SOUND SYMBOLISM AND THE BOUBA-KIKI EFFECT:
UNITING FUNCTION AND MECHANISM IN THE SEARCH FOR LANGUAGE UNIVERSALS

Alan Kirkland Staun Nielsen
BSc. University of Lethbridge, 2009

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirements for the Degree

MASTER OF SCIENCE

Department of Psychology

University of Lethbridge

LETHBRIDGE, ALBERTA, CANADA

ABSTRACT

In contemporary linguistics, the relationship between word form and meaning is assumed to be arbitrary: words are mutually agreed upon symbolic tokens whose form is entirely unrelated to the objects or events that they denote. Despite the dictum of arbitrariness, examples of correspondences between word form and meaning have been studied periodically over the last century, labeled collectively under the banner of *sound symbolism.* To investigate the phenomenon of sound symbolism, a series of experiments was conducted on the classic Bouba-Kiki phenomenon. These experiments not only addressed previous methodological and interpretive issues of earlier research, but also entailed a more fine grained approach that allowed for a clearer delineation of the word characteristics responsible for sound symbolic biases. An interpretation of the findings of these experiments is presented that is both in line with potential functional accounts of sound symbolism and also grounded in probable mechanistic instantiations.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Then, Hermogenes, I should say that this giving of names can be no such light matter as you fancy, or the work of light or chance persons; and Cratylus is right in saying that things have names by nature, and that not every man is an artificer of names, but he only who looks to the name which each thing by nature has, and is able to express the true forms of things in letters and syllables.

The Cratylus- Plato

CHAPTER ONE

INTRODUCTION

1- Overview

"The symbol is arbitrary."

The preceding statement has been the central dogma of linguistics since Ferdinand de Saussure (1983) codified it in the early 20th century. Although de Saussure personified the conventionalist position that the relationship between words and the objects or events that they describe is arbitrary and established by mutually agreed upon usage, thinking about the relationship between objects and designata goes at least as far back as Plato. In *The Cratylus* Plato presents Socrates as taking the naturalist position, according to which words have a natural correspondence to the objects they denote, against Hermogenes, his interlocutor. Through a lengthy discussion Socrates, acting as Plato's mouthpiece, traces the origins of dozens of words to connect them to the objects which they denote. Ultimately, Plato is able to convince Hermogenes that words and objects are related in systematic ways, in many cases through their roots and origins. In some cases, this involves tracing words through many levels of interpretation down to their earliest root, but at that root connections between objects and designata are unmistakable.

Despite Plato's rhetoric, linguistics has operated under the assumption that all language is arbitrary and effectively involves the manipulation of abstract symbols, any of which could stand in for any other given an established history of use in that fashion. Despite this assumption, experimenters in the past century have consistently presented

evidence that demonstrates systematic variation in the acoustical characteristics of words for certain classes of objects or events. Variously, researchers have reported correspondences between pitch and size (Mondloch & Maurer, 2004; Sapir, 1929), pitch and brightness (Marks, 1996), timbre and color (Ward, Huckstep, & Tsakanikos, 2006), timbre and taste (Gallace, Boschin, & Spence, 2011), amplitude and size (Lindauer, 1990) and many other correspondences (see Figure 1.1, which shows the large number of crossmodal connections that have been demonstrated). Broadly, these findings have been labeled under the banner of *sound symbolism*, although linkages between what seem as if they should be completely separate sensory qualities are not limited to the acoustical domain.

One example of sound symbolism is the Bouba-Kiki effect, which describes an experimentally pervasive association between the shape of objects and the acoustic characteristics of the words used to label them. First described by Wolfgang Köhler (1947), the Bouba-Kiki effect involves reliable pairing of unfamiliar objects based on their structure (i.e. curvature of lines, number and acuteness of their angles etc.) with nonsense words based on their acoustic structure (i.e. spectral harshness or plosivity). In Köhler's work, participants consistently matched the nonsense word *'maleme'* to objects with a curved visual form and the nonsense word *'takete'* to objects with a spiky and jagged visual form (see Figure 1.2). This result has been replicated both in other languages unrelated to English (e.g. Davis, 1961), and also with toddlers who are relatively naïve regarding conventional regularities in language (e.g. Maurer, Pathman, & Mondloch, 2006).

Despite consistent demonstrations of the Bouba-Kiki effect and other sound-symbolic associations, the central tenet in linguistics of language arbitrariness has held. In this thesis I provide evidence that the Bouba-Kiki effect, far from being an unimportant linguistic marginalia, is a *bona fide* example of universal sound symbolism that is rooted not only in common neural architecture that is shared between all humans, but that is likely also conserved from other animals who use vocal communication. In fact, the Bouba-Kiki effect and other cross-modal biases may be a natural outcome of the reliable pairing and temporal co-occurrence of otherwise unrelated events, with differential hedonic, affective, and attentional effects based on neural and perceptual configurations modulating which connections are strengthened with experience and which are not. If this argument seems shocking, it is important to consider one connection which we all take for granted- that between smell and taste, which primarily use far-removed portions of the brain but are inextricably linked to one another perceptually (Stevenson & Oaten, 2010). Given the ability of temporal co-occurrence and simple learning rules to create linkages so strong that we often experience powerful taste sensations when presented with a smell (such as describing smells as sweet; Dravnieks, 1985), the idea that other sensory connections might be established in otherwise unconnected modalities is perhaps less surprising. Such an account heavily favors learning as an explanation for sound symbolism, which seems entirely in line with de Saussure and Hermogenes' conventionalist position, but evidence from neuroscience and animal communication suggests that some connections are easier to make than others, providing naïve expectations that can be co-opted in the learning and establishment of language.

Previous work on the Bouba-Kiki effect has been problematic, but I have attempted to improve on previous efforts by couching my interpretation in mechanistically and functionally plausible explanations, and in taking great care in the design of experimental stimuli and the research protocol in which they are embedded.. The dual influences of cognitive neuroscience and animal communication have not only shaped our mechanistic understanding of sound symbolism, but have also provided input on which sorts of sound-symbolic associations are likely.

Far from being a linguistic curiosity, then, the implications of the Bouba-Kiki effect are far reaching and even small sound symbolic biases might have significant influences on processes like language learning and lexicon crystallization. By marrying research into the Bouba-Kiki effect to functional accounts of non-arbitrariness researchers may eventually be able to elucidate what processes are necessary to get human language up and running. The Bouba-Kiki effect may seem to be a very small part of that process, but without carefully exploring low level regularities we cannot know the boundary conditions required for the emergence of language. It may be the case that language, the hallmark of human behaviour, emerges not from complex cognitive modules, but from statistical regularities, learning, and the interaction of small biases with processes of cultural transmission (Saffran, 2003).

1.1-   Language

Maynard-Smith and Szathmary (1997) have described language as one of the major transitions in evolution; a new way of transmitting information that essentially changed the rules of evolution. In this regard, language stands on a pedestal next to evolutionary

transitions like eusociality, sexual reproduction, and even the emergence of DNA as the main carrier of genetic information. To be described in similar terms as the emergence of multicellular organisms and the formation of organelles is significant, but language is perhaps the single hallmark of human behaviour that we share with absolutely nothing else (at least in its most fully robust form), and additionally makes possible almost all other things that are potentially unique to humans (whether they are unique in kind, or merely in degree). Storytelling, economics, education, and the rule of law are but a few examples of language-mediated common goods (Smith, 2010) that are unique to humans.

Hockett (1960) codified the important and unique characteristics of language early on, pointing to thirteen universals that are necessary characteristics of a language system as robust as that possessed by humans. Rather than assuming that each and every portion of language was unique to humans, Hockett attempted to explain which language universals were likely shared with other animals, relying instead on the aggregation of all universals into a single system for producing our unique and flexible communication system. Two of Hockett's universals stand out immediately as most important and perhaps the unique features of human language. Productivity describes the ability of language to be combined and modified in ways that novel sentences can always be produced, effectively creating a system with an infinite number of messages based on only a finite number of communicative tokens (words and letters). Duality of patterning works similarly, but on a smaller scale, with a relatively small number of phonemic units producing a combinatorial explosion in the number of total linguistic tokens (words) available.

Each of the above explorations of human language attempts in one way or another to explain its uniqueness, but leads ultimately to questions of how humans acquired language in the first place. Generally, scientists and philosophers have viewed language as a complex phenomenon, and thus invoked complex answers as to how it arose, using the uniqueness of fully robust language to argue for unique modules and abilities in humans to account for our unique proficiency.

Chomsky (1965) was among the first to seriously address the uniqueness of human language by positing that there must be a distinct and unique explanation that sets language apart from other communication systems and thus explains human language proficiency. Before Chomsky, radical behaviourism held sway as the scientific method for explaining human behaviour- cycles of reinforcement and simple learning rules were assumed to account for even complex behaviours in ways that removed subjectivity and human experience from the picture (Skinner, 1957). According to Chomsky, human language was too complex to be amenable to explanation by the behaviourists; the stimulus from which children would be forced to learn language was too impoverished. Specifically, Chomsky suggested that the statistical learning required by behaviourism could not explain the rapid rate with which language was acquired given the amount of language exposure that children had during learning. To account for this inconsistency, Chomsky posited that language must require some complex cognitive module that was unique to humans and that provided rules for language acquisition- a *universal grammar* that, based on experience with different languages, could change slightly over time, but provided a mostly invariant structure upon which all language could be built.

Pinker and Bloom (1990) followed Chomsky in claiming not only that a unique cognitive module is required for language but also that such a cognitive module could arise only via the process of natural selection; creating unique brain areas in humans which were not shared with other animals but which were directly responsible for language. Pinker and Bloom (1990) make this claim, like Chomsky, based on their perception of the complexity of language, claiming that complex systems can only emerge naturalistically through the action of natural selection and refinement over multiple generations. Because natural selection is a unique process, Pinker and Bloom further argued that language acquisition in children should systematically differ from the course of language evolution in the species- any attempts to analogize the two are inherently misleading.

Thus, two major and very influential thinkers in psycholinguistics and evolutionary psychology have attempted to address the question of the uniqueness of language by positing unique and specialized adaptations, in some cases discounting entirely the influence of learning as the foundation of language. Counter to Pinker and Chomsky, others have tried to explain human language without invoking unique cognitive modules, instead relying on relatively simple, common and widely distributed processes to get language off the ground.

In his theorizing about language evolution, Pagel (2009), much like Smith and Szathmary (1990), has focused on the role of language as a transmitter of information. Specifically, he provided evidence from cross-cultural analysis of certain lexical items to make the case that language evolves laterally, within the course of single generations. In

multiple iterations of communication, Pagel proposes that words develop similarly to Lamarckian characteristics by becoming increasingly faithful transmitters of information over time through change effected by individual agents using the word itself. Word form polymorphisms (e.g. couch vs. chesterfield vs. sofa) arise largely by arbitrary processes to label objects and events, but these polymorphisms are selected for or against based on multiple iterations of information transmission, with those words that for one reason or another denote an object more effectively winning out over other polymorphisms in a process analogous to natural selection. Although Pagel does not provide a full account of what characteristics of words make them more effective for transmission of certain classes of information, he provides compelling evidence that some important, commonly used words do not change as rapidly as words that aren't used as often, suggesting that maximizing faithful information transmission has varying importance depending on the importance of the specific object or event being denoted.

Although Pagel's (2009) account provides a plausible explanation of lateral forms of language transmission, it doesn't address how lexicons emerge in the first place to be subject to modification by cultural transmission. In telling the story of an experiment undertaken by the Egyptian pharaoh Psamethiccus I, the greek historian Herodotus described the linguistic isolation of two boys who were never spoken to but who eventually produced the word *bekos*, Phrygian for bread, which led Psamethiccus to conclude that Phrygian must be the original language of man. Despite this purported finding, conducting such an experiment in a modern setting would never be possible, and from limited cases in which this has occurred due to abuse or neglect (Curtiss, 1977; Harlan, 1976) we have reasons to believe that language does not emerge *de novo* in

isolated individuals. Rather than attempting to experiment on children, Landauer & Dumais (1997: Landauer, Foltz, & Laham, 1998) created a computer network that was designed to learn language through exposure to levels similar to those experienced by children embedded in a linguistic environment. Although early iterations of their model fell short, subsequent versions of their model were able to learn language nearly as proficiently as human children simply by plotting words in semantic space depending on their rates of co-occurence, suggesting that Chomsky's dismantling of behaviourism based on the poverty of the stimulus may have been premature.

Whether one posits the presence of a specific cognitive module or not, each of the explanations of the emergence of language falls short in some way, suggesting that researchers may have been asking the wrong question. Perhaps the more important question is what the boundary conditions really are that are sufficient for the emergence of language. Modularity is certainly one possibility, but it seems to entirely ignore commonalities between humans and other animals that even Hockett (1960) recognized.

1.2- Language vs. Communication

Language writ large is generally seen as a system for the manipulation of arbitrary, conventionally established symbol strings. Further, the belief that language is entirely about the transmission of information between conspecifics makes the belief that it should be instantiated in entirely arbitrary conventionalized symbols more tenable. After all, computers and symbol manipulation systems generally do not require differential salience; any symbol can stand for any object or event so long as the reference is mapped out. The conventional view of language as a system of arbitrary symbol manipulation

10

however completely ignores the potential importance of communication channels, which we do not share with computers or other ideal systems, but which we do share with other animals. Most human language, at least historically, has been transmitted through the vocal apparatus and received via the auditory system, creating at least two possible loci at which differential salience could be established. Despite differences in vocal tract anatomy between humans and other animals which allow greater flexibility of production (Lieberman, 1968), our own vocal and auditory apparatuses have a long evolutionary history, which we share with (at least) most other extant mammals (Fitch 2000; Owren, Amoss, & Rendall, 2011) and that yield many basic commonalities in the form of common vocal sounds (reviewed in Ghazanfar & Rendall, 2008). Furthermore, there is also some obvious continuity in basic mechanisms of sound perception (Ghazanfar & Hauser, 1999).

1.2.1- Animal Communication

Given the clear continuity in some aspects of vocal production and perception between humans and other animals, one would expect that experimenters would consistently search for vestiges of early communication systems in human language. But the opposite is in fact true; researchers have focused largely on looking for evidence of complex, intentional symbolic communication in animals. For example, Seyfarth, Cheney, and Marler (1980) have claimed that the alarm calls of vervet monkeys are truly referential signals for their three main types of predators, and stand as direct analogues to conventionally, agreed upon, human words for snakes, leopards, and aerial predators. Similarly, other primate species have been taught language with some modest success (Patterson & Matevia, 2001; Savage-Rumbaugh & Lewin, 1994; Terrace, Petitto,

Sanders, & Bever, 1979), but have never achieved the language proficiency of even young children, and have required a great deal more direct teaching. Given the ease with which human children pick up language with far less direct coaching but equal linguistic immersion, it seems that for non-human primates Chomsky's poverty of the stimulus argument might hold sway, but for humans the stimulus does not seem to be nearly so impoverished or insufficient to account for learning.

But if not transmitting intentional symbolic information in a way analogous to human communication, what might animals be doing the majority of the time they are producing vocal signals? One possibility is that rather than transmitting information, animals are modulating the behaviour of conspecifics through manipulation of their emotional states. Morton (1977) was the first to clearly delineate what he termed motivation-structural rules, pointing out that harsh sounds are often used in agonistic contexts while mellifluous sounds are used in more prosocial affiliative contexts (Morton, 1994; Rendall, 2003). In many cases, manipulative strategies are selected against in a population (Gneezy, 2005), but research with animals has demonstrated that some sounds produce aversive reactions that are unavoidable. For example, Herzog and Hopf (1983) found that squirrel monkeys were unable to habituate to aversive sounds and would always elicit a startle reaction, while Johnson, Cole, and Stroup (1985) found that starlings were startled both by broadband hisses and their own species typical distress call, but were unable to habituate to the sound of the distress call and continued to be startled when it was presented, even after many iterations. In these species, it is likely that the costs of ignoring potentially important calls are simply too great, so that no amount of false signaling can induce costs that are competitive. The boy who cried wolf

might be annoying, but never so annoying that ignoring his calls would be beneficial to the listener given the possibility that any given signal might be an honest one. Undoubtedly, the inextinguishability of certain responses is instantiated in some way in the perceptual circuitry of various animals, and thus it is no surprise that such manipulation appears to not only be used to modulate the behaviour of conspecifics, but also of predators or prey species. Fenton and Licht (1990) found, for example, that rattlesnakes produce a rattle which is unlikely to be useful for communication between themselves and other rattlers (given that its pitch is outside of the optimal range for their hearing) and interpreted this finding as suggesting that they might rattle to induce a startle response in predators which is sufficient to allow the snakes time to escape or defend themselves when attacked.

