch and brightness (where low aligns with dark and high aligns with bright). The size-brightness correspondence does not contradict the alignment that either feature also shares with pitch. P. Walker, et al. (2015) propose that a core set of feature dimensions are aligned such that cross-activation arises between them, Since, this is argued to occur at an amodal, conceptual level, the same pattern of associations should emerge irrespective of which feature dimensions is used to probe it. This is in part supported by the demonstration of transitivity which can be observed across a wide range of correspondences.

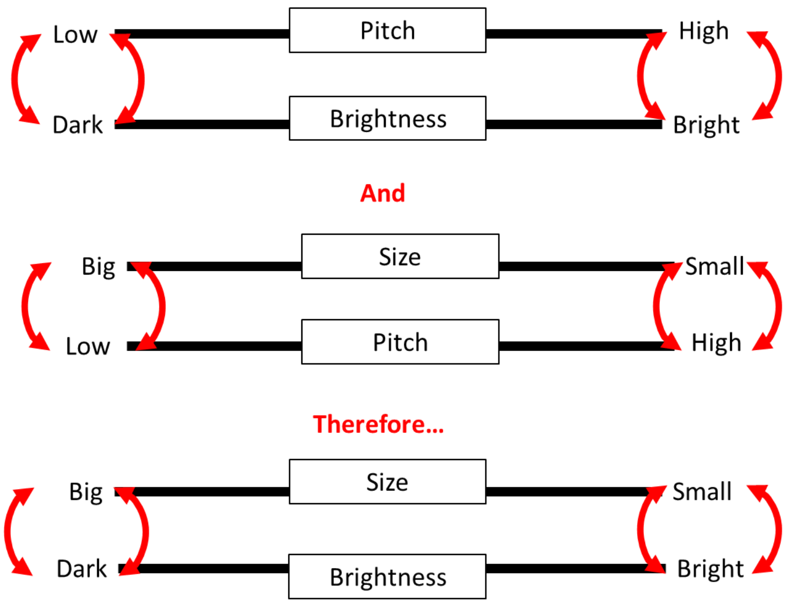



*Figure 1.3.* A depiction of the transitivity of correspondences involving brightness, pitch and size.

### 1.3.3 Bi-directionality

Bi-directionality refers to the demonstration that the cross-talk between any two feature dimensions arise irrespective of which feature induces it. For example, the correspondence between brightness and pitch has been consistently found, irrespective of whether participants are judging the brightness of sounds varying in pitch, or the pitch of colours varying in brightness (Hubbard, 1996; Marks, 1974). In the study by L.Walker, et al. (2012), stimuli varying in size (visually and through touch), pointiness, pitch, and brightness were all presented to participants as contrasting features to be rated. But in addition, they appeared among the set of feature dimensions that each set of stimuli were rated upon (see Figure 1.1). It was observed that the same alignment of this set of feature dimensions emerged irrespective of which feature was the one being presented in contrast and which was the dimension upon which the stimuli were being rated.

The implication of transitivity and bi-directionality is that the interference occurs at a level which is not modality-specific. This is in contrast to a modality-specific account which would describe an association as an aspect of the sensory mechanisms of the particular modality of the feature being probed. For example, Lupo and Barnett-Cowen (2015) argue that the visual system is able to form accurate predictions of an object's stability from its shape and material. This does not predict that notions or perceptions of stability would subsequently result in predictions about shape and material. However, in the case of many sensory feature dimensions such as brightness, pitch, and size, the bi-directionality of observed associations is argued to suggest that the cross-talk is not accessed through specific modalities. Instead it suggests that the cross-talk arises at an abstract semantic level which is available irrespective of modality.

According to P. Walker et al. (2015), transitivity and bi-directionality support the existence of a core set of correspondences arising at a level of connotative meaning. This is supported by work which demonstrates the same alignment of many conceptual attributes, including perceptual characteristics (Eitan and Timmers, 2010; Karwoski, et al., 1942; Osgood, 1960). This potentially means many feature dimensions can be considered to enter into correspondences. However there is a great deal of disparity in the amount of evidence implicating different feature dimensions as entering into correspondences. While some pairs of feature dimensions have been demonstrated to arise in a wide range of perceptual and cognitive tasks (for example, correspondences between pitch with spatial elevation, brightness, and size) other feature dimensions (such as heaviness, thickness and speed of motion) are implicated to a much lesser extent. This begs the question of which feature dimensions would be included in such an alignment.

Another uncertainty with this framework is to what extent implicated feature dimensions can be meaningfully distinguished from one another, as opposed to reflecting a shared underlying core concept. Some obvious examples where implicated dimensions may not capture meaningfully different dimensions are size and thickness (Dolsheid, et al., 2013). Similarly, size and heaviness may be considered as describing the same underlying concept of mass (for example, Eitan and Timmers, 2010 used size and weight seemingly interchangeably to describe a concept of mass). To summarise, despite the success of the framework in summarising and accounting for the evidence of cross-sensory interactions; it is clear that much more work is required to add further depth of understanding to this framework. In order to do so, it is necessary to take a closer look at the way in which specific feature dimensions implicated in this framework enter into correspondences. Of particular interest to the present thesis is heaviness.

### 1.4 Heaviness in the Correspondences Literature

Heaviness is a feature dimension which has been implicated to some degree in the correspondences literature. Alexander and Shansky (1976) demonstrated that colours that were darker in shade were judged to be heavier than those brighter in shade. The association of brightness and heaviness has been subsequently replicated on a number of occasions (L. Walker et al, 2012; P. Walker, Francis, et al., 2010). Other studies have demonstrated that feature dimensions systematically align with heaviness for example stimuli that are bigger, lower in pitch and more rounded tend to be rated as being heavier than their opposites (Eitan and Timmers, 2010; Karwoski, et al; P.Walker and Smith, 1984; 1985). Heaviness has also been included in a speeded-classification paradigm where participants were asked to press large and small sized objects in response to words reflecting opposite poles of feature dimensions. Responses were faster when a large object was used to respond to antonyms *STRONG*, *HEAVY*, *DOWN*, *BOTTOM* and the small object was used to respond to *WEAK*, *LIGHT*, *UP*, *TOP* compared to the other way round.

The way in which heaviness is aligned with size, brightness, pitch and roundedness in these examples is in keeping with the proposed alignment of feature dimensions. This could be interpreted as evidence that heaviness is a feature which can be included in this framework. However, there are two key limitations to the evidence demonstrating that heaviness belongs to this set of correspondences. Firstly, the concept of heaviness has only ever been represented verbally in these examples. Secondly, the majority of examples of correspondences between heaviness and other sensory features have asked people to judge the heaviness of another feature contrast. Therefore, it is unclear to what extent heaviness itself as a physical property would induce the same set of correspondences.

### 1.4.1 Heaviness as a Physical Property

It is argued that correspondences arising at a conceptual or semantic level are accessible via linguistic as well as perceptual exemplifications (Martino and Marks, 1999). This would predict that the same set of associations which emerge when heaviness is represented with verbal labels would emerge if heaviness were presented perceptually. Exploring whether correspondences between heaviness and other feature dimensions are induced by the lifting of objects is of particular interest, because of how weight perception is influenced by cross-sensory contexts.

Heaviness is defined as our perceptual experience of weight, weight being the effect of gravitational pull on an object as a function of mass. (Perceived) heaviness is influenced by a wide range of contextual factors, including other feature dimensions, such as the size and brightness of the lifted objects. These have been found to induce weight illusions where the heaviness of equally weighted objects is influenced by differences in size or brightness. Although lower level interactions between size and weight have been argued to have some role in the size-weight illusion; some research indicates that the size-weight illusion is the result of higher cognitive levels. For example, where the very same object is lifted while participants believe the object to be either larger or smaller, a difference in felt heaviness continues to emerge whereby the object was rated to be heavier when believed to be smaller and lighter when believed to be bigger (Buckingham and Goodale, 2010). In this case, as with the differences in heaviness observed in the brightness weight illusion cannot be explained by lower level interactions between the physical properties involved in lifting. Finally, evidence that the weight illusion persists despite the motor system adapting lift and grip forces to the appropriate level for the actual weights suggests some cognitive elements to the weight illusions (Flanagan, Bittner and Johansson, 2008).

Flanagan, et al. (2008) proposed that a cognitive mapping of size and weight has some role in heaviness perception and explaining size-weight illusions. This seems to parallel the mappings between size and heaviness which would be predicted in the correspondences literature and is consistent with the notion that these mappings may act as priors involved in generating expectations of our environment (in this case, the heaviness of an object). However the perceptual effect of these proposed mappings with heaviness are quite different compared to the way other cross-sensory correspondences have been shown to have perceptual influences. The resulting perception of heaviness is an exaggeration of the violation to expectation, as opposed to an assimilation of predicted and perceived feature values (see Section 1.1.5). This has caused heaviness to be considered “anti-Bayesian” (Ernst, 2009; Brayanov and Smith, 2010).

Given these interesting parallels, and differences between the weight-illusion and correspondences literatures, to explore whether cross-sensory correspondences arise when heaviness is presented physically through variation in weight may have interesting implications and further our understanding of heaviness perception in cross-sensory contexts. What would it mean for cognitive theories of weight illusions, if the proposed mappings of heaviness with dimensions including size, brightness etc. form part of a larger network of feature dimensions which are bi-directional, and transitive in nature? What does this mean for our understanding of heaviness as a conceptual dimension in relation to the felt weight of objects?

### 1.4.2 Bi-Directionality and Heaviness

If heaviness enters into the framework of correspondences proposed by P. Walker and colleagues, it would predict associations between heaviness and other implicated features which are bidirectional (see Section 1.3.3). That the same alignment can be found irrespective of which feature is used to probe it suggests that they are not modality specific associations but are more general and amodal in nature. This in turn suggests they arise at higher levels of processing. However, despite the theoretical importance of bi-directionality, all the examples of correspondence between heaviness and other sensory features have asked people to judge the heaviness of another feature contrast. Examples of this include the heaviness of colours (Alexander and Shansky, 1976) and the connotations of heaviness induced by music or other physical features (Eitan and Timmers, 2010; Karwoski et al., 1942; L. Walker et al., 2012). It is unknown whether objects contrasting in heaviness induce the same associations the other way round.

To emphasise the association in one direction is reasonable in the case of heaviness. Unlike many other feature dimensions, it is often perceived last, after information about brightness, size, or shape etc. has already been received. The tendency to receive the information in this order may result in a one-directional relationship: for example, an ability to anticipate heaviness from brightness, but not brightness from heaviness. Other mappings for example pitch and spatial location may exhibit bi-directionality since the order in which we experience these features may occur equally often in both directions for example hearing something before seeing it is as likely as seeing it before hearing it. Whereas, you will almost always know the size, shape and brightness of an object before lifting it (granted heaviness and pitch may not have this kind of directional relationship).

As mentioned in Section 1.1.5.1, a recent way of understanding correspondences is within the context of cue integration; this argues that cross-sensory correspondences have a role in anticipating events or ‘filling in the blanks’ of our perceptual experience. An assumption is that the correspondences are based to some degree in our experience co-occurrences between different feature dimensions in our environment. From this perspective, and given the asymmetry in how heaviness is experienced in relation to other feature dimensions; it seems less safe to assume that the associations heaviness enters into with other feature dimensions would necessarily be bi-directional. However, if mappings of cross-sensory correspondences, including the mapping between brightness and heaviness, are part of a larger network of cross-sensory correspondences arising at an amodal connotative level, it would predict an association between heaviness and these features which is bidirectional.

## 1.5 The Present Thesis

The aim of the present thesis is to examine if heaviness enters into correspondence with other feature dimensions such as pitch and brightness. Specifically, the aim is to determine if these associations can be *induced* by stimuli contrasting in heaviness, where heaviness is manipulated with lifted objects (as opposed to being represented verbally).

In the first part of the present thesis (Chapters 2-4) cross modal matching tasks are used to determine if the cross-sensory correspondences that have been demonstrated between heaviness and other feature dimensions can be induced by the felt heaviness of lifted objects. In Experiment 1 and 2, participants lift objects that are hidden from view and rate them on scales referring to other feature dimensions including brightness, pitch, size, and pointiness. If correspondences are a result of cross-talk between aligned feature dimensions that are transitive and bi-directional, it would be anticipated that lifted objects varying in heaviness will induce the same associations that have been demonstrated where contrasts in other feature dimensions have been considered to vary in heaviness. On the other hand, the particular way that heaviness is experienced in relation to other feature dimensions within our interactions with objects may mean that it does not itself, when varied, give rise to these associations. In Experiments 3, 4, and 5, the objects varied in both size and weight, this allowed the separate influence of each of these dimensions on judgements of brightness and pitch to be explored more closely.

In the second part of the thesis, the speeded-classification task was used to explore if the same associations between heaviness with brightness and pitch can be induced when lifted objects are used in a task which does not require participants to make explicit judgements about how feature dimensions correspond. In Experiments 6-10, people used objects varying in size and/or weight to tap a touch sensitive surface in response to stimuli varying in brightness (Experiments 6-9) and pitch (Experiment 10). For Experiments 11-15, the objects were held but were not actively used to make a response. Instead, a micro-switch was attached to the objects being held in order to respond to brightness and pitch categories. In Experiment 15, an articulatory suppression task was included in order to begin to explore the potential role of common verbal labels to describe one end of the brightness and heaviness dimensions (the term ‘light’).

### 1.5.1 The relationship between size and heaviness

The close relationship between size and heaviness was particularly important to explore since 1) size has already been well established in the correspondences literature. And 2) size is strongly related to our concept of heaviness. If, as Parise (2016) argues, correspondences reflect the relationship between semi-redundant feature dimensions within our environment, perhaps an intrinsic relationship between size and heaviness within our environment may account for correspondences arising between heaviness and other feature dimensions. Therefore it is necessary to account for the potential confound of a strong size-heaviness relationship in our exploration of correspondences with heaviness.

In an attempt to isolate size and heaviness during both types of experimental design, a set of objects were created which varied in both size and weight independently of one another. This allowed the exploration of heaviness as induced by variation in weight as well as variation in size-weight combinations. A set of nine objects were produced of three different sizes (diameters of 3, 4, and 5 cm, and heights matching these diameters) crossed with three different weights (i.e., 44, 107, and 190 gm) (see Figure 1.4). These objects or subsets of them were used in all the experiments reported. The objects were made from thin-walled (approx. 1mm) aluminium tubing filled with evenly distributed fragments of lead and builder’s expanding foam. The ends of the cylinders were smoothed with a fine layer of epoxy resin, after which the cylinders were painted matt grey. The weights of the cylinders were manipulated by varying the proportion of lead and builder's foam from which they were formed. The target weights were selected in order to produce an object at each size with an equivalent density (2 gm/cm3).

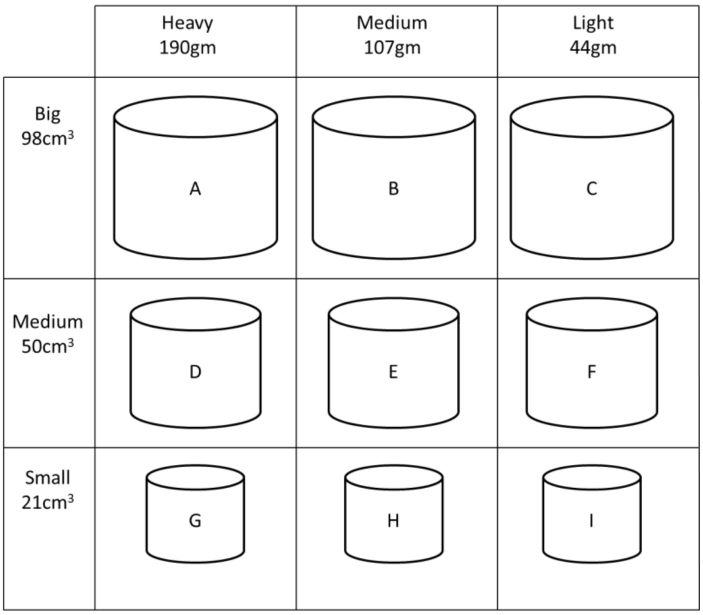



*Figure 1.4.* The set of nine objects used in the series of experiments reported in the present thesis.

# Chapter 2

## 2.1 Introduction

As described in Section 1.4, heaviness has been implicated to some degree in the correspondences literature. It has been demonstrated that lower pitch, larger, darker and more rounded stimuli are expected to be heavier than their opposites (Alexander and Shansky, 1976; Eitan and Timmers, 2010; Karwoski, et al., 1942; P. Walker and Smith, 1984; 1985; L. Walker, et al., 2012). Where associations between heaviness and other feature dimensions have been observed, the direction of the association has been in one direction: people judging the heaviness of stimuli contrasting in terms of another feature dimension. Therefore it is currently unknown whether the associations between heaviness and other feature dimensions are bi-directional in nature. That is to say, whether stimuli contrasting in heaviness, will induce the same pattern of associations as have been demonstrated the other way around.

Cross-sensory correspondences have often been argued to be bi-directional associations (Martino and Marks, 2001; L. Walker, et al., 2012; P. Walker, 2016). That means that the cross-activation between different feature dimensions can occur irrespective of which dimension is being presented. The implication of bi-directionality is that the correspondences between feature dimensions are arising at a level that is abstract and not modality-specific. It is therefore predicted that any feature dimensions which enter into a framework of correspondences at this level will demonstrate the same set of associations irrespective of which feature dimension is the one being presented in contrast and which dimensional associations are being induced. Therefore, in order to establish whether heaviness enters into correspondences at this level, it is necessary to determine if the same pattern of associations are induced by presenting variation in heaviness.

In almost all cases, the heaviness of an object is experienced after information about other feature information such as brightness, size, or shape has already been received. The tendency to receive the information in this order may result in a one directional relationship. We may be able to predict heaviness from features such as brightness, but not necessarily predict brightness from heaviness. We would very rarely have access to heaviness information in the absence of other feature information. In contrast, other mappings, such as that between pitch and spatial elevation may exhibit bi-directional relationships since the order in which we experience these features may occur equally often in both directions for example hearing something before seeing it is as likely as seeing it before hearing it. Might this peculiarity with heaviness mean it does not abide by the same bi-directional pattern of associations observed in other correspondences?

An example of a bidirectional association occurring despite asymmetric functional cuing is found between shape and speech sounds. It has often been shown that visual information about lip movements/shapes affects the perception of speech sounds (McGurk and MacDonald, 1976). The benefit of the cross-modal association in this direction is obvious as it enhances our ability to understand what someone is saying when auditory input is restricted. However Sweeny, Guzman-Martinez, Ortega, Grabowecky, & Suzuki (2012) asked whether the relationship between visual shape and auditory speech sounds is bidirectional. Specifically, exploring whether the presence of different speech sounds can influence judgements about the aspect ratio of ellipses perceived visually. If we were to consider this from a functional perspective, the relationship may seem less useful in this direction, we use lip-shape information to discern speech sounds, the shape itself does not contain the meaning. Nevertheless, it was demonstrated that the perception of a basic ellipse shape was indeed judged to vary as a result of simultaneously presented speech sounds. Ellipses were judged to be thinner and longer when presented with a /woo/ sound and wider and flatter when presented with a /wee/ sound. This suggests that the interaction between speech sounds and shapes reflects a more general association, as opposed to one which is specified by a one-directional relationship.

The aim of the experiments in the present chapter is to determine if the relationship between heaviness and other feature dimensions, implicated in the correspondences literature, are also bidirectional. This is to say, will the same pattern of associations demonstrated in previous work be induced by the felt heaviness of lifted objects? Participants were presented with objects varying in weight and asked to rate them on a set of scales representing size, brightness, pitch and pointiness. In order to isolate heaviness from other feature dimensions, the objects were hidden from view and lifted only by strings. In Experiment 1, the rating scales capture each feature dimension with verbal labels for each extreme as anchors on numerical scales, for example “big” and “small”. In Experiment 2, the scales represented each feature dimension non-verbally, as images varying in the specific feature or as sounds varying in pitch. If, according to P. Walker (2016), these associations are bi-directional in nature, then it is expected that heaviness will induce the same pattern of associations with size, brightness, pitch and pointiness that has been found in the opposite direction: the heavier objects will be judged to be bigger, darker, lower in pitch and more rounded than less heavy objects when presented with verbal labels to describe each dimension (Experiment 1) and when non-verbal representations of each dimension are used (Experiment 2).

## 2.2 Experiment 1

### 2.2.1 Method

#### *2.2.1.1 Participants*

Thirty students from Lancaster University (13 females, 17 males) between the ages of 19 and 55 (mean age = 27.6 years) volunteered to take part in the study for payment or course credit. All participants were right handed by self-report and spoke English as their first language.

#### *2.2.1.2 Apparatus and Design.*

***Objects.*** The three small objects (objects G, H and I) from the set of nine described in Section 1.5.1 were used (see Figure 1.4). The objects were attached to lengths of string (23cm) and positioned on the inside of a box (length 27cm x height 20cm x depth 15cm). The strings were threaded through small holes in the top of the box. The central object was positioned at the midpoint and the other two were 7cm either side of the centre. Participants could see the string but the objects were hidden from view (see Figure 2.1). A layer of sponge was secured to the interior base of the box to ensure that the objects did not make a sound when being placed down after lifting.

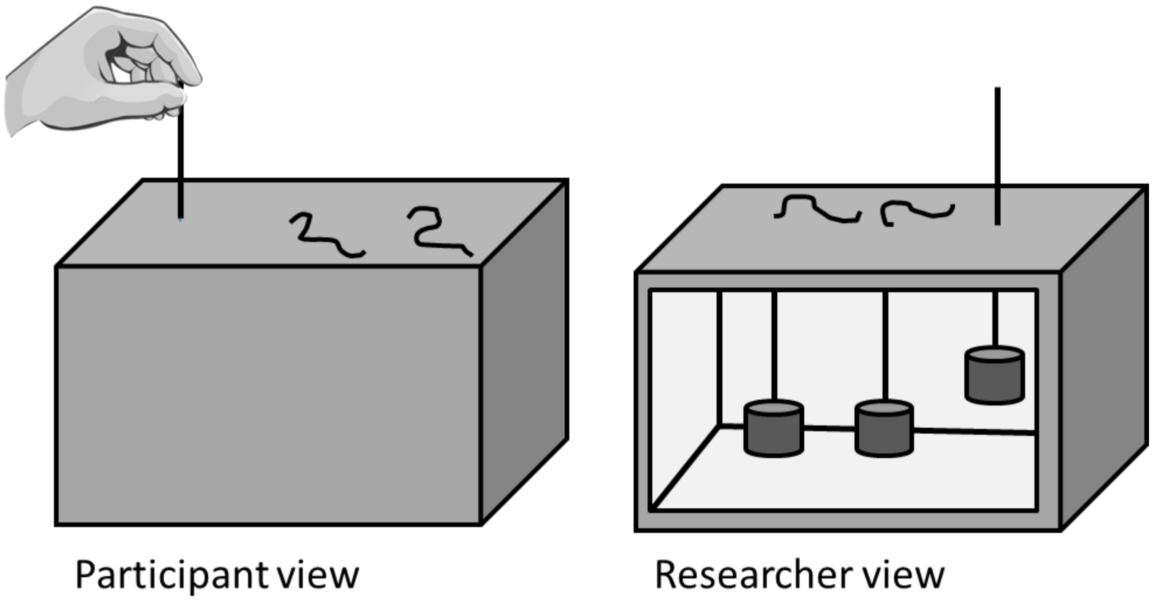



*Figure 2.1.* Representation of the apparatus used in Experiments 1 and 2.

***Scales.*** Four 6-point rating scales were used to elicit participants' judgements about the assumed size (small-big), brightness (bright-dark), pitch (high pitch-low pitch) and pointiness (pointy-rounded) of the objects. The scales ranged from 1 to 6 with anchor labels at each end point, of the form 'very + (extreme)', for example 'very small' to 'very big'. The Likert scales were presented together on a sheet of paper, a separate set of scales presented for each object. The question above each scale was worded as follows "how small or big is this object?" where the two antonyms were written in the same order in the question as presented in on the scale. The left-right positioning of the anchors on each scale was counterbalanced between participants. Counterbalancing was done independently for each question to ensure that extremes across different scales were not aligned in the same way for each participant. The order of the questions was randomised for each participant.

#### *2.2.1.3 Procedure*

Participants sat opposite the researcher, directly in front of the box that held the three objects. They were assured that there were no right or wrong answers to any of the questions, and instructed to use their initial impressions from lifting the objects to make their decisions. Before rating any object they explored all three in turn by lifting each one once. They were instructed to lift the string with the thumb and first finger of their dominant hand. After this initial familiarisation, they were asked to focus on one of the objects at a time and given a set of scales to complete for that object. They could lift the particular object being rated as often as they wished while they completed the scales, but not the others. Participants circled the value for that object on each scale. All participants started with the furthest left object, followed by the centre object then the furthest right object. There were six possible left to right combinations of the three objects, five participants were randomly assigned to each one.

### 2.2.2 Results

In all cases, the ratings on each scale were assumed to reflect a continuous variable (see Norman, 2010). Despite the labels assigned to each end of the scale being counterbalanced during the procedure, for the purposes of analysis scores were recoded such that very small, bright, high pitch, and pointy were indicated with a score of 1 and the opposites (very big, dark, low pitched, and rounded) with a score of 6. This is in accordance with the alignment each of these features has been demonstrated to share with one another (L. Walker, et al., 2012). Figure 2.2 summarises the mean ratings for the heavy, medium and light object on each of the scales.

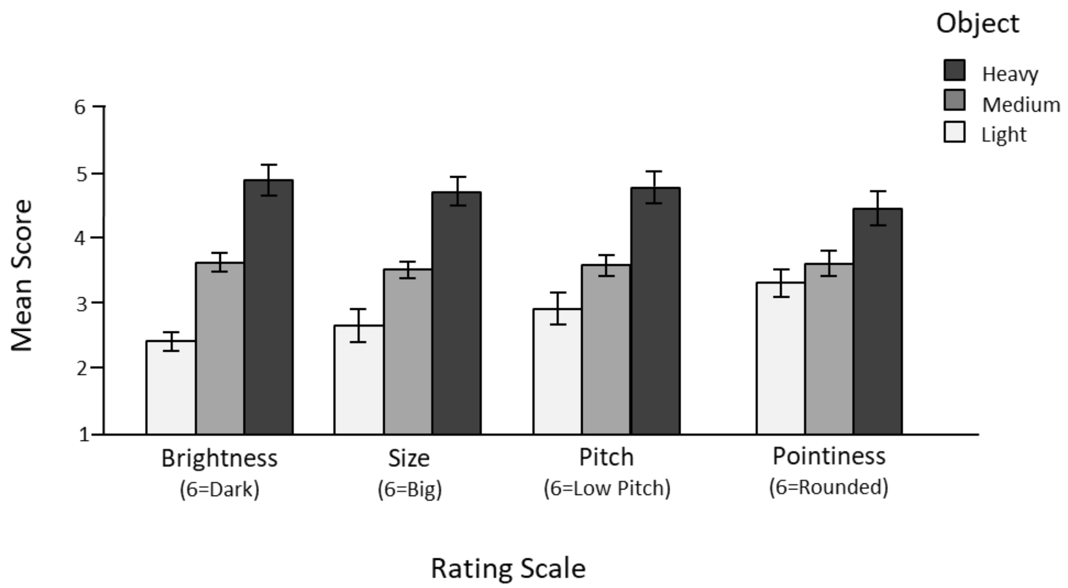



*Figure 2.2.* Mean ratings for the light, medium, and heavy objects on each feature scale. A rating of 6 signifies very -dark, -big, -low pitch, and –rounded. Error bars represent standard error.

Due to the repeated measures design, a linear mixed effects approach was used for analysis using R (R Core Team, 2012) version 3.2.0 and package lme4 (Bates, Maechler &

Bolker, 2014). The same model was fitted for each scale, this included heaviness as an explanatory variable and a random effect of participant. Heaviness was coded as a continuous variable (1=the lightest object, 2=the middle weighted object and 3=the heaviest object). A likelihood ratio-test was used to compare the model including heaviness as a main effect, with a null model that contained only a random effect of participant. 95% confidence intervals were calculated using the *Wald* method with the *confint()* function. Alkaike Information Criterion (AIC) values provided a relative estimate of the amount of information not being captured by a model, balancing goodness of fit with the number of parameters the model contains. A lower AIC value indicates a superior model. Visual inspection of residual Quantile-Quantile (Q-Q) plots did not reveal any departures from normality.

Table 2.1

*Parameter Estimates for Linear Mixed Effects Models Conducted on Each Rating Scale Including Results of Likelihood Ratio Tests for Comparison with Null Model.*

| | | Likelihood ratio test statistics | | |
|---|---|---|---|---|
| Rating scale | Parameter estimate for heaviness (95% CI) | $\chi^2$ | *p* | AIC Heaviness Model (AIC of Null Model) |
| Size | 1.03 [0.75, 1.32] | 39.42 | <.001 | 286.68 (324.10) |
| Brightness | 1.23 [0.99,1.48] | 63.67 | <.001 | 262.29 (323.95) |
| Pitch | 0.93 [0.64, 1.22] | 33.17 | <.001 | 287.24 (318.41) |
| Pointiness | 0.57 [0.27, 0.86] | 12.676 | <.001 | 295.37 (306.05) |

Note. Parameter estimates represent change in score with each step increase in weight. A higher score on each rating scale indicates a rating of bigger, darker, lower pitch and more rounded.

### 2.2.3 Discussion

The results show that as the heaviness of the objects increased, they were rated as larger, darker, lower in pitch and more rounded. This is the first indication that heaviness induces the same associations that have been previously been observed in the opposite direction (Alexander and Shansky, 1976; Eitan and Timmers, 2010; Karwoski, et al., 1942; P. Walker and Smith, 1984; 1985; L. Walker, et al., 2012). Therefore indicating that the cross sensory correspondences between heaviness and these feature dimensions are bi-directional in nature.

However, the use of verbal labels to represent each feature dimension makes it difficult to determine precisely what representations are invoked by variation in heaviness. For example, it is possible that the brightness dimension is interpreted as luminance rather than achromatic colour. This is especially the case for the pointiness dimension. In the present study it was assumed that the antonyms “pointy” and “rounded” reflect the bouba-kiki style dimension of roundedness and pointiness that is often used in the literature (e.g. Parise and Spence, 2012; L. Walker, et al., 2012). However, the antonyms rounded and pointy may be interpreted in several different ways.

Chen, Huang, Woods and Spence (2016) identified different aspects of pointiness: the frequency (number of points), amplitude (length of points) and what they describe as the spikiness (the extent to which edges are pointed or rounded). It is unclear which aspect of pointiness the antonyms rounded and pointy may refer to. It may be that some participants interpreted this scales in different ways. Therefore, in an attempt to clarify this uncertainty, Experiment 2 replicates Experiment 1, but with each scale anchored by non-verbal representations: visual images varying in brightness, size, the three aspects of pointiness, and sounds varying in pitch.

## 2.3 Experiment 2

### 2.3.1 Method

#### *2.3.1.1 Participants*

Thirty students from Lancaster University (24 female and 6 males) between the ages of 18 and 48 (mean age = 19.33 years) volunteered to take part in the study for payment or course credit. All participants except four were right handed by self- report. Twenty-one participants spoke English as a first language. The remaining nine participants spoke the following first languages: Chinese/Cantonese (n=4), Norwegian (n=1), Afrikaans (n=1), Russian (n=1), Italian (n=1) and Malay (n=1).

#### *2.3.1.2 Apparatus and Design*

The objects and apparatus were identical to Experiment 1 aside from the scales which, in this case, were physical representations of the sensory feature being probed and were presented to participants one at a time (as opposed to all together on a sheet). The order of presentation of the questions was randomised for each participant. The direction of the scales was counterbalanced between participants. Counterbalancing was done independently for each scale, so the extremes across different scales were not aligned in the same way for each participant.

***Scales.*** Six 6-point rating scales represented each feature dimensions with non-verbal exemplifications incrementally varied from one end point to the other.

*Pitch*. Pitch was represented with selected keys numbered from one to six on a Casio SA-47H5 Mini-Keys keyboard (the white keys from Middle C to the A above Middle C).

*Size.* The size scale was made up of six square outlines with rounded corners varying from $2cm^2$ to $4.5cm^2$ with the length and width of each shape increased by an increment of .5cm (Figure 2.3a).

*Brightness*. The brightness scale was made up of six squares (2cm) varying in achromatic brightness between white and black presented on a black and white checked background (Figure 2.3b).

*Pointiness*. Pointiness was represented in three ways: Firstly, *pointiness* was represented with traditional 'bouba/kiki' style shapes 3cm in length and 3cm in width with 5 points varying incrementally from completely pointed to rounded (Figure 2.3c). Secondly, the *number of points* was varied with star shapes (3cm in diameter) ranging from 8 to 32 points. Finally, variation in the *length of points* on a 16-pointed star shape 3.5cm in diameter with the length of points ranging from 0.5cm to 1.5cm.

***2.3.1.3 Procedure***

As in Experiment 1, participants were first asked to explore all three objects by lifting each one once. After this, they were asked to focus on one of the objects at a time. While lifting that object they were presented with a scale by the experimenter printed on a sheet of paper, or asked to use the keyboard to make the pitch judgement. They were instructed to select the number on the scale for the stimuli which best fitted their expectation of the object in question and the experimenter recorded each response. Every participant started with the furthest left object, followed by the centre object then the furthest right object. There were six possible left-right combinations of the three objects; five participants were randomly assigned to each one. They could lift the particular object being rated as often as they wished, while they completed the scales but not the others.

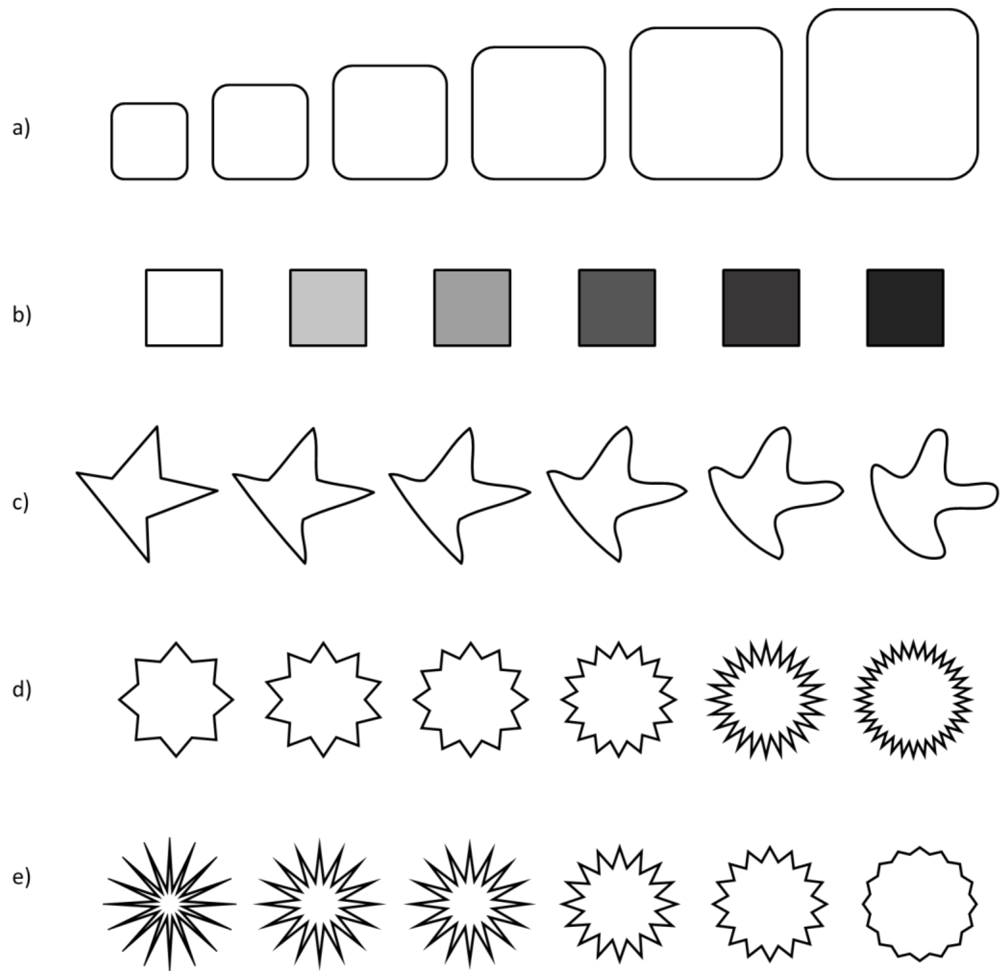



Figure 2.3. The non-verbal scales used in Experiment 2. a) size b) brightness c) pointiness d) number of points e) length of points. The images are reproduced to scale at 50% of actual size.

### 2.3.2 Results

The numbers 1-6 assigned to each end of the scale were counterbalanced during the procedure; for the purposes of analysis, scores were recoded such that a score of 1 marked the brightest, highest pitch, smallest, pointiness, most points and longest points ends of the scale and the opposites (dark, low pitch, big, rounded, fewer points and shorter points) were assigned a score of 6. Figure 2.4 summarises the mean ratings for the heavy, medium and light object on each of the scales.

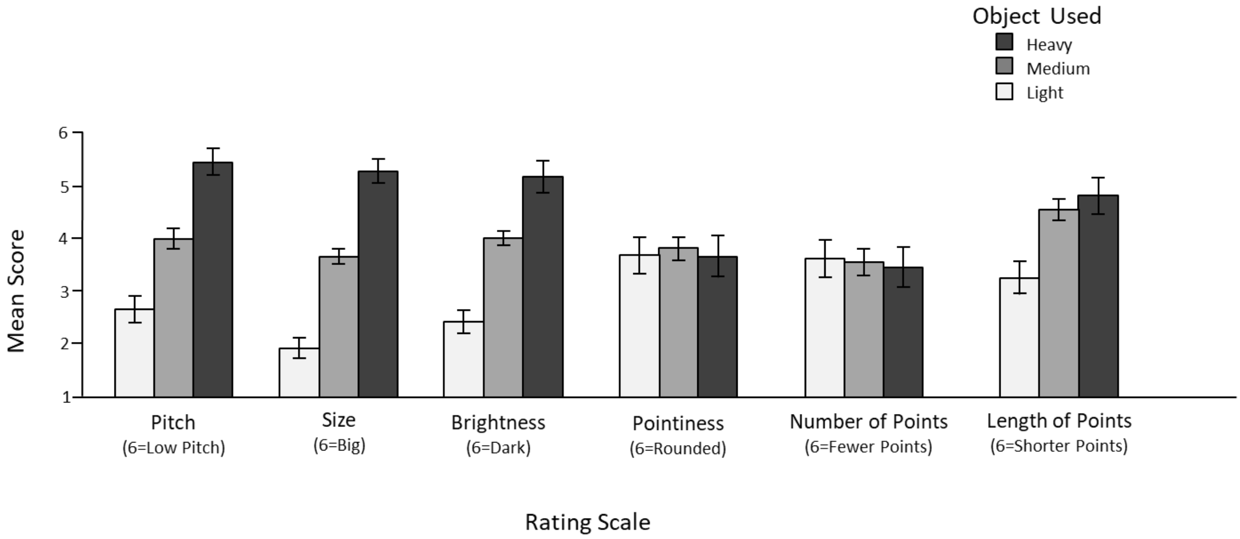



*Figure 2.4.* The mean ratings of the heavy, medium and light objects for each rating scale. A rating of 6 signifies low pitch, big, dark, rounded, fewer points, shorter points. Error bars represent standard error.

The analysis approach was the same as that used in Experiment 1. A model was fitted for each scale which included heaviness as an explanatory variable and a random effect of participant. Heaviness was coded as a continuous variable (1=the lightest object, 2=the middle weighted object and 3=the heaviest object). A likelihood ratio-test was used to compare the model that included heaviness as a main effect with a null model that contained only a random effect of participant. Table 2.2 summarises the analysis for each feature scale including 95% CI and AIC values for the null and comparison models. Visual inspection of Q-Q plots did not reveal any departures from normality.

Table 2.2

*Parameter Estimates for Linear Mixed Effects Models Conducted on Each Rating Scale Including Results of Likelihood Ratio Tests for Comparison with Null Model*

| Rating scale | Parameter estimate for heaviness (CI) | Likelihood ratio test statistics | | AIC Heaviness Model (null) |
|---|---|---|---|---|
| | | $\chi^2$ | p | |
| Size | 1.68 [1.42, 1.94] | 91.20 | <.001 | 270.04 (359.24) |
| Pitch | 1.40 [1.07,1.73] | 50.51 | <.001 | 313.04 (361.55) |
| Brightness | 1.38 [1.08, 1.69] | 57.60 | <..001 | 295.16 (350.76) |
| Length of Points | 0.78 [0.38, 1.18] | 13.654 | <.001 | 345.77 (357.42) |
| Number of Points | -0.083[-0.36, 0.52] | 0.14 | 0.71 | 366.65 (364.79) |
| Pointiness | -0.017 [-0.43, 0.46] | 0.005 | .94 | 365.42 (363.43) |

Note. Parameter estimates represent change in score with each step increase in weight. A higher score on each rating scale indicates a rating of lower pitch, bigger, darker, rounder, fewer points, shorter points.

**2.3.3 Discussion**

The findings demonstrate that heaviness induces associations with size, pitch and brightness in accordance with the alignment demonstrated in Experiment 1. Heavier objects were rated as bigger, darker and lower in pitch compared to the less heavy objects when each of these feature dimensions was represented by non-verbal stimuli varying incrementally between extremes. However, no association was found between pointiness and heaviness where pointiness was represented as variation between bouba/kiki style images. This finding is surprising considering the demonstration in Experiment 1 that heavier objects were rated as more rounded than lighter objects. And that images varying in a similar way have been demonstrated to induce associations with heaviness such that a rounded shape is rated as heavier than the pointier shapes (L. Walker, et al., 2012).