One possible interpretation of regularities in communication based on states of arousal is that animals produce calls which are reflective of their own arousal state. But as Dawkins and Krebs (1978) point out, signaling one's own motivational and emotional states is a potentially costly behaviour, which would only rarely allow an animal to accrue any benefit. Insofar as animals might signal their own motivational state, it should be voluntary, and thus should be seen not as an honest signal of their emotional state, but as a signal of intent designed to modulate the behaviour of another animal. Dunbar (2003) takes a similar tack to Dawkins and Krebs, positing that language has its roots in social grooming. Specifically, as group size increased past a number of conspecifics where direct and intimate relationships were possible with all members, modulation of behaviour to establish relationships and alliances without direct contact

13

was necessary, providing a selective impetus for the emergence of language built upon motivation-structural rules.

1.2.2- Affective Semantics

The theory of affective semantics suggests that animal communication systems are rarely if ever about the transfer of information between conspecifics, but rather are more generally about influencing others, often through relatively low-level perceptual, motivational and emotional systems that guide their behaviour (Bachorowski & Owren, 2008; Owren & Rendall, 1997; Owren, Rendall, & Ryan, 2010; Rendall, Owren, & Ryan, 2009). Even in the most language-like examples of animal communication, such as the alarm calls of vervet monkeys noted above, communication lacks the fundamental intentionality of human language (Cheney & Seyfarth, 1996, 1998), and it is therefore unlikely that signal structures arises by social convention and is entirely arbitrary with respect to the functions they serve. In fact, many common vocalizations show very clear structure-function connections (Owren & Rendall, 2001).

For example, structural similarities exist across a wide range of primate species in the alarm vocalizations produced during encounters with predators. Across species, alarm calls tend to be short, abrupt-onset, broad band calls, preserving acoustic characteristics that function to manipulate the arousal of listeners (Owren & Rendall, 2001), additionally making the calls easy to localize and contrasting to background noise. Accordingly, listeners orient immediately in the direction of calls, which additionally function to prepare the listener to flee, a highly functional initial response in the context of predator encounters. The same basic alarm call structure is evident not only in other mammals

(Owings & Morton, 1998), but also some birds (Marler, 1955), suggesting that the effect of certain acoustic structures is highly conserved. Developmental studies have shown that startle responses to species typical alarm calls are induced even in young without significant experience with either calls or predators (e.g., Seyfarth & Cheney, 1986), suggesting that sensitivity to punctuate sounds reflects widely conserved, low-level brainstem and subcortical processes associated with sound localization, orienting, and autonomic responding.

Evolved sensitivity to certain types of sounds creates opportunity for signalers to influence the attention, arousal, and concomitant behaviour of others through their vocalizations, sometimes in ways that are impossible for perceivers to ignore. Among non-human primates, 'squeaks, shrieks, and screams' capitalize on this potential for manipulation by using sharp signal onsets, dramatic frequency and amplitude fluctuations, and chaotic spectral structures to directly impact the nervous systems of conspecifics (Gil-da-Costa et al, 2004; Rendall et al., 2009). These sounds are numerous and ubiquitous in every primate species studied, as well as many other mammals, birds, and amphibians. Most frequently, vocalizations of this type are produced by infants and juveniles (Owren, Dieter, Seyfarth, & Cheney, 1993) and are not simply a function of immature vocal production systems. Instead, these calls are likely to be especially functional in youngsters who otherwise would be unable to influence the behaviour of older and larger individuals in their groups.

For example, a weanling whose feeding attempts have been repeatedly rejected by its mother cannot physically force its mother to allow nursing, but can produce

protracted bouts of aversive vocalizations that can induce her to allow feeding. Vocal protests in the context of weaning are ubiquitous, including in humans, and they share acoustic characteristics including dramatic amplitude and frequency modulations, high frequency tonal cries and broadband screams. (Owings & Zeifman, 2004; Owren & Rendall, 2001). Perceptual studies on humans confirm the aversive nature of these signals and that they are effective in inducing responses from listeners (Lester & Boukydis, 1985; Bradley & Lang, 2000).

In species with a developed social dominance hierarchy, daily activities involve regular interaction where dominant individuals often antagonize and attack subordinates. In these cases, low-ranking victims of aggression are like juveniles in that they do not pose legitimate physical threats to their dominant counterparts. They can however make themselves unappealing targets for further aggression by producing loud bursts of broadband noise and high-frequency tonal whines and screams. Experienced at close range, the aversive qualities of such screaming may be noxious and difficult to habituate to, effectively testing a dominant's commitment to further aggression (Rendall et al., 2009)

In most primate species an additional class of sounds known as 'sonants and gruffs' exists which is effectively the structural antithesis of squeaks, and screams. In contrast to shrieks, which are spectrally chaotic, sonants tend to be either tonal, harmonically rich calls, or characterized by diffuse broadband spectral structures. Although the chaotic patterning of screams is well suited to inducing affect in listeners, the predictable tonal quality of sonants gives them less inherent affective force. However,

16

the spectra of these calls reveal cues to caller identity as their production is labile based on developmental idiosyncrasies in the operation and anatomy of the larynx and supralaryngeal vocal-tract cavities. Differences in these structural characteristics impart distinctive acoustic features to calls that are associated with either the dynamic action of the vocal folds or relatively static resonance properties of the vocal tract cavities (Ghazanfar et al., 2007; Rendall, Owren, & Rodman, 1998). Many of these kinds of calls are used in face-to-face social interactions, where cues to identity seem redundant. However, providing an additional cue of identity in intensely social groups where social status hinges entirely on personal identity and a history of previous interactions may effectively highlight and reinforce one's history of social influence.

Vocal signals that induce autonomic responses in listeners can serve to highlight salient events and thereby support additional learning about them. The alarm call system of vervet monkeys is one example, with monkeys producing a small number of different vocalizations that are specific to major classes of predators, with each predator requiring a different avoidance response. The alarm calls elicit the different escape responses from adult listeners as though the calls were referential or semantic, carrying information about the predator encountered the way human words function (Seyfarth et al., 1980). Predator-naïve infant vervets do not respond to these calls with differentiated responses to different alarm calls, however they do respond to all calls with a generalized startle reaction based of their shared affect-inducing acoustic features (Seyfarth & Cheney, 1986). With experience, the undifferentiated startle response of infants is shaped and they learn which behavioural response is appropriate for a given call, although it is the basal

affective salience of the call structure that flags this significance prior to subsequent specification.

A more familiar example of the importance of acoustic characteristics of sounds in modulating behavioural propensities is further exemplified by the difference between fire trucks and ice cream trucks. Each of the two vehicles is attempting to signal its presence in an obvious fashion, but does so by using very different signals because the induced affect and behavioural responses sought are very different. Ice cream trucks are melodic and inviting; Fire trucks are loud and foreboding, and in choosing acoustic characteristics to match their purpose their designers ensure that people are more likely to flock towards ice cream trucks and away from fire trucks.

Given that vocalizations can establish conditions for learning by flagging salient events, it follows that signalers can condition listeners to the consequences of their vocalizations if they reliably pair calls of a particular kind with salient, emotion-inducing acts. In primates (and other social species), dominant individuals routinely antagonize subordinates, pairing threat vocalizations with biting and chasing the subordinate. As a consequence, the threat vocalizations become predictors of the associated aggression-induced affect. Subsequently, the dominant can elicit similar negative affect and submission in previous victims by use of calls alone. Conversely, the same dominant individual may produce affiliative sounds before approaching a subordinate with peaceable intent, effectively guarding against defensive behaviour by the subordinate which may be costly even for the dominant individual. Commonly, this type of interaction occurs when dominant females interact with the young infants of subordinate

females, in many cases including grooming by the dominant of the approached female or her infant. In these cases, the subordinate might come to associate the tonal calls with their own emotional response to a grooming episode that often follows approach and calling. In both instances, an animal becomes socially potent by directly impacting the learning processes of a listener to create vocal leverage over its affective states and subsequent behaviour (Owren & Rendall, 1997, 2001).

Just as naïve juveniles are unable to know which specific response to make to a predator call in vervets, human infants are unable to know that a loud siren with particular characteristics signals a fire truck. The acoustic characteristics of these two signals ensure however that their presence will be salient to even naïve listeners, not only inducing behavioural responses that may be simplistically adaptive (in the case of predators, some response is almost certainly better than none; so too for a fire truck barreling down the street) but also flagging the event and motivating attentional states that allow for easy learning of contingencies. Fire trucks are loud and aversive; ice cream trucks are melodious and inviting.

1.2.3- Speech Prosody and Affect Modulation

The theory of affective semantics raises important questions for human communication about what the majority of linguistic acts are all about. Certainly, in the modern age the bias towards understanding language as a system of information transmission is understandable given that we consume massive amounts of information in written form on a daily basis. But, historically this was not the case, and even in written language traits like linguistic flow are important at the very least in making messages

interesting. The informational account of human language also suffers somewhat from the problem of irreducible complexity: undoubtedly, a fully productive language system is very valuable for transmitting information and doing uniquely human things like solving collective action problems and creating surplus goods; but where does that leave the connection between animal and human communication? More likely, it seems, is the possibility that despite the strengths of language for transmitting information, a great deal of human verbal communication is still about modulating the affect of other people.

Swearing serves as a paradigm example of emotionally valenced language that exemplifies the principles of affective semantics. Lancker and Cummings (1999; see also Yardy, 2010) for example found that swear words contain an unusually high number of fricative and plosive consonants (specifically /f/, /sh/ and /k/). Whether one thinks of swear words as an honest reflection of emotional or motivational state (such as when signaling pain or embarrassment), or as being exceptionally effective attention grabbers (such as when directing swears at others), they are undoubtedly effective, motivated at least in part by their acoustical structure (and also prosodic differences in how they are said, for example, generally with increased amplitude). Even in cases where new swear words are lexicalized from existing words, such as in the French *'tabernacle!'*, the words are most often plosive in nature, and in the above case do not represent any threatening object or event, but rather a box in which the Eucharist is stored in a church.

Gross acoustical characteristics are of course far from the only way in which human language can be modulated. Temporal patterning, pitch, and other aspects of speech prosody are all important for both conveying meaning or emotion and influencing

emotions in others; a fact that has always been apparent to musicians and poets but has been strangely ignored by linguists. For example, rising or falling pitch can be used to demarcate certain phrase types (e.g. questions from statements; Hirst & Di Cristo, 1998) and also to signal the relative confidence of the speaker (with confident speakers using low or falling pitch and unconfident ones using high or rising pitch; Bolinger, 1978). Mothers also use a specific form of child-directed speech called motherese (Matychuk, 2005) when communicating with their children; relying on exaggerated emotional valence, rhythm, and pitch modulation to strengthen bonds and influence behaviour. So powerful is the concept of rhythm that Darwin himself (Darwin, 1871) and many others since (e.g. Levitin, 2007) have pointed to traditions of chanting and singing as possible roots of language, describing musical protolanguages that might have served as the foundation for the eventual emergence of more robust communication systems.

Whatever the case, considerations of speech prosody and how it might influence the affective powers of language suggest that communication, even in the refined and lexicalized form used by humans, cannot be only about the transmission of information, and further that Hockett's language universals are likely not exhaustive. The communication systems of both animals and humans are full of non-propositional content that still serves a purpose despite not being a carrier of information. Inducing a listener to pay some *fucking* attention to what you are saying is not simply a function of the mutually agreed upon naughtiness of the word *fuck*, but is contingent upon the affective valence inherent to the acoustic signal of the word itself.

1.3- Natural Structure in Language

Given differential perceptual and hedonic salience, what sorts of structural regularities might we expect to exist in language, and how might they be modulated? Connections between certain types of sounds and events seem necessary; objects moving in relation to the listener produce a Doppler effect, increasing in pitch as they near a subject and decreasing as they move farther away; similarly, amplitude rises and falls off based on distance as sound is dispersed in the environment. All other things being equal, larger objects also produce louder sounds with lower pitches; a relationship that holds generally both within and across species, including humans (reviewed in Fitch & Hauser, 1995; Fitch, 2000; Rendall, Kollias, Ney, & Lloyd, 2005; Ghazanfar & Rendall, 2008). Further, for reasons explained above, certain characteristics of sounds are also related to valenced affective reactions in perceivers (Owren & Rendall, 1997, 2001).

Normal research participants have exhibited a plethora of sound-symbolic correspondences that vary in how surprising they seem to be, but within the population are groups of abnormal individuals for whom existing sound-symbolic relationships are more prevalent or who display cross-modal associations that appear peculiar and do not reflect the perceptual experiences of normal individuals. Cross-modal integration in individuals in these groups is known as synesthesia, and researchers have provided different accounts of how aberrant synaesthetes are compared to the rest of the population, with some claiming that synesthesia relies on cortical wiring that is not present in normal individuals (e.g. Brang, Hubbard, Coulson, Huang, & Ramachandran,

22

2010) and others claiming that at least some forms of synesthesia use cross-modal connections which are common in the normal population (e.g. Ward et al., 2006).

1.3.1- Sound Symbolism in Normal Populations

Magnitude symbolism is perhaps the best known example of sound symbolism in normal populations. Sapir (1929; see also Diffloth, 1994) first described the phenomenon, relying on an experiment in which participants were shown pictures of two objects which were identical in all characteristics other than their size and then asked to choose between two labels for the objects. At rates far higher than chance, participants labeled the larger of two objects (e.g. tables) with words containing low back vowels (e.g. *'mal'*) and the smaller of the two objects with words containing high front vowels (e.g. *'mil'*). Further evidence for size-sound symbolism has been found not only in English (Johnson, 1967), but also in Chinese, Thai (Huang, 1967), Korean (Kim, 1977), and several other languages (Gebels, 1969; Malmberg, 1964; cf Newman, 1933; Newmeyer, 1993). Not only are words denoting size themselves (e.g. *huge* vs. *teeny*) exemplars of magnitude symbolism, but work by Berlin (1994) suggests that the size of individuals within particular linguistic classes is sound-symbolically mediated. In examinations of ethnozoological literature Berlin found that larger fish and birds were typically given names containing low back vowels while smaller fish and birds were typically given names containing high front vowels, and that this association even held across languages.

Far from being the lone case of sound symbolism, magnitude symbolism is only the tip of the iceberg. Researchers have found through experimental investigation or word corpus studies that a number of sound symbolic relationships can be reliably

demonstrated. Roblee and Washburn (1912) found that in the context of nonsense words the plosive consonants /g/, /t/, and /k/ were deemed least agreeable and the sonorant consonants /l/, /m/, and /n/ most agreeable. Similarly, Bar and Neta (2006) found that participants rated curved objects as more likeable than jagged objects, which the authors interpreted to be because jagged objects were seen as more threatening and aggressive, with evidence presented that the viewing of more jagged objects produced greater amygdalar activation associated with fear. Supporting this contention, Aronoff (2006) found that when asked to draw "war masks" participants produced sketches with more jagged features, creating masks with curved features for the purposes of courtship. Aronoff also found that threatening or aggressive roles in ballets were typified by dancers who moved in straight lines and at acute angles, while more affectionate and sympathetic characters moved in curved trajectories.

Aronoff's findings are interesting in that they directly implicate emotional valence as a potential mediator of sound symbolic relationships- a finding which is well supported by a large corpus of experiments. Tarte (1981) found for example that low pitch sounds, presented as both pure tones and low back vowels were strongly associated with large, heavy, slow, and masculine traits, while high tones and high front vowels were associated with smallness, lightness, femininity, and activity. Similarly, Lindauer (1990) found that jagged two dimensional objects were labeled as tough, strong, and unfriendly, whereas curved images were labeled as peaceful, relaxed, and quiet. Both Marks (1996) and Newman (1933) found further evidence for perceptual correlations between the pitch of sounds and their perceptual brightness, with high pitched sounds being rated as brighter

24

than low pitched sounds (see Hinton, Nichols, & Ohala, 1994; and Nuckolls, 1999 for more detailed treatments of other sound symbolic associations).

1.3.2- Sound Symbolism and Synaesthesia

Some of the sound symbolic relationships which hold broadly in the population at large can also be seen in synaesthetes to an exaggerated degree; for example, many synaesthetes experience cross-modal correspondences which seem to be either mediated by their emotions or to have an impact on their emotional state. Callejas, Acosta, and Lupianez (2007) found that in grapheme color synaesthetes (the most common form of synesthesia in which words or letters are associated with certain colors) words presented visually in the same color as the association of the synaesthete are processed more quickly and induce positive feelings, whereas words presented in incongruent colors are processed slowly and induce negative affect. Day (2004) found evidence that although mappings are not exact, certain graphemes and phonemes have noticeable tendencies towards certain colors among synaesthetes, and furthermore, that colored phoneme trends share some similarities with worldwide sound symbolic trends.

Grapheme-color synesthesia is the most common type of synesthesia, but as mentioned cross-modal linkages between the chemical senses like taste and smell are linked in nearly 100% of the population and could easily be labeled as synesthesia. Even within the basic senses there are a number of cross-modal correspondences that seem far more bizarre. Richer, Beaufils, and Poirier (2011) have described a case study of lexical-gustatory synesthesia, where words have associated taste sensations, and, surprisingly, Spence and Gallace (2010) have described similar associations in normal members of the

25

population, who were, for example, more likely to believe that an ice cream labeled

'Frosch' was more creamy and smooth than an ice cream labeled *'Frisch'* (see also Ngo,

Mishra, & Spence, 2011).

Correspondences between the cross-modal associations of synaesthetes and

controls were originally quite surprising, as generally synesthesia was assumed to be a

very rare condition, but recent work by Simner et al. (2006) suggests that the prevalence

of cross-modal experience is approximately one hundred times higher than previously

believed. For types of synesthesia for which no direct analogies have been recognized in

normal members of the population, it may be the case that (as Stalin is credited with

saying) "Quantity has a quality all its own"; in normal members of the population other

cross-modal correspondences might simply be too weak to be easily observed, despite

being present. Spector and Maurer (2009; Maurer & Mondloch, 2004) would likely support

this interpretation of synesthesia, as they have theorized that synesthesia reflects an early

developmental stage which all normal individuals go through, but in synaesthetes for

some reason the neural connections associating the senses are not effectively pruned.

1.3.3- Mechanistic Explanations of Cross-modal Processing

Thanks to advances in brain imaging we now know a great deal more about the

cross-modal associations of synaesthetes and how they might be instantiated. Rouw and

Scholte (2007) found for example that there is increased structural connectivity between

cortical areas associated with grapheme and color processing in grapheme-color

synaesthetes that may underlie their cross-modal experiences. Aleman, Rutten,

Sitskoorn, Dautzenberg, & Ramsey (2001) have provided support for the notion of

aberrant connectivity; demonstrating that in at least one case activation of the visual striate cortex occurred when hearing words, even in the absence of any direct visual stimulation, suggesting that the leakage occurs because of redundancy in processing (see also Brang, Rouw, & Ramachandran, 2011; Burton, Snyder, Diamond, & Raichle, 2006; Muggleton, Tsakanikos, Walsh, & Ward, 2007). Nikolie, Lichti, and Singer (2007; see also Gray et al., 2006) provide evidence suggesting that synesthesia can be mediated not only in low level cortical activation, but also in higher order channels. In an experiment using the Stroop task, Nikolie et al. demonstrated that synesthesia is processed in opponent process channels for color perception; greater interference was found in reaction times and concordance rates when participants were asked to name words printed in colors which were incongruent with their own representations than when they were asked to name congruent words. Furthermore, opponent colors interfered with one another more strongly than did non-opponent colors (thus, a yellow associated grapheme printed in blue was harder for participants to name than one printed in red).

Pervasive evidence that synaesthetic experiences might be mediated by increased cortical connectivity does not however suggest that synesthesia operates in isolation or without mediation. Simner and Hubbard (2006) have demonstrated that different types of synesthesia often co-occur within synaesthetes, and also that in these cases the variants of synesthesia can interact within cognitive tasks, suggesting that synaesthetic associations may be at least partially modulated by linguistic labels affixed to objects and events (Simner, 2006). Further, research has revealed that synaesthetic percepts can be moderated not only by directed attention (Rich & Mattingley, 2010), but also by posthypnotic suggestion (Terhune, Cardena, & Lindgren, 2010), suggesting that cross-

modal percepts are not a mandatory outcome of cross-cortical activation and may be mediated by higher level cognitive processes. Finally, Withhoft and Winawer (2006) provided evidence that in at least one case the specific form of grapheme-color synesthesia was heavily influenced by having colored refrigerator magnets during childhood.

Spector and Maurer (2009)'s account of synesthesia as a state inherent to neural development, coupled with evidence that cross-modal linkages allow for more fluent and efficient processing of information (e.g. Calvert, Hansen, Iversen, & Brammer, 2001; Kovic, Plunkett, & Westermann, 2010) suggest that sound symbolic associations might have broad implications for language learning or stabilization that far from being aberrant are incredibly beneficial when brought to the difficult task of language learning.

1.4-   The Bouba-Kiki Effect

In the broad scheme of sound symbolism, the Bouba-Kiki effect is a rather small piece of a large and complex puzzle, but by better understanding individual examples of sound-symbolism it might be possible to piece together the boundary conditions from which language can be built up. Insofar as the Bouba-Kiki effect is real, it is representative of the associations of a rather wide swath of acoustic space, encompassing not only the range of consonants based on their sonority, but also potentially the vowels based on their locus and the feeling of their articulation. Further, the characteristics of figures with which sonorant and plosive sounds are associated is an open question that hasn't been systematically explored.

1.4.1- The Nature of the Effect

Work on the Bouba-Kiki effect has focused primarily on the vowel content of the words used and possible associations between those vowel sounds and the visual characteristics of the objects presented to the participants. Variously, the findings of Davis (1961), Ramachandran and Hubbard (2001), and Maurer et al. (2006) have been couched in the importance of vowels within the target words. In each case, the authors appealed to the differences in sound or orthographic profiles between the contrasting vowel types. Vowels which are produced with a concurrent rounding of the lips (such as */ah/, /oh/,* and */oo/*) have been demonstrated to be reliably associated with curved line, or rounded images, while vowels that do not involve rounding of the lips (i.e. */uh/, /ay/, /ee/*) have been associated with more jagged and chaotic images. Davis (1961) presented the words '*takete'* and '*uloomu'* along with jagged and curved images similar to those used by Köhler to a population of English naïve children in Tanganyika, an east African territory that was relatively geographically isolated from English speaking populations. Davis demonstrated that even the non-English speakers reliably paired the word '*takete'* with the jagged object and the word '*uloomu'* with the curved object. Subsequently, Ramachandran and Hubbard (2001) produced similar results to Köhler, with both English and Tamil speaking participants matching the word '*Bouba'* to a rounded object and '*Kiki'* to a more jagged nonsense object. In 2006, Maurer et al. replicated these findings using a larger set of image and word pairs (4 pairs of each) and showed that both toddlers and adults paired words containing rounded vowels (e.g. '*maluma'*) to rounded objects and words containing non-rounded vowels (e.g. '*takete'*) to more jagged objects.

An explanation for the rounded vs. non-rounded distinction has been ascribed to both the sensorimotor percept of sound production and the observed roundness of the lips when monitoring another speaker. Ramachandran and Hubbard (2001) proposed that activation of contiguous cortical areas entangled what should be two unique perceptual faculties, for example the motor cortex responsible for articulation of vowel sounds and the areas of the visual cortex associated with perception of rounded edges, in such a way that associations between the two cortices could produce activation even in the absence of one of the two types of stimuli. Although the proposal raised by the two authors that an association between rounded objects and rounded vowels might be present seemed eminently plausible, the reasoning behind an association between non-rounded vowels and jagged objects was not as clearly delineated. Despite this, subsequent authors (e.g. Maurer et al., 2006) have offered up Ramachandran and Hubbard's explanation as the most plausible one by which associations between seemingly nonsense words and images could be associated.

Interestingly, although the findings of Köhler (1947) are amenable to classification based on their vowel content, with the "rounded" word '*maluma*' being associated reliably with curved figures and the "non-rounded" word '*takete*' being associated with jagged figures, Köhler himself never delineated the source of the effect, merely using the findings as evidence of non-arbitrariness that ran counter to the tenets of behaviourism. Köhler's word pairs, much like those of subsequent researchers, also contained a second contrast- between the consonants present in the word pairs. Despite the obvious differences in both the acoustic characteristics and the differences in production which underlie those characteristics, most subsequent research has not only

failed to control the consonant content of words to analyze vowels in isolation, but has unwittingly paired vowels and consonants in uncontrolled ways, leading to confusion about how results should be interpreted.

Tarte and Barritt (1971; Tarte, 1974) were the first to systematically control for the consonant content of words; they constructed a full complement of single syllable CVC words using the consonants /w/, /d/, and /k/ along with the vowels /i/, /a/, /u/ where each word ended with the letter /s/. Rather than using nonsense image stimuli reminiscent of those used by Köhler and others, Tarte and Barritt used simple triangles and ellipses of various sizes and orientations, asking participants to chose which word best fit with a given figure. By far the largest effects were for a persistent matching in both English (1971) and Czech (1974) speakers of words containing the high front vowel /i/ to smaller objects and the low back vowel /u/ to larger objects (consistent with earlier findings by Sapir (1929), Newman (1933) and others regarding size-sound symbolism). They also demonstrated that participants matched triangular shapes more often to the letter /i/ regardless of object size and the letters /a/ and /u/ to the elliptical shapes. Effects for consonants in both the size and image type conditions were found to be non-significant, although a general trend was found where /w/ was associated more often with ellipses and the remaining two consonants more often with triangles. Although the findings of the study provided evidence for further vowel driven sound symbolic symbolism, the differences in stimulus construction served to make the results difficult to analyze, as commensurability between this study and previous ones was difficult to evaluate. The complement of consonants used was rather small, and the contrast between those consonants in terms of sonority was fairly minimal. Additionally, the presence of an /s/ in

the word final position for each word ensured that all were relatively sonorant, perhaps yielding differences in consonant composition that were swamped by the larger and more obvious vowel contrasts. Despite these possible concerns, Tarte and Barritt's findings were likely part of the motivation of future research to selectively explore vowels and ignore possible consonant based confounds.

Westbury (2005) serves as a prime example of a methodology for testing specifically the effect of consonants on sound symbolic relationships like the Bouba-Kiki effect. Westbury both delineated the conceptual and methodological issues of previous studies and bypassed a number of them by moving to an implicit task wherein nonsense words were presented within jagged or curvy frames. Participants were asked merely to identify whether the string of syllables within the frame was a word or a non-word, and Westbury found that strings containing continuant consonants were identified more quickly and accurately within curved frames and strings containing plosive consonants were identified more quickly and accurately within jagged frames. Despite addressing a number of methodological shortcomings, the least of which were transparency and task demand, Westbury's study did not demonstrate the validity of his concerns regarding previous experiments other than demonstrating an effect of consonants that had not previously been shown.

Lack of precision in the creation of experimental stimuli has in fact been the primary impediment to research into the Bouba-Kiki effect. Image construction in previous experiments had also been problematic, with experimenters using both a very

limited range of stimuli, and also failing to quantify what characteristics of the images differed other than a vague percept of curviness or jaggedness.

Word stimuli from previous experiments were typically constructed in such a way that analyzing them based on their vowel content produced identical results to an analysis of their consonant content. Specifically, Köhler's pairing of '*takete*' and '*maluma*', Davis' pairing of '*takete*' and '*uloomu*' and Maurer et al.'s pairings (e.g. '*tuhkeetee*' and '*maaboomaa*') were all perfectly confounded. In each case, the prediction by the authors that objects with a curved visual form should be matched with words that involve rounded vowels follows an identical trajectory to predictions made by other authors (e.g. Westbury, 2005) that relatively more mellifluous consonants (e.g. /*m*/, /*n*/, /*l*/) should be associated with more curved objects and that plosive/stop consonants (e.g. /*t*/, /*k*/, /*p*/) should be associated with jagged objects. Although Ramachandran and Hubbard (2001)'s word choice of '*bouba*' and '*kiki*' (which subsequently provided a name for the effect) avoided this complication somewhat by using only plosive consonants, to claim that there was no difference between consonants in these word pairings would rely on a simplistic categorization wherein the acoustic or production differences of /*b*/ and /*k*/ were claimed to be equivalent.

In addition to concerns about whether the Bouba-Kiki effect is best explained in terms of vowels, consonants, or both, the perceptual channel through which participants interact with nonsense words and objects has been raised as a potential confound. Köhler, Davis, and others have typically presented their figures as two-dimensional representations while allowing their words to be presented both in the auditory domain

(said by the experimenter in clear view of the participants) and visually in written form. Additional visual cues such as the rounding of the experimenter's lips during production of the words may also have served to influence participant responses, enhancing the saliency of for example rounded vs. non-rounded vowels. Other experimenters (e.g. Westbury, 2005) in seeking to standardize their experimental procedures have eliminated a number of these cues. The different results of these differing experimental protocols might thus be invoked as a reason for the diverging results found by researchers. Unfortunately, thus far appeals to problems associated with changing experimental methods of presentation have mostly been self serving- placing the onus of demonstrating equivalence between experimental manipulations on experimenters who are only attempting to make things more systematic (e.g. Westbury, 2005).

Variously, lip shape during production (and the associations this recalls upon merely hearing a word), the orthographic form of letters, and the spectral characteristics of sounds have all been proposed to explain the Bouba-Kiki effect, and the possibility that any or all might be confounded with each other has only been dealt with on a cursory level by ensuring in some experiments that not all sources of variation are immediately available to participants.

The orthographic effect, where the visual characteristics of letters are proposed to be directly related to the objects they are subsequently paired with, is perhaps the thorniest of all possible explanations. Although it seems sensible to suggest that participants might match jagged objects with letters like /k/,/t/ and /A/ and curved objects to letters like /n/, /o/, and /u/, when orthographic effects have been controlled by

researchers reliable sound symbolic effects have still been found. Westbury (2005) for example presented both curved (e.g. &, @, 6) and jagged (e.g. #, 7, >) symbolic tokens (letters, numbers, and other symbols) within either curved or jagged visual frames (using the same procedure wherein he showed implicit evidence for the Bouba-Kiki effect) and showed that the curvature of the frame within which the symbol was embedded did not affect participants' response times or accuracy in determining whether the embedded symbol was a letter or not. Westbury took this as evidence that at least in his task the structure of letters themselves was insufficient to be driving the effects he had observed using full words.

Of course, even in the best of situations it may not be possible to fully control for orthographic effects- insofar as experimenters are correct that sound symbolic relationships between objects and their tokens exist, it is possible that this sort of representation might have leaked over into the creation of letters themselves. Thus, the letter /*k*/ might have the form that it does precisely because of the way it sounds- given multiple candidates for the form of a letter used to represent a plosive, broadband sound with certain acoustic characteristics, it is perhaps unsurprising that a specific visual polymorphism of a chaotic nature was chosen to represent it. The idea that orthographic characteristics of letters might be confounding thus ignores the question of why there might be systematic confounds in the first place.