Of the three pointiness scales used in the present study, heaviness only induced systematic associations with *length of points*. This may suggest that the interpretation of the antonyms “pointy” and “rounded” in Experiment 1 was relating to this aspect of pointiness. This is consistent with Chen, et al. (2016) who found that western participants were more likely to assign labels ‘bouba’ and ‘kiki’ to stimuli on the basis of difference in the length of points, compared to other aspects of pointiness such as the roundedness of edges. Considering the majority of participants in the present study were western, this may explain why an association between heaviness and length of points was the only association found between heaviness and an aspect of pointiness. However, an alternative explanation for the association with this dimension could be that the overall surface area of the exemplars. In this scale, those with shorter points had a larger surface area than those with longer points. Thus, the association demonstrated by this particular aspect of pointiness may be explained in terms of variation in surface area (heavier objects being assigned a shape with a larger surface area than the less heavy object in accordance with a size-heaviness association).

## 2.4 General Discussion

The results from the two experiments in this chapter demonstrate that when presented with objects varying in heaviness, people expect a heavier object to be darker, bigger and lower in pitch than a less heavy object. This is the first indication that heaviness enters into bidirectional correspondences with these sensory features when taken in conjunction with previous findings demonstrating these associations in the opposite direction (Alexander and Shansky, 1976; Eitan and Timmers, 2010; Karwoski, et al., 1942; P. Walker and Smith, 1984; 1985; L. Walker, et al., 2012). As discussed above, the findings in the present chapter are less clear with regard to heaviness and pointiness. In Experiment 1, heavier objects were rated as more rounded than lighter object; a finding is in keeping with previous work (e.g. P. Walker and Walker, 2012) who found that more rounded stimuli were rated as heavier than pointier stimuli. However this was not replicated with the rating scales used in Experiment 2 where shapes varying from rounded to pointy edges were used to represent the pointiness dimension. The demonstrated of alignment in Experiment 1 where verbal labels were used does indicate an abstract association between the two dimensions. Further work is necessary in order to clarify if the stimuli used in Experiment 2 were not appropriate in some way capture a correspondence between pointiness and heaviness.

However, the finding that variation in heaviness induces the same systematic associations found previously with brightness, pitch and size supports the suggestion that heaviness enters into a framework of aligned feature dimensions proposed by L. Walker, et al. (2012). The demonstration that these associations can be induced by contrasts in heaviness adds further support to the claim that correspondences arising at this level enter into bi-directional associations and are amodal, and conceptual in nature. This is of particular interest when taking into consideration that we usually experience heaviness in relation to other object features in one direction: where expectations of heaviness will be induced by variation in size and brightness (less so for pitch) but rarely would we experience heaviness without this prior information. This may have predicted a one-directional correspondence between heaviness and features such as brightness and size, since heaviness would be unlikely to be used as a cue to this feature dimensions. Therefore, the demonstration of the same systematic associations induced by variation in heaviness suggests that the same associations can be accessed irrespective of which sensory dimension is used to probe it (P. Walker and Walker, 2012; P. Walker, 2016).

# Chapter 3

## 3.1 Introduction

Chapter 2 demonstrated that when people rated objects varying in heaviness in terms of other sensory features, a systematic pattern of associations arise. Heavier objects were expected to be darker, lower in pitch and bigger than less heavy objects. This is the first demonstration that heaviness induces associations with other sensory features and therefore enters into correspondences, with these feature dimensions, that are bi-directional in nature. The direction of (and transitivity between) these correspondences is consistent with the notion of cross-talk arising between a set of aligned feature dimensions as proposed by P. Walker and Walker (2012). The findings from Chapter 2 add further support for heaviness being included in this framework.

According to this framework (P. Walker, 2016; P. Walker and Walker 2012) a value on any feature dimension can cross-activate a corresponding position on another feature dimension with which it corresponds. However, although transitivity and bi-directionality may suggest that cross-talk arises between each pair of implicated feature dimensions, it is not the only way that the same observed transitivity may arise. There are several possible organisations which could underlie a network of interrelated feature dimensions where transitivity is observed. For example, a small number of feature dimensions may mediate correspondences observed between a wider range of features. For example, in the previous chapter where heavier objects were rated as darker than less heavy objects, we may conclude that heaviness directly corresponds with brightness. However, given that these features share correspondences with several other common feature dimensions, including pitch and size, can we really be certain that this is a direct correspondence? It may be that heaviness induces associations with another feature which subsequently influences brightness judgments.

Often, this challenge can be addressed when feature values, other than the dimension being manipulated are also available. For example, when participants are presented with stimuli varying in pointiness, features such as size and brightness are also available to participants, but are held constant. In which case, the potential confound of anticipated size or brightness has been ruled out. However, in the case of the heaviness correspondences demonstrated in the previous chapter, no other feature values were available to participants. Therefore, the potential that the correspondences observed were mediated by another feature dimension cannot be ruled out.

The present chapter attempts to address this challenge by investigating whether the associations of heaviness with brightness and pitch demonstrated in Chapter 2, were mediated by the expected size of the objects. That is to say, whether the heaviness of the objects produce an expectation of size, which subsequently drive judgements about brightness and pitch (see Figure 3.1). It must be acknowledged that any other corresponding feature dimension also had the potential to mediate the associations observed in Chapter 2. However, the influence of size is considered in particular for two reasons. Firstly, size and weight are two feature dimensions which are closely integrated in our environment. The size-weight illusion (Buckingham, et al., 2014; Flanagan, et al., 2008; Murray, et al., 1999) demonstrates the influence that size has on expected and perceived heaviness. Therefore, if heaviness is to produce associations with any feature dimension, the most likely would be size. What is more, it is well established that size enters into correspondences with many other feature dimensions in the correspondences literature including brightness, pitch and pointiness (Gallace and Spence, 2006; Mondloch and Maurer, 2004; L. Walker, et al., 2012; P. Walker and Smith, 1985). For this reason, it is perhaps the most likely of any to be a confounding influence on the ratings observed in Chapter 2. And teasing apart heaviness from size is necessary to determine if heaviness can be considered to enter into independent associations with other feature dimensions in this network.

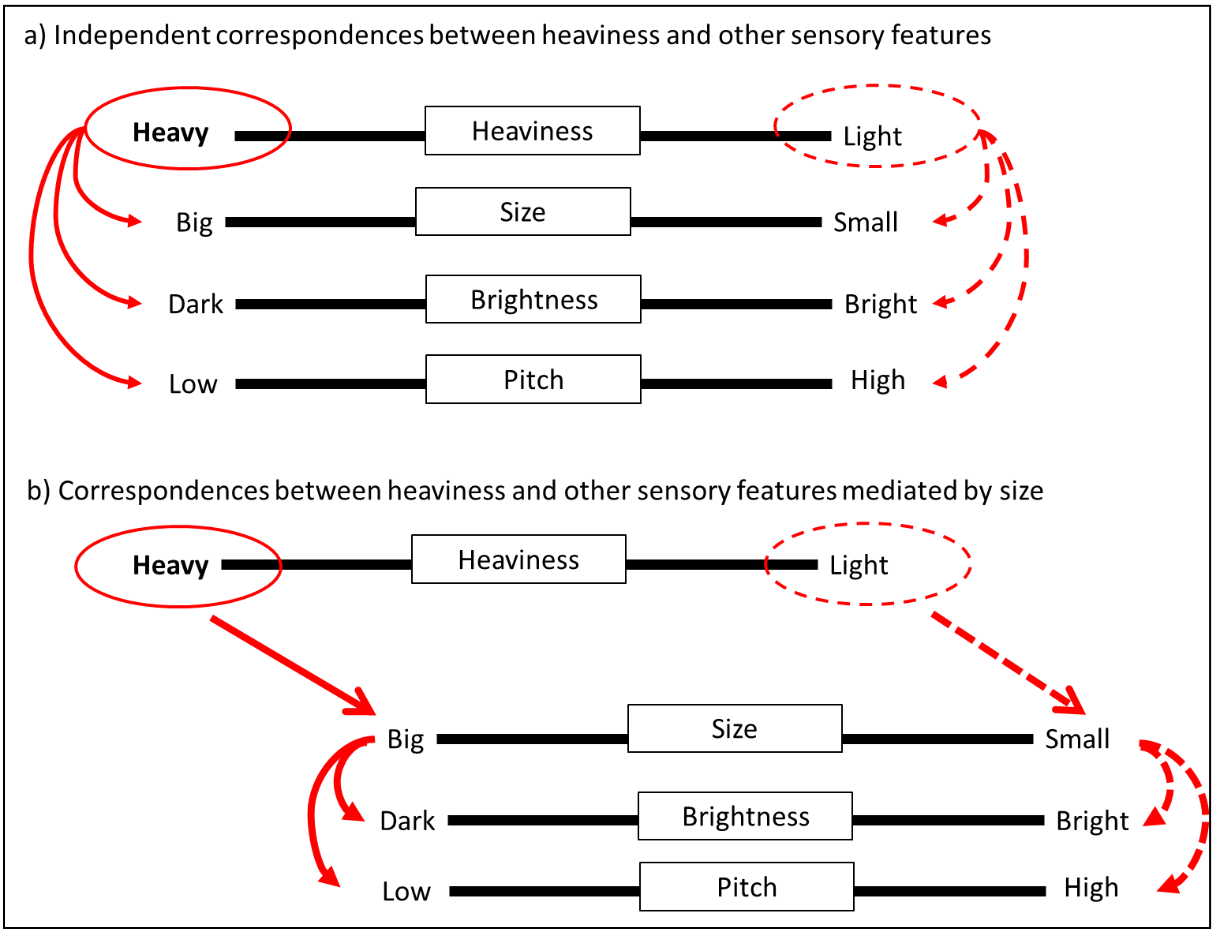



*Figure 3.1.* A depiction of two possible ways heaviness is associated with other feature dimensions. a) heaviness enters into independent correspondences with size, brightness and pitch b) correspondences implicating heaviness are mediated by size.

Therefore, the aim of Experiment 3 is to determine whether the heaviness of objects continues to induce associations with brightness and pitch[2] when the size of the objects is known, and remains constant. To explore this, a similar method is used as in Chapter 2, where objects varying in heaviness are lifted by string and rated on scales. However, in this case, the size of the objects is visible to participants as silhouettes on a translucent screen. If

[2] Unfortunately, size cannot be made available to participants without also revealing an object's shape. Therefore, judgements of pointiness could not be tested in the present experiments.

heaviness enters into correspondence with brightness and pitch independently of size, it is predicted that the same pattern of responses as Experiment 1 and 2 will be found. If the associations of heaviness with brightness and pitch were mediated by the presumed size of the objects, the ratings of the objects should no longer vary when the size of the objects is available but held constant.

## 3.2 Experiment 3

### 3.2.1 Method

#### *3.2.1.1 Participants*

Thirty-four students from Lancaster University (27 females and 6 males) between the ages of 18 and 40 (mean age = 21.5 years) volunteered to take part in the study for payment or course credit. All participants except one were right handed by self-report. Twenty-eight participants spoke English as their first language. The remaining six participants spoke the following first languages: Chinese (n=2), Italian (n=1), Polish (n=1) and Romanian (n=2).

#### *3.2.1.2 Apparatus and Design*

***Apparatus.*** As in Experiments 1 and 2, the three small objects from the set of nine (objects G, H and I, see Figure 1.4) were attached to lengths of string (27cm). The objects were placed on the inside of a box (width 30cm X height 20cm X depth 10cm) and the strings were threaded through small holes in the top. The central string was positioned at the midpoint of the top of the box and the two other strings were 6cm either side of the centre. A layer of sponge was secured to the interior base of the box for the objects to ensure they did not make a sound when being placed back down. A viewing window was cut out of the side facing the participant (26cm X 14cm) and covered with a layer of thick tracing paper. This acted as a screen for a silhouette of the object to be projected onto (see Figure 3.2). The objects were lit from behind with a single point LED light positioned 10cm from the objects and repositioned to be centred behind the object being lifted.

***Scales.*** To elicit judgements about brightness and pitch, the same 6-point rating scales were used as in Experiment 2. The scale representing brightness was presented on a laptop screen placed to the side of the box that housed the objects. Pitch was represented with same 6 keys on a Casio SA-47H5 mini-keys keyboard. On both scales, the numbering was counterbalanced between participants. Counterbalancing was done independently for both scales, so the extremes of each dimension were not aligned in the same way for each participant.

***3.2.1.3 Procedure***

Participants sat directly in front of the box that held the three objects, the room was dimly lit in order for the silhouettes to be clearly visible. They were told they had to rate each of the objects attached to string on a set of scales. Before rating any object they were instructed to explore all three objects, by lifting each one once. This was to give some initial context for the rating of the objects. For all lifts, participants were instructed to lift the object by the end of the string with the thumb and first finger of their dominant hand until it was fully visible on the screen. Next they were asked to focus on one of the objects at a time and given the brightness and pitch scales to rate that object. They could lift the particular object they were rating as often as they wished while they completed the scales, but not the others. Participants indicated where they would place the object on each scale and the experimenter wrote down the responses. Half of participants rated the object on the brightness scale first followed by pitch, for the other half this was reversed. Every participant started with the furthest left object, followed by the centre object then the furthest right object. There were six possible left-right orders for the three objects; six participants completed each one except one combination which was completed by 5 participants.

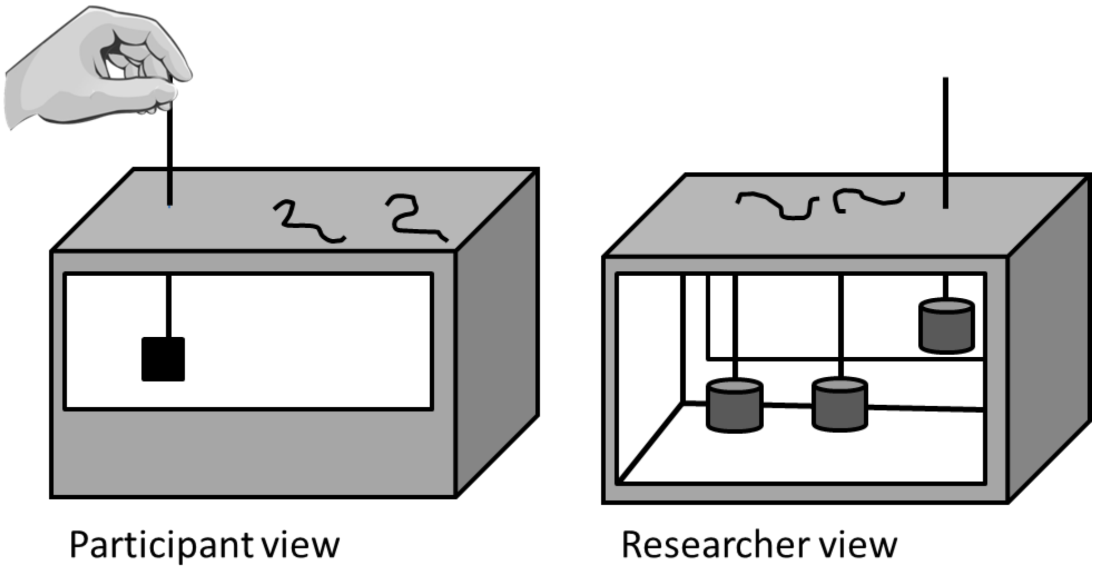



*Figure 3.2*. Representation of the apparatus used in Experiment 3

### 3.2.2 Results

The same analysis approach was used as in Experiments 1 and 2. Although the labels assigned to each end of the scale was counterbalanced, for analysis the scores were recoded such that very bright and high pitch were indicated with a score of 1 and the opposites (very dark and low pitched,) with a score of 6. Figure 3.3 summarises the mean ratings for the heavy medium and light object for each of brightness and pitch.

A model was fitted for brightness and for pitch which included heaviness as an explanatory variable and a random effect of participant. Heaviness was coded such that 1=the lightest object, 2=the middle weighted object and 3=the heaviest. A likelihood-ratio test was used to compare the model that included heaviness as an explanatory variable with a null model that contained only a random effect of participant. Visual inspection of Q-Q plots did not reveal any obvious departures from normality.

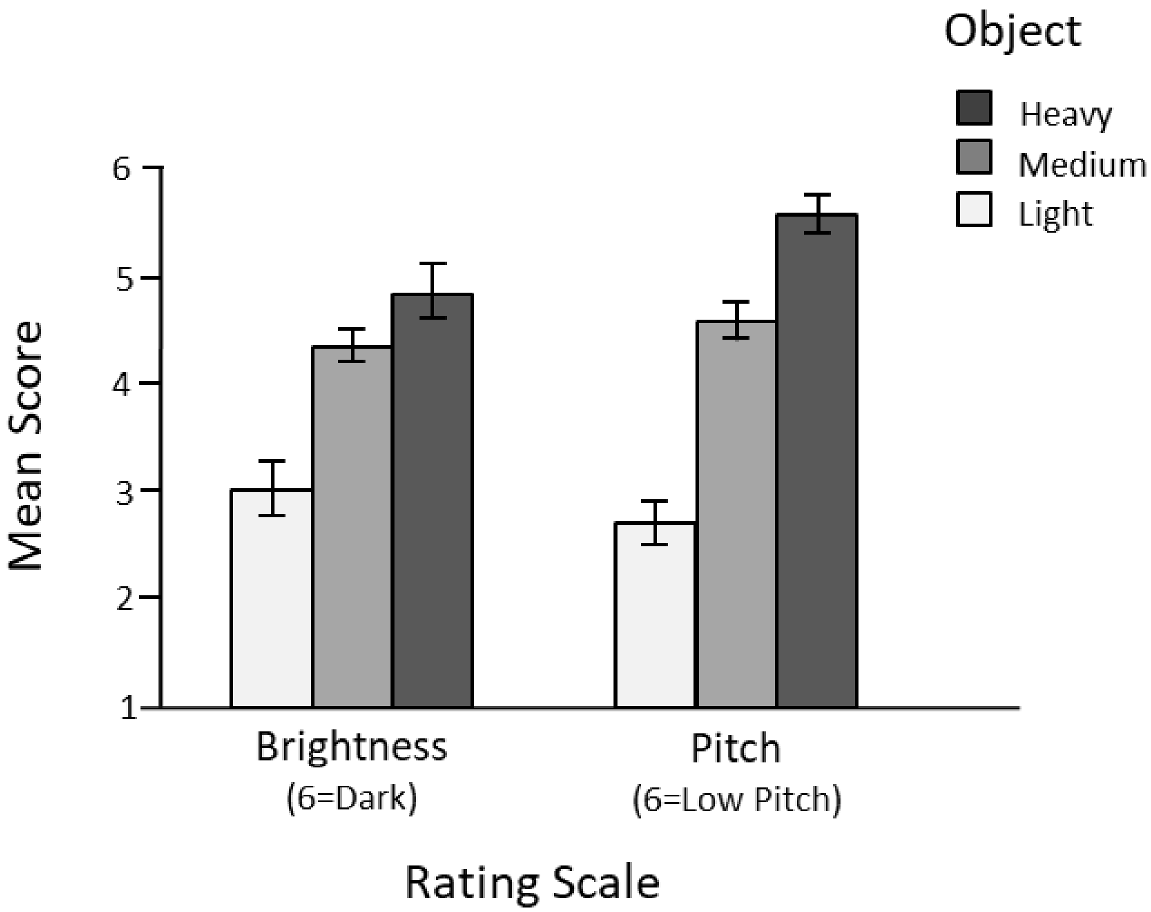



*Figure 3.3.* The mean brightness and pitch ratings for the heavy, medium and light objects. A rating of 6 signifies low pitch, and dark. Error bars represent standard error.

#### *3.2.2.1 Brightness*

The model including heaviness as a fixed effect was preferred to a null model ($\chi^2$=31.91, p<.001) with an AIC of 343.58 compared to the null model which had an AIC of 373.49. It is shown in the parameter estimates that a step increase in heaviness increases the brightness rating (i.e. is judged to be darker) by 0.87 points CI [0.60, 1.13]. Figure 3.3 shows the mean ratings for each object.

#### *3.2.2.2 Pitch*

The model including object as a fixed effect was preferred to a null model ($\chi^2$=108, p<.001) with an AIC of 284.20 compared to the null model which had an AIC of 388.19. It is shown in the parameter estimates that a step increase in heaviness increases the brightness rating (i.e. is judged to be lower pitch) by 1.47 points CI [1.23, 1.71]. Figure 3.3 shows the mean pitch ratings for each object.

### 3.2.3 Discussion

The findings demonstrate that heavier objects are rated as being darker and lower in pitch than less heavy objects, replicating the results of Chapter 2. These associations arose even when the size of the objects was known to remain constant, which suggests that the associations of heaviness with pitch and brightness are independent of the correspondence each one has been previously demonstrated to share with size (Gallace and Spence, 2006; L. Walker, et al., 2012; P. Walker and Walker, 2012). These findings suggest that despite the close relationship between size and heaviness, heaviness is a feature dimension which enters into the proposed network of aligned feature dimensions in its own right, independently of size.

## 3.3 Experiment 4

When people are presented with stimuli varying in size alone, the size of the objects will induce associations of brightness and pitch such that bigger objects are judged to be darker and lower in pitch than smaller objects (P. Walker and Walker, 2012; L. Walker, et al., 2012). Similarly, as has been demonstrated in Experiments 1-3, when objects vary in heaviness alone, the heaviness will induces associations of brightness and pitch such that heavier objects are darker and lower in pitch than less heavy objects. Often, the size and heaviness of a set of objects will be in keeping with this alignment (e.g. a bigger object is heavier than a smaller object), therefore both feature dimensions will be consistent in inducing the same cross-activation of other feature dimensions (e.g. brightness and pitch). However, it is less clear what happens when information from one sensory feature is contradictory to information of another. For example, in cases where a larger object is lighter in weight than a smaller object. How will the occurrence of contradictory feature dimensions influence judgements about features such as brightness and pitch?

This question was explored in Experiment 4. Participants were presented with objects varying in both size and heaviness, in a way which is inconsistent with how each feature is usually aligned with one another, as well as with the associations they induce with other feature dimensions. In this case, in order for associations relating to one dimension (e.g. size) to be in keeping with the systematic pattern of correspondences usually observed, the way that particular feature dimension usually corresponds with the other dimension (heaviness) must be violated.

There are several possible ways that this conflict between size and heaviness may influence judgements of brightness and pitch. Firstly, the two associations may counteract each other resulting in no clear or systematic pattern of ratings. Another possible outcome is that both features influence the response pattern, resulting in all objects being rated with central values on brightness and pitch dimensions. A third possibility is that one feature dimension becomes the basis of judgements about brightness and pitch and the other feature does not influence responses.

### 3.3.1 Method

#### *3.3.1.1 Participants*

Thirty students from Lancaster University (23 females, 7 males) between the ages of 18 and 24 (mean age =19.6 years) volunteered to take part in the study for payment or course credits. All participants except three were right handed by self-report. Twenty-two participants spoke English as their first language. The remaining eight participants spoke the following first languages: Arabic (n=1), Chinese (n=4), Greek (n=1), Malay (n=1) and Slovak (n=1).

#### *3.3.1.2 Apparatus, design and procedure*

The scales, design and procedure were the same as Experiment 3. However in this case, the objects varied in size in contradiction to the objects heaviness (Objects C, E and G in Figure 1.4). String was attached to the objects of a length required to make the total length (object + string) to equal 30cm.

### 3.3.2 Results

The same analysis approach was taken as Experiment 2. Although the labels assigned to each end of the scale was counterbalanced, for analysis the scores were re-coded such that very bright and high pitch were indicated with a score of 1 and the opposites (very dark and low pitched,) with a score of 6. Figure 3.4 summarises the mean ratings for the heavy/small, medium and light/big object on each of the scales.

A model was fitted for each scale which included *object* as an explanatory variable and a random effect of *participant*. *Object* was coded in accordance with heaviness such that 1=the lightest/biggest object, 2=the middle weighted /middle size object and 3=the heaviest/smallest object. A likelihood ratio-test was used to compare the model that included *object* as a main effect with a null model that contained only a random effect of *participant*. Visual inspection of Q-Q plots did not reveal any obvious departures from normality.

#### *3.3.2.1 Brightness*

The model including *object* as a fixed effect was preferred to a null model ($\chi^2$=33.35, p<.001) with an AIC of 280.76 compared to the null model which had an AIC of 312.12. It is shown in the parameter estimates that a step increase in heaviness/ decrease in size increases the brightness rating (i.e. is judged to be darker) by 0.9 points CI [0.63,1.17]. Figure 3.4 shows the mean ratings for each object.

#### *3.3.2.2 Pitch*

The model including object as a fixed effect was preferred to a null model ($\chi^2$=32.101, p<.001) with an AIC of 314.79 compared to the null model which had an AIC of 352.39. It is shown in the parameter estimates that a step increase in heaviness/ decrease in size increases the pitch rating (i.e. is judged to be lower pitch) by 1.07 points CI [0.73,1.40]. Figure 3.4 shows the mean pitch ratings for each object.

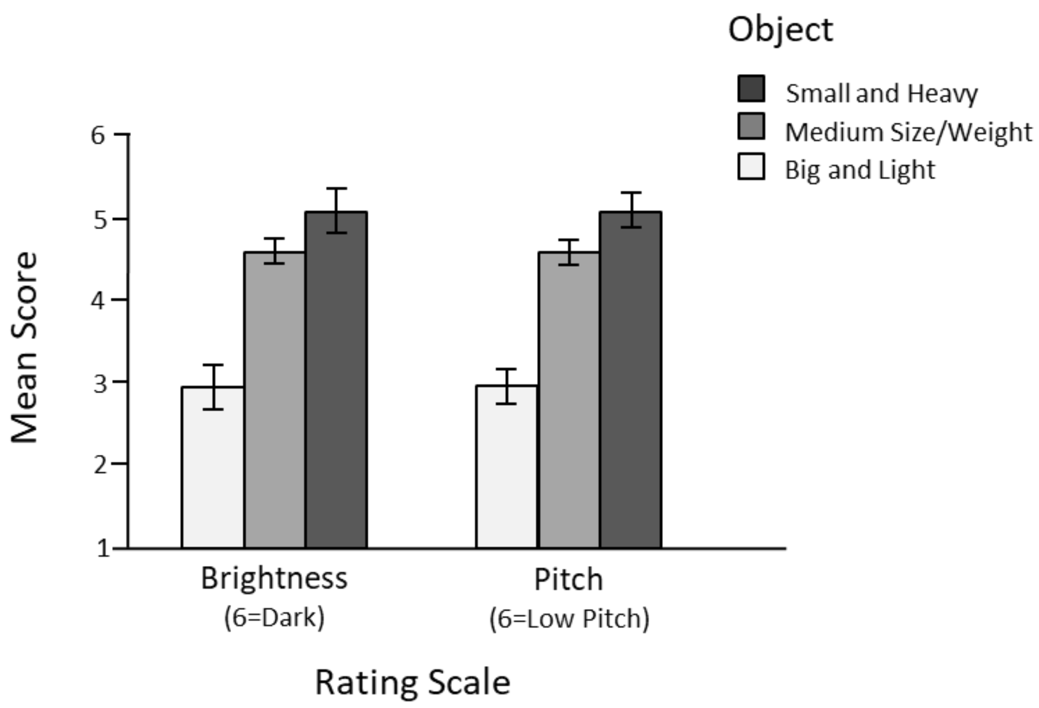



*Figure 3.4.* The mean brightness and pitch ratings of the small/heavy, medium and big/light objects. A rating of 6 signifies low pitch, and dark. Error bars represent standard error.

#### *3.3.2.3 Does size influence the ratings of brightness and pitch?*

The present findings suggest that the heaviness of the objects formed the basis of judgements of brightness and pitch, despite contradictory information about size. However, does the presence of a size difference have any influence on the ratings? To answer this question the brightness and pitch ratings in the present study and in Experiment 3 (where the size of the objects was held constant) were compared in a LME analysis including *experiment* and *object* as explanatory variables, and a random effect of participant. If variation in size had any influence on size ratings, this would result in an interaction between *experiment* and *object*. For both brightness and pitch, a likelihood ratio test compared a model which included an interaction between *experiment* and *object* with a model that did not. For brightness, there was no difference between a model that included an interaction (AIC=623.20) and a model that did not (AIC=621.23) ($\chi^2$ =0.028, p=0.87). Although for pitch, the difference was approaching significance such that the model including an interaction was marginally preferred (AIC=619.04) over the model that did not (AIC= 620.76) ($\chi^2$=3.73, p=0.054). Overall this suggests that addition of variation in the size of the objects in Experiment 4 had little bearing on brightness and pitch judgements compared to Experiment 3 where the objects were all of equal size (see Figure 3.5). There is a marginal difference in pitch judgements between Experiment 3 and 4 such that the pitch ratings were less extreme when there was contradictory size information available (Experiment 4); this effect approaches significance.

a)

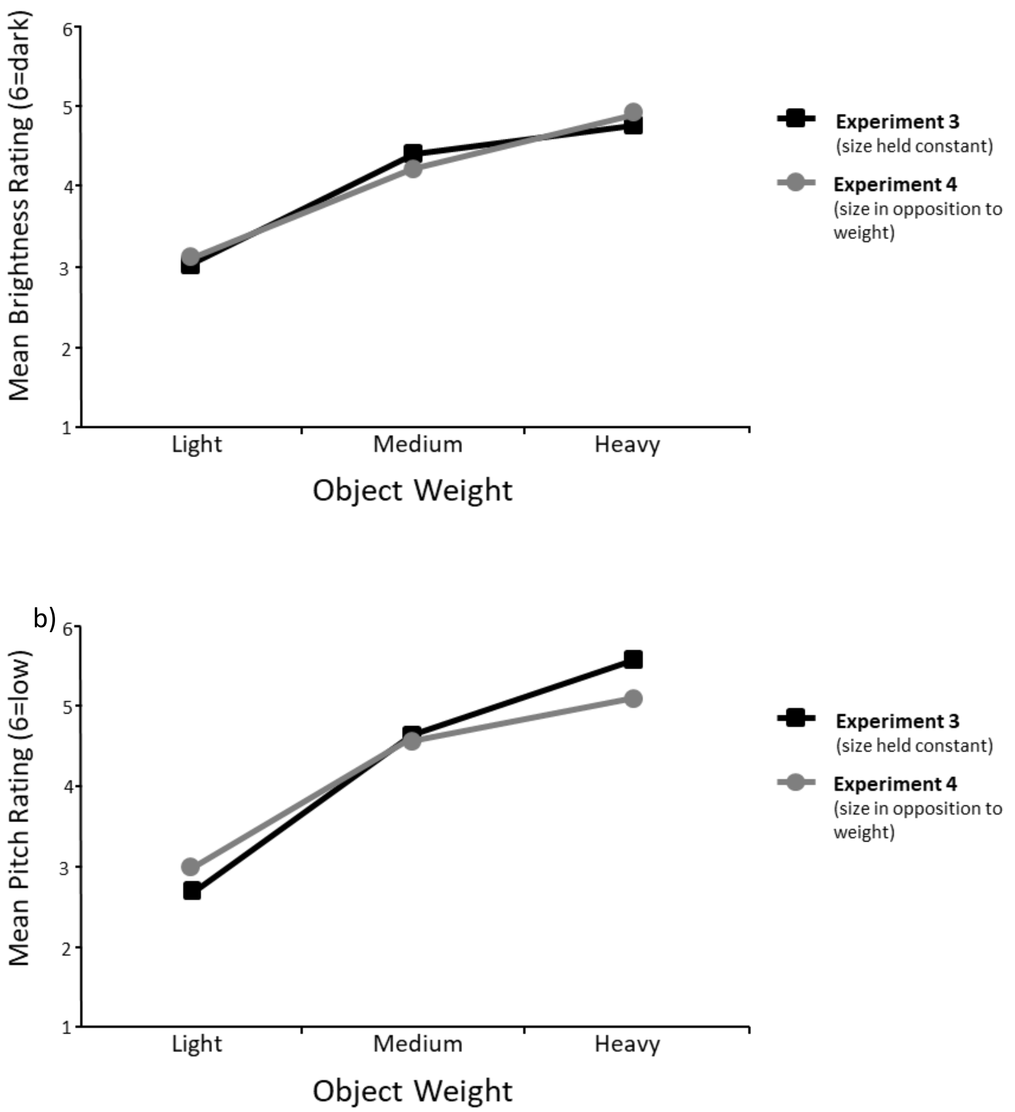


*Figure 3.5*. A comparison between the mean ratings for the objects in Experiments 3 and 4 for a) brightness b) pitch.

### 3.3.3 Discussion

The findings indicate that heavy objects are judged to be darker and lower in pitch, even when the size of these objects predicts an association in the opposite direction. This confirms the findings from Experiment 3 that the associations of heaviness with brightness and pitch are independent of size. When both size and heaviness are presented in conflict, the majority of people use heaviness as the basis for their judgements of brightness and pitch, and not size. It seems that rather than the two features averaging out, or there being a random split in the use of size or heaviness as the basis for judgements, people systematically base ratings of brightness and pitch on heaviness.

The comparison of Experiments 3 and 4 suggests that when contradictory information about size was available, the ratings of brightness were not affected by this; the ratings were based on heaviness in the same way as they were when size was held constant. However, for pitch, the interaction between experiment and object was approaching significance. Interestingly, Figure 3.5b shows that contradictory size information resulted in less extreme pitch ratings. This is interesting because we might have predicted that the presence of contradictory size information would increase the contrast in felt heaviness between the objects (a smaller object would be felt to be heavier than a bigger object that is equally weighted). If ratings were based on the felt heaviness of objects only (i.e. not size) one would expect the difference in pitch ratings between the objects to be more pronounced in Experiment 4 (where there is a presence of contradictory size information) than in Experiment 3. In fact we see that the opposite, pitch ratings are less extreme when contradictory size information is available. This can be understood as there being a separate effect of size working against heaviness, and potentially having an independent influence on pitch ratings. Of course this must be interpreted cautiously because the interaction only approaches significance. Nonetheless, this is an interesting finding worth further consideration.

There are a number of reasons why the heaviness of the objects lifted in this experiment may override their size in entering correspondence with brightness and pitch. It could be explained by the underlying nature or organisation of these particular cross-sensory correspondences. For example, the correspondences heaviness enters into with brightness and pitch may be stronger than that of size. It is also possible to attribute this finding to specific elements of the procedure used in the present experiments. For example, the heaviness of the objects was experienced first; size was only available once an object had been lifted high enough to be viewed in the window. It is possible that the order in which the features became available influenced which feature formed the basis of correspondence judgements. Alternatively, size may have been less influential than it otherwise could have been, because it was only perceived visually. Ellis and Lederman (1993) observed that the size-weight illusion is considerably stronger when size is experienced through touch as opposed to through vision alone. Therefore, were the size of the objects available to participants through touch, this may increase the influence of size of brightness and pitch judgements.

## 3.4 General Discussion

Taken together, the results presented in this chapter demonstrate that the associations of heaviness with brightness and pitch are independent of the associations these features share with size. When objects varying in heaviness are known to be of equal size, the heaviness of the objects continues to correspond with brightness and pitch such that the heavier objects are rated as darker and lower in pitch compared to less heavy objects. When the size of the objects varied in the opposite direction to heaviness, the rating of brightness and pitch corresponded with heaviness, despite the size of the objects predicting brightness and pitch judgements in the opposite direction.

Experiment 4 also indicates some potentially interesting findings about how size and heaviness may independently contribute to judgements of pitch. Although only approaching significance, the findings suggest that the size of the objects may have an independent influence on pitch ratings such that the contradictory size information reduced the rated difference in pitch between the smaller/heavier and larger/lighter object compared to the pitch ratings in Experiment 3. More work is necessary to further understand how judgements of brightness and pitch are influenced by variation in the size and heaviness of lifted objects.

# Chapter 4

## 4.1 Introduction

In Chapter 2, it was demonstrated that heaviness induces associations with size, brightness and pitch, such that heavier objects are expected to be bigger, darker and lower in pitch than less heavy objects. In Chapter 3, heaviness was shown to continue to influence judgements about brightness and pitch, when the size of the objects was also made available visually. Heaviness was shown to induce the same pattern of brightness and pitch judgements when size was held constant (Experiment 3) and when size was varied in contrast with heaviness (Experiment 4). This suggests that the correspondences of heaviness with brightness and pitch are independent of size. Moreover, Experiment 4 suggests that heaviness prevails over conflicting information about size to form the basis of brightness and pitch ratings.

There are a number of possible reasons why the heaviness of the objects formed the basis of brightness and pitch judgements in Experiment 4. Some of these potential explanations reflect specific aspects of the experimental design used in Chapter 3 as opposed to the underlying nature of correspondences involving heaviness and size (See Section 3.3.3). The aim of the present chapter is to replicate the findings from the previous chapters with a method that is different in two crucial ways. Firstly, participants will lift the objects directly by touch (more closely reflecting how we experience the size and heaviness of objects when we lift them in our everyday experience). Secondly, a paired-comparison methodology is used which allows us to further explore how various combinations of size and heaviness influences judgements of brightness and pitch, and the potential influence of the size-weight illusion (see Section 4.1.1).

Asking participants to lift the objects directly creates two key differences between the present experiment and Experiment 4. Firstly, the modality by which size is available to the participant changes from visual to tactile. The potential difference between visual and tactile size has been observed in the size-weight illusion literature, the haptic size-weight illusion is much stronger than the visual one (with visual information having very little additional influence to a haptic size-weight illusion) (Ellis and Lederman, 1993). This modality-effect may be specific to the size-weight illusion, and be a result of additional influence of interactions between haptic size and weight arising at lower level having an on subsequent influence of perceived heaviness (see Buckingham, 2014, *p. 1627*). However, it may also reflect a more general difference in the salience of size when it is available visually compared to felt through touch.

In addition to the change in the modality through which size is presented, the order that size and weight become available to participants in the present study is the reverse of Experiment 4. In Experiment 4, the size of the objects was only available after the objects had been lifted. However, in the present study, the objects will be lifted directly and so the size of the objects is available first, before the objects are lifted and heaviness is felt. In this way, Experiment 5 reflects more closely, the way in which size and weight become available to us lifting objects naturally in our everyday life (having access to size information through vision and again through touch immediately before lifting). If the order in which size and weight became available to participants influenced the judgements found in Chapter 3, the opposite will be found in the present experiment.

Finally, it is anticipated that the paired comparisons design in the present study will obscure the relationship between the objects in such a way as to make the aims of the task more ambiguous to participants. This will reduce participants’ ability to respond in ways that are based on a desire to be consistent or in anticipation of the Experiment’s aims. These differences between Experiment 4 and the present experiment allow the robustness of the findings from Chapter 3 to be tested. And the interactions between size and weight to be further explored.

For the Experiment outlined in the present Chapter, the full set of nine objects outlined in Section 1.5.1 was used. Participants were presented with every possible pairing of the nine objects and rated them in terms of either brightness or pitch. The objects varied orthogonally in both size and weight (see Figure 4.1) which allowed the independent role of both features on judgements of brightness and pitch to be investigated.

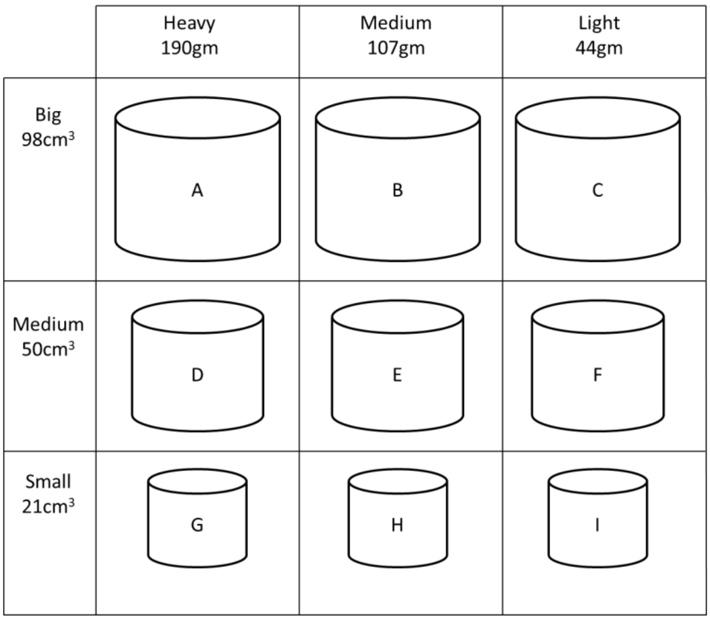



*Figure 4.1.* The set of nine objects used in the present study. The objects were of three values of size each at three values of weight.

In a similar enquiry, Marks (1989) investigated how the combination of two aspects of sound: pitch and loudness contribute to the cross-modal mapping of sounds to luminance (a bright versus a dim light). Prior to this study, it had been demonstrated that luminance corresponds with each of these dimensions separately (Lewkowicz and Turkewitz, 1980; Marks, 1974, 1978; Marks, Szczesuiul, and Ohlott, 1986). However, in this case, participants were presented with sounds varying in each combination of high/low pitch with high/low loudness in order to determine the separable contributions of each dimension to luminance judgements. Participants were asked to match each sound to either a bright or dim light. Marks (1989) considered three possible ways in which the two features may contribute to decisions about a sound's luminance. Firstly, luminance judgements may be based on one feature (either pitch or loudness), irrespective of variation in the other. Secondly, both features may influence judgements of luminance, having an aggregated effect on the final mapping decision. Thirdly, another unitary feature which is a product of both features (e.g. auditory density) may form the basis of the judgements. It was found that pitch and loudness independently influenced the judgements of luminance, and what is more, that pitch was observed to be more influential than loudness in decisions about luminance.