Moving to an auditory presentation of word stimuli in experiments where words are not presented in written form sidesteps this problem somewhat, giving participants access to neither the orthographic forms of letters nor to words as a whole. Unfortunately,

one cannot be sure what sorts of processing participants are actually doing- the auditory presentation of a word might be sufficient to invoke some sort of association to the visual percept of that word, or in the case of nonsense words at least to the best guess a participant might have to what the word would look like. In fact, given explanations of the Bouba-Kiki effect which rely on associations formed between visual percepts and auditory ones, it would be peculiar if auditory presentation of words did not evoke associations with their visual percepts and vice versa.

Presentation of word stimuli in multiple domains has thus demonstrated that the Bouba-Kiki effect is a robust one; there are sound symbolic associations that seem non-arbitrary regardless of the way in which stimuli are presented. Unfortunately, this strength of previous research has also been a weakness- the linkage between auditory and visual elements is unknown, and it cannot be determined from existing experiments whether these results are entirely commensurate with one another, or are merely demonstrating parallel effects. Further, given that processing stimuli from a single domain is likely associated with processing in parallel domains, it may be very difficult to demonstrate the "locus" of the effect, as it may not actually be localized in any conventional sense.

The characteristics of figures that have been used in experiments examining the Bouba-Kiki phenomenon has rarely been given any real attention, which is perhaps surprising given the amount of attention that has been paid to the visual percepts of the letters in the words themselves. Previous authors have all seemed to agree roughly on the differences in jaggedness and curviness between the figures in their pairs, and these

36

differences have indeed been obvious to the naked eye, but what characteristics of these images actually differ other than the vague visual percept of blobiness versus jaggedness has rarely been delineated. That authors are able to produce images that are representative of the Bouba-Kiki effect demonstrates at least that their specific biases seem to be in line with their participants, but the paucity of stimuli used in experiments raises the possibility that only limited stimuli or very obvious differences between stimuli are sufficient to demonstrate the effects in question.

Even in contemporary studies of the Bouba-Kiki effect the number of stimuli has suffered from this problem; Maurer et al. (2006) for example used only four figure and word pairs, while Westbury (2005), significantly improving on this number, still used only twenty pairs. It is difficult to delineate which parts of the figure are selectively driving the effect; is it the curvature of the lines themselves, or the interaction of line segments joining in acute angles which participants are attending to selectively?

1.4.2- Methodology and Transparency

The importance of stimulus control, both in the characteristics of the generation and subsequent presentation to participants, is also implicated in concerns about the robustness of observed sound-symbolic correspondences and its possible real world implications. Early findings by Köhler (1947) and Davis (1961) demonstrated concordance rates of over 90%, and subsequent experiments using similar experimental methodologies have yielded large, but slightly more modest concordance rates around 80% (e.g. Maurer et al., 2006). As protocols have shifted towards making the experimental manipulations of researchers less transparent to participants, concordance

rates have become less and less robust. Westbury (2005)'s interference experiment for example yielded only very small differences in reaction time between conditions. Similarly, Tarte's more tightly regulated stimulus generation yielded responses which differed from one another typically by less than 10% (e.g. the vowel /i/ was chosen ~40% of the time for triangles, with /a/ and /u/ each being chosen around 30% of the time). At least two explanations can be given to explain the reduction in concordance rates as experimental manipulations become less transparent, one of which is potentially problematic.

Liberally interpreted, the reduced concordance rates of less transparent experiments are to be expected; even given a fully formed innate bias, the removal of salient cues could potentially cause performance to fall. For example, given the hypothesis that curved and jagged images are differentially associated with plosive and sonorant consonants, the presentation of only a single (curved or jagged) image could potentially have less of an impact for cueing the association than would the simultaneous presentation of exemplars from both categories. Similarly, as concurrently presented images become increasingly similar to one another and less stylized the associations may become less salient, analogous to monitoring temperature changes on a thermometer-larger changes would be easier to notice than micro fluctuations not as salient to the observer. Alternatively, reduced concordance rates in less transparent experiments may be a function of explicit transparency, with participants effectively "figuring out" what individual experiments are exploring and thereafter answering based on task demands.

One possible answer to this concern about transparency is that regardless of whether the participants are explicitly aware of the experimental manipulations they still respond uniformly, which serves only to push back the effect to explain why certain manipulations and rules for matching seem more obvious than others. Unfortunately, in experiments where the vowel and consonant content of words was confounded this answer does not suffice- whether participants chose to respond based on either consonant or vowel content they would have made the same responses, and thus a response bias claimed to be based on either characteristic could equally well be explained by a bimodal response bias from participants (with half choosing to answer correctly based on vowels, and the other half based on consonants).

In the worst case, experiments plumbing the Bouba-Kiki effect might be overstating its importance- given the possibility that response biases emerge only when experimental manipulations are transparent and can be easily deciphered by participants, it would follow that we should expect very little sound symbolism in wholly robust language, which after all takes place almost exclusively outside of laboratories and which involves far more experiment naïve participants than populations of undergraduate psychology students. The majority of work into the Bouba-Kiki effect has sidestepped this possibility by using only nonsense words, which is curious given the fact that the Bouba-Kiki effect is often cited as an example opposed to the arbitrariness of language. Only Westbury included real words with the same consonant-vowel configuration as his nonsense words and found no effect of frame type on either reaction time or accuracy in the recognition of words (where he found an effect for nonwords).

1.5-   Thesis Outline

In Chapter Two, I report two experiments designed primarily to address concerns regarding stimulus construction and interpretation of the Bouba-Kiki effect. In the first experiment, I present both an improvement in the creation of figures and words (and an overall increase in the number of trials compared to previous experiments) and a re-analysis of previous Bouba-Kiki results based on the consonant characteristics of words. This is accomplished by decoupling previously confounded consonant and vowel characteristics; revealing that previous results were preferentially driven by consonant class of nonsense words and had previously been interpreted incorrectly due to consonant-vowel confounds. In the second experiment, I provide additional evidence of a consonant-driven Bouba-Kiki effect while simultaneously moving to an auditory presentation that ameliorates some of the concerns relating to the orthographic forms of the presented letters.

Chapter Three presents a more systematic examination of the effects of consonants and vowels. In the first experiment, I demonstrate that, in the absence of variation in consonant content, vowels are capable of driving the Bouba-Kiki effect in the same direction as proposed by previous researchers. In the second experiment, in which participants were tasked with constructing words rather than simply choosing between them, I demonstrate an effect for both classes of letters, with consonants modulating associations to jagged images and vowels modulating associations to curved or rounded images. Furthermore, Experiment 2 demonstrates that consonants and vowels might have additive effects in producing word sound/ object shape correspondences. Finally,

evidence from Experiment 2 is suggestive of the influence of other characteristics like ease

of articulation or adherence to conventional regularities in the production of words.

In Chapter Four I report the results of an implicit experiment designed primarily to

deal with concerns regarding task transparency and demand characteristics. Rather than

presenting participants with a task and allowing them to answer based on the task, I

separated them into two categories, testing teaching them a matching rule congruent

with previously observed sound-symbolic biases or one incongruent with those biases and

measuring performance on a recognition task after a brief learning phase. The results of

the experiment in Chapter Four suggest that the Bouba-Kiki effect is a *bona fide* example

of universal sound symbolism that might have important effects for learning and

language stabilization.

Finally, in Chapter Five, I summarize the major findings of the preceding chapters

and discuss potential implications of the Bouba-Kiki effect and sound symbolism more

broadly for the thorny question of language acquisition. I also consider limitations of the

current work and research directions that would be profitable for future work,

highlighting the interface between work describing sound symbolic relationships and

work testing the functional value of those relationships.

Throughout the remainder of the thesis I pay particular attention to the notion of

linguistic arbitrariness and how the information transfer paradigm has influenced

research into sound symbolism and the Bouba-Kiki effect. I specifically stress that, by

combining functional accounts for sound-symbolism in the transmission of non-

propositional content and the modulation of affect with plausible mechanistic accounts of

the instantiation of cross-modal biases, I can arrive at a better understanding of the

boundary conditions that may reveal the importance of sound-symbolic correspondences

for language acquisition.
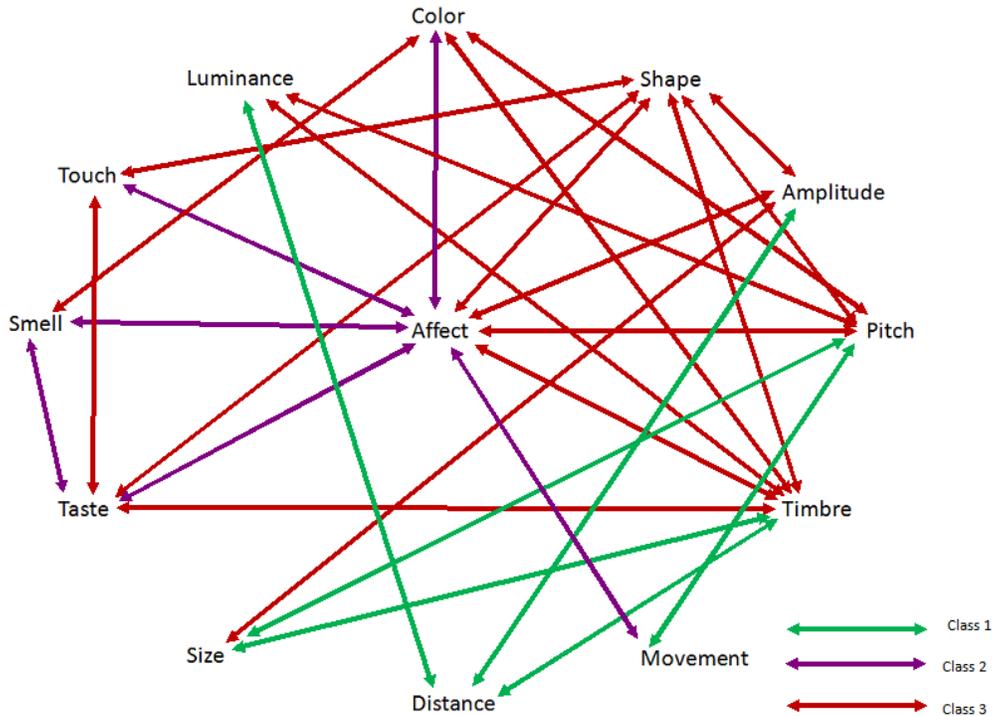
Figure 1.1. Examples of cross-modal connections identified in previous research labeled according to their relative pervasiveness. Class 1 connections are those which seem mandatory given the laws of physics (such as light fading over distance). Class 2 connections are those which don't seem mandatory, but which I otherwise not consider aberrant (e.g. the association between taste and smell). Class 3 connections are those which are typically labeled as aberrant and due to synesthesia (e.g. between shape and pitch), some of which have also been observed in non-synaesthetes.
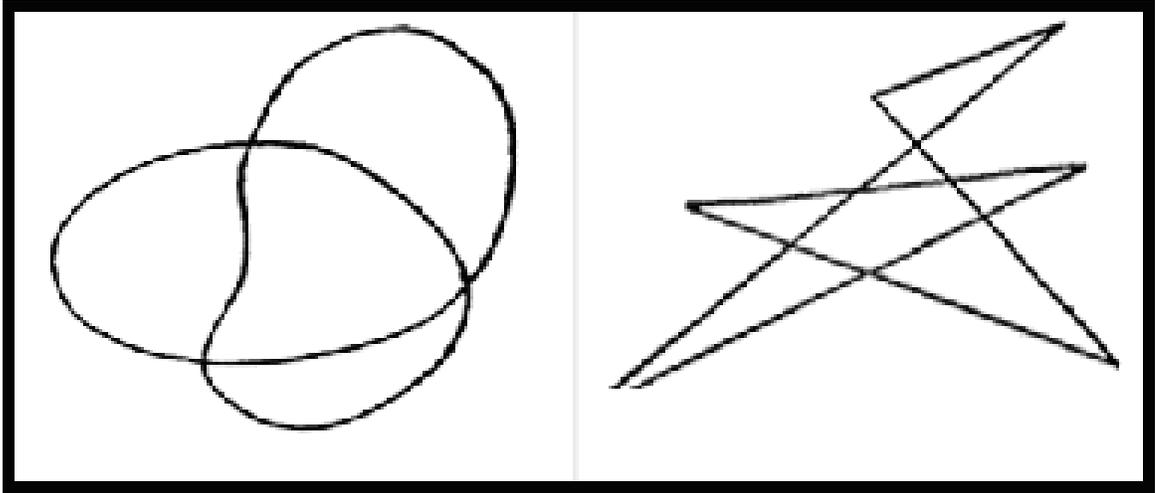
Figure 1.2. Figures used by Köhler in his original study.

CHAPTER TWO

THE SOUND OF ROUND: EVALUATING THE ROLE OF CONSONANTS IN THE CLASSIC TAKETE-MALUMA PHENOMENON

2.1- Introduction

Debate about the potential naturalness of words dates at least to Socratic times. In Plato's *Cratylus* dialogue, Socrates argues for the natural relationship between the structure of words and the things they denote (i.e. their meaning), while his interlocutor, Hermogenes, argues that the relationship is purely conventional. The naturalist and conventionalist positions have been debated many times since, but following Saussure (1949), the dominant view in contemporary linguistics is that *the form of the sign is arbitrary* and established by social convention.

This conclusion is not seriously disputed for the vast majority of words, but many possible exceptions have also been noted. Such exceptions are often considered together under the banner of *sound-symbolism* to indicate that the physical form of words (or of higher-order language constructions) might sometimes bear a non-arbitrary connection to the meanings instantiated (reviewed in Hinton et al. 1994). Among the best known examples is the phenomenon of vowel-specific marking of size diminution and augmentation where real or nonsense words containing high, front vowels (e.g., *bit*, *chip*, *teeny*) are judged more natural for denoting small size, while words containing low, back vowels (e.g. *block*, *hunk*, *chunk*) are judged more natural for denoting large size (Sapir 1929).

This effect has been traced to broader sound-size relationships obligatorily embodied in the physics of sound production generally and more specifically in the physics of voice production in humans and other animals (Morton 1994; Ohala 1994). For example, compared to small individuals, larger individuals also often have larger sound production apparatuses that naturally produce sounds of lower frequency. This basic relationship between body size and sound frequency holds across a variety of animal species and modes of sound production. Importantly, it also applies among primate species, including humans, where sound production is primarily voice-based. In humans and other primates, the vocal folds of the larynx and the vocal-tract cavities that resonate sounds generated by the larynx are both generally larger in larger-bodied individuals. This pattern yields lower frequency voices (with both lower pitch and lower resonances) in, for example, adults compared to children and in men compared to women (reviewed in Fitch & Hauser 1995; Fitch 2000; Ghazanfar & Rendall 2008; Rendall et al. 2005).

Consistent evidence that sound-symbolic relationships like this are actually manifest in the structure of real words is often lacking, controversial, or complicated (e.g. Bentley & Varon 1933; Diffloth 1994; Ultan 1978). As a result, proposals of sound-symbolism are often viewed with skepticism and treated as marginalia in linguistics and psycholinguistics. However, the consistency of people's behaviour in experimental studies of sound-symbolism, even if they involve interpreting nonsense word material, suggests that people might operate with inherent perceptual-cognitive biases that yield naïve expectations about language structure. Hence, it is important to investigate these possibilities further because perceptual-cognitive biases of this sort could serve to

facilitate normative language processing, or even language acquisition, if the biases are commensurate with real language constructions. Alternatively, such biases could serve to frustrate or impede language processing or acquisition if they run counter to the relationships instantiated in real languages. Either way, the biases could influence peoples' language faculties and they must, therefore, feature in the development of a fuller understanding of the psychology of language.

2.1.1- The Takete-Maluma Phenomenon

Another paradigm example of potential naïve language structure expectations involves the '*takete-maluma*' phenomenon. This phenomenon was famously described by Köhler (1947) who reported that subjects reliably matched nonsense words such as *takete* and *maluma* to images of unfamiliar *jagged* and *curved* objects, respectively, suggesting that people might indeed have some naïve expectations about the kinds of words best suited to name certain kinds of objects. Similar effects have been reported in various additional studies and these studies have also extended the observations to include other kinds of images and other nonsense words, such as *bouba* and *kiki*.

For example, Davis (1961) found that participants preferentially matched the nonsense word *takete* with a jagged image and *uloomu* with a curved image, and this was true for both English and Swahili speakers. Westbury (2005) tested the effect using a more indirect, selective interference design and found that participants were quicker to recognize and classify non-words (by comparison to real words) that were presented visually in text form if the consonant class of the non-words matched the shape of the frame within which the words were presented. Specifically, non-words presented within

47

curved frames were recognized more quickly if they contained continuant consonants (e.g., /l/ and /m/) and non-words presented within jagged frames were recognized more quickly if they contained stop consonants (e.g., /t/ and /k/). In another test, Tarte found that speakers of both English (Tarte & Barritt 1971) and Czech (Tarte 1974) consistently matched words containing the vowel /i/ (as in *beat*) to figures that were triangular in shape and they matched words containing the vowel /u/ (as in *rule*) to figures that were elliptical.

Despite results like these, there have been lingering doubts about the robustness of the matching biases involved. For example, there have been concerns about the limited number and variety of word and image stimuli used in many experiments. There have also been concerns that inadvertent selection biases by experimenters might account for some of the effects reported, making them effectively circular (reviewed in Westbury, 2005). At the same time, explanations for the source of the *takete-maluma* phenomenon have also varied considerably and added to the ambiguities. Thus, some studies trace the effect to the consonant content of words (e.g., Westbury, 2005), while others trace it to the vowel content (e.g., Maurer *et al.* 2006; Ramachandran & Hubbard 2001; Tarte & Barritt 1971; Tarte 1974).

Recently, Ramachandran and Hubbard (2001) offered a comprehensive neuropyschological account of the phenomenon. They proposed that the object-shape: word-form matching bias reflects synaesthesia-like co-activation of the motor or somatosensory areas involved in articulating different sounds and the visual areas associated with the perception of differently shaped objects. They argued that such cross-

48

modal linkages yield the natural matching of, for example, the visual percept of a curved, or rounded, object and the motor representation involved in the round-mouth articulation of the /oo/ vowel sound in the word *bouba*. Ramachandran and Hubbard (2001) elaborated this proposal into a broader synaesthetic theory of language origins and consciousness.

Maurer et al. (2006) extended this synaesthetic proposal in experiments conducted on young children (2.5 years old, sometimes referred to as toddlers). Despite relatively little language experience, these children showed the same matching tendencies as adults. Maurer et al. attributed the effect in infants to synaesthesia-like, cross-modal linkages that they propose are a general feature of neural organization in infants. Many of these cross-modal linkages are thought to be gradually pruned during development to yield the relative independence of different sensory systems that characterize most adults, but the cross-modal linkages are argued to be especially functional in infancy for language learning (Maurer & Mondloch, 2004; Mondloch & Maurer, 2004).

This recent emphasis on vowels and on potential cross-modal neural activity in accounts of the *takete-maluma* phenomenon is compelling. However, so too is the possibility that the consonant content of words might play a role (Westbury, 2005). Thus, there are clear spectral density and attack differences between /k/ and /m/ that make /k/ a relatively harsh (or plosive) consonant and /m/ a relatively mellifluous (or sonorant) consonant. These basic differences in spectral structure might naturally tend to imply or conjure '*harsh, fractured, or jagged*' constructs on the one hand and '*smooth connected,*

rounded' constructs on the other, and these effects might characterize other plosive and sonorant consonants.

Part of my impetus for re-evaluating the latter possibility stems from consistent *affective-semantic* relationships observed widely in the communication systems of animals (Rendall & Owren, 2009). Across many primates and other animal species, harsh, noisy and punctuate (i.e., plosive) sounds are associated with situations of high arousal and often also hostility and aggression. In contrast, smoother more harmonic (i.e., sonorant) sounds are associated with situations of lower arousal and also positive affiliation and contact (reviewed in Morton, 1977; Owren & Rendall, 1997, 2001; Rendall & Owren, 2009).

For example, in aggressive confrontations with predators or rival group members, many primate species produce very loud calls, often referred to by investigators as 'barks', 'screeches' or 'screams'. As the latter names for these calls imply, the calls tend to have sharp (abrupt) onsets and to be composed primarily of unstructured, broad-band noise. These calls tend to induce reflexive reactions in listeners preparatory to fight-or-flight type responses based on fairly direct stimulation of autonomic processes regulating whole-body arousal. In contrast, when calmly foraging or seeking close contact with social companions, the same primates produce calls with more gradual onsets and structured tonal or harmonic frequency spectra that tend to be referred to as 'coos' or 'peeps'. These kinds of calls tend not to induce the same sort of dramatic autonomic responses in listeners (reviewed in Owren & Rendall, 2001; Rendall & Owren, 2009).

These and other vocal-affect relationships are continuous with effects also known in human voice production. For example, young infants variously produce either sonorant-type coos or relatively plosive loud and abrasive screeches and screams in situations reflecting, respectively, relative comfort or contentment versus hunger or distress. And calls of these two types tend to induce very different affective and behavioural responses in caretakers and other adult listeners. Similarly, adult voicing patterns can be relatively punctuate, harsh and noisy in situations of high arousal (whether related to fear or anger) compared to during normative relaxed speech (reviewed in Bachorowski & Owren, 2008; Rendall, 2003).

Taken together, there appear to be some very broad relationships between plosive and sonorant sounds and the contexts they mediate that have different social and behavioural salience for listeners. It is possible that these relationships might extend, at least in a limited way, to certain consonant sounds of language that exemplify this same plosive-sonorant distinction and that thus inherit some natural semantic potential.

To test this hypothesis, I undertook a set of experiments which were organized by the following logic: 1. My first objective was simply to replicate the matching biases reported in previous studies of the *takete-maluma* phenomenon, utilizing the same word and image materials and the same testing methods used previously. This seemed a critical first step to be certain I was dealing with the same phenomenon. 2. A second major objective was to address concerns raised about stimulus materials used in previous studies and possible biases in their selection and construction. To address this issue, I developed new methods for generating random image and word materials that eliminate subjectivity

in the selection of experimental stimuli. I used these randomly generated images and words to extend previous work on the *takete-maluma* phenomenon and evaluate its robustness. 3. The third major objective was to test the extent to which the matching biases involved might be influenced specifically by the consonant content of the words.

EXPERIMENT 1: REPLICATING AND EXTENDING THE TAKETE-MALUMA PHENOMENON

The first step in research was an attempt to replicate the word-form: object-shape matching patterns reported in previous studies and extend past research on the influence that consonants in particular might play. To do this, my first experiment replicated the design of previous studies of the *takete-maluma* phenomenon and used similar word and image materials. Thus, participants were shown line drawings of an unfamiliar jagged object and an unfamiliar curved object and were given a choice of two nonsense words to match to them. At the same time, my experiment included two additional conditions designed to address methodological concerns raised about previous studies and to specifically test the role that consonants might play in the word-matching bias. One condition involved simply swapping the consonant and vowel contents of the original word stimuli to test whether participants' matching performance would track the previous vowel content or the consonant content. The other condition involved using an entirely new set of randomly generated images and words in an attempt to eliminate potential biases in experimenters' selection and construction of stimuli.

2.2- METHODS

2.2.1- Participants

Participants were 24 undergraduate students (13 Female, 11 Male) who were enrolled in introductory psychology courses at the University of Lethbridge and received partial course credit for their participation.

2.2.2- Word and Image Stimuli

In Condition 1 of this experiment, which was the direct replication condition, the word and image stimuli came directly from those used in previous studies by Köhler (1947) and Maurer et al. (2006). The word and image stimuli used are shown in Table 2.1 and Figure 2.1, respectively. Some of the stimuli used in Maurer et al. (2006) involved a mixed media presentation that allowed for haptic interaction with the objects. To improve comparability to previous studies, including the original work by Köhler, I used the typical two-dimensional graphical representations of these objects.

For Condition 2, I created a second set of word pairs by swapping the consonant content of the original word pairs used in Condition 1. For example, the classic word pair used originally by Köhler (1947) contrasted *takete* and *maluma*. In Condition 2, this word pair became *maleme* and *takuta* (see Table 2.1 for other examples). This manipulation was undertaken to test the effect that consonants might have using exactly the same word materials and methods that were used in previous studies. If previous interpretations were correct that the vowel content of words is primarily responsible for the preferential matching effects reported, then swapping the consonant content of the words should not

alter the result; subjects' matching responses should remain the same. However, if the consonant content of words has an effect, then subjects' matching responses should often change and follow the consonant (not vowel) content of the original words.

For Condition 3, I created an entirely new set of words to more systematically test the potential role of consonants and to minimize additional possible confounds raised previously. For example, to construct words, I selected as plosive consonants /t/, /k/, and /p/ and as sonorant consonants /l/, /m/, and /n/. The former are established stop-consonants in which voicing (i.e., vibration of the vocal folds) follows release of the consonant, creating a brief pulse of aspirated turbulent noise prior to the onset of stable vocal fold vibration for the following vowel sound. The latter sonorants are continuant consonants which are produced simultaneously with the vocal-fold vibration involved in the following vowel sound which makes the consonant component of the sound relatively vowel-like in structure. This selection of consonants allowed us to test the proposed affective-semantic distinction between relatively noisy, plosive sounds and relatively mellifluous, sonorant sounds. This selection of consonants also minimized potential orthographic confounds in which the visual form of the letters in the words resembles the jagged or curved form of the object shapes. This has been proposed to account for some previous results when for example words containing the letter /K/ are matched to jagged objects while words containing the letter /B/ are matched to rounded objects. However, such orthographic effects should not have been a factor in my experiment, because all three sonorant consonants, which are predicted to be matched to curved object shapes, are actually relatively jagged in their capitalized form *(/L/, /M/, /N/)* and one remains so in lower-case form *(/l/)*. At the same time, one of the plosive consonants, which are

proposed to be matched to jagged object shapes, is patently curved in both upper- and lower-case forms *(/P/, /p/)*. As a result, the orthographic representation of the consonants used in this experiment often worked against the predictions of my hypothesis.

For vowels, I chose not to use /o/ and /i/ because the orthographic form of these two vowels could be too obviously and canonically associated with rounded and jagged images, respectively. I used instead the vowels */a/, /e/,* and */u/* and always in their lower-case forms. Hence, the orthographic forms of these vowels were all similarly curved or rounded and would therefore introduce no systematic bias.

To avoid any possible subjectivity in the construction of words from this limited alphabet, I created word stimuli using Gammadyne Random Word Generator ©. Words were constrained to be four letters long in a two-syllable, consonant-vowel-consonant-vowel (cVcV) sequence, with the first letter capitalized. An additional constraint was that the letter */e/* could not appear in word-final position because that construction alters the pronunciation of the preceding vowel sound and can create a single-syllable word out of a two-syllable one (e.g., *Paku* versus *Pake*). The result was a database of two syllable words that were either entirely plosive (e.g. *Taku*) or entirely sonorant (e.g. *Nemu*) by consonant content (for a complete listing of word stimuli used in Condition 3, see Table 2.1).

For Condition 3, I also created an entirely new set of randomly generated curved and jagged images. To accomplish this, I developed a radially constrained mathematical formula which populated a field of finite size (Figure 2.2-A) with either five or ten sequentially ordered, randomly generated calculus points. The location of these points was determined by creating vectors which departed from the center of a field at random

magnitudes and directions. These points were numbered in the order that they were created, and sequentially labeled points had a second vector calculated from the midpoint of the line connecting them, giving a calculus point for the drawing of simple curves. Individual line segments were defined by their two original points and the generated curvature point (Figure 2.2-B). They were then joined with either straight (Figure 2.2-C) or curved (Figure 2.2-D) lines. Sequential line segments were then connected to one another. The resulting curved and jagged images were identical to one another except for the curvature of the lines used to draw them, which often yielded only very subtle differences between them (Figure 2.1). Importantly, this method also standardized the size of images and so avoided any possible size-related confounds in subjects' matching responses.

Together, these precautions for Condition 3 addressed common concerns about the limited variety of past word and image stimuli and potential biases in experimenters' selection of them.

2.2.3- Experimental Design

The experiment was conducted on computer via a graphical interface created using Runtime Revolution, Version 2.8.1. In each trial, the participant was shown two images side-by-side (one curved, one jagged) with two words below. Participants were asked to label the images with the words. Importantly, and as in previous studies, labeling one image with a given word automatically led to the assignment of the remaining word to the other image. Participants were free to change their labeling. They advanced to the next trial by clicking a button on-screen.

The experiment involved twenty of these forced-choice trials in three conditions: Condition 1 consisted of five trials using the original word and image pairs from previous experiments; Condition 2 consisted of five trials using the original image pairs and the original word pairs in which the consonant content of the words had been swapped; Condition 3 involved ten trials using the new set of random images and random words. The coupling of word and image pairs, and their placement on screen, in each trial was randomized within condition and the ordering of trials belonging to the different conditions was randomized within and across subjects.

2.2.4- Data Analysis

Matching responses for experimental trials were scored for each subject using two different schemes. In the first scheme, subjects' matching performance was scored according to the vowel content of words based on criteria used in previous studies. Hence, a correct score was defined as matching the curved image with the word containing vowels labeled in previous studies as rounded *(/ah/, /oh/, /oo/)*, or matching the jagged image with the word containing vowels labeled in previous studies as un-rounded *(/uh/, /ay/, /ee/)*. In the second scheme, subjects' matching performance was scored according to the consonant content of words as per my alternative hypothesis emphasizing the acoustic quality of consonants. Here, a correct score was defined as matching the curved image with the word containing sonorant consonants *(/l/, /m/, /n/)* or matching the jagged image with the word containing plosive consonants *(/t/, /k/, /p/)*. Because some previous studies were not focused on differences in consonant quality, some of the word pairs I used that replicated past word pairings were complicated in that they contained only

plosive consonants. To address this issue, I categorized the different words within such pairs according to differences in their relative sonority using established differences in the voice-onset-time and spectral density of their consonants (see Table 2.1). Subjects' correct matching scores were averaged across trials within each condition, and individual subject averages were tested for their deviation from chance (50% correct) using one-sample, t-tests.

2.3- Results

In Condition 1, subjects' responses to the original stimuli used by Köhler (1947) and by Maurer et al. (2006) were scored based on both the vowel and consonant matching schemes and the original findings of these authors were replicated (Figure 2.3). Participants chose correctly 82% of the time when trials were scored using the vowel matching scheme (t_{23}=8.35, P<0.01). Participants also chose correctly 82% of the time when trials were scored using the consonant matching scheme (t_{23}=8.35, P<0.01). Note that this pattern of outcomes, in which subjects could show correct matching of original stimuli according to both the vowel content of words and the consonant content of words, is possible because the word stimuli used in previous studies involved inadvertent consonant-vowel associations. That is, vowels that were labeled in previous studies as rounded appeared in words containing sonorant consonants and vowels previously labeled as non-rounded appeared in words containing plosive consonants. This confound is actually apparent in Table 2.1 which shows the vowel- and consonant-type classifications of the original word stimuli and their predicted jagged or curved object-type associations. It is apparent that, in Condition 1, the rounded and unrounded vowel-type categories and

the plosive and sonorant consonant-type categories are perfectly confounded and thus make the same object-type association predictions.

In Condition 2, this confound is broken because the consonant content of words was swapped within pairs. Hence, the vowel and consonant content of the words now make different predictions about object-type associations. In this condition, correct scores based on consonant content remained significant at 79% (t_{23}=5.97, P<0.01), but correct scores based on the vowel matching scheme dropped to 21% (t_{23}=-5.97, P<0.01) confirming that subjects were tracking consonant content.