The same possible outcomes apply to the present study, as potential ways that size and heaviness may interact to influence judgements of brightness and pitch. Firstly, it is possible that one feature (either size or heaviness) may form the basis of brightness and pitch judgements, irrespective of variation in the other feature. Secondly, size and heaviness may have some combined influence on ratings such that the final brightness or pitch rating is an aggregate of the size and heaviness values of the objects. Finally, another unitary feature generated through the combination of size and heaviness, for example density, may form the basis of judgements. Based on the findings from Experiment 4, which demonstrated that heavier objects were rated as darker than lighter objects despite the heavier objects being smaller, it is predicted that the heaviness of the objects will continue to be the main basis for brightness and pitch judgements such that heavier objects will be rated as darker and lower in pitch. It is less clear based on previous findings if, and in what way size variation will influence judgements about brightness and pitch.

### 4.1.1 The influence of Size on Heaviness

One complication which must be taken into account is the influence of size on perceived heaviness. In the previous chapters, there has been little distinction between the weight of an object and its subsequent perceived heaviness. This was because when all else is equal, perceived heaviness systematically maps onto the mass of objects, albeit imperfectly (Weber, 1834/1978). However, the size of objects has an additional influence on perceived heaviness. The size-weight illusion means that the smaller of two equally weighted objects is felt to be heavier than a bigger object (Buckingham, 2014; Murray, et al., 1999). Given this effect, it is expected that a dissociation between the weight of the objects and the perceived heaviness will be observed as a result of variation in size. In order to assess this disparity, a group of participants were asked to rate the same set of objects on a scale measuring perceived heaviness. These ratings were used to explore whether perceived heaviness as opposed to veridical weight is a more meaningful predictor of brightness and pitch ratings. In addition, a preliminary study was conducted to determine how size alone influences judgements of brightness and pitch when the sizes of the objects are felt, but they are not lifted.

## 4.2 Preliminary Study: confirming the correspondence of size alone with brightness and pitch

If size does have an influence on the ratings of brightness and pitch, it is expected to be in accordance with previous demonstrations of size-brightness and size-pitch correspondences, which have shown bigger objects to be rated as darker and lower in pitch than smaller objects (Evans and Treisman 2010; Gallace and Spence, 2006; P. Walker and Smith, 1985; L. Walker, et al., 2012). However, it is only assumed that the size of these particular objects would induce the same pattern of correspondences observed elsewhere. To check this assumption, a group of participants explored objects of the three values of size through touch without lifting, and were asked to give ratings of brightness, pitch and expected heaviness.

### 4.2.1 Method

#### *4.2.1.1 Participants*

Twenty-eight students from Lancaster University (2 males and 26 females) between the ages of 18 and 29 (mean age = 19.30 years) volunteered to take part in the study after being approached in various social and learning spaces on campus. All participants except three were right handed by self-report.

#### *4.2.1.2 Stimuli and Materials*

***Materials.*** One object at each level of size from the set of 9 (i.e., objects A, E, & I in Figure 4.1) were used in the present study. The objects were presented in pairs inside a wooden frame (33cmx33cmx33cm) which had a thick, black curtain on one side to hide the objects from view of the participant. The rating scale being used was placed on top of the frame for participants to refer to.

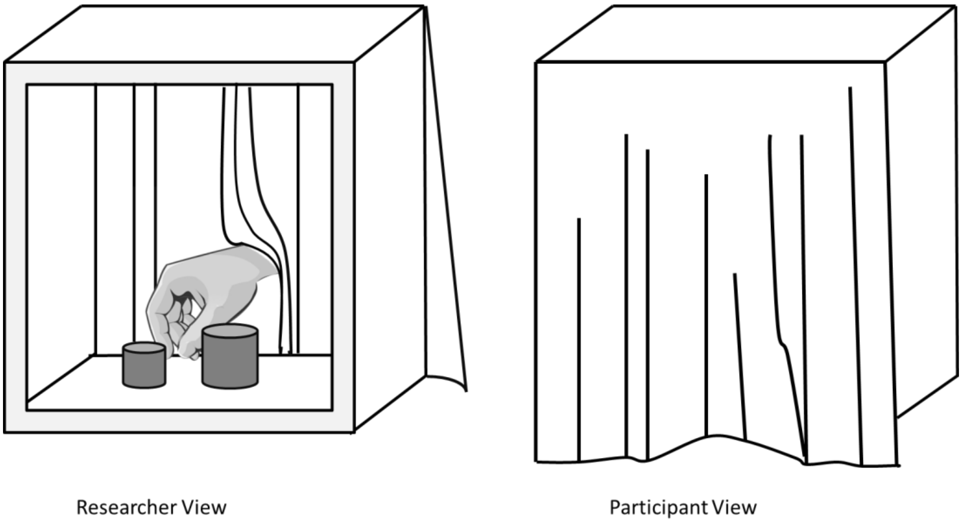



*Figure4.2*. Schematic of the apparatus. The objects remained hidden from view behind a thick black curtain. Each scale was placed in front of the participant on top of the wooden frame.

***Scales.*** Three rating scales were used to elicit participants' judgements about the assumed heaviness, brightness and pitch of the objects. The direction of each scale was randomly determined, separately for each feature and each participant to ensure that all scales were not aligned in the same direction for each person.

*Heaviness.* The heaviness scale ran from1-9 with anchor labels "light" and "heavy" at each end point. For half of participants, "heavy" was assigned to 1 and "light" was assigned to 9, for the other half of participants the reverse assignment was used.

*Brightness.* Brightness was represented with nine squares ($2cm^2$) varying in achromatic brightness between white and black in equal psychological steps selected from the Munsell Book of Color (1976) printed on an olive textured background. Each square was numbered from 1-9 starting with 1 at the left. For half of participants the squares ran from black to white and for the other half this was reversed.

*Pitch***.** Pitch was represented with selected keys numbered from one to nine on a Casio SA-47H5 Mini keys keyboard (the nine white keys from Middle C to the D above Middle C). For half the participants the highest pitch sound was labelled 1 and the lowest pitch sound labelled 9, for the other half this was reversed.

#### *4.2.1.3 Design*

A paired comparisons procedure was used. Participants were presented with all possible pairings of the differently sized objects in both left/right presentations (six pairings in total). They judged each pairing in terms of brightness, pitch and expected heaviness. The presentation order of the six pairings was randomised for each participant. There was one constraint, that the same pairing must not be used within two consecutive presentations. Participants rated all pairings on one scale (pitch, brightness or heaviness) at a time, before moving on to the next. Half of participants completed the pitch ratings first and the other half completed the brightness ratings first. All participants completed the heaviness ratings last.

#### *4.2.1.4 Procedure*

Participants sat opposite the researcher, in front of the wooden frame, inside of which the objects were placed (hidden behind the curtain). The objects were presented side by side on the base of the frame and were secured with tac to ensure the weight of the objects could not inadvertently be felt through lifting or moving. Participants were asked to use the thumb and first two fingers of their dominant hand to explore each object. They were told to explore both objects in each pair before giving any answers. On top of the frame was the scale used to judge the objects. In the brightness condition, they were asked to try to picture each object; in the pitch condition they were asked to imagine the two objects were to come to life and make a sound; in the heaviness condition they were asked to imagine lifting the objects. In all cases, they were asked to indicate which of the two objects was darker/lower in pitch/

heavier[3] and then to give a score for where on the scale each object in the pair would be placed. The researcher wrote down the responses on a score sheet.

### 4.2.2 Results

The data were the scores out of 9 on the brightness, pitch and heaviness scales for each object across all pairings. In all cases, the ratings on each scale were assumed to reflect a continuous variable.  The numbers 1-9 assigned to each end of the scale were counterbalanced during the procedure; for the purposes of analysis, scores were recoded such that a rating of 9 marked the heaviest, darkest, and lowest pitch ends of the scales. Figure 4.3 shows the mean ratings for judged brightness, pitch, and heaviness for each level of object size collapsed across all pairings.

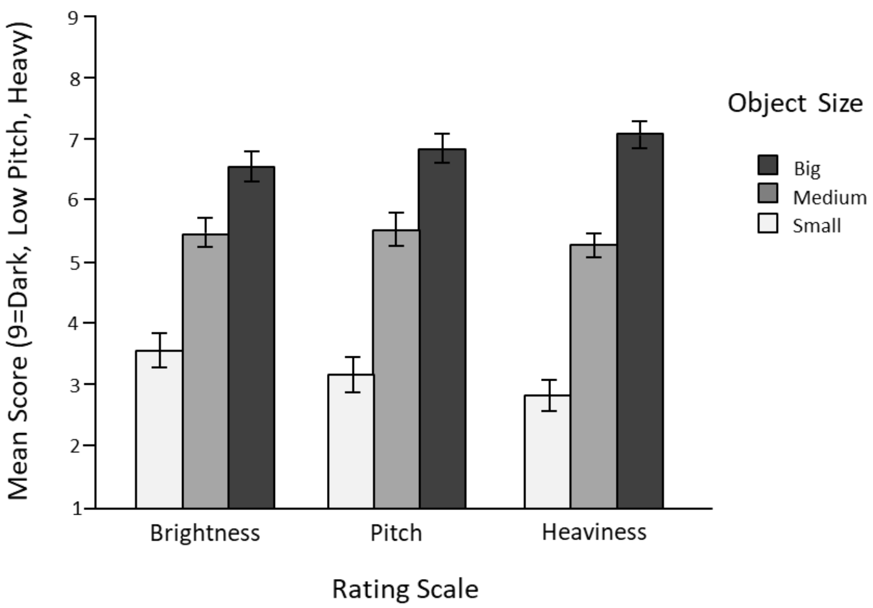



*Figure 4.3.*The mean brightness, pitch and heaviness ratings of the big, medium and small objects. A rating of 9 signifies low pitch, dark and heavy. Error bars represent the standard error.

[3] The questions were worded to refer to this extreme of each feature dimension only, to avoid using the shared label “light”.

For analysis, a linear mixed effects approach was used due to the repeated measures design using *R* (R Core Team, 2012) version 3.2.0 and package *lme4* (Bates, et al., 2014). The same model was fitted for each scale which included size as an explanatory variable and a random effect of participant. Size was coded with values 1 - 3 (such that 1=smallest, 2=medium and 3=biggest). A likelihood ratio test was used to compare the model including size as a main effect with a null model that contained only a random effect of participant. 95% confidence intervals were calculated using the *Wald* method with the *confint()* function. Alkaike Information Criterion (AIC) values provided a relative estimate of the amount of information not being captured by a model, balancing goodness of fit with the number of parameters the model contains. The AIC was compared for models of the same data set, with lower AIC values indicating a superior model. For all analyses of this kind reported below, visual inspection of residual Q-Q plots did not reveal any obvious departures from normality.

#### *4.2.2.1 Brightness*

For the brightness ratings, a likelihood ratio test confirmed that a model including size was preferred to the null model, $\chi^2(1) = 60.77$, $p < .001$, with AIC values of 751.4 and 692.7 for the null and comparison model, respectively. For each step increase in size, objects were judged darker by 1.50 points on the brightness scale, CI [1.16, 1.84].

#### *4.2.2.2 Pitch*

For the pitch ratings, a likelihood ratio test confirmed that a model including size was preferred to the null model, $\chi^2(1) = 79.90$, $p < .001$, with AIC values of 783.76 and 705.86 for the null and comparison model, respectively. For each step increase in size, objects were judged to be lower in pitch by 1.84 points on the pitch scale, CI [1.49, 2.19].

#### *4.2.2.3 Heaviness*

For the heaviness ratings, a likelihood ratio test confirmed that a model including size was preferred to the null model $\chi^2$ (1) = 138.82, $p$ <.001, with AIC values of 777.29 and 640.5 for the null and comparison model, respectively. For each step increase in size, objects were expected to be heavier by 2.21 points on the heaviness scale, CI [1.91, 2.50].

### 4.2.3 Discussion

The effects of size on judgements of brightness, pitch, and expected heaviness replicate previous findings that larger objects are associated with being darker, lower pitch and heavier (L. Walker, et al., 2012). This confirms the assumed association between size and these feature dimensions. The findings of this preliminary study and the findings of previous chapters in the present thesis demonstrate that both size and heaviness have the potential to induce a systematic pattern of associations with brightness and pitch in the present set of objects when varied individually. In Experiment 5, the size and weight of objects are varied orthogonally to determine how the two features in combination may influence judgements about brightness and pitch.

## 4.3 Experiment 5

### 4.3.1 Method

#### *4.3.1.1 Participants*

Eighty-four students volunteered to participate in the study after being approached in various social and learning spaces on Lancaster University campus. Groups of twenty-eight participants each provided judgements of brightness, pitch, or heaviness. Participants in the brightness group (10 males, 18 females) were aged between 19 and 21 (mean age = 20 years) and all except 1 participant were right handed by self-report. Participants in the pitch group

(12 males, 16 females) were aged between 18 and 43 (mean age = 20.86 years) and all except 4 were right hand by self-report. Participants in the heaviness group (8 males, 20 females) were aged between 18 and 25 (mean age = 20.64 years) and all except two were right handed by self-report.

#### *4.3.1.2 Materials*

All nine objects outlined in Section 1.5.1 were used. The objects varied orthogonally at three levels of size and three levels of weight (see Figure 4.1). As in the preliminary study, the objects were presented to participants on a wooden frame (33cmx33cmx33cm) which had a thick, black curtain on one side to hide the objects from view. The same three scales used in the preliminary study were used to elicit judgements about brightness, pitch and heaviness.

#### *4.3.1.3 Design*

Participants were presented with every possible pairing of the nine objects (36 pairings) however only once in one of the two possible left/right positions. Fourteen random order sequences of the pairings were used, constrained such that the same object was not used within two consecutive pairings. Each object appeared equally often on the left and the right hand side and two participants completed each sequence, with opposite left-right positioning of the objects in each pairing. All pairings were presented in one block of trials and participants rated them in terms of either: brightness, pitch or heaviness.

#### *4.3.1.4 Procedure*

The procedure was the same as that adopted in the preliminary study, with the exception that participants rated all objects on one scale only and were able to lift the objects. Participants were instructed to lift the objects with the thumb and first two fingers of their dominant hand. They could lift the object in either order and as often as they wished while answering the questions. They were instructed to lift both objects in each pairing at least once before making any judgements. The same questions were asked as in the preliminary study, with the exception of the heaviness group who rated the perceived heaviness of the objects being lifted as opposed to the expected heaviness. The experimenter repeated the question for each of the first five pairings and then subsequently every fifth pairing or sooner if necessary.

### 4.3.2 Results

The data were the ratings out of 9 for the brightness, pitch and heaviness of the objects within each pair. As in the preliminary study, the scores were recoded such that very bright, high pitch, and light weight were indicated with a score of 1 and the opposites (very big, dark and heavy) with a score of 9.

#### *4.3.2.1 Analysis by pairing type*

The data were initially divided into sub-sets based on four different pairing types: a) congruent size and weight. b) Objects varying in weight while size is held constant. c) Objects varying in size while weight is held constant. d) Incongruent size and weight (see Figure 4.4). These pairing types were first explored independently. In each case a repeated measures Analysis of Variance (ANOVA) was conducted on the brightness and pitch ratings.

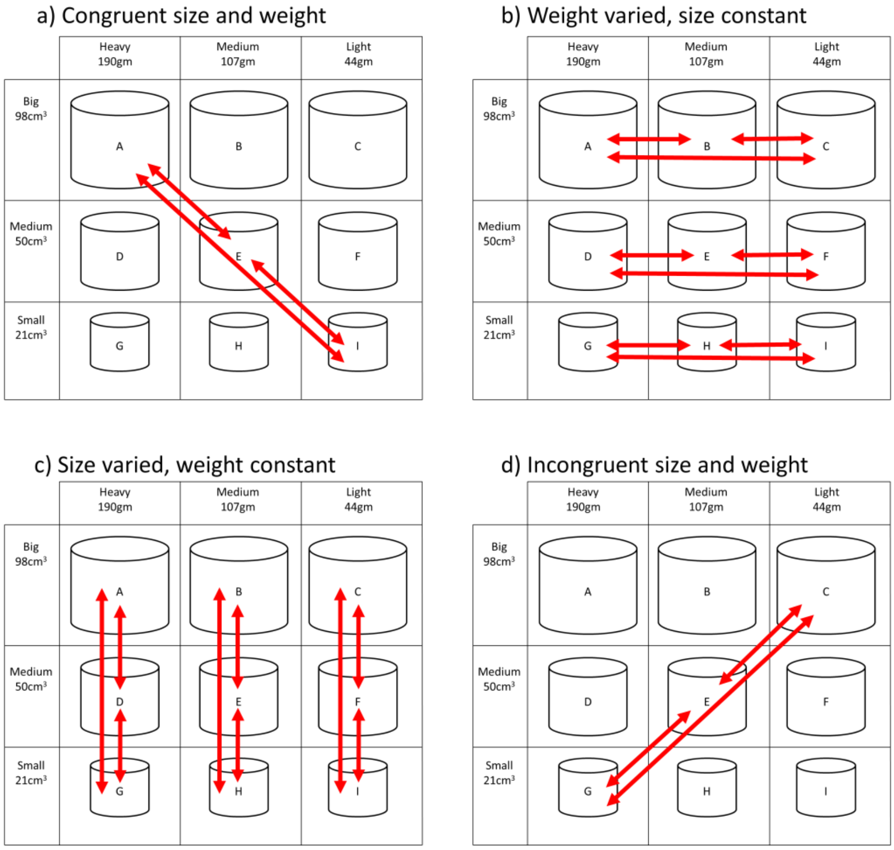



*Figure 4.4.* Each pair of objects belonged to one of these four pairing types. For subset *a* and *d*, only the central diagonals were used in the analysis.

**Congruent size and weight**. The objects A, E and I were the focus of this analysis. The size and weight of these objects were such that the biggest object was the heaviest, the medium sized object was of a medium weight and the small object was the lightest. The size and weight values were such that the three objects were of equal density. In any given pairing of these objects, the larger object was also heavier than the smaller object. The context created by the other object with which any particular object was paired had the potential to influence its perceived size/weight. For example, when the contrast in size or weight of two objects was bigger (i.e. the biggest/heaviest object was paired with the smallest/lightest) the weight and size of the objects could potentially be enhanced. The bigger/heavier object being felt to be even bigger/heavier when paired with the smaller/lighter object compared to when it was paired with the medium object. Similarly, the smallest/lightest object would be felt to be even smaller/lighter when paired with the biggest/heaviest object compared to when it was paired with the medium object. What is more, the identity of the medium object varied from being the lighter object in one pairing (when paired with the heavier object) to being the heavier object in the other (when paired with the lighter object). These differences in relative size/weight were captured in the analysis by a factor which coded whether a particular object was bigger/heavier or smaller/lighter based on which of the other two objects it was paired with. This factor is illustrated in Figure 4.5.

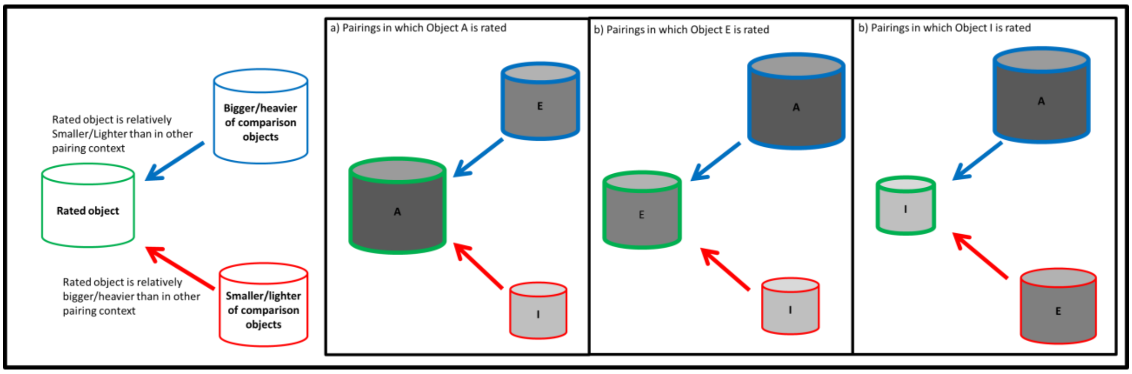



*Figure 4.5*. A demonstration of how each object in a pairing created a difference in context. This was captured in the factor "contextual size/weight". It can be understood as whether the object was in a pairing which made it contextually bigger/heavier (red) or smaller/lighter (blue) than when it was paired with the other comparison object. For example b) Object E is the bigger/heavier object when paired with comparison Object I, but is the smaller/lighter object when paired with comparison object A. Although the relative identity of a) Objects A and c) Object I do not change when paired with each comparison object, the ratings may be influenced in the same way by these differences in comparison object.

A 3x2 repeated measures ANOVA was conducted on the ratings with: Size/Weight (big/heavy, medium, small/light) summarising both size and weight together and Contextual Size/Weight (bigger/heavier and smaller/lighter) capturing whether the object was relatively bigger/heavier or smaller/lighter based on the pairing context it was in. For both brightness and pitch ratings there were significant linear trends across different levels of size/weight such that larger/heavier objects were judged to be darker ($F(1, 27) = 14.79$, $p = .001$, $\eta p^2 = 0.35$) and lower in pitch $F(1, 27) = 36.32$, $p < .0001$, $\eta p^2 = 0.57$). For brightness, there was no significant context effect, which suggests that the objects had a similar rating irrespective of the object with which it was paired. However a context effect was observed for pitch ratings ($F(1, 27) = 7.17$, $p = .01$, $\eta p^2 = 0.21$) indicating that objects were rated as being lower in pitch when they were in a context where they would seem relatively bigger/heavier.

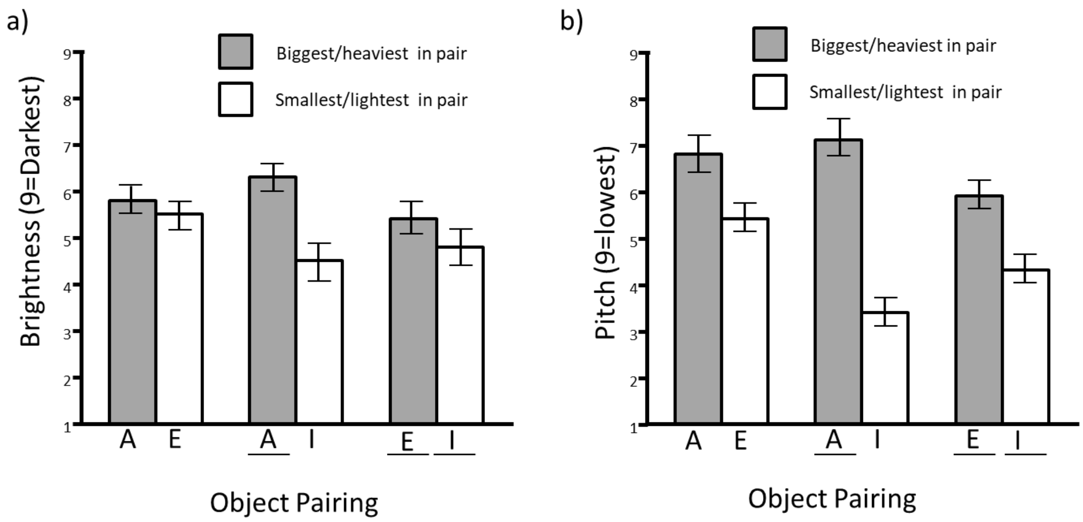



*Figure 4.6.* The mean ratings of a) brightness and b) pitch for the objects within each pairing. The object label is underlined to signify that it is the pairing context within which it is bigger/heavier compared to the pairing that object enters. Error bars represent SEM.

**Weight varied, size constant.** This subset involved all pairs where the size of the objects was equal but the weight varied (see Figure 4.4b). Although the size of the objects was constant within each pairing, there were pairings of this type at three levels of size: as can be seen in Figure 4.4b, pairing involving object A, B and C were bigger than pairings of objects D, E and F and G, H and I. In a similar way to the first subset, each object within a row was rated twice; once with each of the other two objects. The context created by the other object in a pairing had the potential to alter the judged heaviness of an object being rated. In pairings where the contrast is bigger (i.e. the heaviest object is paired with the lightest), the heavier object will be felt to be even heavier (compared to when it is paired with the medium object) and the lighter object would be felt to be even lighter (compared to when it is paired with the medium object). What is more, the identity of the medium object varied from being the lighter object in one pairing (when paired with the heavier object) to being the heavier object in the other (when paired with the lighter object). These differences in relative weight were captured in the analysis by a factor which coded whether a particular object was heavier or lighter compared to when it was paired with the other object.

A 3x3x2 repeated measures ANOVA was conducted with size (biggest, medium, smallest) X weight (heaviest, medium, lightest) X contextual weight (heavier, lighter) as factors. Contextual Weight referring to the pairing within which an object can be considered relatively heavier or lighter. For both brightness and pitch ratings a significant linear trend of weight was found, such that the heavier object was judged to be darker , $F(1, 27) = 22.09$, $p < .0001$, $\eta p^2 = 0.45$, and lower in pitch $F(1, 27) = 43.62$, $p < .001$, $\eta p^2 = 0.62$. There were no significant effects of contextual weight on brightness but there was a significant main effect of contextual weight on judgements of pitch, $F(1, 27) = 27.53$, $p < .0001$, $\eta p^2 = 0.50$. The nature of this effect indicated that when the object being rated was paired with the contextual object most likely to make it seem heavier, the object was rated as being lower in pitch. Interestingly, although size varied between pairs (rather than within pairs), linear trend across different levels of size for judgements of brightness, $F(1, 27) = 3.43$, $p = .07$, $\eta p^2 = 0.11$ and pitch size ($F$ (1, 27) = 3.03, $p$ = .09, $\eta p^2$ = 0.10) were approaching significance. The trend demonstrates that the smaller sized objects were rated darker and lower in pitch than the larger objects. This is the reverse of the expected relationship of size on ratings of brightness and pitch based on the preliminary study. This may indicate the presence of a size-weight illusion. The demonstrated influence of size on judgements within this pairing type indicates that the full set of objects may have formed the relative context for ratings of the objects, rather than the specific pairing being presented.

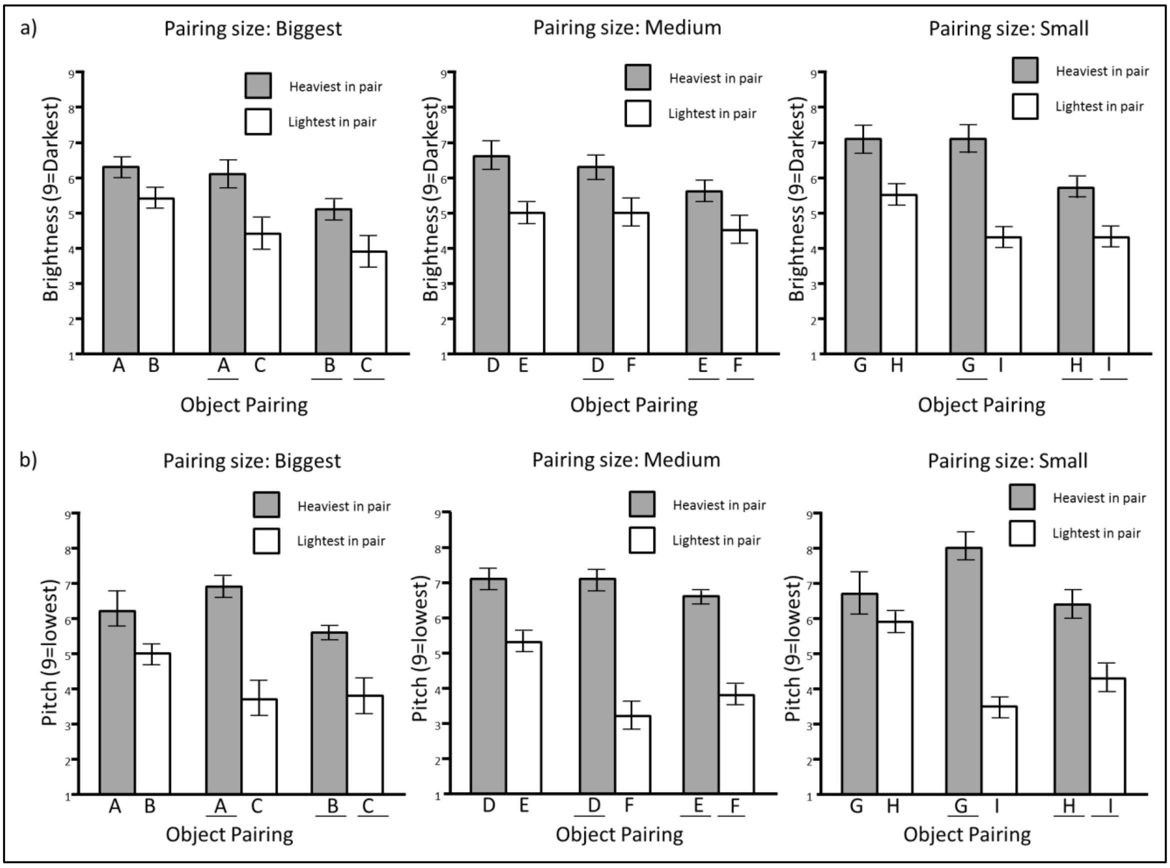



*Figure 4.7.* The mean ratings of a) brightness and b) pitch for the objects within each pairing. The object label is underlined to signify that it is the pairing context within which it is heavier compared to the pairing that object enters.Error bars represent SEM

**Size varied, weight constant.** This subset included all pairings between objects which varied in size but whose actual weights were equivalent. Although the weight of the objects was constant within each pairing there were pairings of this type at three levels of weight. As can be seen in Figure 4.4c pairing involving object A, D and G were heavier than pairings of B, E and H and C, F and I. Once again, each object within a column was rated twice; once with each of the other two objects. The context created by the other object in a pairing had the potential to effect judged size of an object being rated. In pairings where the contrast was larger (i.e. the biggest object is paired with the smallest), the size of the object would be emphasised. The bigger object would seem bigger when paired with the smallest object compared to when it was paired with the medium sized object and the smallest object would seem smaller when paired with the biggest object compared to when it is paired with the medium object. What is more, the identity of the medium object varied from being the smaller object in one pairing (when paired with the bigger object) to being the bigger object in the other (when paired with the smaller object). These differences in relative size were captured in the analysis by a factor which coded whether a particular object was bigger or smaller compared to when it was paired with the other object.

A 3x3x2 repeated measures ANOVA was conducted with weight (heaviest, medium, light) X size (biggest, medium, smallest) X contextual size (bigger, smaller) as factors. There were significant linear trends of size for judgements of brightness ($F(1, 27) = 5.18$, $p = .03$, $\eta p^2 = 0.16$) however for pitch this was not significant at the .05 level ($F(1, 27) = 2.46$, $p = .13$, $\eta p^2 = 0.08$). Smaller objects were judged to be darker than larger objects. This is the reverse as would be expected considering the effect the size of objects had on ratings of brightness and pitch when the objects were not lifted in the preliminary study. The difference in the direction of the relationship between size with brightness and pitch may be due to the influence of size on perceived heaviness. An equivalent analysis of heaviness ratings revealed that smaller objects were judged to be heavier than bigger objects, $F(1, 27) = 199$, $p < .001$, $\eta p^2 = 0.88$. This suggests that the difference in size resulted in a difference in perceived heaviness, despite the objects being of equal weight. This difference may have subsequently determined brightness and pitch ratings, this is further explored in the next section.

Despite the actual weight of objects only varying between pairings, there was a significant linear trend of weight for judgements of brightness, $F(1, 27) = 42.48$, $p < .0001$, $\eta p^2 = 0.61$, and pitch, $F(1, 27) = 35.12$, $p < .0001$, $\eta p^2 = 0.57$. Heavier objects were judged to be darker, and lower in pitch than objects that were lighter in weight. This suggests that the ratings were not only based on the context of the particular pairing but on the full set of objects as a whole. There were also no significant effects of contextual size, except for a significant quadratic interaction between size and contextual size in relation to judgements of brightness, $F(1, 27) = 5.70$, $p = .02$, $\eta p^2 = 0.17$. The nature of this quadratic interaction was such that ratings of brightness were not influenced by the other object in a pairing except for with the middle object for which there was a categorical difference in the identity of the object based on the object with which it was paired.

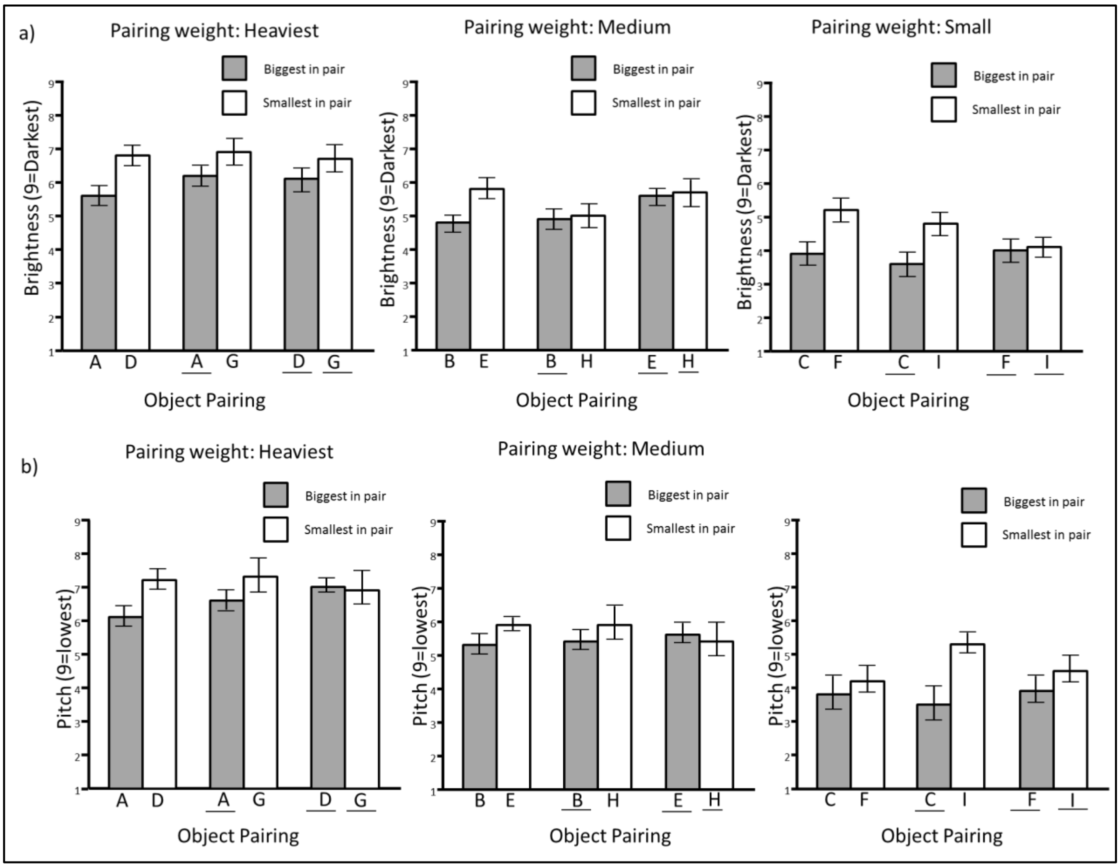



*Figure 4.8.* The mean ratings of a) brightness and b) pitch for the objects within each pairing. The object label is underlined to signify that it is the pairing context within which it is bigger compared to the pairing that object enters. Error bars represent SEM.

**Incongruent size and weight.** For this pairing type, Objects C, E and G were the focus of analysis. The size and weight of these objects contrasted such that the smaller object was heavier than the larger object. Pairings involving these objects were such that the smaller object was heavier than the larger object, irrespective of which two objects were paired. This may be considered similar to the second pairing type as there is a conflict between the size of the object and the perceived heaviness; however in this case, this is the result of an actual difference in object weight. The objects were rated twice in this context; once with each of the other two objects. The context created by the other object in a pairing had the potential to effect the specific rating for a particular object. In pairings where the contrast is bigger (i.e.

the smallest/heaviest object is paired with the biggest/lightest), the size and heaviness of the object would be emphasised in opposite directions (the smaller/heavier object will be felt smaller and heavier when paired with the biggest/lightest object compared to when it is paired with the medium object and the bigger/lighter object would be felt as bigger and lighter when paired with the smallest/heaviest object compared to when it is paired with the medium object). What is more, the identity of the medium object varied from being the smaller/heavier object in one pairing (when paired with the bigger/lighter object) to being the bigger/lighter object in the other (when paired with the smaller/heavier object). These differences in relative size/weight were captured in the analysis by a factor which coded whether a particular object was bigger/lighter or smaller/heavier when compared to when it was paired with the other object.

A 3x2 repeated measures ANOVA was conducted on the ratings with the size and weight of the objects summarised in one factor: size/weight (big/light, medium, small/heavy). The analysis also included a factor of contextual weight (bigger/lighter and smaller/heavier) pertaining to whether the object was in a pairing in which it was relatively bigger/lighter or smaller/heavier in contrast to the other object. For both brightness and pitch ratings a significant linear trend of size/weight was found, such that the smaller/heavier object was judged to be darker $F(1, 27) = 16.16$, $p < .0001$, $\eta p^2 = .374$ and lower in pitch $F(1, 27) = 12.20$, $p = .002$, $\eta p^2 = 0.311$. There was no significant context effect for brightness or pitch ratings which suggests that the objects had a similar rating irrespective of the object with which it was paired.

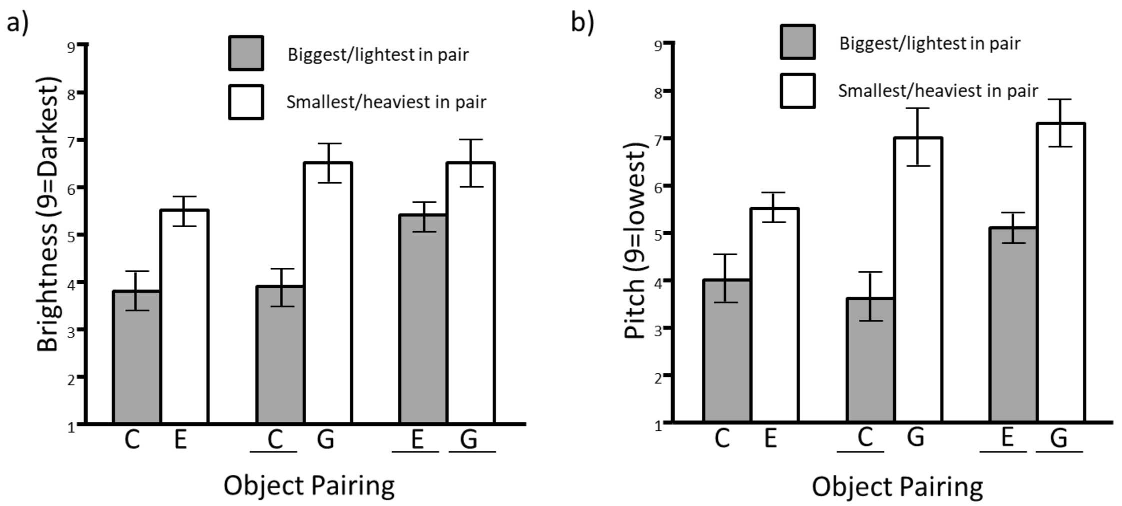



*Figure 4.9.* The mean ratings of a) brightness and b) pitch for the objects within each pairing. The object label is underlined to signify that it is the pairing context within which it is bigger/lighter compared to the pairing that object enters. Error bars represent SEM

#### 4.3.2.2 Interim summary

The results indicate that ratings of brightness and pitch are influenced by heaviness such that heavier objects are rated as being darker and lower in pitch, irrespective of the specific pairing type. This means that ratings were influenced by heaviness irrespective of variation in size. For pairs where the weight of the object was held constant but size varied (Pairing Type C), it seems that the ratings of brightness and pitch were based on the illusionary difference in heaviness caused by the difference in size (the size-weight illusion). No consistent pattern of within-pairing context effects was found. What is more, the size and weight of the objects appeared to have some influence on object ratings even when they were not varied within a specific pair. This suggests that participants used the full set of objects as a context for comparison, rather than the particular objects presented in a pairing. Therefore, an analysis exploring the roles of size and weight across the full set of ratings, irrespective of the pairing type was undertaken to further understand the roles of size and weight in judgements of brightness and pitch.

#### 4.3.2.3 Modelling Contributions of Size and Weight Across All Objects

A linear mixed effects analyses was conducted using *R* (R Core Team, 2012) with the *lme4* package (Bates, et al., 2014), due to the repeated measures design. 95% confidence intervals were calculated using the *Wald* method with the *confint()* function. The ratings on each scale (brightness, pitch and heaviness) were treated as continuous variables. The size and weight of the objects were summarised with values 1 - 3 to indicate the three values for size (1 = small, 2 = medium, 3 = big) and weight (1 = light, 2 = medium, 3 = heavy).

Likelihood ratio tests were conducted to assess the significance of size and weight as explanatory variables compared to null models which in all instances included a random effect of participant. AIC values, which provide a relative estimate of the amount of information not being captured by a model, balancing goodness of fit with the number of parameters the model contains, were compared (a lower AIC value indicating a preferred model). For all analyses, visual inspection of residual Q-Q plots was performed and did not reveal any obvious departures from normality.

**Brightness.** Figure 4.10 shows the mean ratings for judged brightness for each level of object size and object weight. A likelihood ratio test confirmed that a model including size was preferred to the null model, $\chi^2(1) = 34.647$, $p < .001$, with AIC values of 8616.4 and 8583.7 for the null and comparison model, respectively. A model including both size and weight was preferred $\chi^2(1) = 397.74$, $p < .001$, yielding a new AIC value of 8188.0. Estimates from the comparison model that included size and weight reveal how with each step increase in size objects were judged to be brighter(rather than the more usual darker) by 0.32 points on the brightness scale, CI [0.23, 0.42]. With each step increase in weight they were judged to be darker (as expected) by 1.04 points on the brightness scale, CI [0.94, 1.14]. Finally, a likelihood ratio test confirmed that removing size from this comparison model resulted in a significant reduction in its explanatory power, $\chi^2(1) = 42.24$, $p < .001$, with an AIC of 8228.2, confirming that size continued to be influential when weight was included in the model.