Condition 3 used new and randomly generated word and image materials and was a test specifically of the possible influence of different consonant types on subjects' matching choices. Correct scores based on the vowel scheme could not be calculated then because my word generation method controlled for and balanced vowel content across word types which therefore did not preserve the rounded-unrounded vowel distinction used in previous studies. In this condition, according to the distinction between plosive and sonorant consonants, participants chose correctly 81% of the time (t_{23}=7.71, P<0.01; Figure 2.4).

Finally, when the data for Conditions 1 and 2 were collapsed and analyzed together, the consonant content of word stimuli significantly predicted participant choices overall (80%: t_{23}=7.57, P<0.01), while performance based on vowel content was not different from chance (51%; t_{23}=0.85, P=0.20; Figure 2.4).

2.4- Discussion

Results of Experiment 1 replicated the word form/object shape matching patterns reported in previous studies. In Condition 1, using the same methods and materials of previous studies, participants showed the same matching bias reported in those studies. For example, my subjects consistently matched *baamoo* with the curved image and *kuhtay* with the jagged image, as in previous work. However, in Condition 2, when the consonant content of words was simply swapped within pairs, subjects' changed their selections. Subjects still showed a matching bias but it followed the consonant rather than vowel content of the words. For example, the word *kaatoo* was now matched to the jagged image and *buhmay* was matched to the curved image. This consonant-based matching bias was confirmed in Condition 3 using randomly generated words and images. In this condition, subjects matched words containing plosive consonants to jagged images and they matched words containing sonorant consonants to curved images even though the differences between curved and jagged image forms in this condition were extremely subtle.

It is important to emphasize that my results are unlikely to be attributable to orthographic confounds. My selection of vowels specifically omitted /o/ and /i/ that have obvious visual associative value with curved and jagged images, respectively, and used instead only /a/, /e/ and /u/ in their lowercase forms which are all relatively curved in visual appearance. Furthermore, by using consonants in sonorant and plosive categories that had both jagged and curved visual form polymorphisms, depending on their

capitalization, I reduced the possibility of any systematic orthographic biases in the visual form of consonants.

At the same time, there are some potential limitations in my experiment and in the inferences that can be drawn from it. First, my experiment followed previous studies in using a procedure in which both types of images and both categories of words were presented simultaneously for subjects to match. This design is intuitive but risks making the matching task relatively transparent to subjects by allowing them to directly compare both types of images and both categories of words on each trial.

In addition, the simultaneous presentation procedure makes it impossible to discover whether the observed plosive-jagged and sonorant-curved matching outcomes truly reflect two biases (and choices) or really only one, because, on every trial, the subjects second "choice" is dictated automatically by which image and word they first match. In other words, with the simultaneous presentation design used, a one-side matching bias would yield the same evidence as a two-sided matching bias. This ambiguity does not mean that there is, in fact, no underlying matching bias, only that I cannot be sure on the basis of my results whether and to what degree it involves both a plosive-word/jagged-image association and a sonorant-word/curved-image association as opposed to only one of these. The very same problem plagues previous studies that have used this experimental design.

Finally, although my results highlighting the potential role of consonants in subjects' matching preferences were unlikely attributable to orthographic effects, I cannot definitively conclude that it was necessarily the acoustic quality of the different

consonant types that mediated subjects' performance as hypothesized because, strictly speaking, the words were presented only visually in this experiment and not aurally. It is possible that when subjects saw the words on-screen, they tended to pronounce the words to themselves softly or subvocally, but I cannot be sure of this.

As a result I undertook a second experiment to address these issues.

EXPERIMENT 2: AUDITORY TEST OF THE ROLE OF CONSONANT QUALITY IN THE TAKETE-MALUMA PHENOMENON

Experiment 2 paralleled the design of Experiment 1 but involved two methodological changes. One change involved moving to sequential rather than simultaneous presentation of image materials. Thus, on each trial subjects saw only a single object image and subjects were given a choice of two words to match to that object (i.e., a two-alternative forced choice design). This change eliminated the opportunity for direct comparison between curved and jagged images on each trial and made the matching task less transparent. The second change involved shifting the word-dimension of the matching task into the auditory domain where the hypothesized effects arising from the acoustic qualities of different consonants could be tested more definitively. That is, words were presented aurally on each trial rather than visually in text form.

2.5- METHODS

2.5.1- Participants

Participants were 88 undergraduate students (75 Female, 13 Male) who were enrolled in introductory psychology courses at the University of Lethbridge and received partial course credit for their participation.

2.5.2- Image and Word Stimuli

Experimental stimuli once again derived from previous studies by Köhler (1947) and by Maurer et al. (2006). In this experiment, I also included stimuli used in a study by Westbury (2005). In total, 20 images (10 curved, 10 jagged) and 42 pairs of words from these previous studies were used as stimuli in the present experiment. Of the original word pairs, 21 preserved the original vowel and consonant content and 21 involved pairs in which the consonant content of words in a pair had been swapped as in Condition 2 of Experiment 1. Once again, I supplemented this sample of materials with a set of random images and words that I generated according to procedures described for Experiment 1. In total, I used 22 images (11 curved, 11 jagged) from the random image generation technique and 42 pairs of words from the random word generator. In generating words, I again used the stop-consonants /t/, /k/, and /p/ as plosive consonants and the continuants /l/, /m/, and /n/ as sonorant consonants. I used only the vowels /a/, /e/, and /u/ and retained the restriction that the vowel /e/ could not appear in word-final position.

To facilitate aural presentation of word material in this experiment and standardize the pronunciation of all words, I used a commercial text-to-speech program

(Swifttalker©) to generate spoken versions of each word from their associated text forms. Spoken versions of each word were generated in each of two realistic-sounding synthetic voices, one male (David) and the other female (Diane). These synthetic voices were developed using the parameters and standard phoneme pronunciations for American English without any strong regional accent. Hence, the pronunciation of individual word tokens did not vary between David and Diane. For word stimuli derived from previous studies, the original pairing of words was left intact (e.g. *Bouba-Kiki*). For new word stimuli, word pairs were constructed to control and balance vowel and consonant content.

2.5.3- Experimental Design

The experiment was conducted on computer via a graphical interface created using Runtime Revolution, Version 2.9. On each trial, the participant was shown a single image, and two words were played sequentially via headphones. Participants were then asked to choose which word best matched the image using a button labeled either *<word 1>* or *<word 2>*. Participants could replay the words by clicking a button on-screen.

The experiment involved eighty-four of these forced-choice trials: each of the 42 images was presented twice, each time coupled with a different word pair one of which involved words derived from previous studies and one of which involved my new randomly generated words. The coupling of particular word pairs and images was randomized. Which word in a pair was played first and the sex of the synthesized speaker were counterbalanced across trials.

2.5.4- Data Analysis

To preserve the comparability of the results of the two experiments as much as possible, the analysis of data for Experiment 2 paralleled that for Experiment 1. Thus, subjects' matching responses in experimental trials were once again scored based on both the vowel content and the consonant content of words as per Experiment 1. Subjects' correct matching scores according to these two schemes were averaged across trials within each condition, and individual subject averages were tested for their deviation from chance (50% correct) using one-sample, t-tests.

2.6-Results

Condition 1 of this experiment once again replicated the matching patterns reported in previous studies and in Condition 1 of my own first experiment (Figure 2.5). Using the original word and image materials, participants chose correctly 60% of the time based on the vowel matching scheme (t_{87}=6.24, P<0.01). Participants also necessarily chose correctly 60% of the time using the consonant matching scheme (t_{87}=6.24, P<0.01) because the vowel and consonant content of the original word pairs used in this condition were perfectly confounded.

In Condition 2, in which the consonant content of the word stimuli used in previous studies was simply swapped, subjects' performance based on consonant content once again remained significant at 58% (t_{87}=4.72, P<0.01), but their performance based on the vowel matching scheme dropped below chance 42% (t_{87}=-4.72, P<0.01, Figure 2.5).

In Condition 3, using new and randomly generated word and image materials, participants chose correctly 59% of the time based on consonant content ($t_{87}=8.05$, $P<0.01$; Figure 2.6). Finally, when the data for Conditions 1 and 2 were collapsed and analyzed together, the consonant content of word stimuli significantly predicted participant choices overall (59%: $t_{87}=6.95$, $P<0.01$), while vowel content did not (51%; $t_{87}=1.28$, $P=0.10$; Figure 2.6).

Overall, subjects' matching performance based on the consonant content of words was higher for jagged images (63% correct) than for curved images (54% correct: $t_{86}=4.66$, $P<0.01$). Matching performance based on consonant content was also better for the more stylized and differentiated jagged and curved images used in previous studies (63% correct) than for the randomly generated jagged and curved images that I developed and that were only subtly different from each other (55% correct: $t_{86}=4.89$, $P<0.01$).

2.7- Discussion

Results of Experiment 2 corroborated those of Experiment 1 in showing a bias to match unfamiliar jagged and curved images with nonsense words based on differences in their consonant content. In both experiments, words containing plosive consonants were preferentially matched with jagged images, and words containing sonorant consonants were matched with curved images. Results of Experiment 2, which involved only aural presentation of word stimuli, suggest that these consonant-based effects were mediated by differences in the acoustic quality of plosive and sonorant consonants as hypothesized.

Results of Experiment 2 also tended to confirm some methodological concerns about previous studies and my own Experiment 1. These revolved around using

simultaneous presentation of both types of words and images on each trial which allows subjects to make direct comparisons among them. This approach raises questions about the relative transparency of the overall task and also about the possibly one-sided nature of the matching bias needed to account for previous findings. I attempted to address both issues in Experiment 2 by moving to a procedure that involved sequential presentation of experimental stimuli. Consistent with this change in experimental procedure, subjects' correct matching performance also dropped substantially to an average of 58% in Experiment 2 from nearly 80% in Experiment 1. I attribute this drop in performance to the change in experimental procedure which made the task more difficult (less transparent), but it is possible that it also reflects other differences between the two experiments related to the processing of word materials presented either visually or aurally.

At the same time, subjects' responses in Experiment 2 were found to involve an asymmetry in matching bias: the association of plosive words with jagged images was stronger than the association of sonorant words with curved images.

The latter matching asymmetry pertained only to trials involving my newly generated images, however, and reflected low correct matching scores specifically for the curved images. In re-examining images, it is clear that, in general, my curved images are less curved and more jagged than are the stylized types of curved images used in previous studies. This results from the fact that, in my image generation technique, the curved and jagged forms of images were generated from exactly the same set of randomly generated calculus points. This precaution was undertaken purposefully as an attempt to reduce

concerns about inadvertent experimenter biases creeping into the choice of image materials used previously. However, in some cases, my "unbiased" figure generation technique necessarily yielded intersecting lines that resulted in acute interior angles in the final form of the curved version of each image (see Figure 2.1 for examples). As a consequence, and in the context of the larger image set they were exposed to, subjects might have viewed many of my curved images as relatively jagged.

If true, the overall asymmetry in correct matching of jagged and curved images that I report might not reflect a real asymmetry in subjects' word-form/object-shape matching biases so much as an additional image-related methodological factor to be cognizant of in future work as discussed in more detail below.

2.8- General Discussion

I report two experiments testing the classic *Takete-Maluma* phenomenon. In both experiments, subjects preferentially matched nonsense words to unfamiliar jagged and curved images based on the consonant content of words. Importantly, the effects held for classic image and word material used originally by Köhler and by other investigators since, as well as for a new set of randomly-generated images and words created specifically to circumvent concerns about previous biases in the selection of stimulus materials. The latter outcomes with controlled and randomized words and randomized images in which jagged and curved image forms were only subtly different from each other helps to buttress the conclusion that the word-form/object-shape matching bias is a *bona fide* effect, which has often been questioned. They also suggest that the observed matching biases are not leashed to a specific set of experimenter-selected images, nor do they

depend on dramatic differences in image curviness or jaggedness, which has also been suspected. At the same time, the results highlight that the details of object form are certainly important. For example, I found that subjects' correct performance with my curved images was not as strong as it was with my jagged images, likely because the mathematics of my random figure generation technique necessarily produced curved images containing some acute angles that gave them an element of jaggedness. This finding opens the door to additional future work testing how subjects' matching biases are sensitive to such specific details of visual object form. In this respect, the novel figure generation techniques introduced here offer considerable promise because they allow a virtually limitless number, range, and variety of controlled images to be generated for future testing.

Overall, my results are consistent with some previous findings highlighting the role of consonants in the *Takete-Maluma* phenomenon (e.g. Westbury 2005). They also revealed that some of the results obtained in earlier studies that were previously attributed to the vowel content of words reflected a non-arbitrary coupling of consonants and vowels within words. When this inadvertent consonant-vowel association was broken in my experiments, I found that subjects' matching choices tracked the consonant content of words more than they did the vowels. This consonant-based effect was demonstrated to hold whether words were presented visually or aurally. This pattern of findings is consistent with my proposal that the consonant-based effect is likely mediated by differences in the auditory perceptual quality that different consonants have owing to variation in their pattern of spectral density and attack which make them either relatively harsh sounding (plosive) or relatively smooth sounding (sonorant). And, as

outlined in the introduction, these effects may be continuous with vocal-affective relationships observed more widely in the communication systems of many animal species which involve consistently differentiated use of harsh, plosive vocalizations versus smooth, sonorant vocalizations in social behavioural contexts that portend different affective consequences for listeners (reviewed in Owren & Rendall, 1997, 2001).

On the surface of it, this consonant-based account of subjects' matching biases appears to weaken the proposal that previous effects were due to the vowel content of words and to synesthesia-like linkages between the visual percepts of object shape and motor (or somatosensory) activity associated with articulation of different vowel sounds. However, the two accounts need not be mutually exclusive. Thus, although consonants appeared to trump vowels in influencing subjects' performance in my experiments overall, it is likely that the vowel content of words accounted for some of the variation in subjects' responses. In addition, my experiments used only a limited selection of consonants. It is therefore possible that, in the context of different consonants, vowels might play a greater role in matching biases (e.g. Maurer et al., 2006; Ramachandran & Hubbard, 2001; Tarte & Barritt, 1971; Tarte, 1974) and that, when they do, their effects might be mediated by cross-modal linkages of exactly the sort proposed. It is also possible that the relative influence of vowels and consonants might vary depending on whether word stimuli are presented either visually or aurally, or maybe more pertinently on whether they are processed either visually or aurally by subjects no matter how they are canonically presented to subjects.

Furthermore, although the consonant account that I emphasize does not require a synaesthetic linkage between visual and somatosensory systems *per se*, it does nevertheless point to some type of communication across, or integration among, sensory domains in as much the auditory quality of consonant sounds is proposed to interact with the visual system to produce regularities in the matching of auditory and visual percepts.

It is important also to note that the consonant-based effects observed here do not mean necessarily that plosive consonants actually conjure specific semantic constructs like '*harsh, jagged, fractured*', or that sonorant consonants conjure the opposite semantic constructs ('*smooth, connected, rounded*'). Indeed, plosive and sonorant consonants might have no very specific semantic associations in any absolute sense. Instead, the consonant effects might be rooted in processes that are very low-level, peripheral and pre-semantic (Westbury, 2005).

Nevertheless, whatever their proximate reality, the effects involved could yield naïve expectations about language structure at some level, and these are intriguing to consider for their, as yet, relatively unstudied influences on language processing and learning (e.g., Maurer et al., 2006). For example, an increasing body of work confirms the importance of object shape in childrens' early word learning (reviewed in Samuelson & Smith, 2005). Although not yet systematically studied, it is possible that this process could be facilitated further, or, alternatively, that it could be hampered to some degree, depending on the extent to which the form of real object words is either commensurate with, or contradicts, any such naïve expectations about language structure.

Table 2.1. Word Stimuli for Experiment 2.1

Letters in brackets below each word indicate the word's vowel-type classification from previous studies, their consonant-type classification from the current study, and the image type associations that are predicted to follow from these classifications. R=Rounded vowel; U=Unrounded vowel; St =plosive consonant; So= Sonorant consonant; J=jagged image; C=Curved image. Note that, in Condition 1, the vowel and consonant classifications are fully confounded and thus make exactly the same predictions about the image-type associations for the words. In Condition 2, this confound is broken by swapping the consonant content of words within pairs and the predicted image-type associations for the words are now different based on their vowel and consonant content. In Condition 3, the confound is eliminated by using randomly generated word stimuli.

Condition 1 (Word pairs from Köhler and Maurer et al.)	Condition 2 (Consonant-swapped word pairs from Condition 1)	Condition 3 (Randomly generated word pairs)	
Kuhtay (U:J) (St:J) Baamoo (R:C) (So:C)	Buhmay (U:J) (So:C) Kaatoo (R:C) (St:J)	Paka (St:J) Lana (So:C)	Peka (St:J) Mela (So:C)
Tuhkeetee (U:J) (St:J) Maabooma a (R:C) (So:C)	Muhbeemee (U:J) (So:C) Taakootaa (R:C) (St:J)	Kepa (St:J) Muna (So:C)	Tatu (St:J) Lalu (So:C)
Kaykee (U:J) (St:J) Boobaa (R:C) (So-C)	Baybee (U:J) (So:C) Kookaa (R:C) (St:J)	Tuka (St:J) Nema (So:C)	Keka (St:J) Lulu (So:C)
Teetay (U:J) (St:J) Gogaa (R:C) (So:C)	Geegay (U:J) (So:C) Totaa (R:C) (St:J)	Kapa (St:J) Munu (So:C)	Kuta (St:J) Luna (So:C)
Takete (U:J) (St:J) Maluma (R:C) (So:C)	Maleme (U:J) (So:C) Takuta (R:C) (St:C)	Tapa (St:J) Namu (So:C)	Petu (St:J) Mana (So:C)

73

Curved	Jagged	Curved	Jagged
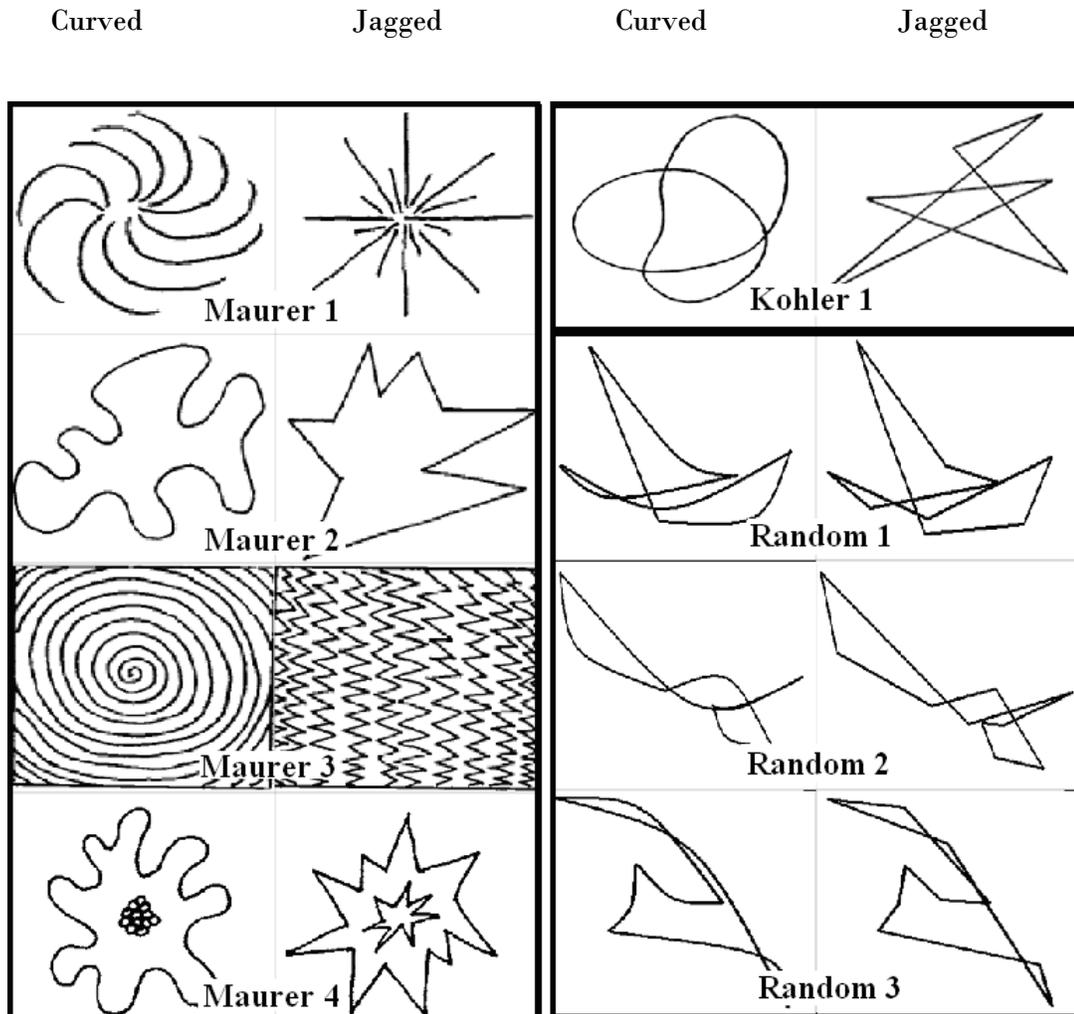
Figure 2.1. Image stimuli used in Experiment 2.1.

Side-by-side comparisons of curved and jagged image pairs from previous studies by

Maurer *et al.* (2006) and Köhler (1947) that were used in Conditions 1 and 2, and a sample

of the new, randomly generated curved and jagged images used in Condition 3. Note that

the latter curved and jagged images are very subtly different, varying only in the

curvature of the lines linking calculus points.
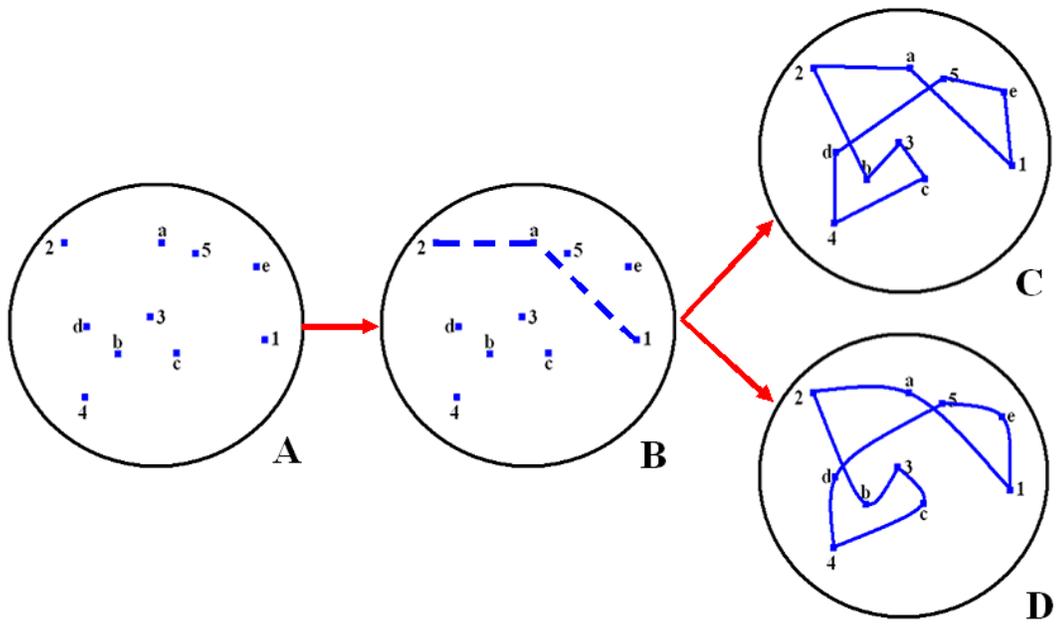
Figure 2.2. Schematic illustrating the process used to create curved and jagged versions of

the same image based on a set of randomly generated calculus points.
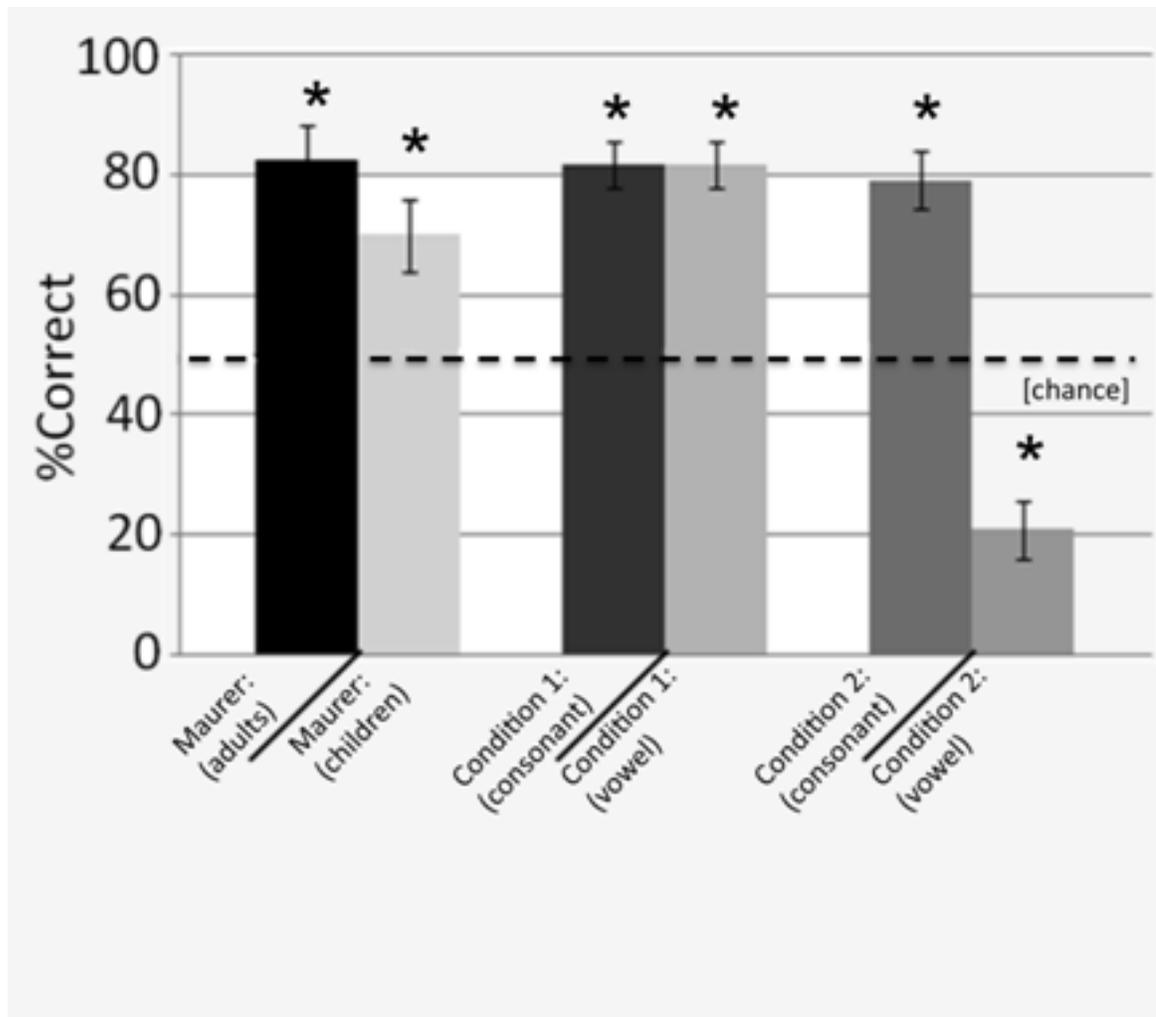
Figure 2.3. Mean (+-SEM) correct matching performance by adults and children in Maurer *et al.* (2006) compared to subjects' correct matching performance in Conditions 1 and 2 of Experiment 1 when scored by either the vowel or the consonant content of words. Effects shown are significantly different from chance, 50% (*P*<0.01).
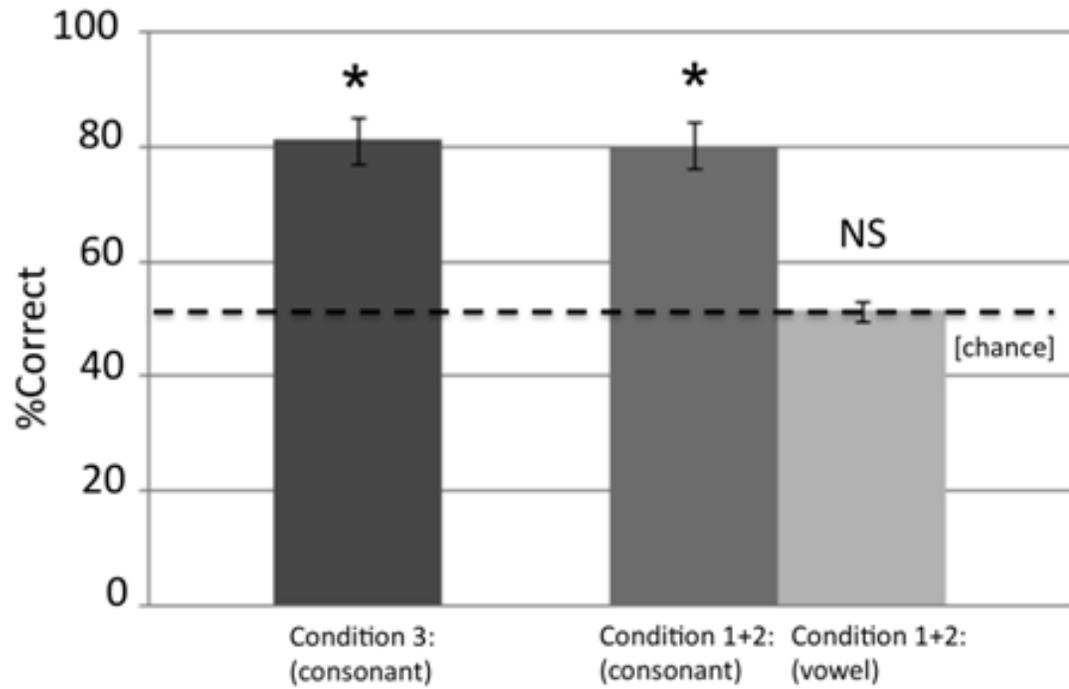
Figure 2.4. Mean (+-SEM) correct matching performance in Experiment 1 in Condition 3 and in Conditions 1 and 2 combined when scored by either the vowel or the consonant content of words. Effects shown are significantly different from chance ($P<0.01$) except where noted (NS).
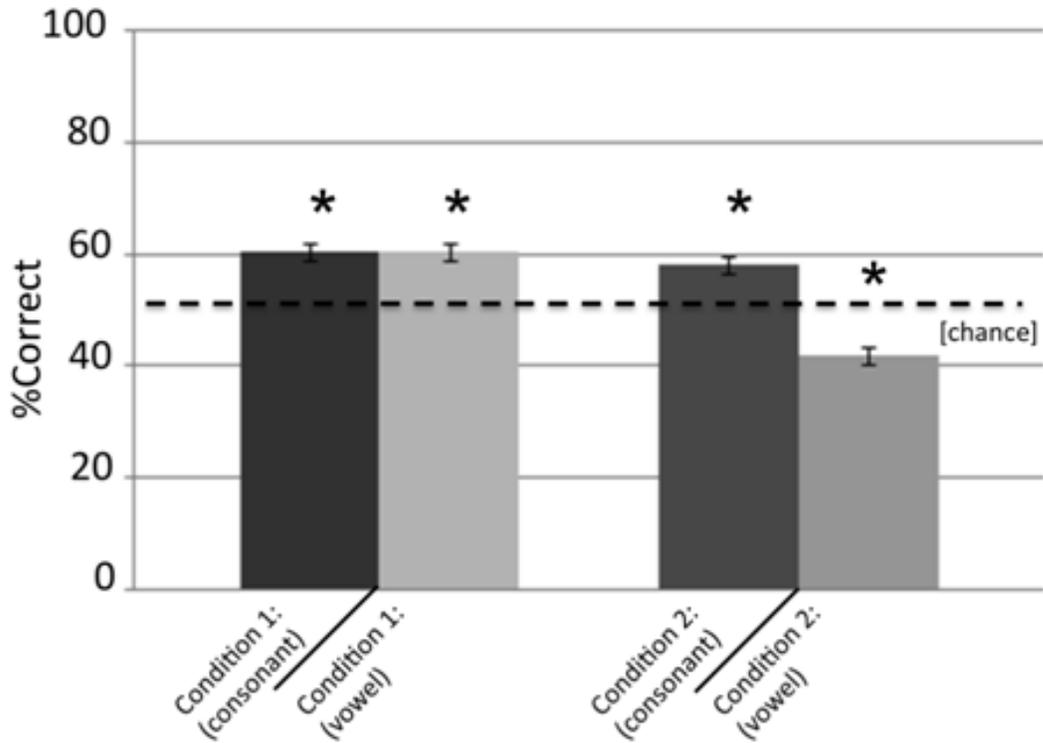
Figure 2.5. Mean (+-SEM) correct matching performance Conditions 1 and 2 of Experiment 2 when scored by either the vowel or the consonant content of words. Effects shown are significantly different from chance, 50% (*P*<0.01).
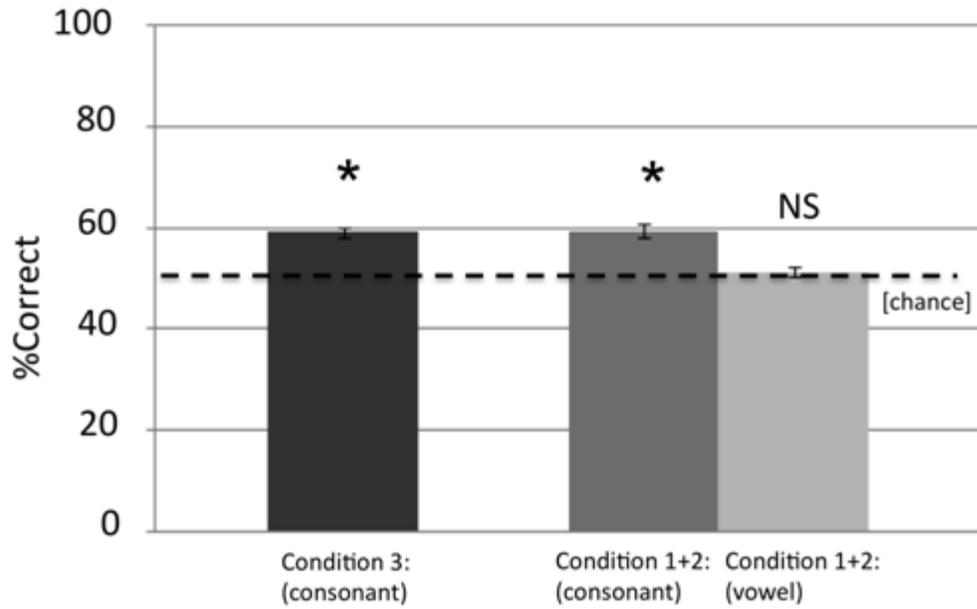
Figure 2.6. Mean (+-SEM) correct matching performance in Experiment 2 in Condition 3 and in Conditions 1 and 2 combined when scored by either the vowel or the consonant content of words. Effects shown are significantly different from chance ($P<0.01$) except where noted (NS).