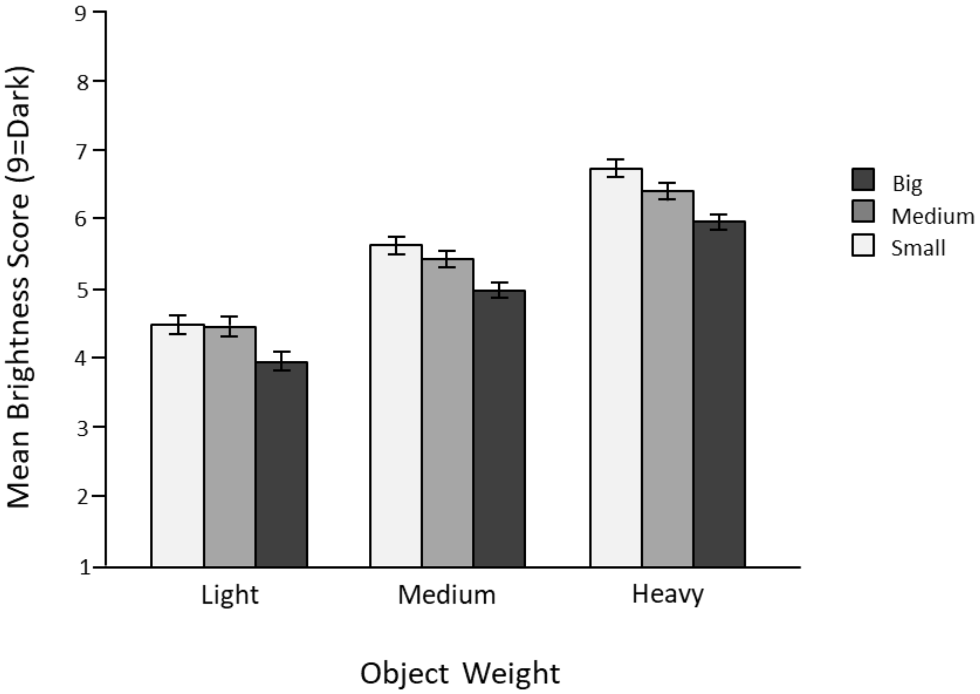



*Figure 4.10.* The mean brightness ratings of the light, medium and heavy objects at each level of size. Error bars represent the standard error.

**Pitch.** Figure 4.11 shows the mean ratings for judged pitch for each level of object size and object weight. A model including size was preferred to the null model, $\chi^2(1) = 16.76$, $p < .001$, with an AIC value of 9331.0 and 9316.2 for the null and comparison model, respectively. Adding weight to the model significantly and considerably increased its explanatory power, $\chi^2(1) = 722.54$, $p < .001$, yielding a new AIC value of 8595.7. Estimates from the comparison model that included both size and weight reveal how with each step increase in size objects were judged to be higher in pitch, rather than the usual lower in pitch, by 0.27 points on the pitch scale, CI [0.16, 0.38]. With each step increase in weight they were judged to be lower in pitch (as expected) by 1.62 points on the pitch scale, CI [1.51, 1.73]. Finally, a likelihood ratio test confirmed that removing size from this comparison model resulted in a significant reduction in its explanatory power, $\chi^2(1) = 24.06$, $p < .001$, with an AIC of 8617.7, confirming that size continued to be influential when weight was included in the model.

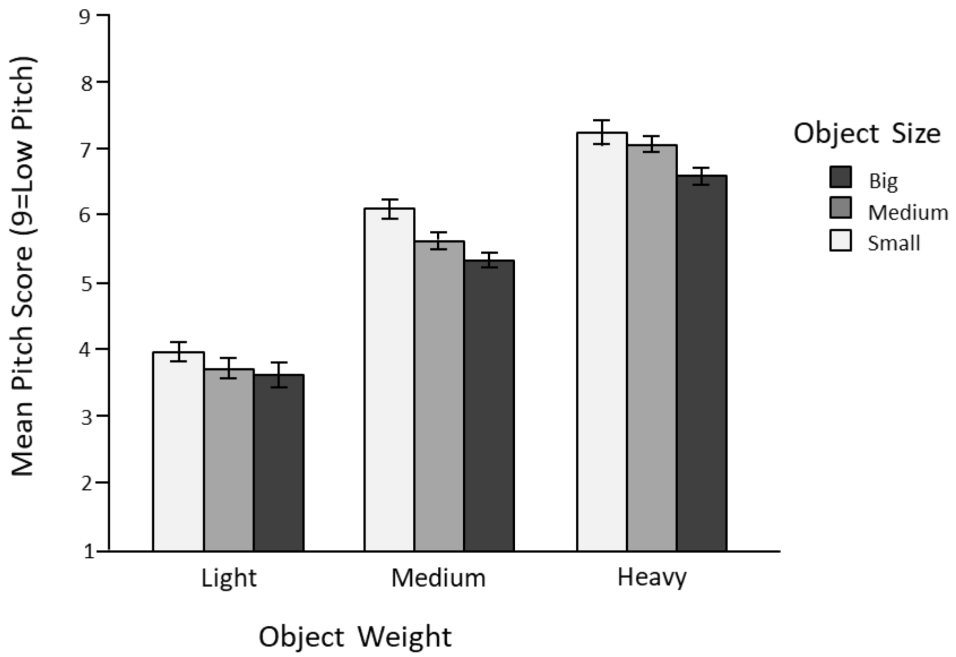



*Figure 4.11.* The mean pitch ratings of the light, medium and heavy objects at each level of size. The Error bars represent the standard error.

In the analysis of the pairing type where objects varied in size but were of equal weight (Figure 4.4c ) it was demonstrated that variation in size induced a mis-perception of weight known as the size-weight illusion. Smaller objects of equal weight were rated as feeling heavier than the bigger objects. To further explore the influence of size on perceived heaviness, the ratings of heaviness given by the additional group of participants were analysed in the same way as brightness and pitch ratings.

Figure 4.12 shows the mean ratings for judged heaviness for each level of object size and object weight. A model including size was preferred to the null model, $\chi^2$ (1) = 168.71, $p$ <.001, with AIC values of 9535.4 and 9368.7 for the null and comparison model, respectively. Adding weight to the model significantly increased its explanatory power, $\chi^2$(1) = 2098.4, $p$ <.001, yielding a new AIC value of 7272.3. Estimates from the comparison model that included size and weight revealed how, with each step increase in size, objects were judged to reduce (rather than increase) in perceived heaviness by 0.885 points on the heaviness scale, CI [0.81, 0.96]. With each step increase in weight their perceived heaviness was judged to increase by 2.40 points on the heaviness scale, CI [2.32, 2.48]. Finally, a likelihood ratio test confirmed that removing size from this comparison model resulted in a significant and considerable reduction in its explanatory power, $\chi^2$(1) = 450.76, $p$ <.001, with an AIC value of 7721.1, confirming that size continued to be influential when weight was included in the model.

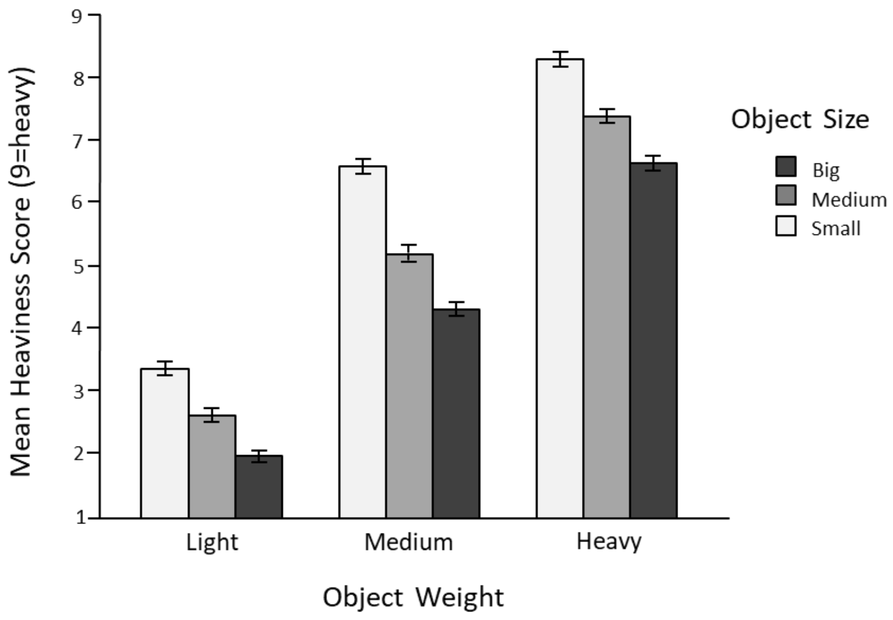



*Figure 4.12.* The mean heaviness ratings of the light, medium and heavy objects at each level of size. Error bars represent the standard error.

**Replacing Weight with Judged Heaviness.** The mean heaviness rating for each object across all participants in the perceived heaviness group replaced object weight as an independent variable, in a repeat of the above analyses of the brightness and pitch ratings.

***Brightness.*** A model including heaviness was preferred to the null model, $\chi^2$ (1) = 425.26, $p$ < .001, with AIC values of 8616.4 and 8193.1 for the null and comparison model, respectively. Adding size to the model did not enhance its explanatory power, $\chi^2$(1) = 1.0537, $p$ = 0.305, consistent both with the confidence interval for the estimated effect of size straddling zero, CI [-0.049, 0.16], and with an unchanged AIC value (which went only slightly from 8193.1 to 8194.1 by adding size to the model). Estimates linked to the comparison model that incorporated heaviness alone reveal how with every point increase on the heaviness scale, objects were now judged to be darker by 0.420 points on the brightness scale, CI [0.38, 0.46].

***Pitch.*** A model including heaviness was preferred to the null model, $\chi^2$ (1) = 719.01, $p$ < .001, with AIC values of 9331.0 and 8613.9 for the null and comparison model, respectively. Adding size to the model did enhance its explanatory power, $\chi^2$(1) = 31.187, $p$<.001, yielding a new AIC value of 8584.8. Estimates from the comparison model including both size and heaviness reveal that for every step increase in size the objects were now judged to be lower in pitch by 0.327 points on the pitch scale, CI [0.21, 0.44], which is as expected on the basis of the results from the passive touch condition, and correspondences more generally. Finally, with every point increase in perceived heaviness the objects were judged to be lower in pitch by .675 points on this scale, CI [0.63, 0.72].

## 4.4 General Discussion

The preliminary study confirmed that for the present set of objects, size induces the same association with brightness, pitch and heaviness that has been found previously (P.Walker and Smith, 1985; P. Walker and Walker, 2012; L. Walker, et al., 2012). When the objects were not lifted, bigger objects were rated as darker, lower in pitch and heavier than smaller objects. When the objects were lifted, the separate influences of size and heaviness on ratings of brightness and pitch could be examined. It was found that heavier objects were rated as darker, and lower in pitch than less heavy objects. This replicates the pattern of results found in Chapters 2 and 3. This adds further support to the claim that heaviness enters into bidirectional associations with these feature dimensions. What is more, when size and heaviness both vary, heaviness appears to have had a stronger influence on ratings of brightness and pitch than size does. This replicates the finding from Experiment 4 which demonstrated that ratings of brightness and pitch were based on the heaviness of the objects despite conflicting values of size. This replication, despite differences in the modality through which size was presented and the order in which size and weight were available to participants, suggests that neither of these aspects of the experimental design had an influence on the dominance of heaviness over size.

The analysis suggests that the pattern of brightness and pitch ratings can be explained by two processes: 1) the influence size has on perceived heaviness and 2) the dominance of heaviness over size in determining judgements of brightness and pitch.

### 4.4.1 The Influence of Size on Perceived Heaviness

Although perceived heaviness has been demonstrated to map onto the weight of lifted objects when all other feature dimensions are held constant (Weber, 1834/1978); our perceptual experience of heaviness is changed when the size of objects also varies. The size-weight illusion demonstrates that for two objects of equal weight, the smaller object is perceived to be heavier than the larger object (Buckingham, 2014;Murray, et al., 1999). This influence of size on perceived heaviness can be observed in the present findings. Figure 4.12 demonstrates that smaller objects are judged to be heavier than larger objects of the same weight.  The effect of size on perceived heaviness is in the opposite direction to the expected heaviness of objects of different sizes when they are not lifted, as was observed in the preliminary study (see Figure 4.3).

When size and weight were the explanatory variables in a model of brightness and pitch ratings in Experiment 5, the observed influence of size on brightness and pitch judgements was the reverse of the size-brightness and size-pitch associations found in the preliminary study. Smaller objects were rated as darker, and lower in pitch than larger objects. The potential for this finding to reflect the role of size on perceived heaviness was tested. When perceived heaviness was used as an explanatory variable in the model, replacing the variable weight, model fit was improved for both brightness and pitch ratings. For brightness ratings, size no longer had any additional contribution to model fit. However for pitch, size continued to improve model fit, although now in the opposite direction (i.e. increase in size leading to judgements of lower in pitch). This is the direction of the association of size with pitch that has been observed in the preliminary study and elsewhere in the correspondences literature (Evans and Treisman 2010; Gallace and Spence, 2006; P. Walker and Smith, 1985; L. Walker, et al., 2012). Interestingly, this difference in the role of size for pitch and brightness judgments reflects the finding which approached significance in Experiment 4 (see Section 3.3.2.3).

This pattern of findings suggests that the role of size in influencing judgements of brightness and pitch was primarily through its influence on perceived heaviness; which subsequently influenced judgements of brightness and pitch. For brightness it is suggested that this is the only influence size had on ratings. However, for pitch, the findings suggest that although size primarily influenced ratings via its impact on perceived heaviness, it also had an additional independent influence on pitch ratings.

### 4.4.2 Dominance of Heaviness Over Size

It was demonstrated that the size of the objects was secondary to that of heaviness in influencing brightness and pitch judgements (the heaviness of the objects, more so than size, predicting the extent to which the objects are rated to be darker or lower in pitch). This finding is consistent with the results of Experiment 4 which also demonstrated that judgements about brightness and pitch were based on the heaviness of objects despite contradictory information about size. Marks (1989) observed a similar pattern of results when exploring the independent contributions of pitch and loudness when people were asked to map sounds varying in these features onto lights varying in luminance. In that case, pitch was found to be more dominant than loudness in determining judgements of luminance. By essentially pitting two feature dimensions against one another, to determine how they each influence judgements of other feature dimensions provides an interesting opportunity to further probe the organisation of cross-sensory correspondences. In order to gain a clearer idea about this, much more work is required. However, for now it is interesting to acknowledge several possible explanations for why heaviness in particular may have a larger influence on brightness and pitch compared to size.

Dominance, or primacy, of one feature dimension over another in influencing cross-sensory judgements may either reflect the nature of the cross-talk between different pairs of associated feature dimensions. Or alternatively, reflect the extent to which one particular dimension is more salient or dominant in a particular set of circumstances. In this case, the associations shared between heaviness with brightness and pitch may be stronger than those shared between size with brightness and pitch. Differences in strength may reflect the extent to which two feature dimensions correlate with one another in the environment or are reinforced by correlates in language. Or heaviness may be a much more salient or prominent feature compared to size in general, or when objects are being lifted.

### 4.4.3 Unitary versus Separate Feature Dimensions

Another important question is whether a different unitary feature which emerges as a result of size and weight, e.g. object density, formed the basis of pitch and brightness judgements. A similar possibility was acknowledged by Marks (1989) who explored whether the associations of pitch and loudness each with luminance may in fact be caused by a unitary feature, auditory density, a dimension which can be understood as a product of the relationship between the other two features (pitch and loudness). However, the findings from Marks (1989) suggested that auditory density was not an adequate explanation for luminance judgements. This is in keeping with the notions of parallel feature dimensions proposed by

Karwoski et al. (1942) and by P. Walker and Walker (2012) (see also P.Walker, 2016). Similarly, there are several aspects of the present analysis which suggest density can be ruled out as a possible explanation for brightness and pitch judgements. Firstly, a model in which the density values for each object were included as an explanatory variable did not improve model fit of judgements of brightness and pitch[4] compared to a model containing size and weight as separate dimensions. Furthermore, as can be demonstrated by pairing sub-set A, objects of the same density were rated as varying in brightness and pitch. If density formed the basis of brightness and pitch ratings, it would be expected that these objects would have been rated as equal in brightness and pitch.

Attempts to understand heaviness perception and particularly the size-weight illusion, have often appealed to explanations of perceived heaviness capturing a unitary feature dimension which can be understood as a function of both size and weight such as density (Ross and Di Lollo 1970; Stevens and Rubin 1970), or rotational inertia, the resistance to motion around a rotational axis, which is dependent on mass distribution (Amazeen and Turvey, 1996) rather than it resulting from an interaction between size and weight as two independent feature dimensions. Despite face validity, there is evidence in the weight-illusion literature which suggests that density is not a satisfactory explanation for perceived heaviness (Buckingham, 2014). What is more, Buckingham (2014) argues that it seems unreasonable to argue that one feature is mis-represented, while another related feature is accurately

[4] The density values for each object were used as an explanatory variable in a model which once again included a random effect of participant. For brightness, although a density model was preferred to a null model ($\chi^2$=425.26, p<.001); a comparison of the AIC values suggests that the Density Model (AIC=8385.5) was not preferred to a model which included both size and weight (AIC=8188.0). The same pattern of findings was found with pitch and heaviness also.

perceived. To explore the possibility that perceived heaviness was a miss-attribution of object density as heaviness in the present study, a model of perceived heaviness which included the object density as an explanatory variable was compared to a model including size and weight[5]. However, the density model was not preferred. This suggests that perceived heaviness is not in fact a representation of density.

Evidence from the weight-illusion literature argues that the perceived heaviness of objects is in some ways dependent on the relationship between size and weight as separate dimensions as opposed to an alternative unitary property. For example, Flanagan et al. (2008) demonstrated that the size-weight illusion can be removed and reversed through extensive exposure to objects with an inverse size-weight relationship. This suggests that the relationship between size and weight (that bigger suggests a heavier object and smaller suggests a lighter object) has a key role in influencing the perceived heaviness of objects in the size-weight illusion. If the size-weight illusion could be explained as an accurate perception of another property, e.g. a unitary dimension such as density, or rotational inertia, it would not have altered with exposure to objects that have a different size-weight relationship, since the test objects continued to have the same physical properties as before. This therefore suggests that violation of size-based expectation has some role in the size-weight illusion. As such, it indicates that it is meaningful to describe size and weight as separable dimensions. Furthermore, the demonstration that size continued to have an influence on pitch judgements in addition to its effect on perceived heaviness adds further support to the idea of them being separable dimensions.

[5] The AIC values suggest that a model including size and weight as predictors of perceived heaviness was preferred to a model with density as a predictor variable with AIC=7272.3 and 8361.0 for the size and weight and density models respectively.

## 4.5 Summary of the Thesis Thus Far

The aim of the thesis thus far has been to determine if the heaviness of lifted objects induce associations with other sensory feature dimensions implicated in the correspondences literature. Through a series of rating scale tasks, it has been demonstrated that the felt heaviness of objects are associated with being lower in pitch, darker, bigger and rounder (when presented as a verbal antonym to “pointy”) than less heavy objects. Brightness and pitch associations can be induced by heaviness independently of size and what is more, the heaviness of objects overrides conflicting information about the object’s size to form the basis of judgements about brightness and pitch. When taken alongside previous findings, the demonstration that heaviness induces cross-model correspondences with these feature dimensions suggests that the associations are bi-directional in nature (Alexander and Shansky, 1976; L. Walker, et al., 2012). These findings are consistent with a theoretical framework of correspondences arising at a conceptual level that are amodal and abstract in nature. What is more, these findings confirm that heaviness can be included among the set of feature dimensions argued to correspond at this level of processing. The second half of the thesis aims to add further support for the correspondences of heaviness with brightness and pitch by exploring if they give rise to congruity effects in speeded-classification tasks.

# Chapter 5

## 5.1 Introduction

Over the previous chapters, it has been demonstrated that the felt heaviness of lifted objects induces systematic associations with both brightness and pitch such that heavier objects are rated as darker, and lower in pitch, than less heavy objects. The pattern of these associations is consistent with the alignment predicted by the framework put forward by P. Walker, and Walker (2012). According to this framework, cross-sensory correspondences emerge as a result of cross-talk between relative extremes of various feature dimensions at a conceptual level. The findings from the previous rating scale tasks suggest that heaviness can be considered among dimensions implicated in this network of feature dimensions.

Although rating scale studies have been used to demonstrate the occurrence and direction of many cross-sensory correspondences (Eitan and Timmers, 2010; P. Walker and Smith, 1984, 1985: L. Walker, et al., 2012); on their own, they are somewhat limited in what they can tell us. Since participants are free to apply cognitive flexibility in their approach to aligning dimensions, they may be vulnerable to demand characteristics. For example, participants can impose a superficial, systematic method of responding or make judgements in anticipation of the purpose of the research. It has been demonstrated in some cases that feature dimensions that are found to correspond in tasks requiring explicit mapping between dimensions do not necessarily produce compatibility effects in information processing or perceptual judgement tasks, for example an association between hardness and brightness has been observed in a rating task where people are asked how these dimensions might go together despite these dimensions not spontaneously influencing responses in a perceptual judgement task (Ernst, 2007; Parise, 2016). Therefore, to establish if explicit mapping of dimensions reflect any underlying correspondences, it is important to corroborate findings from rating scale tasks with evidence from tasks which are less vulnerable to demand characteristics. For that reason, the series of experiments reported in the following chapters explores whether the findings from the first half of the thesis are supported by evidence from the speeded classification task.

As described in Section 1.1.4, the speeded classification task has often been used to test for interactions between feature dimensions in the correspondences literature (see Marks, 2004 and Spence, 2011 for reviews). In the speeded classification task, participants are presented with stimuli simultaneously varying along two feature dimensions, for example, visual images varying in brightness and auditory tones varying in pitch (Marks, 1987). Participants are asked to classify one of the features while ignoring the other (e.g. pressing response keys to classify the visual stimuli as ‘bright’ or ‘dark’ while ignoring the simultaneously presented sound, and its pitch). Despite one feature being irrelevant to the task demands, it is often found that responses are influenced by the compatibility between the incidental dimension and the to-be-classified dimension in accordance with cross-sensory correspondences. Many of the correspondences demonstrated between features in rating scale tasks have been found to also induce congruity effects in the speeded classification task. For example, participants respond faster when high pitch sounds are presented with stimuli that are brighter (Marks,1987; Martino and Marks, 1999; Malera, 1989), pointier (Marks, 1987), smaller (Evans and Treisman, 2010; Gallace and Spence, 2006) and positioned higher in space (Ben-Artzi and Marks,1995; Bernstein and Edelman, 1971; Evans and Treisman, 2010; Patching and Quinlan, 2002).

The incidental feature does not have to vary orthogonally with the to-be-classified feature to induce a congruity effect. On several occasions, response keys varying on the basis of one feature dimension have been demonstrated to induce congruity effects (Lidji, et al., 2007; Rusconi, et al., 2006; P. Walker and Smith, 1985). For example, P. Walker and Smith (1985) asked participants to press large and small sized objects in response to words reflecting opposite poles of various feature dimensions. Responses were found to be faster when a large object was used to respond to the words *STRONG, HEAVY, DOWN, BOTTOM* and the small object was used to respond to their respective antonyms *WEAK, LIGHT, UP, TOP* compared to the opposite assignment.

Similarly, a size-brightness correspondence has been observed when small and large objects are used as response keys to classify visual stimuli as bright and dark (P. Walker and Walker, 2012; L. Walker and Walker, 2016). Participants were presented with circles on a computer screen and asked to classify them according to whether they were brighter or darker than the grey background upon which they were presented. Participants were instructed to press the key held in either their left or right hand to indicate ‘darker’ or ‘brighter’ circles. The size of the object being used in each hand varied between each block of trials, such that for half of trials the big object was being used for classifying stimuli as dark and small object for bright (congruent trials), and for the other half of trials, the small object was being used for dark and the large object for bright (incongruent trials). It was found that responses were faster when the size of the keys was congruent with the classification being made compared to incongruent, according to the size-brightness correspondence where dark aligns with big and small aligns with bright.

The present study adapts the method used by P. Walker and Walker (2012) by incorporating variation in heaviness. This is to determine whether variation in heaviness has the potential to be an influencing feature dimension in a speeded-classification task. The same brightness classification task was used, but instead of pressing down on objects varying in size to make a response, objects were held in the participant’s hands and used to tap a touch sensitive surface. The objects varied in size but also in heaviness such that the larger object was heavy and the small object was light. Since this is the first attempt to induce a congruity effect with objects being used actively to tap a touch sensitive surface to register a classification decision, both the size and heaviness of the objects were available to potentially induce congruity effects with brightness in accordance with size-brightness and heaviness-brightness correspondences. Since size and heaviness are aligned with each other in a way that is consistent with how both features have been demonstrated to align with brightness. It is predicted that the introduction of heaviness in this case will not change the congruity effect as found by P. Walker and Walker (2012), responses being faster when a bigger/heavier object is used to classify stimuli as dark and a smaller/lighter object is used to classify stimuli as bright compared to the other way round.

## 5.2 Experiment 6

### 5.2.1 Method

#### 5.2.1.1 Participants

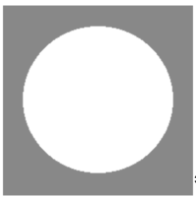
Twenty-nine students from Lancaster University (18 females and 11 males) aged between 18 and 27 (mean age = 21.17 years) volunteered to take part in the study for payment or course credit. All except 4 participants were right handed by self-report. Eighteen participants spoke English as their first language. The remaining 11 participants spoke the following first languages: Bulgarian (n=2), Chinese/Cantonese/Mandarin (n=4), Greek (n=1), Hindi (n=3) and Malay (n=1).

#### 5.2.1.2 Materials

**Stimuli for classification.** The experiment was conducted using PsyScript 2.0 experiment software on a 20in computer screen (Apple A1038, 1,680 x 1,050 cinema back-lit LCD display, controlled by a dual 2 GHz, PowerMac G5). The visual stimuli were four of the six circles used in the study by P. Walker and Walker (2012)[6]. The circles were 4.5cm in diameter and varied in brightness from black to white (340, 150, 42, 2 cd/m$^2$) (see Table 5.1). They were presented in the centre of the screen on a mid-grey background (90 cd/m$^2$). Two brightness values were darker than the mid-grey background and two values were brighter, forming a factor of categorical brightness. More than one brightness value was used in each category to ensure a distinction between absolute or relative brightness mappings in the event of a congruity effect involving categorical brightness. A secondary factor of *within-category brightness* was produced as a result of having two levels in each category, this referred to the darker or lighter brightness value within each category. Although an effect of within-category brightness is not predicted, the potential influence of this factor was explored within analysis.

Table 5.1

*The circles used in the brightness speeded classification task, originally used by P. Walker and Walker (2012)*

| | Categorical Brightness | |
|---|---|---|
| Within-Category Brightness | Dark | Bright |
| Dark |  * |  |
| Bright |  |  * |

[6] It was necessary to reduce the length of the experiment because holding the objects for prolonged periods of time became uncomfortable. Removing two of the levels of brightness used by P. Walker and Walker (2012) (the middle value within each brightness category) meant there were a reasonable number of trials in each condition (congruent vs incongruent) for each level of brightness.

Note: those marked with * signify High level of brightness contrast compared to the mid grey background.

**Response keys.** The biggest/heaviest and smallest/lightest objects from the full set of nine (i.e**.** Objects A and I, see Figure 5.1) were used to tap the touch sensitive surfaces in order to make a response. The touch sensitive surfaces were two circular metal disks (4cm in diameter) mounted onto wooden frames of 1.0 cm depth. The touch-sensitive surfaces were interfaced to the computer keyboard, allowing the classification decision, and the speed with which it was made, to be recorded when tapped by the object. The touch sensors were covered with a layer of thick felt material to dampen the impact sound of the objects when making contact with the sensors. In order to completely mask any sound from the objects, a soundtrack was played to the participant through headphones from a laptop during each block of trials. The track was the sound of random computer keyboard typing which varied in rhythm and tempo in a non-systematic way. This particular soundtrack was used to prevent participants synchronising their responses in time to a set predictable beat. During the procedure the objects and sensors remained hidden from view, by being covered with thick black material.

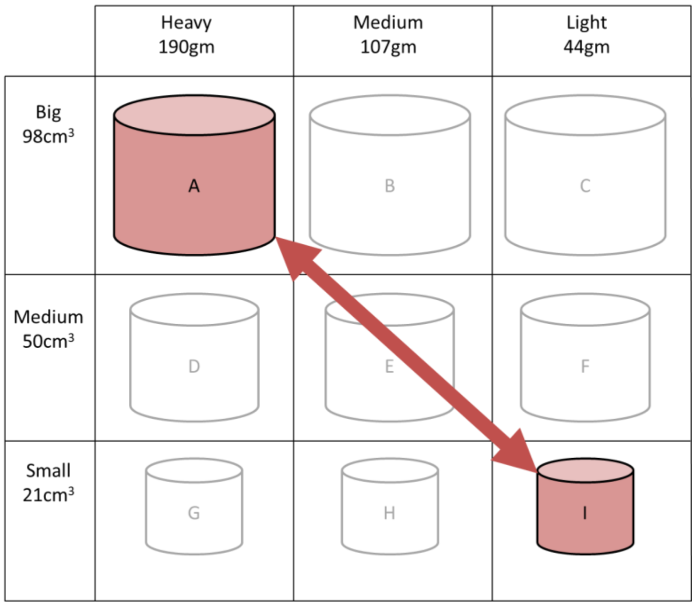



*Figure 5.1.* The objects from the full set of nine used as response keys in Experiment 6

### *5.2.1.3 Design and Procedure*

Participants completed four blocks of 36 trials. On each trial a circle at one of four brightness levels was displayed on the computer screen. Participants responded, as quickly as possible, indicating whether the circle on the screen was brighter or darker than the mid-grey background, by tapping one of the two cylinders (one being held in each hand) on the touch sensitive surface. Half the participants were instructed to tap with their left hand for bright circles and their right hand for dark circles. The other half of participants did this the opposite way. The left-right assignment to bright or dark responses remained the same for each participant throughout the task. A small label at the bottom of the screen reminded the participant which hand they were to use to respond to ‘bright’ and which for ‘dark’. The objects being used in the left and right hand were alternated between blocks. The starting left and right assignment of each object was counterbalanced independently of the assignment of response hands to bright or dark responses.

Participants held an object in each hand underneath the material that was covering the objects and touch sensors such that the corner edge of each cylinder could be used to make contact with the touch sensor. Participants were asked to ensure that the objects were not being rested on the surface of the sensor and to hold them approximately 1.5 cm above the appropriate surface. The difference in size and heaviness between the two objects was not mentioned by the researcher. When required, the participants would tap the objects, using the corner edge, onto the touch sensor before returning to the initial starting position. Participants were allowed to lean on the table with their lower arms in order for them to maintain the position for the duration of the task. Before each block, participants were given the opportunity to practice tapping the objects on the sensors. During this practice session, they received feedback about the response having been registered, the task proper commenced when both participant and researcher were satisfied with the accuracy of responses.

When the object made contact with the sensor the circle disappeared. Response time was measured from the onset of the stimulus to the time a response was registered. After a response was registered, there was a two second interval before the next circle was presented, in this time only the mid-grey background was displayed. Within each block of trials, each circle was shown 9 times. The order in which the circles were presented was randomised across sets of 12 trials, so that within each set each brightness level appeared 3 times. Each block of trials took approximately 2 minutes. Between each block, participants moved to another desk for a two minute break. During this time, they had a word search to complete. At this point the researcher swapped the objects on the left and right sensors, with the effect that participants performed each proceeding block of trials with the opposite object-brightness response mapping. This was not mentioned by the researcher, however, some participants reported being aware that the objects had been switched.

During each block, participants listened to a soundtrack through headphones. It was explained to participants that the soundtrack was to mask any general external sounds and that it was not necessary for them to pay any attention to it. The volume of the sound track was determined prior to the first block of trials to ensure it was at a comfortable level.

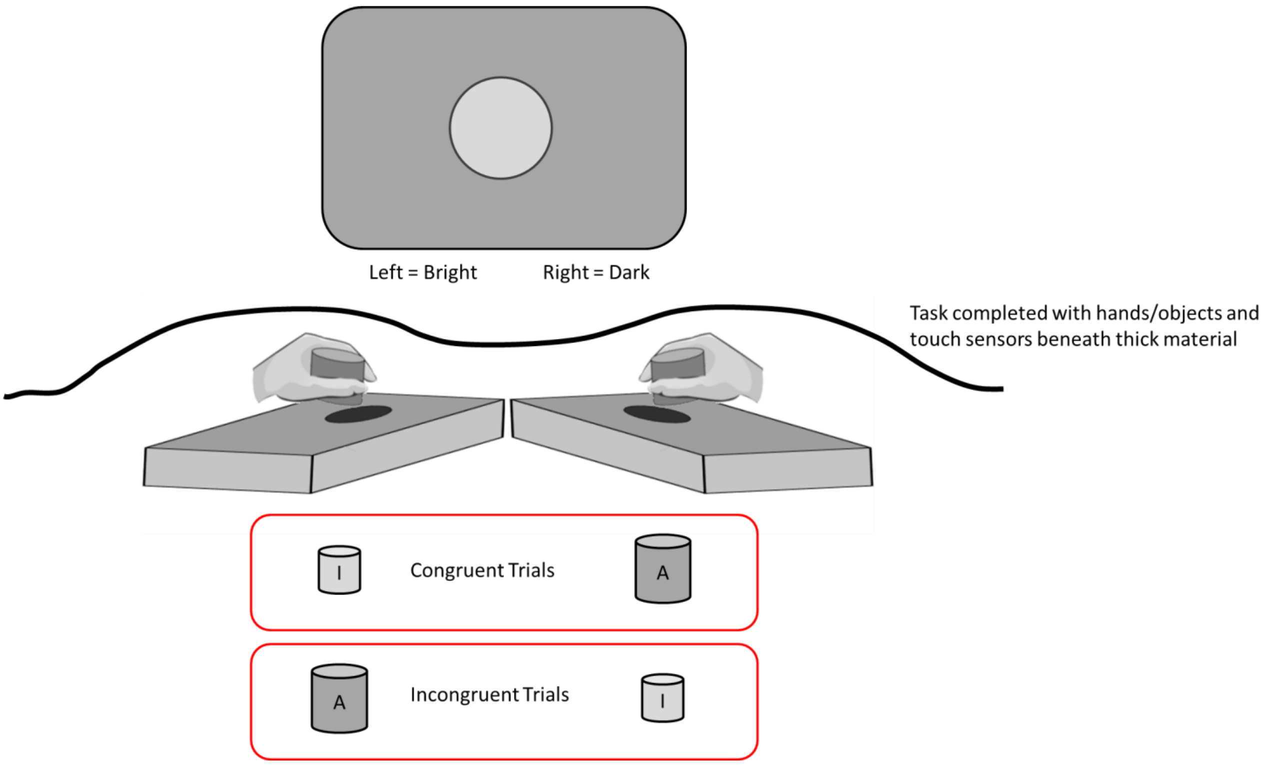



*Figure 5.2.* A depiction of the experimental apparatus. Participants were asked to classify a series of circles presented on screen as either bright or dark with their left and right hand. The objects and sensors remained hidden from view beneath thick material. For half of trials the objects were congruent: the smaller/lighter object was used to classify stimuli as bright and the bigger/heavier object to classify stimuli as dark. The other half of trials were incongruent and the objects were used to make the opposite brightness classifications.

### 5.2.2 Results

The data were the accuracy levels and the correct response times (RTs) for the brightness classifications. Out of 4176 trials, 4099 of these were correct responses (98.16%).

RTs less than 300ms and more than 2.5 standard deviations above each participant's mean RT were excluded from analysis leaving 3964 observations (94.92%). The accuracy and RTs for *Bright* and *Dark* classifications made with the *Big/Heavy* and *Small/Light* objects are summarised in Table 5.2.

*Table 5.2.*

*Mean RTs (SEM in parentheses) and accuracy levels according to the object used and categorical brightness.*

| | | Brightness | |
|---|---|---|---|
| Object Used | | Dark | Bright |
| Big Heavy | Mean RT | **757 (12)** | 850 (15) |
| | Accuracy | **98.7%** | 98.0% |
| Small Light | Mean RT | 798(13) | **797 (13)** |
| | Accuracy (%) | 97.9% | **98.1%** |

Note: Bold typeface signifies congruent conditions

#### ***5.2.2.1 Response Speed***

Due to the repeated measures design, a linear mixed-effects analysis was performed using *lme4* (Bates, et al., 2014) in R version 3.2.0 (R Core Team, 2012). In order to resolve the skew of the RT distribution, a transformation of 1/RT was performed. The product of this transformation is interpreted as *response speed* (responses/second). A series of intercept only models, which included a random effect of *participant,* were developed to assess the explanatory variables that contributed to *response speed*. The contributions of any main effects and interactions were determined using Likelihood-Ratio Tests (LRTs) which compared a model including the particular variables with a model that did not. Alkaike Information Criterion (AIC) values provided a relative estimate of the amount of information not being captured by a model, balancing goodness of fit with the number of parameters the model contains. The AIC for alternative models were compared, with lower AIC values indicating a superior model. 95% confidence intervals were calculated using the *Wald* method with the *confint()* function.

The aim of the analysis was to determine if an interaction between *object* (*big/heavy* Vs *small/light*) and *categorical brightness (dark* Vs *bright*) significantly improves model fit. For ease of interpretation this was included as a main effect of *congruence (congruent trials* Vs *incongruent trials)* along with main effects of *categorical brightness* and *object*. The LRT demonstrates that a model including *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2(1) = 21.365$, $p<.001$) with an AIC value of 5223.8 and 5204.4 for the null and comparison model respectively. The parameter estimates of this basic model suggest that congruent trials had a speed which was faster on average by 0.068 responses/second (SE=0.02), CI [0.039, 0.096].

The contribution of the following additional variables was assessed along with interactions between variables which may also have theoretical implications: *trial*, *response hand* (*left* Vs *right*), *brightness contrast* (*high* Vs *low)* and *within-category brightness* (Whether the brightness value is the *dark* or *bright* value of the two sharing a categorical brightness label, see Table 5.1). A model including these additional explanatory variables was preferred to the basic model according to the LRT ($\chi^2(4) = 277.7$, $p<.001$) with an AIC value of 4934.7. An inspection of the CIs demonstrated that *within-category brightness* CI [-0.019, 0.036] and *response hand* CI [-0.05, 0.003] did not have an effect since the values straddle zero. LRT confirmed their removal did not change the explanatory power of the model ($\chi^2(2) = 3.38$, $p$=.18, AIC = 4934.1) therefore the model which excluded these variables was preferred. An interaction of *within-category brightness with object* also failed to improve model fit ($\chi^2(2) = .99$, $p$=.61, AIC=4937.1). As did interactions of *response hand* with *object* ($\chi^2(2) = 4.64$, $p$=.10, AIC=4933.5), *categorical brightness* ($\chi^2(2) = 3.07$, $p$=.22, AIC=4935.1) and *brightness contrast* ($\chi^2(2) = 5.39$, $p$=.07, AIC=4932.7).

Parameter estimates indicate that the speed of responses increased with *trial*. Responses were also faster when *brightness contrast* was *high* i.e. when the brightness value of the circle was further from the mid-grey background (black and white as opposed to dark grey or light grey). An interaction between *object* and *brightness contrast* failed to improve model fit ($\chi^2(1) = 0.056$, $p$=.81, AIC=4936.1). The LRT for the final model demonstrates that the inclusion of *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2(1) = 22.20$, $p$<.001) with an AIC value of 4954.3 and 4934.1 for the model excluding and including *congruence* respectively. The parameter estimates of this final model suggest that incongruent trials were slower than congruent trials by 0.067 responses/ second (SE=0.01), CI [0.039, 0.094]. Figure 5.3 depicts the congruity effect between *Categorical brightness* and *object* and Table 5.3 summarises the parameter estimates of the final model. Visual inspection of residual Q-Q plots did not reveal any obvious departures from normality. In addition inspection of residual plots demonstrated that the assumptions of linearity and homoscedasticity were also upheld.

Table 5.3

*Summary table for the final model for Experiment 6*

| Fixed effects (criterion) | Estimated coefficient | SE | Wald confidence intervals 2.5% | Wald confidence intervals 97.5% | t-value |
|---|---|---|---|---|---|
| (Intercept) | 1.49 | 0.054 | 1.38 | 1.59 | 27.59 |
| Trial | 0.001 | 0.0002 | 0.001 | 0.002 | 7.88 |
| Brightness Contrast (Low) | -0.21 | 0.014 | -0.24 | -0.18 | -14.97 |
| Object (Small/Light) | 0.006 | 0.014 | -0.022 | 0.034 | 0.42 |
| Categorical Brightness (Dark) | 0.066 | 0.014 | 0.038 | 0.094 | 4.69 |
| Congruence (Incongruent) | -0.067 | 0.014 | -0.094 | -0.039 | -4.72 |

| Random effects | Name | Std. Dev | Variance |
|---|---|---|---|
| Participant | (Intercept) | 0.073 | 0.27 |
| Residual | | 0.20 | 0.44 |

| AIC | BIC | Loglik | Deviance |
|---|---|---|---|
| 4934.1 | 4984.4 | -2459.1 | 4918.1 |

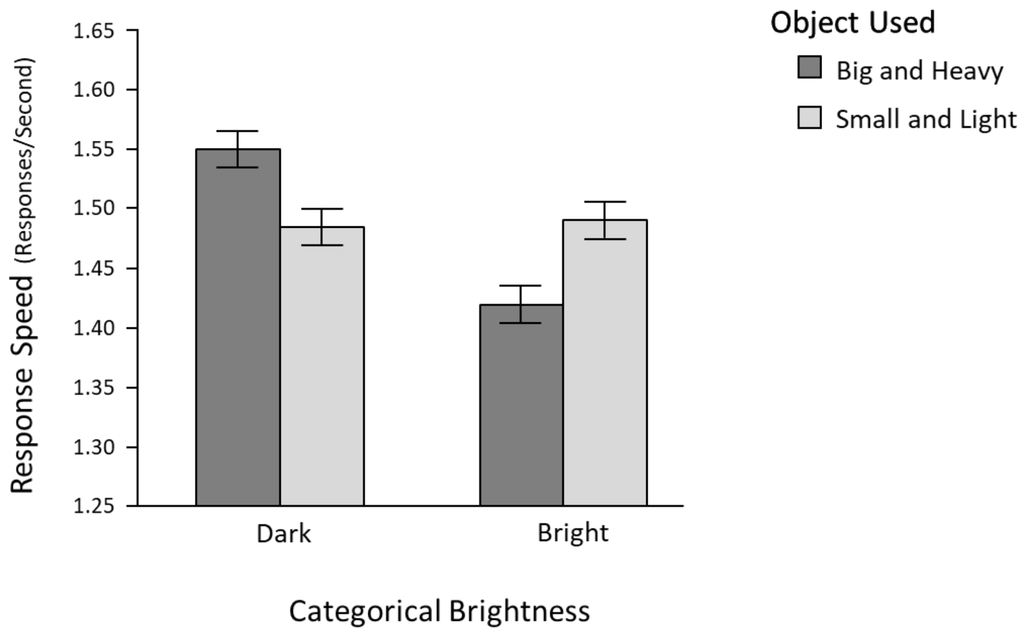



*Figure 5.3.* Mean response speed (responses/second) for responses to the dark and bright circles with each object. Error bars refer to the standard error of the mean.

#### *5.2.2.2 Response Accuracy*

The overall response accuracy was 98.16%. The numbers of incorrect responses were too few to warrant the same form of analysis of error rates as was used on RTs. As an alternative, the number of correct responses for congruent and incongruent trials were analysed using a related-samples Wilcoxon Signed Ranks Test. This demonstrated no significant difference between congruent and incongruent trials based on accuracy ($Z$=-1.214, $p$=.225). This confirms that the difference in speed for congruent and incongruent trials did not result in a speed/accuracy trade off. Furthermore, it suggests that congruity did not influence response accuracy.

### 5.2.3 Discussion

The aim of the present study was to determine if objects being used to tap a touch sensitive surface could be used as response keys in a speeded classification task. It was demonstrated that the object being used did affect the speed with which classifications about the brightness of visual stimuli can be made. Classifications were faster when the *small/light object* was used to classify stimuli as *bright* and the *big/heavy* object was used to classify stimuli as *dark* compared to when they were used the opposite way. The direction of this congruity effect is in keeping with what was predicted based on the brightness-size correspondence demonstrated by P. Walker and Walker (2012) and the brightness-heaviness correspondence demonstrated thus far in the present thesis. This suggests that heaviness can be incorporated as a task-irrelevant feature in a speeded classification task.

#### *5.2.3.1 Interpretation of the Final Model*

There were three attributes of the visual stimuli which had the potential to influence the speed of responses: *within-category brightness*, *brightness contrast* and *categorical brightness*. It was found that *within-category brightness* did not contribute to the model nor did an interaction of *within-category brightness* with *object.* This indicates that the relative position of a brightness value within a brightness category did not influence responses; suggesting that brightness was defined at the level of the classification decision being made in the task. Furthermore, the lack of interaction between *within-category brightness* and *object* means that the size and/or weight of an object did not interact with whether the particular stimulus it was being used to classify was brighter or darker (than the other level of brightness being classified by that object). This again suggests that the congruity effect interferes at the decisional/response selection level. *Brightness contrast* contributed to the model such that responses were faster for the more extreme values of brightness (which were in greater contrast with the mid-grey background) compared to the intermediate values. This is likely to reflect the influence contrast has on the time taken to determine whether the circle is brighter or darker than the mid-grey background (greater contrast being easier to determine). The lack of interaction between *brightness contrast* and *object* indicates that a congruity effect did not arise on the basis of intensity, which may have been predicted since both mass and contrast can be considered protothetic dimensions (having poles which reflect *more* and *less* of a quality). As well as the congruity effect between *categorical brightness* and *object*, inspection of the means and parameter estimates indicate that responses were faster for dark classifications compared to bright classifications. This finding was also observed by P. Walker and Walker (2012) and is argued to account for the asymmetry in the congruity effect which can be seen in Figure 5.3. There was no interaction between response hand and any attribute of the visual stimuli or object. This suggests that neither the dimensions of size (or heaviness) nor brightness mapped onto a spatial left-right dimensions.

P. Walker and Walker (2012) argue that the congruity effect they observed between brightness and size demonstrates that an interaction of these features occurs at a higher level of processing, between abstract semantic representations of these two feature dimensions. In that study, as in the present experiment, each stimulus being classified as ‘bright’ or ‘dark’ could have one of a number of absolute brightness values. Therefore, in order to be classified, its relative brightness (compared to the mid-grey background) needed to be established. L. Walker and Walker (2016) further demonstrate that it is the categorical identity of brightness, as opposed to the absolute values of brightness that is responsible for the congruity effects observed by showing that the same levels of brightness can correspond with the larger or smaller object depending on its relative brightness. Another feature of the task which suggests the correspondence arises at this level is that the object response keys were being perceived continuously and simultaneously throughout the task. Therefore, one or other object becomes more or less salient only once the visual stimulus is classified, placing the interference at a “post-categorical” level of processing (L. Walker and Walker, 2016).

According to P. Walker and Walker (2012) the visual stimulus results in activation of an abstract, amodal representation of brightness. Cross-talk can then occur between this representation of brightness and corresponding extremes of other feature dimensions with which brightness is aligned, including size. The cross-activation of brightness with size makes the bigger object more salient when the visual stimulus is dark and the smaller object more salient when the visual stimulus is bright (see Figure 5.4). As a result, the response times are faster when the more salient object is also the correct response key in the task (congruent trials) compared to when it is not (incongruent trials). Since there were no baseline trials, it is uncertain whether the observed congruity effect is a result of facilitation, interference or a combination of both. However, the presence of a congruity effect demonstrates that cross-talk between the dimensions has an influence on task performance, which, none the less, points to an alignment of brightness with size and/or weight at a conceptual level. In the case of the present study, the objects varied in size and heaviness. Since both of these features have been demonstrated to align with brightness in the same direction, it is unclear whether the congruity effect between object and brightness is due to the dimension of size or heaviness.

#### *5.2.3.2 Summary*

The aim of this chapter was to replicate the object-brightness congruity effect found by P. Walker and Walker (2012) with the introduction of heaviness as a property of the response keys. The congruity effect was successfully replicated when objects varying in size and heaviness were used to tap a touch sensitive surface in order to make a response. This indicates that heaviness as a feature dimension can be available as an incidental feature of response keys in a speeded classification task paradigm with the potential to interact with the classification decision being made. For the current study it cannot be determined whether the congruity effect was induced by the heaviness or size of the object (or both) because both the size-brightness and proposed heaviness-brightness correspondence align in the same direction. In Chapter 6, the size and heaviness of the object response keys are varied independently to determine the separable contributions of size and heaviness.

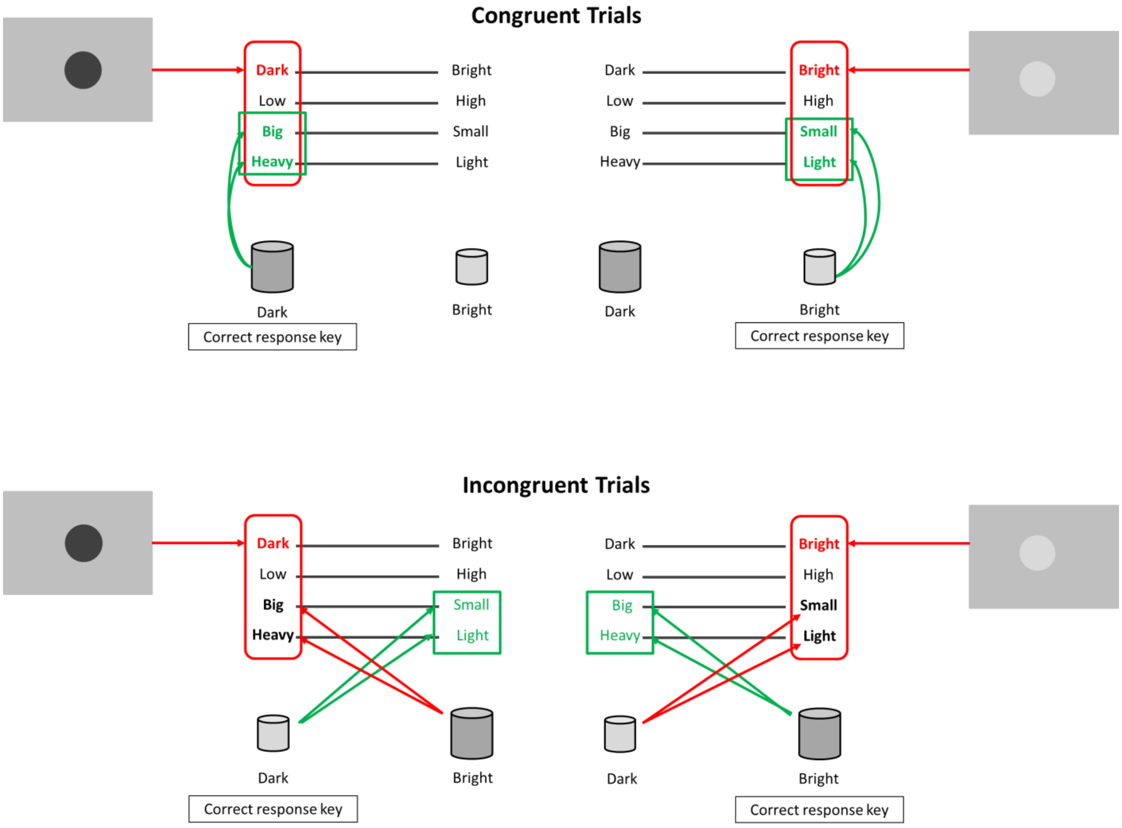



*Figure 5.4.* A diagram of the congruity effect as explained by cross-activation of amodal representations of brightness with size/heaviness. The brightness of a circle induces an amodal representation of one or other extreme of the brightness dimension. The alignment of feature dimensions subsequently results in the cross activation of other extremes of corresponding feature dimensions including extreme of size and/or heaviness, which are subsequently activated. In congruent trials this is consistent with the appropriate object to be pressed in response to brightness classification, in incongruent trials it is not.

# Chapter 6

## 6.1 Introduction

In Chapter 5, a method for including lifted objects as an incidental feature in a speeded classification task was explored. It was found that when two objects were used as response keys to classify stimuli varying in brightness, an interaction between the object being used and the categorical brightness was found. Responses were faster when the small/light object was used to classify stimuli as bright and the big/heavy object was used to classify stimuli as dark, compared to the other way around. The congruity effect is in keeping with what may be predicted by both a size-brightness correspondence (P. Walker and Walker, 2012) and the proposed heaviness-brightness correspondence being explored in the current thesis. Since, size and heaviness were varied in correlation with one another; it remains uncertain whether the size or heaviness of the objects (or both) was responsible for inducing a congruity effect with brightness. Nonetheless, this indicates that lifted objects being used to tap a touch sensor have the potential to give rise to congruity effects which may provide further evidence of a brightness-heaviness correspondence.

The aim of the experiments reported in the present chapter, is to determine the extent to which size and heaviness may each contribute to a congruity effect with brightness. The heaviness and size of the objects used as response keys are varied across three experiments, adopting the same brightness speeded classification task used in Experiment 6. In Experiment 7, the pair of objects used as response keys vary in weight (and therefore heaviness) while size is held constant. In Experiment 8, the objects are varied in size while weight is held constant. And finally, in Experiment 9, both size and weight are varied in a way that contradicts the natural alignment between these two feature dimensions (the larger object was lighter in weight than the smaller object).

## 6.2 Experiment 7

The present experiment aims to determine if the correspondence between heaviness and brightness that has been observed consistently across Chapters 2-4can be found to have an involuntary influence on task performance using the same brightness speeded classification task introduced in Chapter 5. To this aim, participants were asked to classify stimuli as bright or dark (compared to the mid grey background) by using two objects as response keys. The pairs of objects used by participants as response keys in this case were of equivalent size but varied in weight. Based on the nature of the alignment between brightness and heaviness found in the first half of the thesis, it is predicted that responses will be faster when the heavier object is being used to classify visual stimuli as dark and when the lighter object is being used to classify visual stimuli as bright compared to the other way round.

### 6.2.1 Method

#### *6.2.1.1 Participants*

Sixty students from Lancaster University volunteered to take part in the study for payment or course credits. The data for three participants were excluded from analysis. One participant was removed because of an equipment malfunction and the other two because half of responses were incorrect[7]. Of the remaining 57 participants (46 females and 15 males) between the ages of 18 and 48 (mean age= 19.14 years), 8 participants were left handed by self-report. Thirty-seven of the participants spoke English as their first language. The remaining 20 spoke the following first languages: Afrikaans (n=1), Chinese/ Cantonese/ mandarin (n=12), Hungarian (n=1), Igbo (n=1), Italian (n=1), Lithuanian (n=1), Malay (n=1), Norwegian (n=1) and Russian (n=1).

---

[7] It is suspected that these participants based their answers on the object's heaviness as opposed to which object was in their left or right hand.

### *6.2.1.2 Materials, Design and Procedure*

The materials, design and procedures were the same as Experiment 6. However, the experiment differs with regard to the objects being used as response keys. Three pairs of objects made up of the heaviest and lightest object at each level of size from the full set of nine objects were used (i.e., object pairs A & C, D & F and G & I, see Figure 6.1). One group of participants completed the task with each pairing.

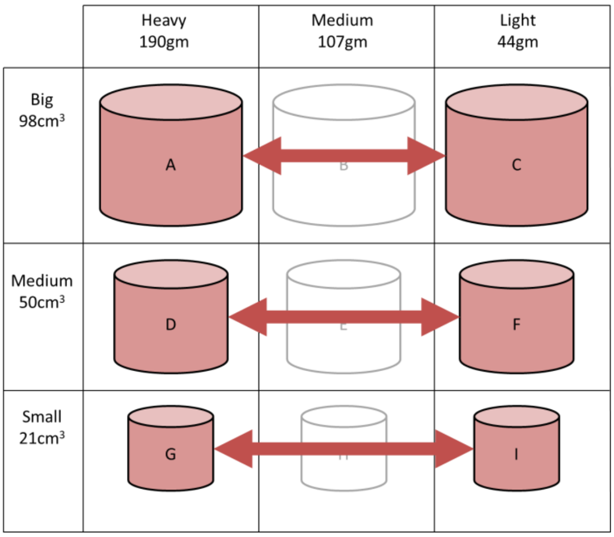



*Figure 6.1.* The objects from the full set of nine used as response keys in Experiment 7

## 6.2.2 Results

The data were the accuracy levels and the correct response times (RTs) for the brightness classifications. Out of 8208 trials, 8120 of these were correct responses (98.9%). RTs below 300ms or 2.5SD above the mean were excluded from analysis leaving 7875 observations (95.9%). Table 6.1 summarises the accuracy and RTs for *Bright* and *Dark* classifications made with the *Heavy* and *Light* objects.

Table 6.1

*Mean RTs (SEM in parentheses) and accuracy levels according to the object used and categorical brightness*

| Object Used | | Brightness | |
|---|---|---|---|
| | | Dark | Bright |
| Heavy | Mean RT(SE) | **778 (9)** | 890 (12) |
| | Accuracy | **99.1** | 98.6% |
| Light | Mean RT(SE) | 847 (11) | **828 (10)** |
| | Accuracy (%) | 98.6% | **99.4%** |

Note: Bold typeface signifies congruent conditions

#### *6.2.2.1 Response Speed*

For analysis, the same approach as Experiment 6 was adopted. In order to resolve the skew of the RT distribution, a transformation of 1/RT was performed. The product of this transformation is interpreted as *response speed* (responses/second). The main aim of the analysis was to determine if an interaction between *object* (*heavy* Vs *light*) and *categorical brightness (dark* Vs *bright*) significantly improves model fit. For ease of interpretation this was included as a main effect of *congruence (congruent trials* Vs *incongruent trials)* along with main effects of *categorical brightness* and *object*. The LRT demonstrates that a model including *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2(1) = 67.67$, $p<.001$) with an AIC value of 9721.2 and 9655.5 for the null and comparison model respectively. The parameter estimates of this basic model suggest that congruent trials had a speed that was faster on average by 0.08 responses/second (SE=0.01), CI [0.06, 0.10].

The contribution of the following additional variables were assessed along with interactions between variables which may have theoretical implications: *trial*, *response hand* (*left* Vs *right*), *within-category brightness (*which is whether the brightness value is the *dark* or *bright* value sharing a categorical brightness level, see Table 5.1), *brightness contrast* (*high* Vs *low)*and *pair size.* A model including these additional explanatory variables was preferred to the basic model according to the LRT ($\chi^2(5) = 594.25$, $p<.001$) with an AIC value of 9071.3. An inspection of the CIs demonstrated that *within-category brightness* CI [-0.02, 0.02] and *pair size* CI [-0.02, 0.18]  did not contribute to the model since the values straddle zero. LRT confirmed their removal did not change the explanatory power of the model ($\chi^2(2) = 2.48$, $p=.29$, AIC=9069.7) therefore the model which excluded *within-category brightness* and *pair size* was preferred. An interaction of *object* and *within-category brightness* also failed to improve model fit ($\chi^2(2) = 0.053$, $p=.97$, AIC=9073.7)

The parameter estimates indicate that the speed of responses increased with *trial*. Responses were also faster when the *brightness contrast* was *high* i.e. when the brightness value of the circle was further from the mid-grey background (black and white as opposed to dark grey or light grey). An interaction between *object* and *brightness contrast* failed to improve model fit ($\chi^2(1) = 53$, $p=.47$, AIC=9071.2). *Response hand* improved model fit such that responses were faster when the left hand was used compared to the right hand. *Response hand* also interacted with *object* ($\chi^2(1) = 14.912$, $p<.001$, AIC=9056.8), but not with *categorical brightness* ($\chi^2(1) = 1.025$, $p=.3114$, AIC=9070.7) nor *brightness contrast* ($\chi^2(1) =.0048$, $p=0.945$, AIC=9071.7). The interaction between *response hand* and *object* was such that responses were faster when the right hand was holding the heavy object and the left hand was using the light object compared to the other way around.

The LRT for the final model demonstrates that a model including *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2(1) = 75.45$, $p<.001$) with an AIC value of 90.56.8 and 9130.3 for the models including and excluding *congruence* respectively. The parameter estimates of this final model suggest that incongruent trials were slower than congruent trials by 0.083responses/second (SE=0.02), CI [0.06, 0.10]. Figure 6.2 depicts the nature of the congruity effect between *categorical brightness* and *object* and Table 6.2 summarises the parameter estimates of the final model. Visual inspection of residual Q-Q plots did not reveal any obvious departures from normality. In addition inspection of residual plots demonstrated that the assumptions of linearity and homoscedasticity were also upheld.

Table 6.2

*Summary table for the final model for Experiment 7*

| | | | **Wald confidence intervals** | | |
|---|---|---|---|---|---|
| **Fixed effects (criterion)** | **Estimated coefficient** | **SE** | **2.5%** | **97.5%** | **t-value** |
| (Intercept) | 1.41 | 0.042 | 1.32 | 1.49 | 33.22 |
| Trial | 0.002 | 0.0001 | 0.002 | 0.002 | 15.71 |
| Brightness Contrast (Low) | -0.18 | 0.010 | -0.20 | -0.16 | -18.97 |
| Object (Light) | 0.039 | 0.013 | 0.013 | 0.05 | 2.93 |
| Categorical Brightness (Dark) | 0.065 | 0.010 | 0.046 | 0.084 | 6.80 |
| Response Hand (Right) | 0.006 | 0.013 | -0.02 | 0.03 | 0.47 |

| | | | | | |
|---|---|---|---|---|---|
| Response Hand (Right): Object (Light) | -0.074 | 0.020 | -0.11 | -0.036 | -3.86 |
| Congruence (Incongruent) | -0.83 | 0.010 | -0.102 | -0.06 | -3.86 |

| Random effects | Name | Std. Dev | Variance |
|---|---|---|---|
| Participant | (intercept) | 0.089 | 0.299 |
| Residual | | 0.179 | 0.423 |

| AIC | BIC | Loglik | Deviance |
|---|---|---|---|
| 9056.8 | 9126.6 | -4518.4 | 7865 |

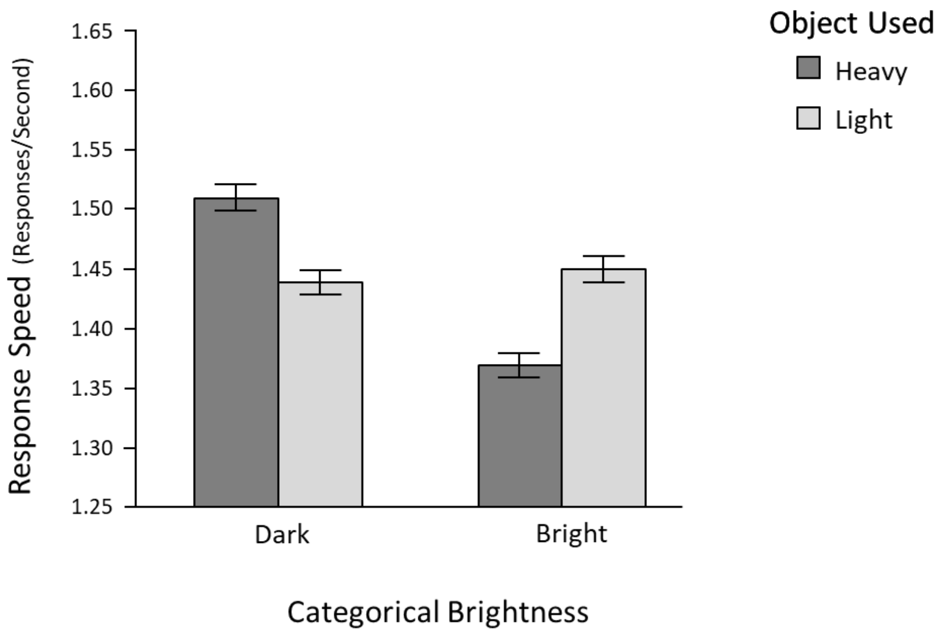



*Figure 6.2*. Mean response speed (responses/second) for dark and bright responses with each object weight. Error bars refer to the standard error of the mean.

#### *6.2.2.2 Response Accuracy*

The overall response accuracy was 98.9%. The percentage correct responses for congruent and incongruent trials were 99.39% and 98.75% respectively. The numbers of incorrect responses were too few to warrant the same form of analysis of error rates as was used on RTs. As an alternative, the number of correct responses for congruent and incongruent trials were analysed using a related-samples Wilcoxon Signed Ranks Test. This revealed a significant effect of congruence on accuracy ($Z$=-2.751, $p$=.006), indicating that responses were more accurate on congruent trials compared to incongruent trials. This suggests the difference in speed for congruent and incongruent results was not the result of a speed/accuracy trade off.

### 6.2.3 Discussion

The aim of the present study was to determine whether the heaviness of objects used as response keys could influence to speed with which people make classification responses for stimuli varying in brightness. The findings demonstrate a congruity effect between the objects and categorical brightness such that responses were faster and more accurate when the heavier object was being used to classify stimuli as dark and the lighter object was being used to classify stimuli as bright, compared to the other way around. The direction of this congruity effect is in keeping with the congruity effect observed in Experiment 6 and the alignment of brightness and heaviness dimensions observed in the first half of the present thesis. For the most part, the final model replicates that of Experiment 6. However, the findings depart from those of Experiment 6 with an interaction between *object* and *response hand*. The nature of this interaction is that responses were faster when the heavy object was used in the right hand and the light object was used in the left hand compared to the other way around. Participants may have had a preference for this hand-object assignment because the heavier object could be held more comfortably in the right (in the majority of cases- dominant) hand compared to the left hand.

#### *6.2.3.1 The Heaviness-Brightness Congruity Effect*

The findings are the first demonstration of a congruity effect involving the felt heaviness of lifted objects. As outlined in Chapter 5, the congruity effect found between brightness and heaviness suggests alignment of these two features at a high cognitive level. According to the interpretation from Section 5.2.3.1, the visual stimulus activates an abstract amodal representation of brightness. The alignment of brightness with other feature dimensions results in the cross-activation of related extremes on other feature dimensions. The present findings suggest that heaviness can be included among aligned dimensions. This means that when a visual stimulus is classified as *dark* this induces associations of *heavy* and similarly when a visual stimulus is classified as *bright* it induces association of *light (less heavy)*. This results in faster responses when the associated extreme of heaviness is the appropriate object to be used to make a correct response. As mentioned in Chapter 5, since there are no base-line trials, it is uncertain whether the congruity effect observed is a result of facilitation, interference or a combination of both. However, the presence of a congruity effect provides support for an interaction between the dimensions of brightness and heaviness which subsequently has an influence on task performance.

This is the first demonstration of a congruity effect between heaviness and another feature dimension in which heaviness is represented through variation in the weight of lifted objects. This adds support to the findings from Experiments 1-5 of a correspondence between brightness and heaviness, in which heavy is aligned with dark and less heavy corresponds with bright. Since in the present study, the heaviness of the objects used as response keys were incidental to the task demands, it suggests that this alignment between brightness and heaviness can influence performance without the conscious deliberation of the participant suggesting that there is an extent to which the mapping between these dimensions is involuntary.

## 6.3 Experiment 8

A size-brightness congruity effect has been demonstrated to emerge where objects varying in size are used as response keys in such a way that the objects are held and pressed to make a response but are not lifted (P. Walker and Walker, 2012; L. Walker and Walker, 2016). In Experiment 6 it was demonstrated that when objects of different sizes were lifted and used actively to make a response, and also varied in heaviness (such that the bigger object is also heavier) the same congruity effect was demonstrated. However, it was uncertain whether the size or heaviness of the objects (or both) was responsible for the congruity effect that emerged. The aim of the present experiment is to determine if size continues to induce a size-brightness congruity effect when the weight of the objects continues to be available to participants through lifting, but is held constant. If size and heaviness are both distinct feature dimensions, entering into correspondence with brightness independently, then it is predicted that the size of lifted objects will continue to enter into correspondence with brightness when the objects are lifted but are of equal weight.

### 6.3.1 Method

#### *6.3.1.1 Participants*

Forty students from Lancaster University volunteered to take part in the study for payment or course credits. The data from three participants were removed from analysis due to equipment malfunction during their performance. Of the remaining participant (31 females and 6 males) between the ages of 18 and 20 (mean age= 18.56 years), all except 6 were right handed by self-report. Thirty of the participants spoke English as their first language. The remaining seven had first languages which included Chinese/ Cantonese (n=5), German (n=1) and Romanian (n=1).

#### *6.3.1.2 Materials, design and procedure*

The materials, design and procedures were the same as Experiment 6. However, the experiment differs with regard to the objects being used as response keys. Two pairs of objects made up of the biggest and smallest object at two levels of heaviness from the full set of nine objects (i.e., object pairs A & G and C & I, see Figure 6.3). One group of participants completed the task with each pairing.

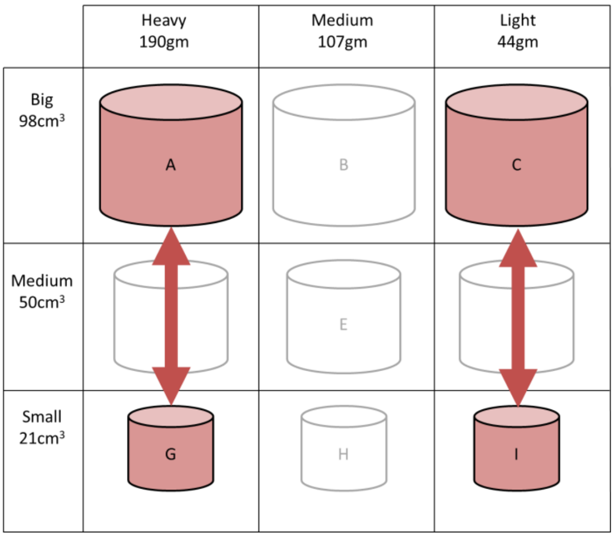



*Figure 6.3.* The objects from the full set of nine used as response keys in Experiment 8

### 6.3.2 Results

The data were the accuracy levels and the correct response times (RTs) for the brightness classifications. Out of 5328 trials, 5243 were correct responses (98.4%). RTs below 300ms or above 2.5SD above the mean were excluded from analysis leaving 5094 observations (95.6%). Table 6.3 summarises the accuracy and RTs for *bright* and *dark* classifications made with the *big* and *small* objects.

Table 6.3

*Mean correct response times (SEM in parentheses) and accuracy levels (%) according to object used and categorical brightness.*

| Object Used | | Brightness | |
|---|---|---|---|
| | | Dark | Bright |
| Big | Mean RT(SE) | **876 (15)** | 874 (13) |
| | Accuracy | **98.5%** | 97.9% |
| Small | Mean RT(SE) | 862 (14) | **897 (14)** |
| | Accuracy (%) | 98.5% | **98.7%** |

Note: Bold typeface signifies congruent conditions

#### *6.3.2.1 Response Speed*

The analysis approach was the same general strategy used in Experiments 6 and 7. In order to resolve the skew of the RT distribution, a transformation of 1/RT was performed. The product of this transformation is interpreted as *response speed* (responses/second).The main aim of the analysis was to determine if an interaction between *object* (*big* vs *small*) and *categorical brightness (dark* vs *bright*) significantly improves model fit. For ease of interpretation this was included as a main effect of *congruence (congruent trials* vs *incongruent trials)* along with main effects of *categorical brightness* and *object*. Congruent trials were defined as those where the big object was used to correctly classify stimuli as dark and the small object was used to correctly classify stimuli as bright. Incongruent were those of the opposite object/brightness mapping. The LRT demonstrates that a model including *congruence* as a main effect is not preferred to a model in which it is not included ($\chi^2(1) = 1.18$, $p$=.28) with an AIC value of 6077.3 and 6078.1 for the null and comparison model respectively. The parameter estimates also reinforce the lack of a significant interaction since the confidence intervals are not of the same sign CI [-0.011, 0.037]. The congruence term was removed and an exploration of additional explanatory variables was conducted.

The contribution of the following additional variables was assessed along with interactions between variables which may have theoretical implications: *trial*, *response hand* (*left* Vs *right*), *within-category brightness (*which is whether the brightness value is the *dark* or *bright* value of the two sharing a categorical brightness level see Table 5.1), *brightness contrast* (*high* Vs *low)* and *pair weight.* A model including these additional explanatory variables was preferred to the basic model according to the LRT ($\chi^2(5) = 473.51$, $p<.001$) with an AIC value of 5613.8. An inspection of the CIs demonstrated that *within-category brightness* CI [-0.042, 0.003] and *object* CI [-0.026, 0.020] did not improve the model as a main effect since the values straddle zero. LRT confirmed their removal did not change the explanatory power of the model ($\chi^2(2) = 2.96$, $p=.23$, AIC=5612.7) therefore the model which excluded *within-category brightness* and *object* was preferred. An interaction of *object* and *within-category brightness* also failed to improve model fit ($\chi^2(3) = 4.10$, $p=.0.171$, AIC=5613.7). As did an interaction between *object* and *response hand* ($\chi^2(2) = 2.05$, $p=.0.36$, AIC=5214.7); and *object* and *brightness contrast* ($\chi^2(2) = 0.877$, $p=.0.645$, AIC=5215.8). Table 6.4 summarises the final model, the parameter estimates indicate that the speed of responses increased with *trial*. The visual stimuli influenced the response speed in two ways. A main effect of *categorical brightness* demonstrates that responses were faster when classifying the stimuli as dark, as has been demonstrated in the previous two experiments. Responses were also faster when the *brightness contrast* was *high* i.e. when the brightness value of the circle was further from the mid-grey background (black and white as opposed to dark grey or light grey). It was demonstrated that the *response hand* improved model fit as a main effect; responses were faster for responses with the left hand. *Response hand* did not interact with *categorical brightness* ($\chi^2(1) = 0.089$, $p$=.766, AIC=5614.6) nor *brightness contrast* ($\chi^2(1)$ =2.83, $p$=0.09, AIC=5611.9). The reintroduction of *congruence* as an explanatory variable (along with a main effect of *object*) failed to improve model fit in this final model ($\chi^2(2)$ =.467, $p$=0.79, AIC=5616.3). Therefore the final preferred model did not include *congruence*. Visual inspection of residual Q-Q plots did not reveal any obvious departures from normality. In addition inspection of residual plots demonstrated that the assumptions of linearity and homoscedasticity were also upheld.

Table 6.4

*Summary table for the final model for Experiment 8*

| | | | **Wald confidence intervals** | | |
|---|---|---|---|---|---|
| **Fixed effects (criterion)** | **Estimated coefficient** | **SE** | **2.5%** | **97.5%** | **t-value** |
| (Intercept) | 1.19 | 0.066 | 1.06 | 1.32 | 18.14 |
| Trial | 0.002 | 0.0001 | 0.002 | 0.002 | 14.54 |
| Brightness Contrast (Low) | -0.18 | 0.012 | -0.200 | -0.155 | -15.35 |
| Categorical Brightness (Dark) | 0.049 | 0.012 | 0.027 | 0.072 | 4.26 |
| Response Hand (Right) | -0.071 | 0.012 | -0.094 | -0.048 | -6.10 |
| Pair weight (light) | 0.32 | 0.09 | 0.14 | 0.50 | 3.52 |

| **Random effects** | **Name** | **Std. Dev** | **Variance** |
|---|---|---|---|
| Participant | (intercept) | 0.077 | 0.277 |
| Residual | | 0.170 | 0.413 |

| **AIC** | **BIC** | **Loglik** | **Deviance** |
|---|---|---|---|
| 5612.7 | 5665.0 | -2798.4 | 5612.7 |

#### *6.3.2.2 Response Accuracy*

The overall accuracy of responses was 98.4%. The percentage correct responses for congruent and incongruent trials were 98.61% and 98.20% respectively. The numbers of incorrect responses were too few to warrant the same form of analysis of error rates as was used on RTs. As an alternative, the number of correct responses for congruent and incongruent trials for each participant were compared using a related-samples Wilcoxon Signed Ranks Test. This demonstrated no significant difference between congruent and incongruent trials based on accuracy ($Z$=-1.561, $p$=.118). This confirms that the difference in speed for congruent and incongruent trials did not result in a speed/accuracy trade off. Furthermore, it suggests that congruity did not influence response accuracy.

### 6.3.3 Discussion

The aim of the present experiment was to determine if the size of lifted objects would induce a size-brightness congruity effect when the weight of the objects is available to participants but is held constant. The findings indicate that there was no reliable congruity effect between the size of the objects and brightness classifications. Aside from the lack of congruity effect, the findings replicate the effect of *trial*, *categorical brightness* and *brightness contrast* exhibited in Experiments 6 and 7. A main effect of *response hand* was also demonstrated such that responses were faster with the left hand compared to the right hand. This is somewhat counter-intuitive since one might expect a preference for the dominant hand. Unlike in Experiment 7, there was no interaction between *response hand* and *objec*t. Interestingly, a main effect of *pair weight* was found, despite the weight of the objects only varying between participants. Responses were faster when the lighter pair of objects were used compared to the heavier pair of objects. This may reflect the influence of weight on the ease with which the objects were held and used.

The lack of congruity effect between *object* and *categorical brightness* is surprising. Especially since a congruity effect has been demonstrated to arise between brightness and size where objects varying in size alone are used as response keys, but are not lifted (P. Walker and Walker, 2012; L. Walker and Walker, 2016). And the congruity effect found in Experiment 6 between brightness and the objects being used as response keys which included variation in object size (in this case the size and heaviness of the objects varied such that the larger object was heavier than the smaller object). The lack of congruity effect in the present study is difficult to reconcile with the explanation of these previous congruity effects. According to the interpretation provided to explain the previous congruity effects (e.g. Section 5.2.3.1), the presence of a congruity effect depends on the objects representing relative extremes on a feature dimension which belongs to a set of associated dimensions, in this case *size*. It is uncertain why the objects in the present experiment would not induce the same abstract connotations of size upon which the previous congruity effects are argued to result from.

One possibility is that, despite varying in size, the two objects in the present experiment may not have been categorised or identified in terms of their relative size. There are a couple of possible reasons why the objects may not have been categorised in terms of size. Perhaps the size difference between the big and small objects used in the present study (5cm and 3cm in diameter and height for the big and small objects respectively) was not enough to provide a salient enough difference. Certainly compared to the objects used by P. Walker and Walker (2012) which were spheres with diameters of 2.5cm and 7.5cm, the size difference is much less. But then again, L. Walker and Walker (2016) demonstrated that a middle size object of 5cm can induce a congruity effect when paired with each of the other two original sized objects. The size difference was still larger than in the present study, however was much closer. Alternatively, since the objects were being used actively to tap a touch sensor, the size of the objects may not be of any relevance to how the objects were being used. It seems unlikely that the way that the objects were used in this particular experiment meant that size was not a salient enough feature dimension, since in the study by P. Walker and Walker (2012) the size of the objects did not have an influence on how the response was made, and yet size did induce a congruity effect in that case. Furthermore, in all cases of congruity effects, the interfering feature is incidental to the task demands.

Another possible explanation would be if the size-brightness correspondence is mediated by heaviness. We know that objects varying in size induce an association of heaviness such that larger objects are expected to be heavier than smaller objects (L. Walker et al., 2012). What is more, the findings from the rating tasks in Experiment 4 and 5 indicate that the heaviness of objects overrides conflicting size information to form the basis of brightness judgements. If this were the case, the congruity effect demonstrated in the work by P. Walker and Walker (2012) can be explained in terms of the anticipated heaviness of the objects based on their size. This also explains the lack of congruity effect in the present experiment, since in this case size does not act as a cue to heaviness; because the heaviness of the objects is also available.

A final possibility worth considering is that the objects were not felt to be of equal heaviness. The size-weight illusion causes a larger object to feel lighter in weight than an equally weighted smaller object. As demonstrated in Experiment 5, smaller objects of the same weight were rated as heavier (and also darker) than larger objects; including the objects used in the present experiment. If, in the present study, the heaviness of the objects was felt to be different, it may be that the opposing size and heaviness of the objects counteracted one another neutralising any potential for either a size-brightness or heaviness-brightness congruity effect to arise. Experiment 9 explores further whether a congruity effect may arise when the size and heaviness of the objects are in opposition; where the larger object is lighter in weight than the smaller object.

## 6.4 Experiment 9

In Experiments 4 and 5, it was demonstrated that the heaviness of lifted objects formed the primary basis for judgements about brightness, despite the objects being rated varying in both size and heaviness. The aim of the present Experiment is to determine if a dominance of a heaviness-brightness correspondence over a size-brightness correspondence can also be observed in the form of a congruity effect in a speeded classification task. To explore this, the same brightness speeded classification task was conducted. However the objects being used as response keys varied in size and heaviness such that the smaller object is heavier than the bigger object. There are a number of possibilities for how these objects may influence response speed for brightness classifications. Firstly, in accordance with the findings from Experiment 4 and 5, it may be that the heaviness of the objects will override size, to produce a congruity effect such that responses are faster when the heavier (but smaller) object is being used to indicate that stimuli are dark and when the lighter (but bigger) object is used to classify stimuli as bright compared to the other way around. Alternatively, the conflict may lead to no clear congruity effect. Another possibility is that a congruity effect may be induced based on the size of the objects. However, considering the results of Experiment 8, which demonstrated that the size of objects did not induce a congruity effect with brightness when the objects did not vary in weight; it is unlikely that the size of the objects will result in a congruity effect that is in line with the size-brightness correspondence in the present experiment.

### 6.4.1 Method

#### *6.4.1.1 Participants*

Eighteen students from Lancaster University volunteered to take part in the study for payment or course credit. One participant was not including in analysis due to equipment problem. The remaining 17 (11 females and 6 males) were aged between 18 and 20 (Mean=18.7 years). All except one were right handed by self-report. All except four spoke English as a first language. The remaining participants spoke the following first languages: Chinese (n=3) and Catalan (n=1).

#### *6.4.1.2 Materials, design and procedure*

The materials, design and procedures were the same as Experiment 6,7 and 8. However, the experiment differs with regard to the objects being used as response keys. One pair of objects made up of the heaviest/smaller and lightest/biggest from the full set of nine objects were used (Objects C and G see Figure 6.4).

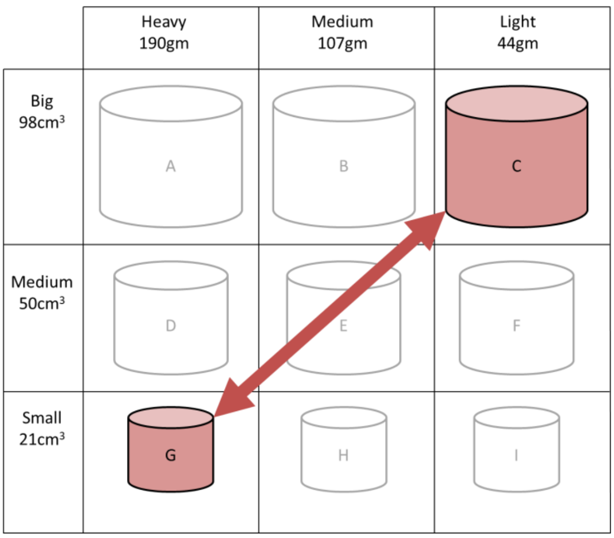



*Figure 6.4*. The objects from the full set of nine used as response keys in Experiment 9

### 6.4.2 Results

The data were the accuracy levels and the correct response times (RTs) for the brightness classifications. Out of 2448 trials, 2394 of these were correct responses (97.8%). Any RTs less than 300ms and more than 2.5 SD above each participant's mean RT were excluded from analysis leaving 2311 observations (94.4%). Table 6.5 summarises the accuracy and RTs for *bright* and *dark* classifications made with the *small/heavy* and *big/light* objects.

Table 6.5

*Mean RTs (SEM in parentheses) and accuracy levels (%) according to the object used and categorical brightness.*

| | | Brightness | |
|---|---|---|---|
| Object Used | | Dark | Bright |
| Big/Light | Mean RT(SEM) | 733(15) | **749 (15)** |
| | Accuracy | 97.7% | **97.7%** |
| Small/Heavy | Mean RT(SEM) | **718 (14)** | 775 (16) |
| | Accuracy (%) | **98.7%** | 97.1% |

Note: Bold typeface signifies congruent conditions in accordance with a heaviness-brightness correspondence

#### *6.4.2.1 Response Speed*

The analysis approach was the same as that used for all previous speeded classification tasks thus far. In order to resolve the skew of the RT distribution, a transformation of 1/RT was performed. The product of this transformation is interpreted as *response speed* (responses/second). The main aim of the analysis was to determine if an interaction between *object* (*big/light* Vs *small/heavy*) and *categorical brightness (dark* Vs *bright*) significantly improves model fit. For ease of interpretation this was included as a main effect of *congruence (congruent trials* Vs *incongruent trials)* along with main effects of *categorical brightness* and *object* where congruent trials were based on the heaviness-brightness correspondence therefore when the large/light object was used to classify stimuli as bright and the small/heavy object used to classify stimuli as dark. The LRT demonstrates that a model including *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2(1) = 7.35$, $p$=0.007) with an AIC value of 3133.1 and 3127.8 for the null and comparison model respectively. The parameter estimates of this basic model suggest that incongruent trials were slower than congruent trials by 0.053 responses/second (SE=0.020), CI [0.015, 0.091].

The contributions of the following additional variables were assessed along with interactions between variables which may have theoretical implications: *trial*, *response hand* (*left* Vs *right*), *within-category brightness (*which is whether the brightness value is the *dark* or *bright* of the two sharing a categorical brightness level, see Table 5.1), and *brightness contrast* (*high* Vs *low).* A model including these additional explanatory variables was preferred to the basic model according to the LRT ($\chi^2(4) = 145.44$, $p$<.001) with an AIC value of 2990.3. An inspection of the CIs demonstrated that *within-category brightness* CI [-0.028, 0.056] and *response hand* CI[-0.037,0.037] did not have an effect since the values straddle zero. LRT confirmed their removal did not change the explanatory power of the model ($\chi^2(2) = 0.216$, $p$=0.898, AIC=2986.5) therefore the model which excluded *within-category brightness* and *response hand* was preferred. An interaction of *object* and *within-category brightness* also failed to improve model fit ($\chi^2(2) = 1.67$, $p$=.43, AIC=2988.9). *Response hand* did not interact with *object* ($\chi^2(2) = 2.14$, $p$=.344, AIC=2986.5), *categorical* *brightness* ($\chi^2(2) = 0.767$, *p*=.68, AIC=2989.8) nor *brightness contrast* ($\chi^2(2) = 1.861$, *p*=.394, AIC=2988.7).

The parameter estimates indicate that the speed of responses increased with *trial* which suggests a practice effect. Responses were faster when the *contrast* was *high* i.e. when the brightness value of the circle was further from the mid-grey background (black and white as opposed to dark grey or light grey). An interaction between *object* and *contrast* failed to improve model fit ($\chi^2(1) = 0.618$, *p*=.432, AIC=2987.9. The LRT demonstrates that a model including *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2(1) = 7.135$, *p*=.008) with an AIC value of 2991.7 and 2986.5 for the models excluding and including *congruence* respectively. The parameter estimates of this final model suggest that incongruent trials were slower than congruent trials by 0.050 responses/second (SE=0.019), CI [0.013, 0.090]. Figure 6.5 depicts the nature of the congruity effect and Table 6.6 summarises the parameter estimates of the final model. Visual inspection of residual Q-Q plots did not reveal any obvious departures from normality. In addition inspection of residual plots demonstrated that the assumptions of linearity and homoscedasticity were also upheld.

Table 6.6

*Summary table for the final model for Experiment 9*

| Fixed effects (criterion) | Estimated coefficient | SE | Wald confidence intervals 2.5% | Wald confidence intervals 97.5% | t-value |
|---|---|---|---|---|---|
| (Intercept) | 1.54 | 0.077 | 1.39 | 1.70 | 19.99 |
| Trial | 0.002 | 0.0002 | 0.001 | 0.002 | 7.68 |
| Brightness Contrast (Low) | -0.18 | 0.019 | -0.22 | -0.14 | -9.59 |
| Object (Small/Heavy) | -0.02 | 0.019 | -0.06 | 0.016 | -1.12 |
| Categorical Brightness (Dark) | 0.052 | 0.019 | 0.015 | 0.089 | 2.78 |
| Congruence (Incongruent) | -0.05 | 0.019 | -0.087 | -0.013 | -2.67 |

| Random effects | Name | Std. Dev | Variance |
|---|---|---|---|
| Participant | (intercept) | 0.089 | 0.30 |
| Residual | | 0.21 | 0.45 |

| AIC | BIC | Loglik | Deviance |
|---|---|---|---|
| 2986.5 | 3032.5 | -1485.3 | 2970.5 |

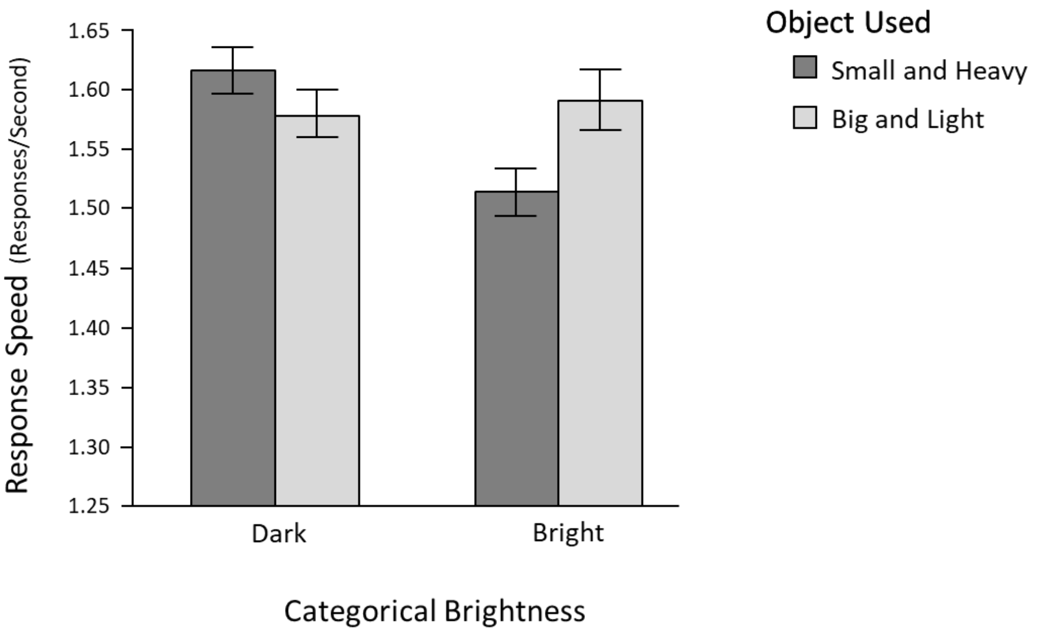



*Figure 6.5.* Mean response speed (responses/second) for dark and bright responses with each object. Error bars refer to the standard error of the mean.

#### *6.4.2.2 Response Accuracy*

Overall accuracy was 97.8%. The percentage correct responses for congruent and incongruent trials were 98.2% and 97.4% respectively. The numbers of incorrect responses were too few to warrant the same form of analysis of error rates as was used on RTs. As an alternative, the number of correct responses for congruent and incongruent trials for each participant were compared using related-samples Wilcoxon Signed Ranks Test. This did not reveal a significant effect of congruency on accuracy score ($Z$=-1.217, $p$=.224). This confirms that the difference in speed for congruent and incongruent trials did not result in a speed/accuracy trade off. Furthermore, it suggests that congruity did not influence response accuracy.

### 6.4.3 Discussion

The present study demonstrates that when objects varying in size and weight are used as response keys in a brightness speeded classification task, where the differences in size and weight contradict one another such that the smaller object is heavier than the bigger object, a congruity effect can occur. The nature of this effect is that responses were faster when the smaller (but heavier) object was used to classify stimuli as dark and the bigger (but lighter) object was used to classify stimuli as bright compared to when they were the other way around. Alongside the congruity effect, the final model replicated main effects observed in previous experiment, including a practice effect, such that responses became faster as trials progressed; a main effect of *brightness contrast*, such that responses were faster when the brightness values were further from the mid-grey background and a main effect of *categorical brightness* which has been argued to be responsible for asymmetry found in the *object* x *categorical brightness* congruity effect.

The findings observed suggest that a heaviness-brightness correspondence overrides the potential for a size-brightness correspondence to induce a congruity effect. The primacy of heaviness is in keeping with the findings from the rating scale studies (Experiment 4 and 5) which suggested that heaviness formed the basis for brightness ratings despite conflicting information about object size. As discussed in Section 4.4.2 there are a number of potential reasons why heaviness may override size in this way. It may be indicative of heaviness being a more dominant or salient feature compared to size. Or, it could reflect in some way the difference in nature of association between heaviness and brightness compared to size and brightness. The findings are also consistent with the possibility that the size-brightness correspondence is in fact mediated by a brightness-heaviness correspondence. The fit of these possible explanations would be an interesting avenue to explore. What is important about the present findings is that, by confirming the same patterns as those observed in the rating scale tasks from Experiment 4 and 5,  it suggests that the preference or dominance of a heaviness-brightness correspondence over a size-brightness is unlikely to be a result of an explicit decision making process.  Instead it suggests that reflects an underlying attribute of the psychological processes involved.

The present findings shed some light on the lack of size-brightness congruity effect found in Experiment 8 by ruling out the possibility that conflicting size and weight information impeded any potential congruity effects from arising. The objects used in Experiment 8 varied in size, while weight was held constant. It was unclear whether a lack of congruity effect could be attributed to conflict between the size and perceived heaviness of the objects, subsequently cancelling one another out. The present findings show that conflicting information alone does not prevent a congruity effect from arising. Although, it is potentially the case that the difference in heaviness is required to be over a certain threshold to induce a congruity effect in the presence of contradictory information about object size.

## 6.5 General Discussion

The three experiments in the present chapter explored how the size and weight of felt objects influence the speed of responses in a brightness classification task. When taken together with Experiment 6 (Chapter 5), the pattern of findings, suggests that a brightness-heaviness congruity effect is induced irrespective of variation in size. When objects vary in weight, responses are faster when the heavier object is used to classify stimuli as dark and when the light object was used to classify stimuli as bright compared to the other way around. This brightness-heaviness congruity effect was found when size varied in correlation with heaviness (Experiment 6); when size was held constant (Experiment 7) and when size was varied in contrast to heaviness (Experiment 9). A congruity effect was not demonstrated when heaviness was held constant despite the objects varying in size (Experiment 8).

The findings confirm the alignment of brightness and heaviness found in the first half of the thesis, such that heavier is aligned with darker and lighter with brighter. It also demonstrated that the heaviness-brightness correspondence continued to arise despite contradictory information about object size (Experiment 9). This is in keeping with the findings of Experiments 4 and 5 which demonstrated that judgements about brightness are based primarily on the perceived heaviness of the objects; despite size and heaviness both have the potential to influence performance.

# Chapter 7

## 7.1 Introduction

The speeded-classification tasks reported in Chapters 5 and 6 demonstrated that when two objects varying in heaviness are used as response keys for classifying stimuli as bright and dark, responses were faster when the heavy object was used for classifying stimuli as dark and the light object for classifying stimuli as bright compared to the other way around. This finding provides additional evidence in support of the associations found to arise in Chapters 2-4 between brightness and heaviness such that heavier objects are rated as darker than less heavy objects. As discussed in Section 5.1 (Chapter 5), the speeded-classification task provides considerable support for associations between dimensions found in rating scale tasks, because they demonstrate that the associations can arise without conscious deliberation from the participant about how they should be aligned.

The aim of the present chapter is to determine if a congruity effect between pitch and heaviness, in line with the pitch-heaviness correspondence demonstrated in the first half of the thesis (that heavier objects are expected to be lower in pitch than less heavy objects), can also be observed using the speeded-classification task. As in Experiment 7, the present study uses objects varying in heaviness as response keys during a speeded-classification task. However, in this case, participants are asked to respond by classifying sounds as either high or low in pitch. It is predicted that the heaviness of response keys will interfere with the classifications of pitch in a similar way to that found with brightness. Responses will be faster when the heavy object is being used to respond to sounds that are lower in pitch and the less heavy object to respond to sounds that are higher in pitch compared to the other way around.

## 7.2 Experiment 10

### 7.2.1 Method

#### *7.2.1.1 Participants*

Forty students from Lancaster University (25 females, 15 males) between the ages 18 and 32 (mean age=20.13) volunteered to take part in the study for payment or course credit. All except seven participants were right handed by self-report, and one participant did not disclose whether they were left or right handed. Twenty-one participants spoke English as their first language. The remaining 19 participants spoke the following first languages: Bulgarian (n=1), Chinese/Cantonese (n=10), Hindi (n=1), Italian (n=1), Malay (n=1), Norwegian (n=1), Polish (n=2), Russian (n=1) and Urdu (n=1).

#### *7.2.1.2 Materials*

***Stimuli for classification.*** The experiment was conducted using PsyScript version 2.0 on a dual 2 GHz, PowerMac G5 with a 20in monitor (Apple A1038, 1,680 x 1,050 cinema back-lit LCD display).Two sine-wave tones were developed using Audacity software, with frequencies of 3520hz and 220hz for the high and low pitch tones respectively. The amplitude of the high pitch sound was reduced to the level where it was judged to be equivalent in loudness to the low pitch sound by the researcher. The sounds were presented through headphones for 2 second durations on each trial.

The visual display was of a mid-grey background upon which a black and white checked question-mark was simultaneously presented with each sound in the centre of the screen and remained until a response was received.

***Response Keys.*** Two pairs of objects made up of the heaviest and lightest object at two size levels from the full set of nine objects (i.e. object pairs A & C and D & F) were used, see Figure 7.1. One group of participants completed the task with each pairing. The touch sensitive surfaces were the same as those used in Experiments 6-9: two circular metal disks (4cm in diameter) mounted onto wooden frames of 1.0 cm depth. The touch-sensitive surfaces were interfaced to the computer keyboard, allowing the classification decision, and the speed with which it was made, to be recorded when tapped by the object.  The sensors were covered with a layer of thick felt material to dampen the impact sound of the objects making contact with the sensors. During the procedure the objects and sensors remained hidden from view by being covered with thick black material.

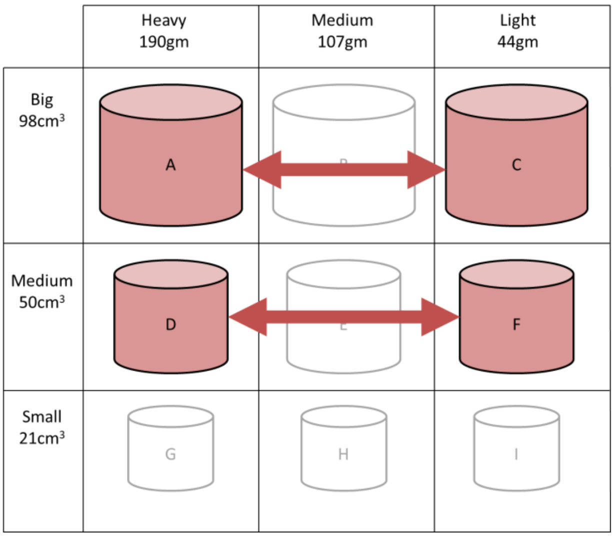



*Figure 7.1.* The objects from the full set of nine used as response keys in Experiment 10

#### *7.2.1.3 Design and Procedure*

Participants completed four blocks of 36 trials. On each trial a sound was presented along with a question mark prompt in the centre of the screen. Participants were instructed to respond as quickly and as accurately as possible to whether each sound was high or low in pitch by tapping the object held in the right or left hand against the touch-sensitive surface. Half of the participants were instructed to tap with their left hand for high and their right hand for low, and half were assigned to the opposite mapping. The left-right assignment of high or low pitch remained the same for each participant during the study. A small label at the bottom of the screen reminded the participant which hand they were to tap for ‘high’ and which for ‘low’.

Participants held an object in each hand underneath the material that was covering the objects and touch sensors such that the corner edge of each cylinder could be used to make contact with the touch sensor. Participants were asked to ensure that the objects were not being rested on the surface of the sensor and to hold them approximately 1.5 cm above the appropriate surface. The difference in heaviness between the two objects was not mentioned by the researcher. When required, the participants would tap the objects, using the corner edge, onto the touch sensor before returning to the initial starting position. Participants were allowed to lean on the table with their lower arms in order for them to maintain the position for the duration of the task. Before each block, participants were given the opportunity to practice hitting the objects on the sensors. During this practice session, they received feedback about the response having been registered, the task proper commenced when both participant and researcher were satisfied with the accuracy of responses.

When the object made contact with the touch sensitive surface, the question mark disappeared and the response time was recorded as the time elapsing since stimulus onset. There was a 2 second interval before the next sound was presented, in which time the screen displayed the mid-grey background only. The order in which the sounds were presented was randomised across sets of 12 trials, so that within each set each sound appeared 6 times. Each block of trials took approximately 2-3 minutes with a two minute break between each block. During the break, participants moved to another desk and completed a word search. At this point the researcher swapped the objects used on the left and right sensors. This was not mentioned by the researcher but some participants reported being aware that the objects had been swapped.

### 7.2.2 Results

The data were the accuracy levels and the correct response times (RTs) for the pitch classifications. Out of 5616 trials, 5585 of these were correct responses (99.4%). Any RTs less that300ms and more than 2.5SD above the mean were excluded from analysis leaving 5392 observations (96.0%). Table 7.1 summarises the accuracy and RTs for *High* and *Low* classifications made with the *Heavy* and *Light* objects.

Table 7.1

*Mean RTs (SEM in parentheses) and accuracy levels (%) according to the object used and pitch categorization*

| | Pitch to be Categorised | |
|---|---|---|
| Object Used | Low | High |
| Heavy | **713 (9)** | 738(10) |
| | **99.5%** | 99.2% |
| Light | 756 (10) | **746 (11)** |
| | 99.6% | **99.5%** |

Note: Bold typeface signifies congruent conditions

#### *7.2.2.1 Response Speed*

The analysis approach was the same general strategy as that used in Experiment 7-9. A linear mixed-effects analysis was performed using *lme4* (Bates, et al., 2014) in R version 3.2.0 (R Core Team, 2012). In order to resolve the skew of the RT distribution, a transformation of 1/RT was conducted. The product of this transformation is interpreted as *response speed* (Responses/Second). A series of intercept only models, which included a random effect of *Participant*, were developed to assess the explanatory variables which contributed to *response speed*. The contributions of any main effects and interactions were determined using Likelihood-Ratio Tests (LRTs) which compares a model including the variable with a model which does not. Alkaike Information Criterion (AIC) values provided a relative estimate of the amount of information not being captured by a model, balancing goodness of fit with the number of parameters the model contains.  The AIC for alternative models were compared, with lower AIC values indicating a superior model. 95% confidence intervals were calculated using the *Wald* method with the *confint()* function.

The main aim of the analysis was to determine if an interaction between *object* (*heavy* Vs *light*) and *pitch (low* Vs *high*) significantly improves model fit. For ease of interpretation this was included as a main effect of *congruence (congruent trials* Vs *incongruent trials)* along with main effects of *pitch* and *object*. The LRT demonstrates that a model including *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2(1) = 9.56$, $p<.001$) with an AIC value of 7137.2 and 7129.6 for the null and comparison model respectively. The parameter estimates of this basic model suggest that incongruent trials were slower than congruent trials by 0.04 responses/second (SE=0.01), CI [0.064, 0.014].

The contribution of the following additional variables was assessed along with interactions between them which may have theoretical implications: *trial, pair size* and *response hand* (*left* Vs *right*). A model including these additional explanatory variables was preferred to the basic model according to the LRT ($\chi^2(3) = 209.0$, $p<.001$) with an AIC value of 6926.6. An inspection of the CIs demonstrated that *pair size* CI [-0.15, 0.14] and *response hand* CI [-0.046, 0.003] did not have an effect since the values straddle zero. LRT confirmed their removal did not change the explanatory power of the model ($\chi^2(2) = 3.06$, $p=.217$, AIC=6925.7). Therefore the model which excluded *pair size* and *response hand* was preferred. An interaction between *pitch* and *response hand* did not improve model fit ($\chi^2(2)$ =

3.175, $p$=0.204, AIC=6926.5). However an interaction between *object* and *response hand* did ($\chi^2(2) = 14.803$, $p<.001$, AIC=6914.9). The nature of this interaction was that responses were faster when the heavy object was being used in the right hand and the light object was being used in the left hand compared to the other way around.

The LRT demonstrates that a model including *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2(1) = 10.45$, $p<.001$) with an AIC value of 6923.3 and 6914.9 for the models excluding and including *congruence* respectively. The parameter estimates of this final model suggest that incongruent trials were slower than congruent trials by 0.040 responses/second (SE=0.01), CI [0.016, 0.064]. This also confirms that the effect of *congruence* remains over and above the contributions of the additional explanatory variables included. Table 7.2 summarises the parameter estimates of the final model and Figure 7.2 depicts the congruity effect between *object* and *pitch*. Visual inspection of residual Q-Q plots did not reveal any obvious departures from normality. In addition inspection of residual plots demonstrated that the assumptions of linearity and homoscedasticity were also upheld.

Table 7.2

*Summary table for the final model for Experiment 10*

| Fixed effects (criterion) | Estimated coefficient | SE | Wald confidence intervals 2.5% | 97.5% | t-value |
|---|---|---|---|---|---|
| (Intercept) | 1.43 | 0.042 | 1.35 | 1.51 | 33.99 |
| Trial | 0.002 | 0.0002 | 0.002 | 0.002 | 14.49 |
| Response Hand (Right) | 0.021 | 0.017 | -0.014 | 0.055 | 1.18 |
| Object (Light) | 0.023 | 0.017 | -0.01 | 0.057 | 1.35 |
| Pitch (Low Pitch) | -0.002 | 0.012 | -0.03 | 0.022 | -0.15 |
| Response Hand (Right): Object (Light) | -0.08 | 0.025 | -0.133 | -0.036 | -3.43 |
| Congruence (Incongruent) | -0.04 | 0.012 | -0.064 | -0.036 | -3.43 |

| Random effects | Name | Std. Dev | Variance |
|---|---|---|---|
| Participant | (Intercept) | 0.056 | 0.24 |
| Residual | | 0.205 | 0.452 |

| AIC | BIC | Loglik | Deviance |
|---|---|---|---|
| 6914.9 | 6974.2 | -3448.4 | 5383 |

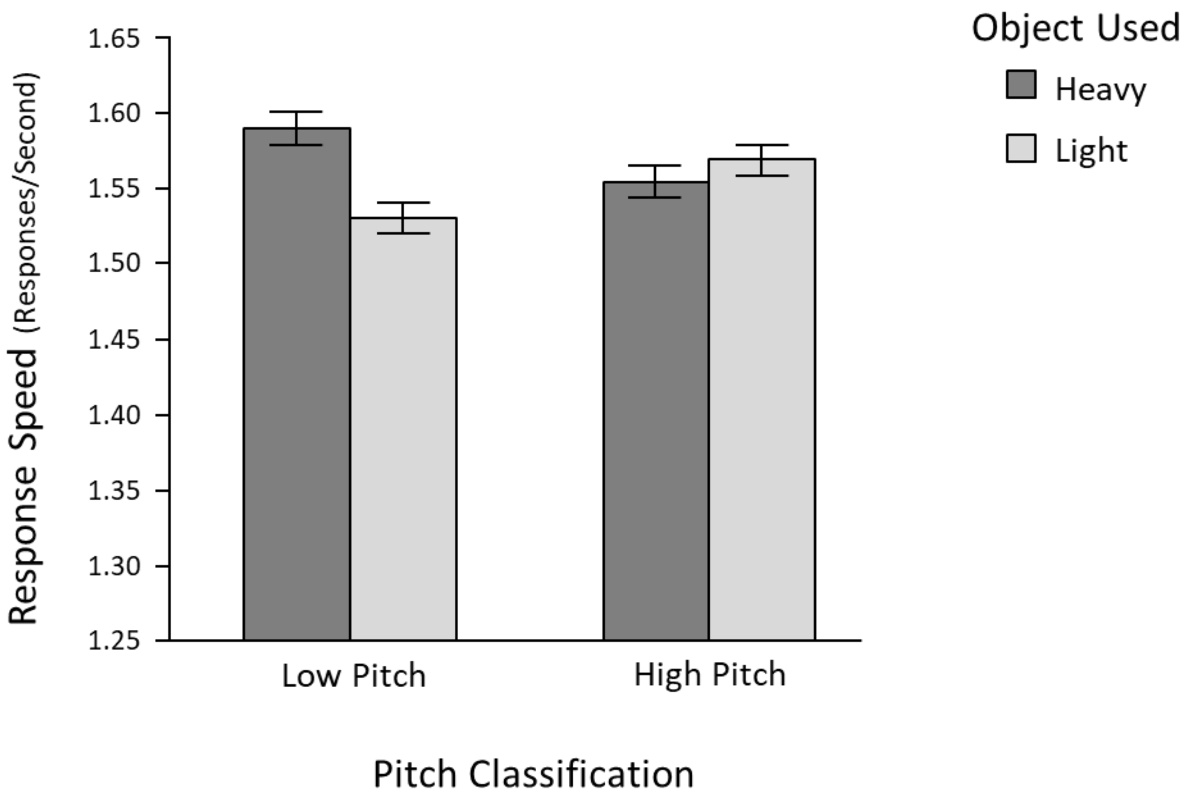



*Figure 7.2.* Mean response speed (responses/second) for low and high pitch responses with each object. Error bars refer to the standard error of the mean

#### *7.2.2.2 Response Accuracy*

The overall response accuracy was 99.4%. The percentage correct responses for congruent and incongruent trials were 99.50% and 99.39% respectively. The numbers of incorrect responses were too few to warrant the same form of analysis of error rates as was used on RTs. As an alternative, the number of correct responses for congruent and incongruent trials were analysed using related samples Wilcoxon signed rank test indicated no significant difference ($Z$=-.277, $p$=.782). This confirms that the difference in speed for congruent and incongruent trials did not result in a speed/accuracy trade off. Furthermore, it indicates that congruity did not influence response accuracy.

### 7.2.3 Discussion

The findings from the present experiment demonstrate an interaction between the pitch-classification being made and the heaviness of the objects being used as response keys. Responses were faster when the heavier object was used to classify sounds as low pitch and the light object was used to classify sounds as high pitch compared to the other way round. This is consistent with the findings from the rating scale tasks (Experiment 1-5) which demonstrated that heavier objects are judged to be lower in pitch than less heavy objects. It is also consistent with findings that lower pitch sounds are judged to be heavier than high pitch sounds (P. Walker et al, 2012).

As with the brightness-heaviness congruity effect, the congruity effect demonstrated in the present experiment suggests an underlying alignment of pitch and heaviness dimensions which subsequently results in response speeds being faster when the two extremes are matched in the task (the *heavy* object is being used to respond to *low pitch* classifications and the *light* object is being used to respond to *high pitch* classifications) compared to the other way around. However, unlike in the brightness speeded-classification task, there are only two values of pitch: one high and one low. The relative pitch category is still likely to be what is important when making the classification (as opposed to absolute values of pitch) since the selected tone frequencies were arbitrary. However, the lack of additional pitch values in both the high and low pitch category, mean this cannot be stated as definitively as it could be in the speeded-classification tasks for brightness.

The final model in the present experiment is consistent with findings from the previous speeded classification tasks in other respects. Firstly, there was a main effect of *trial* which, as with Experiments 6-9 is argued to indicate a practice effect. In addition, as in Experiment 7, the final model of the present study included an interaction between *handedness* and *object*. This interaction indicates that responses were faster when the *light* object is held in the *left* hand and the *heavy* object is held in the *right* hand compared to the other way around. This is potentially due to participants having a preference for this hand/object assignment because the heavier object can be held more comfortably in right (in most cases, dominant ) hand. This interaction between object and response hand has only been demonstrated with object pairs that are of equal size but vary in weight, which may suggest that the size and weight of held objects interact to make an object more or less comfortable to hold, thus altering the extent to which the hand an object is held in influences response speed.

One weakness in the present study is that although measures were taken to ensure that the impact sound of the objects could not be heard, it cannot be ruled out as a possible explanation for the findings. This is because, unlike in the brightness-classification task, an additional auditory mask could not be played to further hide the sound, as this would interfere with the auditory stimuli being presented. Therefore, the sound that the objects made when being tapped had the potential to influence the present findings. When tapped, the heavier object made a lower pitch sound than the less heavy object. Therefore, it is possible that the association between the pitch of the sound being presented and the object being used was based on the objects impact sound. That is to say, what appears to be a pitch-heaviness congruity effect could in fact be a pitch-pitch congruity effect.

This potential confounding variable raises an interesting matter. There is a physical relationship between weight and pitch such that, all else being equal, objects with more mass have a resonant frequency that is lower in pitch. This is because, when struck, a heavier object will vibrate at a slower rate for the same amount of energy applied to it. This physical relationship between pitch and heaviness is an example of what Parise (2016) describes as semi-redundant feature dimensions. Which means that one feature dimension can go some way to explain or predict the other when experienced in our environment (see Section 1.2.1).

Attempting to fully explore this with regards to pitch and heaviness lies outside the scope of this thesis, however, is worth acknowledging. Chapter 8 presents a set of additional speeded-classification tasks which attempts to address the potential confound of the impact sound that occurred within the present experiment. In Chapter 8, some of the speeded-classification tasks from Chapter 6 and the present experiment are replicated, with a method in which the objects are not actively used to make a response. This removes the potential of the objects making a confounding impact sound.

# Chapter 8

## 8.1 Introduction

Chapters 6 and 7 demonstrated that when objects are used as response keys to tap a touch sensitive surface, the heaviness of the objects interacted with the categorisation of brightness (Chapter 6) or pitch (Chapter 7) being made. Responses were faster when the heavier object was used to classify stimuli as dark or low pitch and the less heavy object was used to classify stimuli as bright or high pitch compared to the other way around. This is the first demonstration that cross-sensory correspondences of heaviness with both brightness and pitch have the potential to influence performance when heaviness is available to the participants through the manipulation of objects used as response keys in a speeded classification task. In these experiments, the heaviness of the objects was not only available to participants as an enduring incidental feature of the task, but had relevance with regard to the ability to make a response. The present chapter attempts to explore whether the interaction of heaviness with brightness and pitch was a result of the availability of the heaviness as a property of the object response keys; or whether it was necessary for the objects to be actively used to make a response in order for heaviness to enter into the congruity effects that were observed.

Perceived heaviness has been demonstrated to be closely related to how an object is used. For example, people are more sensitive to differences in heaviness when the objects are being wielded rather than when they are passively placed in the hand (Weber, 1834/1978). Furthermore, one school of thought suggests that our experience of an object’s heaviness is best understood to be the perception of ease with which an object is used (Wagman, 2015; Shockley, Carello, & Turvey, 2004). This illustrates the potential importance that the active use of the objects had on perception of the objects heaviness and its subsequent interaction with classification decisions. The influence of the size and heaviness of the objects on participant's ability to make a response can be observed in the final models of some of the previous experiments. For example, in Experiments 7 and 10 an interaction between the object and response had was found, which is likely to have been a result of the effect of the size/weight combinations on the ease with which the objects could be used to make a response when held in each hand.

In the present chapter, the speeded classification tasks were adapted so that heaviness was available to participants but did not interfere with the way responses were made. Small button-switches were mounted onto pairs of objects which were then held passively during the task. Participants made a classification response simply by pressing the button mounted on the object in either their left or right hand.  The buttons were identical, which meant that the size and weight of the objects being used did not affect how easy it was to make a response. It also removed any potential confound from impact sounds, an issue discussed in Section 7.2.3. The present series of experiments, aims to replicate the key findings from Chapters 6 and 7 with this new response method.

## 8.2 Experiment 11

Before conducting the brightness and pitch speeded-classification tasks, it is necessary to determine whether the difference in heaviness of the objects remains salient enough to have to potential to induce a congruity effect, despite the objects being held passively. If not, a lack of congruity effect in the brightness and pitch speeded classification tasks may simply be a result of no perceived difference between the objects that are used. To address this concern, the present study uses the objects used in the remaining experiments to respond to the words "heavy" and "light" presented on a screen. If a congruity effect is found between the words being classified and the objects being used to classify them, it would confirm that the heaviness of the objects continues to have the potential to induce a congruity effect, despite being held passively.

### 8.2.1 Method

#### *8.2.1.1 Participants*

Twenty students from Lancaster University (16 female, 4 males) between the ages of 18 and 32 (mean age = 20.2 years) volunteered to take part in the study for payment or course credit. All apart from one of the participants were right handed by self-report. Twelve participants spoke English as their first language. The remaining eight participants spoke the following first languages: Arabic (n=1), Chinese (n=3), German (n=1), Indonesian (n=1), Lithuanian (n=1) and Portuguese (n=1).

#### *8.2.1.2 Materials*

**Stimuli for classification.** The experiment was conducted using PsyScript (version 2.3.0) on a 27in (2560 x 1440) computer screen (Apple Thunderbolt LED backlit display controlled by an Intel Core i7 2.6GHz Mac mini Server). The words “Heavy” and “Light”, written in red Arial font (260 X 116 pixels), were presented in the centre of the screen on a mid-grey background (90 cd/$m^2$).

**Response keys.** The two cylinders used were the heaviest and lightest of the middle sized objects used in Experiment 6 (Objects D and F of the full set of nine objects, see Figure 8.1). Micro-switch buttons (6mm diameter, 4mm deep, and with a 2mm gap needing to be closed) were attached to the centre of one end of each object which participants used to make their responses. A thick black material covered the objects so they were hidden from view throughout the study.

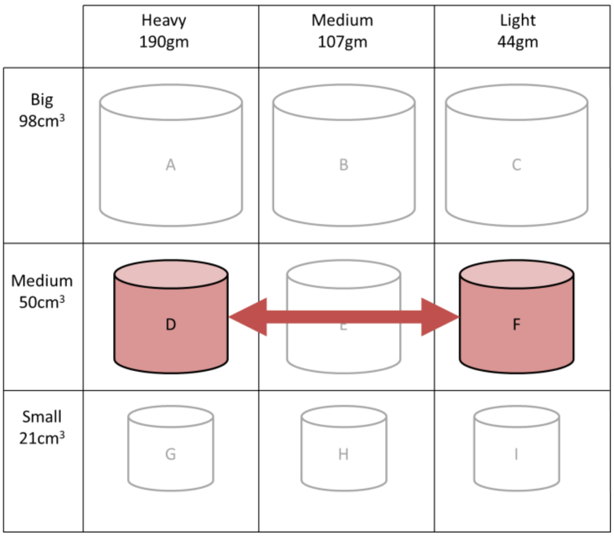



Figure 8.1. The objects from the full set of nine used as response keys in Experiment 11

### *8.2.1.3 Design and Procedure*

Participants completed four blocks of 36 trials. On each trial the word ‘heavy’ or ‘light’ was displayed until a response was given. Participants responded as quickly as possible by pressing the switch on top of one of the two cylinders (one being held in each hand). Half the participants in each group were asked to press with their left hand for ‘heavy’ and their right hand for ‘light’; the other half of participants did this the other way. The left-right assignment of ‘heavy’ or ‘light’ remained the same for each participant throughout the study. A small label at the bottom of the screen reminded the participant which hand (left or right) they were to press for ‘heavy’ and which for ‘light’. The object being held in each hand was alternated between blocks. The starting left and right assignment of each object was counterbalanced.

Participants were asked to hold an object in each hand underneath the material so that the thumb was able to press the button on the top of each object. They were instructed to position themselves so that they were leaning on the table with the elbow and forearm with their hands slightly raised of the table so that the objects were fully lifted off the table in a position that was comfortable for them to maintain for the duration of the block of trials. The difference in weight of the two objects was not mentioned by the researcher.

After each response, there was a two second interval before the next word was presented, in this time only the mid-grey background was displayed. Within each block of trials, each word was shown 18 times. The order in which the words were presented was randomised across sets of 12 trials, so that within each set of twelve trials each word appeared 6 times. Each block of trials took approximately 2 minutes. Between each block, participants moved to another desk for a two minute break. They had a word search to complete during this time. At this point the researcher swapped the left-right position of the two objects. This was not mentioned by the researcher, however, some participants reported being aware that the objects had been switched.

### 8.2.2 Results

The data were the accuracy levels and the correct response times (RTs) for the brightness classifications. Out of 2880 trials, 2835 of these were correct responses (98.44%). Any RTs below 150ms or more than 2.5SD above the mean were excluded from analysis leaving 2762 observations (95.90%). Table 8.1 summarises the accuracy and RTs for *heavy* and *light* classifications made with the *heavy* and *light* objects.

Table 8.1

*Mean RTs (SEM in parentheses) and accuracy levels (%) according to the object used and word categorisation*.

| | Word to be Categorised | |
|---|---|---|
| Object Used | Heavy | Light |
| Heavy | **503 (6)** | 540(5) |
| | **98.61** | 98.47 |
| Light | 541 (6) | **508(4)** |
| | 98.06 | **98.61** |

Note: Bold typeface signifies congruent conditions

#### *8.2.2.1 Response Speed*

The analysis approach was the same general strategy as that used in Experiments 6-10. In order to resolve the skew of the RT distribution, a transformation of 1/RT was performed. The product of this transformation is interpreted as response *speed* (responses/second). The main aim of the analysis was to determine if an interaction between *object* (*heavy* Vs *light*) and *heaviness classification (heavy* Vs *light*) significantly improves model fit. For ease of interpretation this was included as a main effect of *congruence (congruent trials* Vs *incongruent trials)* where congruent trials were those where the heavy object was used to classify visual stimuli as heavy and the light object was used to classify visual stimuli as light and incongruent trials being the other way around. The LRT demonstrates that a basic model including *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2(1) = 67.90$, $p<.001$) with an AIC value of 2346.1 and 2280.2 for the null and comparison model respectively. The parameter estimates of this basic model suggest that incongruent trials were slower than congruent trials by 0.11 responses/ second (SE=0.014), CI [0.087, 0.14].

The contribution of *trial* and *response hand* (*left* Vs *right*) were assessed along with interactions between variables which may have theoretical implications. A model including these additional explanatory variables was preferred to the basic model according to the LRT ($\chi^2(2) = 12.26$, $p<.001$) with an AIC value of 2271.9. An inspection of the CIs demonstrated that *response hand* did not have an effect since the values straddle zero, CI [-0.006, 0.047]. LRT confirmed its removal did not change the explanatory power of the model ($\chi^2(1) = 2.26$, $p=.0.13$, AIC=2272.2) therefore the model which excluded *response hand* was preferred. An interaction of *object* and *response hand* also failed to improve model fit ($\chi^2(2) = 2.31$, $p=.32$, AIC=2273.9). However an interaction of *category heaviness* and *response hand* did improve model fit ($\chi^2(2) = 6.55$, $p=.038$, AIC=2269.6).

Parameter estimates indicate that the speed of responses increased with *trial*. The explanatory influence of *response hand* changed when considered alongside the interaction of *response hand* X *categorical heaviness*. A closer look at this interaction demonstrates that responses were faster when the left hand was used to classify stimuli as light and the right hand was used to classify stimuli as heavy compared to the other way around. In the final model, LRT demonstrates that a model including *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2(1) = 68.32$, $p<.001$) with an AIC value of 2335.9 and 2269.6 for the models excluding and including *congruence* respectively. The parameter estimates of this final model suggest that incongruent trials were slower than congruent trials by 0.11 responses/second (SE=0.01), CI [0.09, 0.14]. Figure 8.2 depicts the congruity effect between *categorical heaviness* and *object*. Table 8.2 summarises the parameter estimates of the final model. Visual inspection of residual Q-Q plots did not reveal any obvious departures from normality. In addition inspection of residual plots demonstrated that the assumptions of linearity and homoscedasticity were also upheld.

Table 8.2

*Summary table for the final model for Experiment 11*

| **Fixed effects (criterion)** | **Estimated coefficient** | **SE** | **Wald confidence intervals 2.5%** | **Wald confidence intervals 97.5%** | **t-value** |
|---|---|---|---|---|---|
| (Intercept) | 1.93 | 0.072 | 1.79 | 2.07 | 26.82 |
| Trial | 0.0005 | 0.0002 | 0.0002 | 0.0008 | 3.17 |
| Response Hand (Right) | 0.24 | 0.10 | 0.04 | 0.43 | 2.37 |
| Object (Light) | -0.01 | 0.014 | -0.039 | 0.014 | -0.91 |
| Categorical Heaviness (Light) | 0.20 | 0.10 | 0.003 | 0.39 | 1.99 |
| Cat. Heaviness (Light): Resp. Hand (Right) | -0.43 | 0.20 | -0.82 | -0.04 | -2.19 |
| Congruence (Incongruent) | -0.11 | 0.014 | -0.14 | -0.087 | -8.32 |

| **Random effects** | **Name** | **Std. Dev** | **Variance** |
|---|---|---|---|
| Participant | (Intercept) | 0.048 | 0.22 |
| Residual | | 0.13 | 0.39 |

| **AIC** | **BIC** | **Loglik** | **Deviance** |
|---|---|---|---|
| 2269.6 | 2322.9 | -1125.8 | 2251.6 |

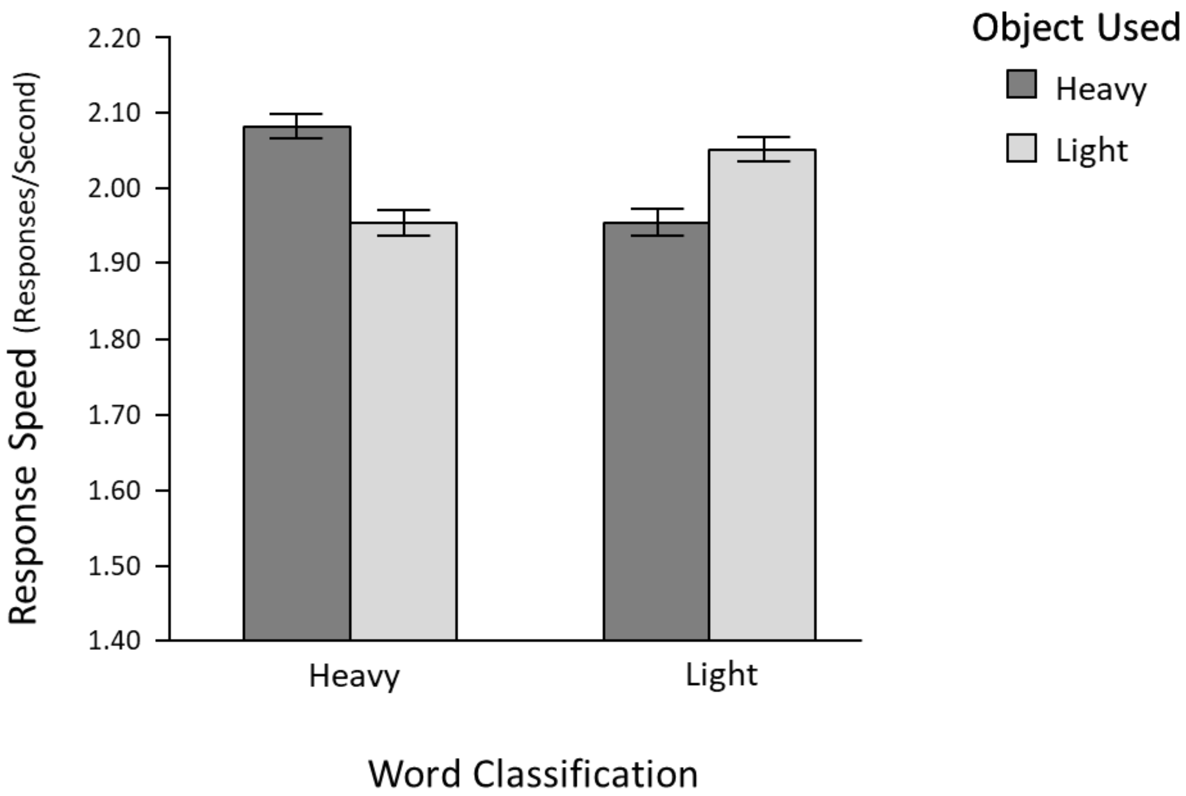



Figure 8.2. Mean response speed (responses/second) for the heavy and light responses with each object. Error bars refer to the standard error of the mean

***8.2.2.2 Response Accuracy***

The overall response accuracy was 98.44%. The mean percentage correct responses were 98.61% and 98.26% for congruent and incongruent trials respectively. The numbers of incorrect responses were too few to warrant the same form of analysis of error rates as was used on RTs. As an alternative, the number of correct responses for congruent and incongruent trials were analysed using a related-samples Wilcoxon Signed Ranks Test. This indicated that this was not a significant difference in congruency ($Z$=-.524, $p$=.600). This confirms that the difference in speed for congruent and incongruent trials did not result in a speed/accuracy trade off. Furthermore, it suggests that congruity did not influence response accuracy.

### 8.2.3 Discussion

The findings demonstrate that the heaviness of objects held passively can interfere with response speed when used to classify the words “heavy” and “light”. Responses were faster when the button on the heavy object was used to respond to the word “Heavy” and when the button on the light object was used to respond to the word “Light” compared to the other way round. This finding confirms that the objects can enter into associations based on their difference in heaviness despite only being held passively. There was also found to be an interaction between the word being classified and response hand which is difficult to interpret. Responses were faster when the right hand was being used to classify stimuli as heavy and the left hand to classify stimuli as light compared to the opposite way. It is unclear why this interaction may arise.

## 8.3 Experiment 12

Having confirmed that the difference in the heaviness of the objects continues to have the potential to induce congruity effects when they are used to make a response while being held passively; the present experiment aims to replicate the speeded classification task used in the Experiment 7 where objects of the same size but varying in heaviness were used as response keys to classify visual stimuli varying in brightness. It is predicted that the same brightness-heaviness interaction found in Experiment 7 will continue to be observed despite the change in the action required to make a response. Responses being faster when the heavier object is used to classify stimuli as dark and the light object is used to classify the stimuli as bright compared to the other way around.

### 8.3.1 Method

#### *8.3.1.1 Participants*

Twenty-one students from Lancaster University volunteered to take part in the study for payment or course credit. Two participants were not included in the analysis because half of responses were incorrect[8]. The remaining 19 participants (18 female, 1 male) were aged between 18 and 21 (mean age =18.68 years). Fifteen participants spoke English as their first language. The remaining three spoke Chinese as their first language. All participants except two were right handed by self-report.

#### *8.3.1.2 Materials*

**Stimuli for classification.** The experiment was conducted using PsyScript (version 2.3.0) on a 27in (2560x1440) computer screen (Apple Thunderbolt LED backlit display controlled by an Intel Core i7 2.6GHz Mac mini Server). The visual stimuli were the same as those used for the brightness classification tasks in Chapters 5 and 6. That is, Four circles, 4.5cm in diameter varying in brightness from black to white (340, 150, 42, 2 cd/m$^2$). They were presented in the centre of the screen on a mid-grey background (90 cd/m$^2$).

**Response Keys.** The response keys were the same objects used in Experiment 11 (see Figure 8.1).

#### 8.3.1.3 Design and Procedure

Participants completed four blocks of 36 trials. On each trial a circle at one of four brightness levels was displayed on the computer screen. Participants responded, as quickly as possible, indicating whether the circle on the screen was brighter or darker than the mid-grey

[8] It is suspected that these participants based their answers on the objects heaviness as opposed to which object was in their left or right hand.

background, by pressing the button on one of the two cylinders (one being held in each hand). Half the participants were instructed to press the object in their left hand for bright circles and their right hand for dark circles. The other half of participants did this the opposite way. The left-right assignment of bright or dark responses remained the same for each participant throughout the task. A small label at the bottom of the screen reminded the participant which hand they were to use to respond to ‘bright’ and which for ‘dark’. The objects being used in the left and right hand were alternated between blocks. The starting left and right assignment of each object was counterbalanced independently of the left-right assignment of bright or dark responses.

Participants were asked to hold an object in each hand underneath the material so that the thumb was able to press the button on the top of each object. They were instructed to position themselves so that the objects were fully lifted off the table in a way that was comfortable for them. In the majority of cases this involved participants leaning on the table with the elbow and forearm with their hands slightly raised of the table. The difference in the weight of the objects was not mentioned by the researcher.

After each response, there was a two second interval before the next circle was presented in this time only the mid-grey background was displayed. Within each block of trials, each circle was shown nine times. The order in which the circles were presented was randomised across sets of 12 trials, so that within each set of twelve trials each circle appeared 3 times. Each block of trials took approximately 2 minutes. Between each block, participants moved to another desk for a two minute break. They had a word search to complete during this time. At this point the researcher swapped the left-right position of the two objects. This was not mentioned by the researcher, however, some participants reported being aware that the objects had been switched.

### 8.3.2 Results

The data were the accuracy levels and the correct response times (RTs) for the brightness classifications. Out of 2736 trials, 2692 of these were correct responses (98.39%). Any RTs below 150ms and more than 2.5SD above a participant's mean RTs were excluded from analysis leaving 2616 observations (95.61%). Table 1 summarises the accuracy and RTs for *dark* and *bright* classifications made with the *heavy* and *light* objects.

Table 8.3

*Mean RTs (SEM in parentheses) and accuracy levels (%) according to the object used and categorical brightness*

| | Brightness | |
|---|---|---|
| Object Used | Dark | Bright |
| Heavy | **518 (8)** | 542 (9) |
| | **98.1%** | 97.7% |
| Light | 528(7) | **526 (8)** |
| | 98.8% | **99.0%** |

Note: Bold typeface signifies congruent conditions

#### *8.3.2.1 Response Speed*

The analysis approach was the same general strategy as that used in the previous speeded classification tasks. In order to resolve the skew of the RT distribution, a transformation of 1/RT was performed. The product of this transformation is interpreted as *response speed* (responses/second).The main aim of the analysis was to determine if an interaction between *object* (*heavy* Vs *light*) and *categorical brightness (dark* Vs *bright*) significantly improves model fit. For ease of interpretation this was included as a main effect of *congruence (congruent trials* Vs *incongruent trials)* along with main effects of *categorical brightness* and *object*. The LRT demonstrates that a basic model including *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2(1) = 4.42$, $p$=.036) with an AIC value of 3750.9 and 3748.5 for the null and comparison model respectively. The parameter estimates of this basic model suggest that incongruent trials were slower than congruent trials by 0.040 responses/second (SE=0.020), CI [0.0027, 0.077].

The contributions of the following additional variables were assessed along with interactions between variables which may have theoretical implications: *trial*, *response hand* (*left* Vs *right*), *within-category brightness*, and *brightness contrast* (*high* Vs *low).* A model including these additional explanatory variables was preferred to the basic model according to the LRT ($\chi^2(4) = 177.69$, $p$<.001) with an AIC value of 3578.8. An inspection of the CIs demonstrated that *within-category brightness* did not have an effect since the values straddle zero, CI [-0.046, 0.026]. LRT confirmed its removal did not change the explanatory power of the model ($\chi^2(1) = 0.286$, $p$=0.59, AIC=3577.1) therefore the model which excluded *within-category brightness* was preferred. An interaction of *object* and *within-category brightness* also failed to improve model fit ($\chi^2(2) = 1.15$, $p$=.56, AIC=3579.9).

Parameter estimates indicate that the speed of responses increased with *trial*. Responses were also faster when *contrast* was *high* i.e. when the brightness value of the circle was further from the mid-grey background (black and white as opposed to dark grey or light grey). An interaction between *object* and *contrast* failed to improve model fit ($\chi^2(1) = 0.33$, $p$=.56, AIC=3578.7). It was demonstrated that the *response hand* improved model fit as a main effect; responses were faster when the left hand was used. *Response hand* did not interact with *object* ($\chi^2(1) = 0.13$, $p$=.72, AIC=3578.9), nor *brightness contrast* ($\chi^2(1) = 0.80$, *p*=.37, AIC=3578.3). However it was approaching significance with *categorical brightness* ($\chi^2$(1) = 3.41, *p*=.065, AIC=3575.7) it was not included in the final model.

The LRT demonstrates that a model including *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2$(1) = 9.40, *p*<.001) with an AIC value of 3584.5 and 3577.1 for the models excluding and including *congruence* respectively. The parameter estimates of this final model suggest that incongruent trials were slower than congruent trials by 0.057 responses/second (SE=0.02), CI [0.02, 0.92]. Table 8.4 summarises the parameter estimates of the final model and Figure 8.3 depicts the congruity effect between object and categorical brightness. Visual inspection of residual Q-Q plots did not reveal any obvious departures from normality. In addition inspection of residual plots demonstrated that the assumptions of linearity and homoscedasticity were also upheld.

Table 8.4

*Summary table for the final model for Experiment 12*

| | | | **Wald confidence intervals** | | |
|---|---|---|---|---|---|
| **Fixed effects (criterion)** | **Estimated coefficient** | **SE** | **2.5%** | **97.5%** | **t-value** |
| (Intercept) | 2.07 | 0.10 | 1.87 | 2.27 | 20.42 |
| Trial | 0.002 | 0.0002 | 0.0014 | 0.0022 | 8.00 |
| Brightness Contrast (Low) | -0.20 | 0.018 | -0.23 | -0.16 | -10.73 |
| Response Hand (Right) | 0.039 | 0.018 | 0.003 | 0.075 | 2.12 |
| Object (Light) | -0.011 | 0.018 | -0.047 | 0.025 | -0.63 |
| Categorical Brightness (Dark) | 0.036 | 0.018 | -0.0004 | 0.072 | 1.94 |

| | | | | | |
|---|---|---|---|---|---|
| Congruence (Incongruent) | -0.057 | 0.018 | -0.092 | -0.020 | -3.07 |

| Random effects | Name | Std. Dev | Variance |
|---|---|---|---|
| Participant | (Intercept) | 0.18 | 0.43 |
| Residual | | 0.22 | 0.47 |

| AIC | BIC | Loglik | Deviance |
|---|---|---|---|
| 3577.1 | 3629.9 | -1779.5 | 3559.1 |

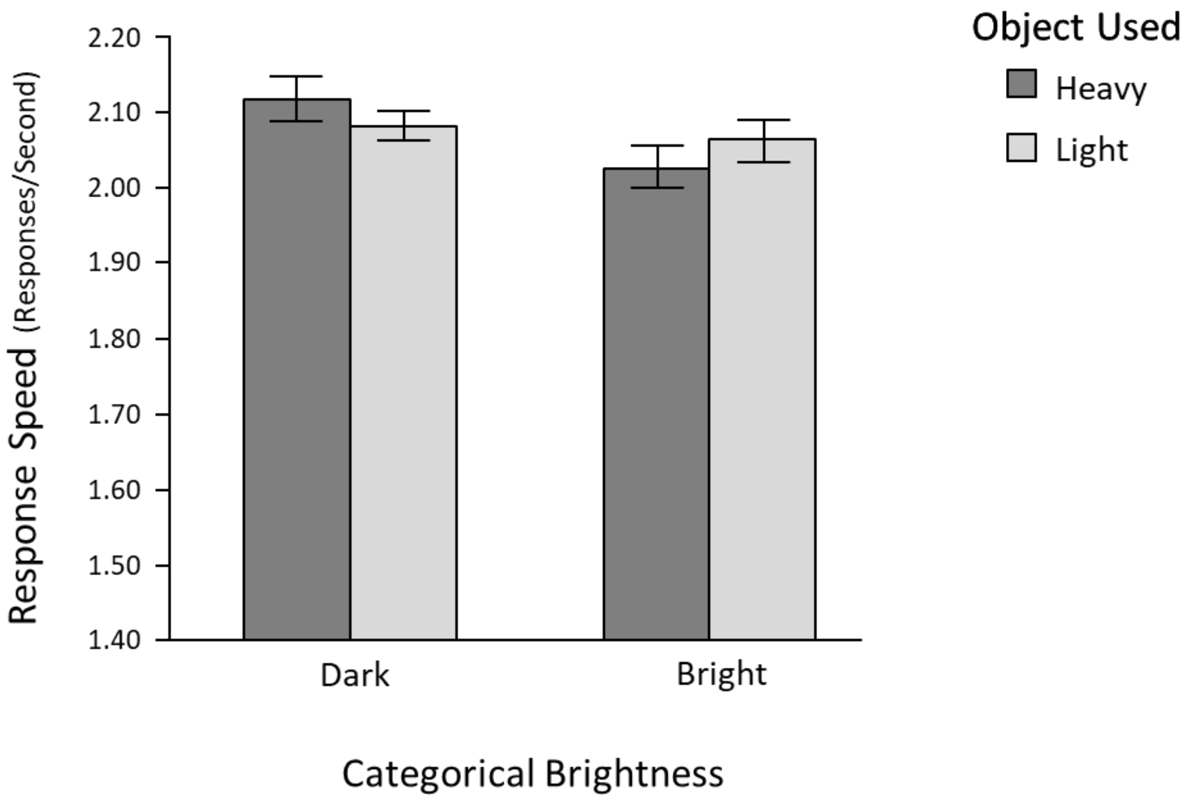



*Figure 8.3.* Mean response speed (responses/second) for the dark and bright responses with each object. Error bars refer to the standard error of the mean.

#### *8.3.2.2 Response Accuracy*

The overall response accuracy was 98.39%. The percentage correct responses for congruent and incongruent trials were 98.54% and 98.26% respectively. The numbers of incorrect responses were too few to warrant the same form of analysis of error rates as was used on RTs. As an alternative, the number of correct responses for congruent and incongruent trials were analysed using a related-samples Wilcoxon Signed Ranks Test. This indicated that there was not a significant difference in congruency ($Z$=-.771, $p$=.441). This confirms that the difference in speed for congruent and incongruent trials did not result in a speed/accuracy trade off. Furthermore, it suggests that congruity did not influence response accuracy.

### 8.3.3 Discussion

The findings of the present experiment replicate the congruity effect found in Experiment 7. Responses were faster when the heavy object was used to classify visual stimuli as dark and the less heavy object to classify visual stimuli as bright, compared to the other way around. This demonstrates that the heaviness of the objects being used as response keys can interfere with the speed in which brightness classifications are made, even when the objects are held passively. The replication of the congruity effect between heaviness and brightness, despite the change in the way the objects were used add support to the suggestion that the congruity effect is a result of the interaction between these dimensions at an amodal and conceptual level. It demonstrates that the specific actions or the way in which the objects are being used are inconsequential, but instead suggests it is the activation of a concept of heaviness which is necessary. This is in keeping with the idea that the congruity effect should arise irrespective of the nature in which the dimensions are probed or activated (P. Walker and Walker, 2012; P.Walker, 2016).

Alongside the congruity effect, the final model in the present experiment replicated effects which were observed in Experiment 7. There was a main effect of *trial* such that responses became faster as the experiment progressed. In addition, there was a main effect of *categorical brightness* such that responses were faster when classifying stimuli as dark. Responses were also faster when the *brightness contrast* was high (i.e. when the circles were more different than the mid-grey background). The present findings depart from that of Experiment 7 as no interaction was demonstrated between *object* and *response hand*. This supports the suggestion made in Experiment 7 that the interaction between these two variables was due to the objects being easier to use when in one left/right hand assignment compared to another.

## 8.4 Experiment 13

In Experiment 9, a heaviness-brightness congruity effect was observed when objects varying in heaviness were used to tap a touch sensitive surface in order to make a response. This congruity effect was found despite the objects being used also varying in size, in such a way that the size of the object predicted a congruity effect in the opposite direction to that of heaviness. Responses were faster when a heavier (but smaller) object was used to classify stimuli as dark and when a lighter (but bigger) object was used to classify stimuli as bright compared to the other way around. As discussed in Section 6.4.2 there are a number of possible explanations for the dominance of heaviness over size, one is that the active use of the objects makes heaviness a more salient feature, resulting in the primacy of a heaviness-brightness correspondence over a potential size-brightness correspondence.

The adapted speeded classification task introduced in this chapter offers an opportunity to explore whether the dominance of heaviness over size can be explained by the active use of the objects to make a response. Therefore, in the present experiment, the same objects used in Experiment 9 had micro-switches mounted to the top of them in a similar way to the objects used in Experiments 11 and 12. These were used in the same brightness speeded classification task, to determine whether the heaviness of the objects will continue to override contradictory size information when the objects are held passively.

### 8.4.1 Method

#### *8.4.1.1 Participants*

Twenty students from Lancaster University volunteered to take part in the study for payment or course credit. One participant was not included in the analysis because of an they . The remaining 19 participants (14 female, 5 male) were aged between 19 and 31 (mean age =25.05 years). All participants except two were right handed by self-report. Eleven of the participants spoke English as their first language. The remaining eight participants spoke the following first languages: Chinese/ Cantonese/ mandarin (n=3), Dutch (n=1), German (n=1), Malay (n=2) and Romanian (n=1).

#### *8.4.1.2 Materials, design and procedure*

The materials, design and procedure were the same as Experiment 12. However, the experiment differs with regard to the objects being used as response keys. A pair of objects made up of the biggest (but lightest) and smallest (but heaviest) objects from the full set of nine objects (i.e., object C and G see Figure 8.) had the same type of microswitches used in Experiments 11 and 12 attached.

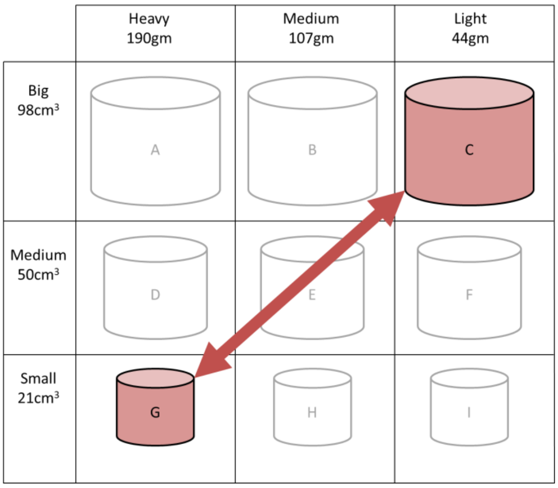



Figure 8.4. The objects from the full set of nine used as response keys in Experiment 13

### 8.4.2 Results

The data were the accuracy levels and the correct response times (RTs) for the brightness classifications. Out of 2736 trials, of these were correct responses (97.26%). Any RTs below 150ms or above 2.5SD a participant's mean response time were excluded from analysis leaving 2598 observations (94.96%). Table 8.5 summarises the accuracy and RTs for *dark* and *bright* classifications made with the *heavy* and *light* objects.

Table 8.5

*Mean RTs (SEM in parentheses) and accuracy levels (%) according to the object used and categorical brightness*

| | Brightness | |
|---|---|---|
| Object Used | Dark | Bright |
| Small/Heavy | **542(8)** | 592 (10) |
| | **97.5%** | 96.9% |
| Big/Light | 586(9) | **560(7)** |
| | 97.7% | **96.9%** |

Note: Bold typeface signifies congruent conditions in accordance with the heaviness-brightness correspondence

#### *8.4.2.1 Response Speed*

The analysis approach was the same general strategy as that used as in the previous speeded classification task. In order to resolve the skew of the RT distribution, a transformation of 1/RT was performed. The product of this transformation is interpreted as *response speed* (responses/second).The main aim of the analysis was to determine if an interaction between *object* (*small/heavy* Vs *big/light*) and *categorical brightness (dark* Vs *bright*) significantly improves model fit. For ease of interpretation this was included as a main effect of *congruence (congruent trials* Vs *incongruent trials)* along with main effects of *categorical brightness* and *object*. The LRT demonstrates that a model including *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2(1) = 31.30$, $p<.001$) with an AIC value of 3108.0 and 3078.7 for the null and comparison model respectively. The parameter estimates of this basic model suggest that incongruent trials were slower than congruent trials by 0.095 responses/second (SE=0.017), CI [0.062, 0.128].

The contributions of the following additional variables were assessed along with interactions between variables which may have theoretical implications: *trial*, *response hand* (*left* Vs *right*), *within-category brightness (*which is whether the brightness value is the *dark* or *bright* value of the two sharing a categorical brightness level, see Table 5.1), and *Brightness contrast* (*high* Vs *low).* A model including these additional explanatory variables was preferred to the basic model according to the LRT ($\chi^2(4) = 328.97$, $p<.001$) with an AIC value of 2757.7. An inspection of the CIs demonstrated that *within-category brightness* did not have an effect since the values straddle zero, CI [-0.052, 0.010]. LRT confirmed its removal did not change the explanatory power of the model ($\chi^2(1) = 1.83$, $p=.18$, AIC=2727.6) therefore the model which excluded *within-category brightness* was preferred. An interaction of *object* and *within-category brightness* also failed to improve model fit ($\chi^2(2) = 2.06$, $p=.357$, AIC=2759.5).

Parameter estimates indicate that the speed of responses increased with *trial*. Responses were also faster when the *contrast* was *high* i.e. when the brightness value of the circle was further from the mid-grey background (black and white as opposed to dark grey or light grey). An interaction between *object* and *contrast* failed to improve model fit ($\chi^2(1) = 0.86$, $p=.354$, AIC=2758.7). It was demonstrated that the *response hand* improved model fit as a main effect; responses were faster when the left hand was used. *Response hand* did not interact with *object* ($\chi^2(1) = 2.77$, $p=.10$, AIC=2756.8), *categorical brightness* ($\chi^2(1) = 0.51$, $p=.474$, AIC=2759.1) nor *brightness contrast* ($\chi^2(1) = 0.48$, $p=.49$, AIC=2759.1).

The LRT demonstrates that a model including *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2(1) = 34.09$, $p<.001$) with an AIC value of 2789.7 and 2757.6 for the models excluding and including *congruence* respectively. The parameter estimates of this final model suggest that incongruent trials were slower than congruent trials by 0.093responses/second (SE=0.016), CI [0.062, -0.12]. This also confirms that the effect of *congruence* remains over and above the contributions of the additional explanatory variables included. Table 8.6 summarises the parameter estimates of the final model and Figure 8.5 depicts the congruity effect of categorical brightness and object. Visual inspection of residual Q-Q plots did not reveal any obvious departures from normality. In addition inspection of residual plots demonstrated that the assumptions of linearity and homoscedasticity were also upheld.

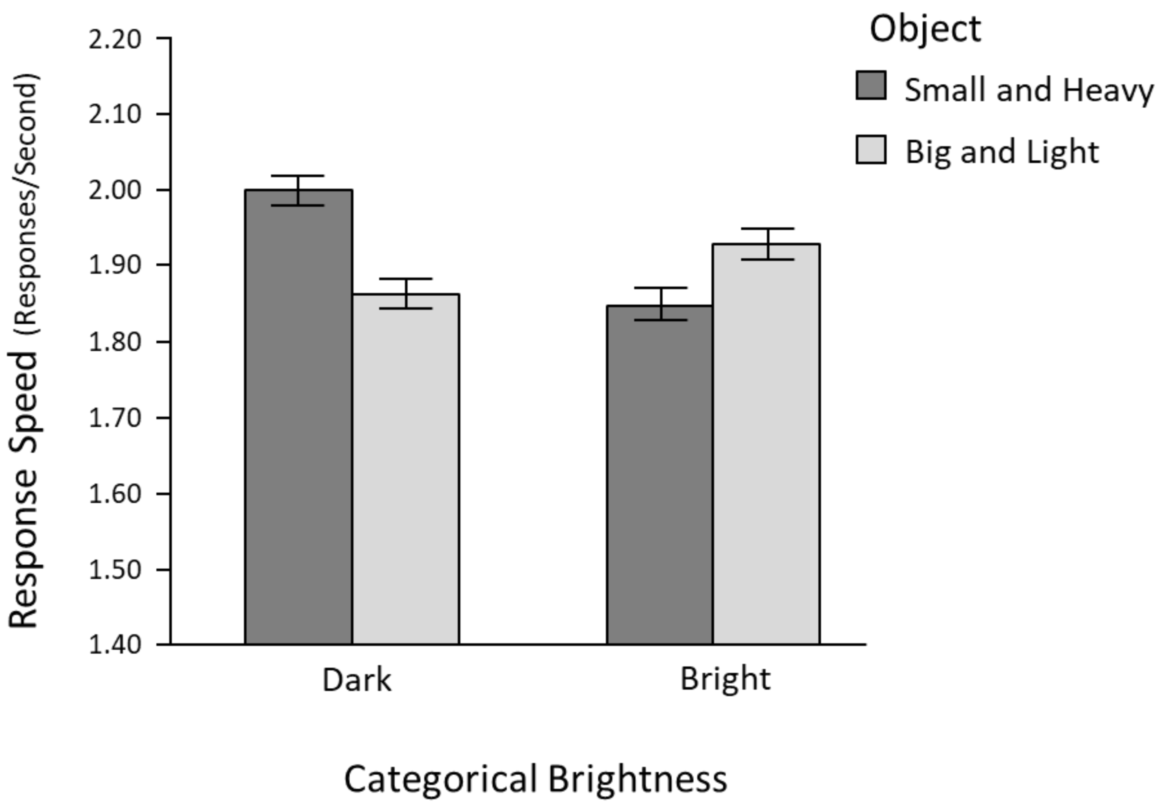



Figure 8.5. Mean response speed (responses/second) for the dark and bright responses with each object. Error bars refer to the standard error of the mean

Table 8.6

*Summary table for the final model for Experiment 13*

| | | | **Wald confidence intervals** | | |
|---|---|---|---|---|---|
| **Fixed effects (criterion)** | **Estimated coefficient** | **SE** | **2.5%** | **97.5%** | **t-value** |
| (Intercept) | 1.92 | 0.08 | 1.75 | 2.08 | 23.07 |

| | | | | | |
|---|---|---|---|---|---|
| Trial | 0.002 | 0.0002 | 0.0018 | 0.0025 | 11.29 |
| Brightness Contrast (Low) | -0.24 | 0.016 | -0.27 | -0.21 | -14.91 |
| Response Hand (Right) | 0.042 | 0.016 | 0.011 | -0.062 | 2.66 |
| Object (Small/Heavy) | -0.039 | 0.016 | -0.070 | -0.008 | -2.45 |
| Categorical Brightness (Dark) | 0.038 | 0.016 | 0.0074 | 0.069 | 2.43 |
| Congruence (Incongruent) | -0.093 | 0.016 | -0.12 | -0.062 | -5.86 |

| **Random effects** | **Name** | **Std. Dev** | **Variance** |
|---|---|---|---|
| Participant | (Intercept) | 0.12 | 0.35 |
| Residual | | 0.16 | 0.40 |

| **AIC** | **BIC** | **Loglik** | **Deviance** |
|---|---|---|---|
| 2757.6 | 2810.3 | -1369.8 | 2739.6 |

#### *8.4.2.2 Response accuracy*

The overall response accuracy was 97.26%. The percentage correct responses for congruent and incongruent trials were 98.15% and 97.30% respectively. The numbers of incorrect responses were too few to warrant the same form of analysis of error rates as was used on RTs. As an alternative, the number of correct responses for congruent and incongruent trials were analysed using a related-samples Wilcoxon Signed Ranks Test. This demonstrated no significant difference between congruent and incongruent trials based on accuracy ($Z$=-1.214, $p$=.225). This confirms that the difference in speed for congruent and incongruent trials did not result in a speed/accuracy trade off. Furthermore, it suggests that congruity did not influence response accuracy.

### 8.4.3 Discussion

The aim of the present experiment was to determine if the congruity effect observed in Experiment 9 would be replicated when the objects were not actively used to make a response. The nature of the congruity effect demonstrates that responses were faster when the heavy (but smaller) object was used to classify stimuli as dark and when the lighter (but bigger) object was used to classify stimuli as bright compared to the other way round. This indicates that the presence of a heaviness contrast gives rise to a dominant brightness-heaviness correspondence over a potential brightness-size correspondence irrespective of whether the object is being used actively to make a response or being held passively. This suggests that the active use of the objects alone is not responsible for the dominance of heaviness over size observed in these experiments.

Alongside the brightness-heaviness congruity effect, the final model replicates that of Experiment 9 in several ways. As will all previous classification tasks, a main effect of trial was observed. There was also a main effect of brightness contrast such that responses were faster when the contrast is high (i.e. when the brightness level is further from the mid-grey background). There was also a main effect of categorical brightness such that responses were faster when classifying circles as dark compared to bright. The model in the present study departs from that of Experiment 9 with a main effect of response hand. Responses were faster with the right hand compared to the left hand, which is consistent with the findings from the previous speeded classification tasks reported in this chapter. Since this has only been observed consistently in the tasks where the objects were held passively, it may reflect an asymmetry in left hand/right hand reaction times that would not have been revealed in the tasks involving more complex movements to make a response.

## 8.5 Experiment 14

The Experiments in the present chapter thus far have replicated the key findings of a brightness-heaviness congruity effect that was initially observed in Chapter 6, when objects were used actively to tap a touch sensitive surface in order to make a response. The aim of the final experiment in this series is to determine if the pitch-heaviness congruity effect observed in Chapter 7 (Experiment 10) can also be replicated in a task where the objects are held passively. This adaptation, allows one of the key limitations of Experiment 10 to be addressed (see Section 7.2.3); the failure to mask the impact sounds of the objects being used to tap the touch sensors to make a response. Therefore the present study uses the same pitch classification task used in Experiment 10 but with the objects that have been used in Experiment 11 and 12. It is predicted that the same pitch-heaviness congruity effect, responses being faster when the heavier object is used to classify sounds as low pitch and when the lighter object is used to classify stimuli as high pitch compared to the other way.

### 8.5.1 Method

#### *8.5.1.1 Participants*

Twenty students from Lancaster University (13 females, 7 males) aged between 18 and 27 (mean age 20.3 years) volunteered to take part in the study for payment or course credit. Sixteen of the participants spoke English as their first language. The remaining four participants spoke the following first languages: Cantonese (n=2), German (n=1) and Swedish (n=1). All except two were right handed by self-report.

#### *8.5.1.2 Materials*

Two frequency sounds, one of a high pitch (3520hz) and one of a low pitch (220hz) were used. The amplitude of the high pitch sound was reduced to a level that was equivalent to the low pitch sound. The sounds were played through headphones. The experiment was run on PsyScript (version 2.3.0) on a 27in (2560x1440) computer screen (Apple Thunderbolt LED backlit display controlled by an Intel Core i7 2.6GHz Mac mini Server). Along with the presentation of each sound a question mark appeared in the centre of the screen.  The objects used to make a response were the same as those used in Experiment 1 and 2.

#### *8.5.1.3 Design and Procedure*

Participants completed four blocks of 36 trials. On each trial a sound was presented along with a question mark prompt in the centre of the screen. Participants were instructed to respond as quickly and as accurately as possible to whether each sound was high or low in pitch by pressing the switch on the object in their left or right hand. Half the participants in each group were asked to press with their left hand for high and their right hand for low. The other half of participants were assigned to the opposite mapping. The left-right assignment of high or low pitch remained the same for each participant during the study. A small label at the bottom of the screen reminded the participant which hand they were to use for ‘high’ and which for ‘low’.

Participants were asked to hold an object in each hand underneath the material and position their hand so they can press the button on top of each object with their thumbs. They were instructed to lean with their forearms on the table and to raise their hands slightly so that the objects were lifted in a position that was comfortable. The difference in heaviness between the two objects was not mentioned by the researcher. The object being held in each hand was alternated between blocks of trials. Within each group, the starting left and right assignment of each object was counterbalanced. There were three groups of participants, each using a different size of objects.

On each trial a sound was presented along with a question mark prompt in the centre of the screen. The sound lasted for two seconds but participants could respond from the onset of the sound. The question mark remained on the screen until a response was made. Participants responded as quickly as possible, to whether each sound was high or low pitch. Participants responded by pressing the switch on one of two cylinders (either the right hand to indicate the sound was low and left hand for high, or vice versa). There was a two second interval before the next sound was presented, in this time only the mid-grey background was displayed. The order in which the sounds were presented was randomised across sets of 12 trials, so that within each set of twelve trials each sound appeared 6 times. Each block of trials took approximately 2 minutes. Between each block, participants moved to another desk for a two minute break. They had a word search to complete during this time. At this point the researcher swapped the left-right position of the two objects.

**8.5.2 Results**

The data were the accuracy levels and the correct response times (RTs) for the pitch classifications. Out of 2880 trials, 2849 of these were correct responses (98.92%). Any RTs below 150ms or above 2.5SD of the mean for each participant were excluded from analysis leaving 2764 observations (95.97%). Table 8.7 summarises the accuracy and RTs for *low* and *high* classifications made with the *heavy* and *light* objects.

Table 8.7

*Mean RTs (SEM in parentheses) and accuracy levels (%) according to the object used and pitch categorisation*

| | Pitch to be Categorised | |
|---|---|---|
| Object Used | Low | High |
| Heavy | **422(7)** | 413 (6) |
| | **99.03** | 98.61 |
| Light | 416(5) | **406 (5)** |
| | 99.03 | **99.03** |

Note: Bold typeface signifies congruent conditions

#### ***8.5.2.1 Response Speed***

The analysis approach was the same general strategy as that used in Experiment 7. In order to resolve the skew of the RT distribution, a transformation of 1/RT was performed. The product of this transformation is interpreted as *response speed* (responses/second).The main aim of the analysis was to determine if an interaction between *object* (*heavy* Vs *light*) and *pitch (low* Vs *high*) significantly improves model fit. For ease of interpretation this was included as a main effect of *congruence (congruent trials* Vs *incongruent trials)* along with main effects of *pitch* and *object*. The LRT demonstrates that a model including *congruence* as a main effect is not preferred to a model in which it is not included ($\chi^2(1)$ =2.53, $p$=.011) with an AIC value of 5071.6 and 5072.2 for the null and comparison model respectively. An inspection of the confidence intervals for the interaction term straddle zero which confirms that there is no congruency effect CI[-0.08,0.0083]. The congruence term was removed and an exploration of additional explanatory variables was conducted.

The contributions of *trial* and *response hand* (*left* Vs *right*) were assessed along with interactions between variables which may have theoretical implications. A model including these additional explanatory variables was preferred to the basic model according to the LRT ($\chi^2(2) = 59.69$, $p<.001$) with an AIC value of 5016.5 An inspection of the CIs demonstrated that *pitch* CI [-0.076, 0.012], *response hand* CI [-0.015, 0.07] and *object* CI [-0.028, 0.060] did not improve the model as an effect since the values straddle zero. LRT confirmed their removal did not change the explanatory power of the model ($\chi^2(3) = 417$, $p=.24$, AIC=5014.7) therefore the model which excluded *pitch* and *object* was preferred. An interaction of *pitch* and *response hand* also failed to improve model fit ($\chi^2(3) = 3.75$ $p=.0.29$, AIC=5016.9). As did an interaction between *object* and *response hand* ($\chi^2(3) = 2.12$, $p=.55$, AIC=5018.5. The final model included a main effect of trial only. Parameter estimates indicate that the speed of responses increased with *trial*. An LRT demonstrates that a model including *congruence* along with object and pitch as a main effects failed to improve model fit when reintroduced ($\chi^2(3) = 5.06$, $p=0.17$) with an AIC value of 5014.7 and 5015.6 for the models excluding and including *congruence* respectively.

#### *8.5.2.2 Response Accuracy*

The overall response accuracy was 98.92%. The percentage correct responses for congruent and incongruent trials were 97.88% and 96.64 % respectively. The numbers of incorrect responses were too few to warrant the same form of analysis of error rates as was used on RTs. As an alternative, the number of correct responses for congruent and incongruent trials were analysed using a related-samples Wilcoxon Signed Ranks Test. This indicated that there was not a significant difference ($Z$=--0.256., $p=0.79$). This confirms that the difference in speed for congruent and incongruent trials did not result in a speed/accuracy trade off. Furthermore, it suggests that congruity did not influence response accuracy.

### 8.5.3 Discussion

Unlike in Experiment 10, no congruity effect was found between the heaviness of the objects and the pitch of sounds being classified. This suggests that the congruity effect observed in Experiment 10 was mediated by the impact sound of the objects. The implication of this explanation is that a mapping between heaviness and pitch does not support correspondence based congruity effects. Although this may be the case, it may be premature to dismiss this possibility as a result of these experiments alone. There are a number of aspects of the task procedure which may have resulted in the failure to observe a congruity effect. Firstly, the task is different to the brightness classification task since only two values on the to-be-classified dimension were used, and no referent sound was used to provide a comparison for the stimuli. In a pitch classification task conducted by Lidji, et al. (2006), a congruity effect was demonstrated between the position of response keys and the pitch of sound being made. This task used four levels of pitch and included a reference tone played at the start of the procedure for comparison. Having said that, congruity effect was observed by Marks (1974) when only two values of pitch were used, however in this study, the incidental feature dimension was visual (bright/dim light and black/white surfaces) therefore is less comparable to the task used in the present experiment. Therefore, further work is required to dismiss other potential experimental factors for the lack of congruity effect.

Alternatively, it may be that the heaviness-pitch correspondence is dependent on the active use of the objects. This would also account for the congruity effect observed in Experiment 10 but not observed here. In order to explore this further, it would be interesting to conduct a classification task where the objects are used actively to make a response but do not make an impact sound. For example, by having participants make a response by moving hand held objects that are picked up by motion detectors. Or if something can be added to the objects or the surface being tapped to dampen the sound further.

## 8.6 General Discussion

The experiments in the present chapter used objects with switches mounted on top as response keys for a series of speeded classification tasks. Experiment 11 confirmed that, despite being held passively, the heaviness of the objects was salient enough to induce a congruity effect when being used to classify the words “Heavy” and “Light”. When these objects were used in a brightness classification task, the same heaviness-brightness congruity effect was found as in Chapter 6. Responses were faster when a heavier object was used to classify stimuli as dark and when a less heavy object was used to classify stimuli as bright compared to the other way around. Furthermore, the same brightness-heaviness congruity effect continued to arise where objects varied in both size and heaviness such that the smaller object was heavier than the larger object, providing further evidence of a dominance of heaviness over size.

By ensuring that the heaviness of the objects had no physical bearing on the way responses were executed (e.g. by influencing how easy they were to move or tap), the present experiments suggest that the congruity effects between brightness and heaviness, observed in Chapter 6, cannot be explained in terms of the action required to make a response. Consequently, the present findings provide further evidence in support of a brightness-heaviness mapping.  When the present findings are taken together with those of Chapter 6, the conceptual basis for correspondences with heaviness is revealed by demonstrating that the same mapping can arise when different forms of object-related actions are used to make a response. This suggests that the heaviness-brightness mapping is not dependent on sensorimotor interactions with the object, but that the association is based on a representation of heaviness which arises irrespective of the way objects are interacted with. The same cannot be said of the correspondence between heaviness and pitch, since the congruity effect between heaviness and pitch, demonstrated in Experiment 10 was not replicated in

Experiment 14 when the objects were held passively. As mentioned in Section 8.5.3, more work is required to inspect the possible explanations for the lack on congruity in this case

# Chapter 9

## 9.1 Introduction

In the rating scale tasks (Chapters 2-4) and in speeded classification tasks (Chapters 5, 6 and 8) a correspondence between heaviness and brightness has been demonstrated. Participants rated heavier objects as darker than less heavy objects. In addition, participants responded faster when a heavy object was used to classify stimuli as dark and a less heavy object was used to classify stimuli as bright compared to the other way around. One possible basis for the association between brightness and heaviness is the common verbal label ‘light’ that is used in English to mean both: less heavy and bright. The possible confounding influence of the common labelling was acknowledged in the design process of the experiments. The term ‘bright’ was used instead of ‘light’ for all procedures involving brightness. The term ‘light’ was also avoided where possible when referring to the less heavy objects[9]. In addition to these precautions, the associations between pitch and heaviness (where no common labelling occurs) were also tested. However, avoiding the term ‘light’ does not necessarily prevent participants from themselves internally labelling the less heavy objects or indeed the brightness stimuli as ‘light’. What is more, a pitch-heaviness congruity effect was not found in Experiment 13, where objects varying in heaviness were held passively and used to classify stimuli as high and low pitch, which in turn casts some doubt on the pitch-heaviness congruity effect found in Experiment 10. Therefore, the present chapter is concerned with the role the common label may potentially have in the observed brightness-heaviness congruity effect found in the speeded classification tasks reported in the preceding chapters.

---

[9] In Experiment 5, one group of participants rated the objects in terms of heaviness on a scale which used “light” and “heavy” as anchors. However, they were not asked to rate the objects in terms of brightness. In the preliminary task described in the same chapter however, the participants did rate objects in terms of heaviness on the scale including ‘light’ as an anchor and also rated the objects on a scale of brightness, which is a limitation.

There are two ways in which the shared label ‘light’ may be influencing the presence of a brightness-heaviness correspondence. Firstly the general presence of a shared label in our vernacular may reinforce an association between brightness and heaviness. It has been demonstrated that the presence of shared verbal labels in different languages can modulate the observed correspondences between pitch and other feature dimensions. For example, Dolsheid et al., (2013) asked speakers of Dutch (who use common verbal labels for pitch and spatial elevation) and speakers of Farsi (who use common verbal labels for pitch and thickness) to reproduce the pitch of sounds that were presented simultaneously with lines varying in thickness or vertical location. It was found that Farsi speakers were influenced by the thickness of visual stimuli, but not visual height when reproducing the pitch of sounds; whereas, Dutch speakers were influenced by visual height, but not thickness. Given that preverbal infants demonstrated sensitivity to mappings of pitch with both thickness and height (Dolsheid, et al., 2014); it is argued that the presence of a verbal overlap in a language strengthens the presence or use of one mapping rather than another. This may suggest that the presence of a common label for brightness and heaviness in English may strengthen the mapping between these features.

Alternatively, it may be that the shared verbal label is having an immediate influence on task performance. For example, it may be that the participants are recruiting articulatory processes either consciously or sub-consciously as a strategy for completing the speeded-classification task. There is evidence which implicates articulatory processes in the speeded-classification task, for example, the bigrams *HI* and *LO* produce congruity effects when presented simultaneously with sounds varying in pitch (Melara and Marks, 1990). In addition several studies have demonstrated congruity effects when words are used to represent various feature dimensions (Gallace and Spence, 2006; Martino and Marks, 1999; P. Walker and Smith, 1984; 1985). This suggests that online verbal labelling/articulatory processes can give rise to observed congruity effects. In the case of the present studies, there is the potential for the congruity effect between brightness and heaviness to be explained in this way, given the shared label ‘light’ meaning less heavy and the bright.

Unlike the shared labels of ‘high’ and ‘low’ for pitch and spatial elevation, the shared label for brightness and heaviness only occurs on one pole of the dimensions. Therefore, if there is an effect of common labelling, this would predict that it would emerge in an asymmetrical way: congruent trials would be especially fast where the common label is present compared to congruent trials which do not involve a common label. To investigate this, the results from the two brightness speeded-classification tasks (Experiment 7 and 12) where the objects varied in heaviness alone were explored[10]. The speed of correct responses for congruent trials that involved the shared label ‘light’ (where the less heavy object was being used to classify stimuli as bright) were compared with congruent trials which did not (where the heavy object was used to classify stimuli as dark). For Experiment 12, in which brightness classifications were made by pressing switches attached to objects varying in heaviness, there was no significant difference found between the speeds of these types of trials. For Experiment 7, there was a significant difference between congruent trials involving the common label ‘light’ and congruent trials which did not. However, this was in the opposite direction to what may be anticipated if the common label was having an influence. The mean speed of trials where the heavy object was used to classify dark stimuli was faster than where the light object was used to classify stimuli as bright. This does not support an effect of common labelling. The aim of the present chapter is to further explore whether

[10] A LME model on the congruent trials in Experiment 7 including trial type (light/bright or heavy/dark) as a main effect was preferred to null model $\chi^2=13.292$, $p<.001$ parameter estimates indicate that responses were faster by .05 when heavy object was used for dark congruent trials compared to when light objects were used for bright congruent trials. The same analysis was done on Experiment 12 but in this case trial type did not improve model fit $\chi^2=2.25$, $p=.134$

online verbal labelling may explain the results found in the brightness speeded-classification task.

One way that the role of verbal labelling has been explored is using the verbal-suppression task which is designed to prevent on-line labelling strategies influencing task performance. For example, Connell, et al., (2013) investigated how cues to spatial elevation (an arm moving up or down) influenced the judged pitch of notes being sung. Meanwhile participants were asked to hold in memory a string of letters, by silently repeating them. Since the effect of spatial elevation cues on pitch judgements were found while this additional verbal task was being used, it was argued that the effect of spatial elevation of judged pitch could not be explained by verbal mediation (See Dolsheid et al., 2013 for another example). In both cases, the interaction between dimensions was observed despite the additional task which increased load on verbal systems, suggesting that the initial task does not require verbal mediation. In a similar way, the aim of the present study is to replicate the congruity effect between brightness and heaviness demonstrated in Experiment 12, with the addition of an articulatory suppression task. This is to prevent participants from using online verbal labelling strategy while classifying the brightness stimuli, and thus indicate if the common verbal label has an in-task role in producing a brightness-heaviness congruity effect.

## 9.2 Experiment 15

### 9.2.1 Method

#### *9.2.1.1 Participants*

Twenty students from Lancaster University volunteered to take part in the study for payment or course credits. Participants (14 females and 6 males) were aged between18 and 34 (mean age =22.5 years).  Twelve of the participants spoke English as their first language.

The remaining eight participants spoke the following first languages: Bulgarian (n=1), Chinese (n=2), French (n=1), German (n=1), Polish (n=1), Turkish (n=1) and Vietnamese (n=1). All participants except one were right handed by self-report.

#### *9.2.1.2 Materials, Design and Procedure*

The materials, design and procedures were the same as those used in Experiment 12, with the addition of an articulatory suppression task which was to be completed concurrently with the speeded classification task.

***Articulatory Suppression Task.*** For each block of trials, participants were given a string of five randomly generated letters to remember. Immediately before the commencement of each block, the researcher read aloud the letter string twice and asked the participant to recite it back to them to confirm they had heard it correctly. Participants were instructed to repeat the letter string to themselves silently throughout the duration of the block and were asked to repeat the sequence aloud to the researcher when the block was completed. A new letter string was given on each block of trials which contained none of the previously used letters.

### 9.2.2 Results

#### *9.2.2.1 Accuracy of articulatory suppression task*

Responses were marked as correct if all letters were recalled in the correct order. Overall accuracy was high, 97.5%. On only two occasions was recall incorrect, with one participant reversing the order of the two adjacent letters, and another participant replacing two letters with letters that had not appeared in the initial sequence.

#### *9.2.2.2 Speeded Classification Task*

The data were the accuracy levels and the correct response times (RTs) for the brightness classifications. Out of 2736 trials, 2674 of these were correct responses. Any RTs below 150ms or above 2.5SD above the mean were excluded from analysis leaving 2602 observations (95.10%). Table 9.1 summarises the accuracy and RTs for *dark* and *bright* classifications made with the *heavy* and *light* objects.

Table 9.1

*Mean RTs (SEM in parentheses) and accuracy levels (%) according to the object used and categorical brightness*

| | Brightness | |
|---|---|---|
| Object Used | Dark | Bright |
| Heavy | **655(12)** | 710 (14) |
| | **98.2%** | 96.8% |
| Light | 675(11) | **687 (14)** |
| | 97.7% | **98.2%** |

Note: Bold typeface signifies congruent conditions

***Response speed.*** The analysis approach was the same general strategy as that used in previous classification tasks. In order to resolve the skew of the RT distribution, a transformation of 1/RT was performed. The product of this transformation is interpreted as *response speed* (responses/second).The main aim of the analysis was to determine if an interaction between *object* (*big/heavy* Vs *small/light*) and *categorical brightness (dark* Vs *bright*) significantly improves model fit. For ease of interpretation this was included as a main effect of *congruence (congruent trials* Vs *incongruent trials)* along with main effects of *categorical brightness* and *object*. The LRT demonstrates that a model including *congruence* as a main effect is preferred to a model in which it is not included ($\chi^2(1) = 19.03$, $p<.001$) with an AIC value of 3896.2 and 3879.2 for the null and comparison model respectively. The parameter estimates of this basic model suggest that incongruent trials were slower than congruent trials by 0.087 responses/second (SE=0.02), CI [0.047, 0.12].

The contributions of the following additional variables were assessed along with interactions between variables which may have theoretical implications: *trial*, *response hand* (*left* Vs *right*), *within-category brightness (*which is whether the brightness value is the *dark* or *bright* values sharing a categorical brightness level, see Table 5.1), and *brightness contrast* (*high* Vs *low).* A model including these additional explanatory variables was preferred to the basic model according to the LRT ($\chi^2(4) = 331.35$, $p<.001$) with an AIC value of 3555.8. An inspection of the CIs demonstrated that *within-category brightness* did not have an effect since the values straddle zero, CI [-0.064, 0.008]. LRT confirmed its removal did not change the explanatory power of the model ($\chi^2(1) = 2.33$, $p=.127$, AIC=3556.1) therefore the model which excluded *within-category brightness* was preferred. An interaction of *object* and *within-category brightness* also failed to improve model fit ($\chi^2(2) = 1.08$, $p=.30$, AIC=3556.7).

Parameter estimates indicate that the speed of responses increased with *trial*. Responses were also faster when the *contrast* was high i.e. when the brightness value of the circle was further from the mid-grey background (black and white as opposed to dark grey or light grey). An interaction between *object* and *contrast* failed to improve model fit ($\chi^2(1) = 1.07$, $p=0.30$, AIC=3557.1). It was demonstrated that *response hand* improved model fit as a main effect; responses were faster when the right hand was used. *Response hand* did not interact with *brightness contrast* ($\chi^2(1) =2.57$, $p=.11$, AIC=3555.6). However, the model was improved by an interaction between *response hand* and *object* ($\chi^2(1) = 4.54$, $p=.033$, AIC=3553.6), as well as *response hand* and *categorical brightness* ($\chi^2(1) = 7.04$, $p$=0.008, AIC=5551.1). The nature of the two interactions involving *response hand* were such that responses were faster when the left hand was used to classify stimuli as dark and the right hand to classify stimuli and bright compared to the other way around in a very symmetrical way. Additionally, responses were faster when the heavy object was held in the left hand and the light object was being used in the right hand compared to the opposite way. An interaction between *response hand* and *congruence* ($\chi^2(1) = 3.13$, $p$=.07, AIC=3555.0) approached significant but was not preferred to a model which included both interactions separately. A model including all three interactions is reported to be the final model.

The LRT demonstrates that a model including *congruence* is preferred to a model in which it is not included ($\chi^2(1) = 25.09$, $p$<.001) with an AIC value of 3571.7 and 3548.6 for the models excluding and including *congruence* respectively. The parameter estimates of this final model suggest that incongruent trials were slower than congruent trials by 0.093 responses/second (SE=0.02), CI [0.06, 0.13]. Table 9.1 summarises the parameter estimates of the final model and Figure 9.1 depicts the nature of the congruity effect between *categorical brightness* and *object*. Visual inspection of residual Q-Q plots did not reveal any obvious departures from normality. In addition, inspection of residual plots demonstrated that the assumptions of linearity and homoscedasticity were also upheld.

***Response accuracy.*** The overall response accuracy was 97.73%. The percentage correct responses for congruent and incongruent trials were 98.15% and 97.30 % respectively. The numbers of incorrect responses were too few to warrant the same form of analysis of error rates as was used on RTs. As an alternative, the number of correct responses for congruent and incongruent trials were analysed using a related-samples Wilcoxon Signed Ranks Test. This indicated that there was not a significant difference (z=-1.002, p=.316). This confirms that the difference in speed for congruent and incongruent trials did not result in a speed/accuracy trade off. Furthermore, it suggests that congruity did not influence response accuracy.

Table 9.2

*Summary table for the final model for Experiment 15*

| Fixed effects (criterion) | Estimated coefficient | SE | Wald confidence intervals 2.5% | Wald confidence intervals 97.5% | t-value |
|---|---|---|---|---|---|
| (Intercept) | 1.51 | 0.11 | 1.29 | 1.72 | 13.60 |
| Trial | 0.002 | 0.03 | 0.0016 | 0.0024 | -0.84 |
| Brightness Contrast (Low) | -0.29 | 0.18 | -0.33 | -0.26 | -15.83 |
| Response Hand (Right) | 0.50 | 0.16 | 0.19 | 0.81 | 3.18 |
| Object (Light) | -0.022 | 0.03 | -0.073 | 0.029 | -0.84 |
| Categorical Brightness (Dark) | 0.48 | 0.16 | 0.18 | 0.79 | 3.08 |
| Cat. Brightness (Dark): Resp. Hand (Right) | -0.91 | 0.31 | -1.52 | -0.30 | -2.92 |
| Object (Light): Resp. Hand (Right) | 0.079 | 0.04 | 0.006 | 0.15 | 2.13 |
| Congruence (Incongruent) | -0.093 | 0.02 | -0.13 | -0.056 | -5.02 |

| Random effects | Name | Std. Dev | Variance |
|---|---|---|---|
| Participant | (Intercept) | 0.11 | 0.34 |
| Residual | | 0.22 | 0.47 |

| AIC | BIC | Loglik | Deviance |
|---|---|---|---|

| | | | |
|---|---|---|---|
| 3548.6 | 3613.1 | -1763.3 | 3526.6 |

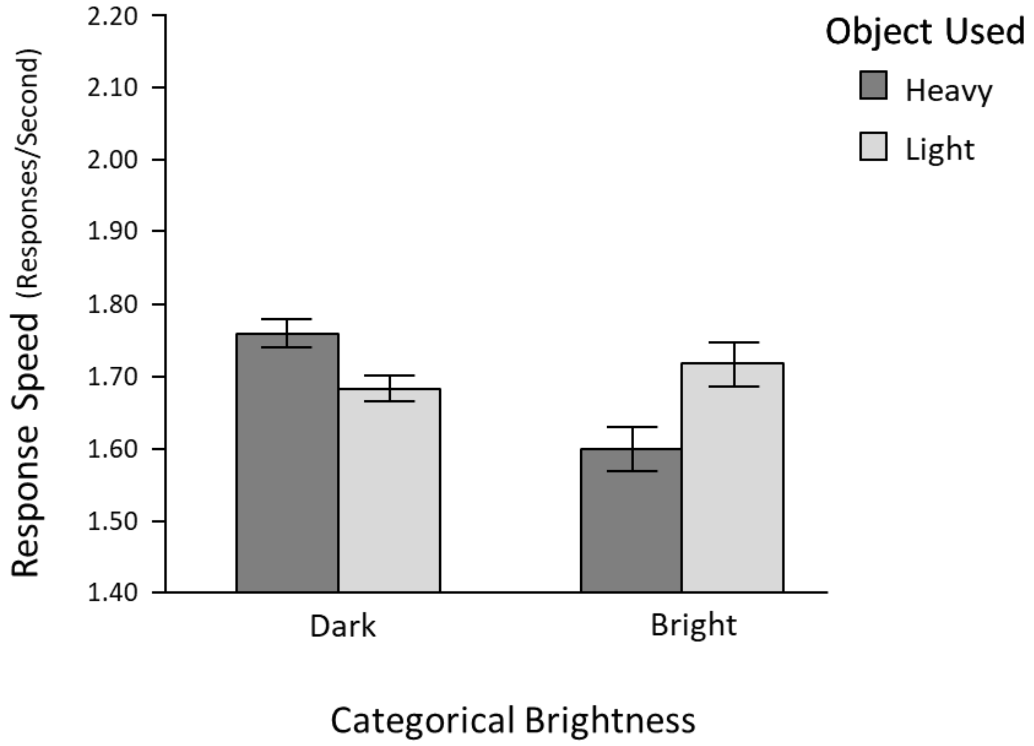



Figure 9.1. Mean response speed (responses/second) for the dark and bright responses with each object. Error bars refer to the standard error of the mean

### 9.2.3 Discussion

The findings replicate the congruity effect found between brightness and heaviness in Experiment 12, despite the addition of an articulatory suppression task. Participants were faster when the heavy object was used to classify stimuli as dark and the light object was used to classify stimuli as bright compared to the other way around. This suggests that the congruity effect found between brightness and heaviness in the previous chapters is unlikely to be based on online verbal labelling of the features and the common label ‘light’. This suggests that the correspondence is not actively mediated by articulatory processes.

The speed of response times are slower than those observed in Experiment 12, indicating the increased demands of this task from adding the verbal suppression task. The final model for the present experiment replicates the final model of Experiment 12 in most respects. In addition to the observed congruity effect between Categorical Brightness and Object, responses in both cases were faster with trial, which suggests a practice effect. Responses were faster when classifying the dark stimuli, and when there was a high contrast between the brightness of the circles and the background. One key difference between the present study and the findings of Experiment 12 is the interactions of response hand with both object used and categorical brightness. The role of handedness is quite difficult to interpret, especially since it has not been observed to influence response times in a consistent manner across the previous classification tasks. This is the first case that an interaction between brightness and response hand has improved model fit. The lack of a consistently demonstrated interaction between response hand and object may in part be due to differences in the size and weight of objects being used in different tasks. In Experiments 7 and 10 an interaction between object and response hand was found. This was suggested to be a result of some objects being easier to use in the right (usually dominant) hand.

#### *9.2.3.1 The role of a common verbal label in correspondences*

As was described in Section 9.1, there are two ways in which the presence of common verbal labels may influence the occurrences of cross-sensory correspondences. Although the present experiment suggests that the congruity effect is not verbally-mediated, it does not rule out the role that the common label between heaviness and brightness in English may have had in promoting or in strengthening the correspondence more generally. Further work is required to determine whether the correspondence between brightness and heaviness is influenced by the occurrence of a common label in English. This may involve exploring whether speakers of languages in which the dimensions do not share a common label still produce a congruity effect. Unfortunately the number of none native English speakers in the present work was not sufficient to analyse whether a difference was present in the current work, furthermore, all participants were acquainted with English as a second language if it was not their first.

# Chapter 10

## 10.1 Overview of the Thesis

In the theoretical framework described by P. Walker and colleagues (P. Walker, 2016; P. Walker and Walker, 2012), a core set of feature dimensions are proposed to be aligned such that cross-talk arises between relative extremes on each implicated feature dimension. Since this cross-talk is considered to arise at an amodal, conceptual level, the same pattern of associations should emerge irrespective of which feature dimensions is used to probe it. The present thesis aimed to investigate further the extent to which heaviness can be considered to enter into this network.

Two experimental paradigms were used to explore heaviness in cross-sensory correspondences. Firstly, a series of rating-scale tasks were used to determine if heaviness enters into systematic associations with pointiness, brightness, pitch and size; four feature dimensions which have often been demonstrated to enter into a range of cross-sensory correspondences. The key question was whether these systematic associations can be induced by variation in heaviness.

Next, a series of speeded-classification tasks were conducted to determine if the associations of heaviness with pitch and brightness could be observed in a task which does not rely on participants making explicit decisions about their alignment. The initial speeded classification tasks (Chapters, 5, 6, and 7) required participants to tap a touch sensitive surface with objects varying in size and/or weight in order to make a response. In the experiments reported in Chapters 8 and 9, the procedures were modified. Rather than using the objects to make a response, participants responded by pressing a switch attached to each object. This allowed the effect of using the objects actively versus passively to be explored.

Finally the potential role of verbal mediation in the brightness-heaviness correspondence was considered with the inclusion of a verbal-suppression task in Chapter 9.

Interactions between the size and weight of the objects were explored in both rating and speeded classification tasks. This allowed the separable (or combined) contributions of these dimensions to cross-sensory correspondences to be explored.

## 10.2 The Key Findings

In Chapter 2, it was observed that heavier objects (when hidden from view and lifted only by string) were rated as being bigger, darker and lower in pitch compared to less heavy objects. The extent to which heaviness induced associations with pointiness was unclear. In Experiment 1, where pointiness was represented with verbal labels, heavier objects were rated as being more rounded (vs. pointy) than lighter objects. However, this was not upheld in any reliable way with scales which represented pointiness non-verbally. The key finding was that the heaviness of lifted objects did induce the associations with size, brightness and pitch which would be predicted by the alignment of feature dimensions proposed by P. Walker et al. (2012).

Chapter 3 demonstrated that the correspondences heaviness induced with brightness and pitch could not be accounted for by the anticipated size of the objects. Heavier objects continued to be rated as darker and lower in pitch when size was made available visually; both when the size of the objects was held constant (Experiment 3) and when the size of the objects was varied in an opposing direction to heaviness (Experiment 4). Experiment 5 further explored how the size and weight of objects being lifted directly influenced brightness and pitch ratings and confirmed the same findings. The perceived heaviness of the objects was the primary contributor to brightness and pitch ratings. However, for pitch, unlike brightness, the sizes of the objects were also demonstrated to contribute to final rating score independently to the influence of heaviness.

The second half of the thesis aimed to determine if the same alignment of heaviness with brightness and pitch could give rise to a congruity effect in a speeded classification task. In Chapter 5 and 6 a heaviness-brightness congruity effect was observed when objects varying in weight were used to respond to circles varying in brightness. Responses were faster when a heavier object was being used to classify circles are dark and a lighter object was used to classify circles as bright compared to the other way. This brightness-heaviness congruity effect was found when size varied in correlation with heaviness (Experiment 6); when size was held constant (Experiment 7) and when size was varied in contrast to heaviness (Experiment 9). A congruity effect was not demonstrated when heaviness was held constant despite the objects varying in size (Experiment 8).

The heaviness-brightness congruity effect continued to be observed when the objects were held passively and responses were made with switches attached to differently weighted objects (Experiments 12 and 13). Finally, it was tested whether the mapping between brightness and heaviness could be accounted for by online verbal labelling and the shared label ‘light’ used to refer to both feature dimensions. The congruity effect continued to be observed despite the addition of an articulatory suppression task to the experimental procedure (Experiment 15).

The evidence for a pitch-heaviness congruity effect was much less clear. A heaviness-pitch congruity effect was found in Experiment 10, where differently weighted objects were used actively to tap touch sensors in order to respond to sounds varying in pitch. Responses were faster when the heavier object was used to classify sounds as low pitch and the light object was used to classify stimuli as high pitch compared to the opposite way. However, this was not replicated when the objects were held passively (Experiment 14).

## 10.3 Implications of the Present Work

### *10.3.1 Support for a framework of aligned feature dimensions*

For the most part, the findings of the present thesis are consistent with the proposed framework of aligned feature dimensions put forward by P. Walker and colleagues (P. Walker and Walker, 2012; P. Walker, 2016). Firstly, by confirming that the heaviness of lifted objects can induce the same alignment of feature dimensions that have previously been observed in the opposite direction (Alexander and Shansky, 1976; L. Walker, et al.2012), the findings are consistent with the argument that these correspondences are bi-directional in nature. This is particularly notable in the case of heaviness, due to the asymmetry with which it is experienced in relation to other feature dimensions in our every-day experience. Heaviness would very rarely act as a cue to other feature dimensions such as brightness and size. Because of this, one might have suspected that it would not induce the same correspondences that are observed in the opposite direction.

Furthermore, the observed mappings of heaviness with brightness, pitch, size and pointiness (in Experiment 1) are consistent with the proposed transitivity between these dimensions according to previous findings. For example, heavier objects were demonstrated to be aligned with darker and lower in pitch which is consistent with a pitch-brightness correspondence such that dark is aligned with low pitch and bright with high pitch. Both bi-directionality and transitivity are argued to indicate that these dimensions arise at a conceptual level. As a result these findings confirm heaviness as a dimension which enters into correspondences at this level.

### *10.3.2 Understanding Heaviness Perception*

Since the present work involved lifted objects, one consideration was whether another unitary dimension, a product of the combination of an objects size and weight (for example, density) may form the basis of the object-brightness/pitch mappings. The findings from the rating scale tasks (Chapter 4) and the speeded classification tasks (particularly those in Chapter 5 and 6) provide evidence that density does not account for the observed cross-modal mappings of heaviness with brightness and pitch. In Experiment 5, it was found that the density of the objects was not a preferred explanatory variable for the judgements of brightness and pitch, when compared to models including either the size and weight or the perceived heaviness of the objects. Furthermore, it was demonstrated that the perceived heaviness of the objects was better explained by size and weight as separable dimensions rather than by values of the object's density. The pattern of findings in Chapters 5 and 6 also cannot be easily explained by another unitary feature dimension. In Experiment 6, an object-brightness congruity effect was observed despite the objects being of equal density. In Experiment 8 no object-brightness congruity effect was observed. In this pair of objects, the difference in size would still have created a difference between the objects in terms of density, despite the weight of the objects being held constant and yet, no congruity effect was observed. These findings are consistent with weight illusion literature suggesting that the size-weight illusion cannot be completely explained in terms of lower level interactions between physical properties of the objects. Instead, the size-weight illusion is likely to have some basis in higher level cognitive mappings between the more general or conceptual representations of size and heaviness (Buckingham, 2014; Buckingham and Goodale, 2010).

The present work provides additional evidence that heaviness as induced by lifted objects does enter into mappings at this conceptual level. Firstly, as mentioned above, this is done by demonstrating that heaviness enters into cross-modal mappings with dimensions such as brightness in a way that is bi-directional and transitive in nature. Secondly, by showing that the same brightness-heaviness congruity effect was observed irrespective of the way the objects were held or used. This suggests that the congruity effect is not dependent on specific sensorimotor input, but rather relies on the conceptual representation of heaviness, irrespective of how this is produced. Furthermore, the findings from Experiment 5 demonstrated that judgments of brightness and pitch were best explained by the perceived heaviness of the objects as induced by the size-weight illusion, as opposed to the weight of the objects. This gives a further indication that weight illusions involve the same types of cross-sensory mappings which underpin correspondences. Therefore, by confirming that correspondences can be induced by the felt heaviness of lifted objects, the present work provides a bridge between these two areas of work which have previously been quite independent.

## 10.4 Limitations and recommendations for future work

### *10.4.1 Heaviness in correspondence with other feature dimensions*

The series of experiments primarily focussed on the correspondence between heaviness and brightness, and secondarily with heaviness and pitch. However, cross-sensory correspondences are argued to arise between any feature dimensions implicated in this network. Some of the findings in the present thesis do not provide conclusive evidence of the correspondences heaviness enters into with other feature dimensions, for example with pointiness (Chapter 2) and with pitch in a speeded classification paradigm (Chapter 8). More work is necessary to further explore potential explanations for these findings, to determine if they are the result of experimental limitations or reflects something about the nature of the underlying correspondence (or lack thereof).  Furthermore, the implication of heaviness in this network, predicts that heaviness will enter into correspondences with other feature dimensions and further more potentially predict a wide range of weight illusions.

### 10.4.2 Hierarchy of corresponding feature dimensions

By varying the size and weight of the objects used in this work independently, it was possible to explore the separable contributions of size, weight and heaviness in entering into cross-sensory correspondences. Through this exploration, it was revealed that the perceived heaviness of objects consistently formed the basis of associations with brightness and pitch in rating scale tasks, and with brightness in speeded-classification tasks. This was observed in some cases despite the size of the objects having the potential to induce opposing cross-modal mappings. This has a number of potential interpretations; firstly it may be the result of heaviness-based correspondences being more dominant in some way compared to size-based correspondences. Alternatively, it may tell us something about preferences for certain forms of sensory information. Finally, it may suggest that the size-brightness correspondence is mediated by heaviness, a position considered also by P. Walker and Walker (2012). This type of exploration goes beyond the initial framework of establishing mappings between pairs of feature dimensions by providing more detail about the nature of the underlying associations within more complex, multi-sensory contexts.

## 10.5 Concluding remarks

The present work confirms that cross-sensory correspondences between feature dimensions can be accessed through the manipulation of felt objects varying in heaviness. Heaviness was demonstrated to map consistently with brightness, and to some degree with pitch, in a way that was predicted by a network of feature dimensions aligned at the level of connotative meaning.

**References**

Alexander, K. R. & Shansky, M. S. (1976). Influence of hue, value, and chroma on the perceived heaviness of colors. *Perception and Psychophysics, 19(1), 72-74*

Amazeen, E.L., Turvey, M. (1996). Weight perception and the haptic size–weight illusion are functions of the inertia tensor. *Journal Experimental Psychology: Human Perception and Performance, 22, pp.213–232.* doi:10.1037/0096-1523.22.1.213

Bates D., Maechler, M., Bolker, B. & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, *<URL: http://CRAN.R-project.org/package=lme4>*.

Bee, M. A., Perrill, S. A., & Owen, P. C. (2000). Male green frogs lower the pitch of acoustic signals in defense of territories: A possible dishonest signal of size? *Behavioral Ecology, 11, 169-177*

Ben-Artzi, E. & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception and Psychophysics, 57,* 1151-1162.

Berstein, I. H., Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology,* 87, 2, 241-247

Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size: Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG and psychophysics study. *NeuroImage, 59, 663–67*2. http://dx.doi .org/10.1016/j.neuroimage.2011.06.09

Bond, B., & Stevens, S. S. (1969). Cross-modality matching of brightness to loudness by 5-year-olds. *Perception & Psychophysics, 6, 337*

Brayanov, J. B. & Smith, M. A. (2010). Bayesian and "Anti-Bayesian" biases in sensory integration for action and perception in the size-weight illusion. *Journal of Neurophysiology, 103,* pp. 1518-1531

Buckingham, G. (2014). Getting a grip on heaviness perception: a review of weight illusions and their probable causes. *Experimental Brain Research, 232,* 1623-1629.

Buckingham, G. & Goodale, M. A. (2010) Lifting without Seeing: The Role of Vision in Perceiving and Acting upon the Size Weight Illusion. *PLoS ONE 5(3): e9709. doi:10.1371/journal.pone.0009709*

Carello, C., Anderson, K. L. & Kunkler-Peck, A. J. (1998). Perception of object length by sound. *Psychological Science,* 9,3,211-214

Chen, Y. C., Huang, P. C, Woods, A., Spence, C. (2016). When “Bouba” equals “kiki”: Cultural commonalities and cultural differences in sound-shape correspondences. *Scientific Reports,* 6, 26681

Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception, 41,* 339-353.

Collier, W. G. & Hubbard, T. L. (2001). Judgements of happiness, brightness, speed and tempo change of auditory stimuli varying in pitch and tempo. *Psychomusicology, 17,* 36-55.

Connell, L., Cai, Z., G., & Holler, J. (2013). Do you see what I’m singing? Visuospatial movement biases pitch perception. *Brain and Cognition, 81,1,124-130*

Cytowic, R. E. (1993). The man who tasted shapes. New York: G. P. Putnam’s Sons.

Day, S. (2005). Some demographic and socio-cultural aspects of synesthesia. In L. C. Robertson & N. Savig (Eds.)*, Synesthesia: Perspectives from cognitive neuroscience (pp. 11–33). New York: Oxford University Press*.

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science*, 25(6), 1256-1261.

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science, 24,* 613-621.

Eitan, Z. & Timmers, R. (2010). Beethoven’s last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition, 114,* 405-422.

Ellis, R.R., Lederman, S.J. (1993) The role of haptic versus visual volume cues in the size–weight illusion*. Perception and Psychophysics, 53, pp. 315–324*

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. In G. Knoblich, I. M. Thornton, M. Grosjean, & M. Shiffrar (Eds.), Human body perception from the inside out (pp. 105–131). Oxford: Oxford University Press.

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch, J. Vis. 7, 7. DOI:10.1167/7.5.7

Ernst, M., O. (2009). Perceptual learning: inverting the size-weight illusion. *Current Biology, 19, 1, R24*

Evans, K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10,* 1–12.

Flanagan, J. R., Bittner, J. P. and Johansson, R. S. (2008). Experience can change distinct size–weight priors engaged in lifting objects and judging their weights, Curr. Biol. 18, 1742– 1747

Gallace, A. & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception and Psychophysics, 68,* 1191-1203.

Grassi, M. (2005). Do we hear size or sound: Balls dropped on plates. Perception & Psychophysics, 67, 274–284.

Guzman-Martinez, E., Ortega, L., Grabowecky, M., Mossbridge, J., & Suzuki, S. (2012). Interactive coding of visual spatial frequency and auditory amplitude-modulation rate. *Current Biology, 22, 383-388*. doi: 10.1016/j.cub.2012.01.004

Harrington, F. H. (1987). Aggressive howling in wolves. *Animal Behaviour, 35,1, pp. 7-12*

Helmholtz, H. V. (1909). Handbuch der Physiologischen Optik (Handbook of Physiological Optics). Voss, Hamburg, Germany.

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *The American Journal of Psychology, 109, 219–238.*

Karwoski, T. F., Odbert, H. S., & Osgood, C. E. (1942). Studies in synesthetic thinking: II. The role of form in visual responses to music. *The Journal of General Psychology, 26,* 199-222.

Klapetek, A., Ngo, M. K., & Spence, C. (2012). Does crossmodal correspondence modulate the facilitatory effect of auditory cues on visual search? *Attention, Perception, and Psychophysics, 74:1154–1167*

Köhler, W. (1929). Gestalt psychology. New York: Liveright.

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory–visual intensity matching. *Developmental Psychology, 16, 597–607.*

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial Associations for Musical Stimuli: A Piano in the Head? Journal of Experimental Psychology: Human *Perception and Performance, 33, 5, pp. 1189–1207*

Lunghi, C. & Alais, D. (2013). Touch interacts with vision during binocular rivalry with a tight orientation tuning. PLoS ONE, 8(3), e58754. doi: 10.1371/journal.pone.0058754

Lupo, J., Barnett-Cowan, M. (2015). Perceived object stability depends on shape and material properties. *Vision Research, 109, 158-165*

Maeda, F., Kanai, R., & Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Current Biology, 14, R990–R991.*

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. The American Journal of Psychology, 87, 173–188.

Marks, L. (1978). The unity of the senses: Interrelations among the modalities. New York: Academic Press

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 384–394.

Marks, L. E. (2004). Cross-modal interactions in speeded classification. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), Handbook of multisensory processes (pp. 85–105). Cambridge: MIT Press.

Marks, L. E., Hammeal, R. J., & Bornstein, M. H. (1987). Perceiving similarity and comprehending metaphor. *Monographs of the Society for Research in Child Development, 52(1, Whole No. 215), 1–102.*

Marks, L. E., Szczesiul, R., & Ohlott, P. (1986). On the cross-modal perception of intensity. Journal of Experimental Psychology: *Human Perception and Performance, 12, 517-534.*

Martino, G. & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception, 28,* 903-923.

Maurer, D., & Maurer, C. (1988). The world of the newborn. New York: Basic Books.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264, 746-748*

Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 69-79.

Melara, R. D. & Marks, L. E. (1990) Processes underlying dimensional interactions: Correspondences between linguistic and non-linguistic dimensions. *Memory and Cognition, 18,* 477-495.

Miller, A., Werner, H., & Wapner, S. (1958). Studies in physiognomic perception: V. Effect of ascending and descending gliding tones on autokinetic motion. *Journal of Psychology, 46, 101–105*

Mondloch, C. J. & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience, 4,* 133-136.

*Munsell Book of Color* (1976). Baltimore, Maryland: Macbeth.

Murray, D. J., Ellis, R. R., Bandomir, C. A, Ross, H. E. (1999). Charpentier (1891) on the size–weight illusion. *Perception and Psychophysics 61, pp. 1681-1685.*

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Science Eduation, 15, pp.* 625-632.

Osgood, C., E. (1952). The nature and measurement of meaning. Psychologcial Bulletin, 49,3, 197-237

Osgood, C., E. (1960). The cross-cultural generality of visual–verbal synesthetic tendencies. Behavioral Science, 5, 146–169.

Parise, C. (2016). Crossmodal Correspondences: Standing Issues and Experimental Guidelines. Multisensory Research, 29 (1-3), 7-28

Parise, C., & Spence, C. (2009). “When birds of a feather flock together”: Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS ONE, 4, e5664.*

Parise, C. V., Knorre, K. & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *PNAS, 111,* 6104-6108.

Parise, C., V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. *Experimental Brain Research, 220, pp. 319–333*

Patching, G. R. & Quinlan, P. T. (2002). Garner and congruence effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 755-775.

R Core Team (2012). R: A language and environment for statistical computing. *R Foundation for Statistical Computing.* Vienna: Austria.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—A window into perception, thought and language. *Journal of Consciousness Studies, 8,3 –34.*

Ramachandran, V. S., & Hubbard, E. M. (2003, May). Hearing colors, tasting shapes. *Scientific American, 288, 43*

Root, R. T., & Ross, S. (1965). Further validation of subjective scales for loudness and brightness by means of cross-modality matching. *The American Journal of Psychology, 78, 285–289*.

Ross, J., Di Lollo, V. (1970). Differences in heaviness in relation to density and weight. *Perception and Psychophysics,7, pp. 161–162*. doi:10.3758 /BF03208648

Rouw, R., & Scholte, H. S. (2007). Increased structural connectivity in grapheme–color synesthesia. *Nature Neuroscience, 10, 792–797.*

Rusconi, E., Kwan, B., Giordano, B. L., Umilta `, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition, 99, 113–129*.

Shockley, K., Carello, C., & Turvey, M. T. (2004). Metamers in the haptic perception of heaviness and moveableness. *Perception & Psychophysics, 66, 731–742.*

Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology, 24, 99 –142.*

Spector, F., & Maurer, D. (2009). Synesthesia: A New Approach to Understanding the Development of Perception. *Developmental Psychology, 45(1), 175–189*

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73, 971–995. http://dx.doi.org/10 .3758/s13414-010-0073-7

Spence, C., Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition, 22, 245-260*

Stevens, J. C., & Marks, L. E. (1965). Cross-modality matching of brightness and loudness. *Proceedings of the National Academy of Sciences, 54, 407–411*.

Stevens, J.C., Rubin, L. L. (1970). Psychophysical scales of apparent heaviness and the size–weight illusion. *Perception and Psychophysics, 8, pp. 225– 230*. doi:10.3758/BF03210210

Sweeny, T. D., Guzman-Martinez, E., Ortega, L., Grabowecky, M., Suzuki, S. (2012). Sounds exaggerate visual shape. *Cognition, 124, 194-200*

Wagman. J. B. (2015). As easy to move as a feather: perception of lightness as ease to move. *Journal of Motor Behaviour, 47(4), pp. 340-342).*

Wagner, S., Winner, E., Cicchetti, D., & Gardner, H. (1981). “Metaphorical” mapping in human infants. Child Development, 52, 728–731.

Walker, L. & Walker, P. (2016). Cross-sensory mapping of feature values in the size-brightness correspondence can be more relative than absolute. *Journal of Experimental Psychology: Human Perception and Performance,42,* 138-150.

Walker, L., Walker, P., & Francis, B. J. (2012). A common scheme for cross-sensory correspondences across stimulus domains. *Perception, 41(10),* 1186-1192.

Walker, P. (2012). Cross-sensory correspondences and crosstalk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright. *Attention, Perception, & Psychophysics, 74,* 1792-1809.

Walker, P. (2016). Cross-sensory correspondences: A theoretical framework and their relevance to music. *Psychomusicology: Music, Mind and Brain, 26,* 103-116.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants’ sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, *21*, 21-25.

Walker, P., Francis, B. J., & Walker, L. (2010). The brightness-weight illusion: Darker objects look heavier but feel lighter. *Experimental Psychology, 57,* 462-469.

Walker, P. & Smith, S. (1984). Stroop interference based on the synaesthetic qualities of auditory pitch. *Perception, 13,* 75–81.

Walker, P. & Smith, S. (1985). Stroop interference based on the multimodal correlates of haptic size and auditory pitch. *Perception, 14,* 729-736.

Walker, P. & Walker, L. (2012). Size-brightness correspondence: Crosstalk and congruity among dimensions of connotative meaning. *Attention, Perception, & Psychophysics, 74(6),* 1226-1240.

Walker, P. Walker, L., & Francis, B. J. (2015). The size-brightness correspondence: Evidence for crosstalk among aligned conceptual feature dimensions. *Attention, Perception & Psychophysics, 77,* 2694-2710.

Walsh, V. (2003). A theory of magnitude: Common cortical metrices of time, space and quality. Trends in Cognitive Sciences, 7, 483 –488.

Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound–colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex, 42, 264–280*

Ward, J., Simner, J., (2003). Lexical-gustatory synaesthesia: Linguistic and conceptual factors. *Cognition 89, 237-261*

Weber, E. H. (1978). The sense of touch (H. E. Ross, Ed. & Trans.). London: Academic Press. (Original work published 1834)