CHAPTER THREE

THE BOUBA-KIKI EFFECT:

THE DIFFERENTIAL AND COMBINATORIAL EFFECTS OF CONSONANTS AND VOWELS

3.1- Introduction

The central dogma of linguistics concerning connections between word form and meaning for the majority of the 20th century has been that no such connections exist. Following de Sassure (1983), the majority of linguists adhered to the dictum that "the symbol is arbitrary", according to which the relationships between words and objects or events which they denote are established only by social convention. Counterexamples that oppose the doctrine of arbitrariness have periodically been raised throughout the 20th century, labeled collectively under the banner of sound-symbolism to highlight that the acoustical structure of words might bear a non-arbitrary connection to their meanings (reviewed in Hinton et al., 1994; Nielsen & Rendall, 2001). The Bouba-Kiki effect, where nonsense words and unfamiliar curved and jagged objects are preferentially paired with one another based on regularities in the phonemic structuring of words is one such example. In short, participants match words containing sonorant and mellifluous sounds with curved and rounded images and words containing harsh and plosive phonemes with jagged images. Research into this effect has yielded consistent results, but due to problems in experimental design, these results were often difficult to interpret. One major problem in interpreting research into the Bouba-Kiki effect has been stimulus construction, which hasn't been controlled in a way to systematically establish what word

characteristics preferentially drive the sound-symbolic associations. The primary purpose of the two experiments presented here was to address this problem and in doing so clearly delineate the differential impact of vowels and consonants on producing sound-symbolic correspondences.

Among the best known examples of sound symbolism is magnitude symbolism, where real and non-sense words containing high front vowels (e.g. /i/ are judged as more natural for denoting small size (e.g. *tiny, wee*) while words containing low, back vowels (e.g. /ah/ and /uh/ are judged more natural for denoting large size (e.g. *chunk, huge, gargantuan*) both in English (Johnson 1967; Sapir 1929) and a variety of other languages (e.g. Huang 1969). Not only are words denoting size non-arbitrarily structured, but within word classes smaller members of a category are typically named with high front vowels and larger members with low back vowels. Berlin (1994) for example found pervasive magnitude symbolism in ethno-zoological literature; small fish and birds were typically named with high front vowels and larger fish and birds with low back vowels.

Experimental demonstrations of sound symbolism have revealed biases which are both large and broadly consistent across languages. The consistency with which participants associate sounds with object characteristics thus suggests the possibility that perceptual-cognitive biases exist which yield naïve expectations about language structure. Regularities in both speech production and perception are well known; for example, larger individuals generally have larger sound production apparatuses that produce not only lower frequencies and formant, but that are also more capable of producing pitch distortions that are perceived as spectral harshness. This basic

relationship between size and harmonics holds across a number of production modalities and applies directly to the larynx and vocal tract cavities of human, which yields lower frequency voices (both in pitch and formants) in adults compared to children and men compared to women (Fitch, 1995; Ghazanfar & Rendall, 2008; Morton, 1994; Rendall et al., 2005). Given these regularities, it is curious that the idea that humans are able to pick up on these contingencies and use them productively is rarely mooted.

Sound symbolic relationships discovered in animal communication have been a rich source of input into sound symbolism in communication systems more generally, contributing to a resurgence of interest in the systematicity of symbols based on their acoustic characteristics. The Bouba-Kiki phenomenon, first studied by Wolfgang Köhler (1947) has been one focus, with broad steps being taken towards shoring up concerns both about experimental methodology employed by previous researchers, in elucidating mechanistic explanation of the neurological instantiation of sound symbolism, and finally in providing functional accounts of what value sound symbolic relationships might have for language. The stereotypical Bouba-Kiki experiments (Davis, 1961; Köhler, 1947; Maurer, Mondloch, & Pathman, 2006) have involved the presentation of two side-by-side line drawings: one with a jagged angular shape and one with a curvier more cloud-like shape (see Fig 3.1) along with two contrasting nonsense words (e.g. '*takete*' and '*maluma*'- see Table 3.1).

Using this paradigm, researchers have found that words like '*takete*' are consistently associated with jagged figures and *'maluma'* with curvier figures, although the interpretation of which characteristics of nonsense words are responsible for this

association has been contentious. Many researchers have focused on the vowel characteristics of words for explaining the Bouba-Kiki effect, pointing to the round mouth articulation required to produce the vowel sounds of /ah/ and /oo/ in 'maluma' compared to the less rounded vowels of /eh/ or /i/ in 'takete'. Ramachandran & Hubbard (2001) have provided the most comprehensive explanation of this phenomenon, proposing that object-shape:word-form biases are mediated by synesthesia-like co-activation of the motor or somatosensory areas involved in articulating different sounds along with the visual areas associated with the perception of differently shaped objects. Thus, curved object percepts and the motor representation of round mouth articulation might be linked to each other in non-arbitrary ways. Maurer et al., (2006) extended this proposal to experiments with both adults and toddlers, demonstrating that both show matching responses in the direction predicted by Ramchandran & Hubbard, (2001)'s account of cross-modal connection. Methodologically, Maurer et al., (2006) took an important step towards improving research- moving from the single stimulus set of many early experiments to four sets of paired words and images, attempting to ensure that previously observed effects were not stimulus bound.

Both the arguments of Ramachandran & Hubbard (2001) and the experimental findings of Maurer et al., (2006) are compelling; so too however is the possibility that the consonant content of words might play a role in the Bouba-Kiki effect. There are, for example, very clear differences between /t/ and /k/ and /m/ and /l/ which make the first two relatively harsh and the second two relatively mellifluous. Westbury (2005) provided the first direct test of a consonant driven explanation of the Bouba-Kiki effect, demonstrating that participants were slower to classify letter strings as words or non-

words depending on the congruency of the letter strings and the frame (either curved or spiky- see Figure 3.1) in which they were presented. Westbury explained these connections based on a matching between the chaotic structure of jagged images and the spectral chaos of plosive letters.

The implicit nature of Westbury (2005)'s experiment was both its greatest strength, because it demonstrated that the effect was not constrained by demand characteristics of the experiment which were transparent to the participants, but also a weakness, in that it made directly comparing Westbury's findings to those of Köhler (1947) and others difficult; rendering the possibility of reconciling how one protocol can find vowel based effects while another finds consonant based effects difficult. In response to this, Nielsen & Rendall (2011) conducted a pair of experiments designed to more closely examine not only the findings of previous researchers, but also to extend those finding to a much larger corpus of both word and image stimuli. In the first condition of their experiment Nielsen & Rendall (2011) used the original stimuli of both Köhler (1947) and Maurer et al., (2006) and found the same results as previous researchers- participants tracked vowel content of nonsense words in labeling the stylized figures present in those experiments. Given the predictions by the previous authors with regards to vowel: image associations it was impossible to determine on the basis of the results of the first condition whether participants were actually tracking vowel content (e.g. /u/ with round shapes) or consonant content (e.g. /m/ with round shapes). Because the consonant and vowel content of the original word pairs was perfectly confounded, both theories yielded identical predictions in all possible trials. Given the confound present between consonant and vowel characteristics, a second subset of trials was required to break this confound (see

Nielsen & Rendall, 2011 for details). In this condition, Nielsen & Rendall (2011) found reliable matching of words to figures based on their consonant content, but the observed vowel-driven bias was completely inverted. Taken collectively, these results were demonstrative of the power of consonant-content for driving sound symbolic associations of word-form: object shape.

In their second experiment, in which single images (rather than pairs) were presented with pairs of nonsense words to make experimental manipulations less transparent, Nielsen & Rendall (2011) once again found a reliable association between the consonant classes of words and the images with which they were associated; although the concordance rates were much lower than in previous experiments. Nielsen & Rendall (2011) interpreted this drop in performance as being due to the removal of transparency, but it is also possible that by removing systematic variation in vowel structure within word pairs that they diluted whatever residual effect vowel configuration might have on participants' matching biases.

Westbury (2005) was the first to explicitly point out that the demarcation of consonants and vowels as being absolutely responsible for the Bouba-Kiki effect is likely a false dichotomy- real words (at least in most languages) contain variation in both their consonant and vowel structure, and if anything the perfect entanglement of the vowel and consonant content of classical stimuli is suggestive of an interaction of vowels and consonants in invoking sound-symbolic associations (which may explain why previous researchers seemed unable to recognize the systematic variation of their stimuli in consonant content). Two experiments are presented here to test this hypothesis. In the

first experiment, the simultaneous presentation design of classical Bouba-Kiki experiments was replicated in an attempt to demonstrate that when consonants are invariant participants are indeed able to track variation in vowel content in the direction established by Maurer et al. (2006) and other researchers. In the second experiment a less transparent protocol wherein participants were asked to construct words for curved and jagged figures from a set of syllables which varied systematically in their phonemic content was used. Collectively, these two experiments seek to demonstrate not only that vowels are a valid target for sound-symbolic research, but also that the interactions of vowels and consonants are potentially complex, and complicated even further by their interface with construction of words which embody both sound-symbolic biases and constraints from productive and perceptual modalities.

EXPERIMENT 1: CONSONANTS AND VOWELS

3.2- METHODS

3.2.1- Participants

Participants were 22 undergraduate students (14 female, 8 male) between the ages of 17 and 24 who were enrolled in introductory psychology courses at the University of Lethbridge and received course credit for their participation in the experiment.

3.2.2- Word and Image Stimuli

The image stimuli used in this experiment were a combination of those used in previous studies by Köhler (1 pair of images) and Maurer et al. (4 pairs of images), and newly generated images (27 pairs). The stimuli originally used by Maurer et al. involved

mix media presentation that allowed for haptic interaction with the objects, but for standardization and ease of administration I used the typical two-dimensional graphical representations of these objects provided by the authors. The newly created images I used relied on a radially constrained mathematical formula for their generation. This procedure populated a field of a finite size (See Figure 2-1) with either five or ten sequentially ordered, randomly generated calculus points. The location of these points was determined by creating vectors which departed from the center of the field at random magnitudes and directions (constrained by the overall radius of the field). These points were sequentially numbered and subsequently a second vector was generated from the midpoint of the line connecting sequential points, yielding a calculus point for rendering simple curves. Individual line segments were defined by their two original points and the generated curvature point (Figure 2-2). These point fields were then joined with either curved (Figure 2-3) or straight (Figure 2-4) lines. The resulting images were identical to one another except for the curvature of their lines, which produced only subtle differences between the two images in a set (Figure 1). The potential benefits of this figure generation method are twofold: it allows for the generation of an effectively infinite number of image sets which are not constrained or produced by experimenter induced biases, increasing the ease with which Bouba-Kiki experiments which do not rely on a small number of trials or stimuli can be conducted; and it also standardizes image sizes, ensuring that not only are all images size constrained, but also that individual images within pairs are roughly equal in size (measured either as edge length or surface area) to one another.

Word stimuli were created specifically for this experiment using /t/, /k/, and /p/ as my plosive consonants and /m/, /n/, and /l/ as my sonorant consonants, as was done in my

previous research into the Bouba-Kiki effect. In keeping with the predictions of Maurer et al. and others, I used the phonemes */oo/* and */oh/* as my rounded vowel sounds and */ee/* and */ay/* as my non-rounded vowels sounds. All vowels were presented in the roughly phonetic form they are presented here to ensure that differences in assumed pronunciation were minimized. All words were constrained to two syllables in alternating consonant-vowel order, yielding six letter words (e.g. *Looloh* vs. *Laylee*). Across participants the entire set of the possible 144 word permutations allowed by these consonants and vowel phonemes were used, but within individual subjects the specific word permutations were sampled randomly.

3.2.3- Experimental Design

Experiment 1 was conducted on computer via a graphical interface created using Runtime Revolution v. 2.8.1. On each trial, participants were shown two images side-by-side (one curved, one jagged) from the same data set (i.e. the images were nearly identical in all but the curvature of the lines joining their calculus points) with two words below. Participants were asked to label the images with the words. As in previous studies, labeling one image with a given word automatically led to the assignment of the remaining word to the other image. Participants were free to change their labeling before advancing to the next trial by clicking a button on screen. The paired words, analogously to the images, differed only on a single dimension from each other- their vowel content, with one word in each pair being made up of non-rounded vowels (henceforth jagged-associated words) and one word being made up of rounded vowels (henceforth curve-associated words). The consonant construction of words, though invariant within trials,

covered all of the general consonant configurations (plosive-plosive, plosive-sonorant, sonorant-plosive, sonorant-sonorant) and sampled randomly from the larger corpus of 144 possible word configurations. Each participant completed thirty-two of these forced-choice trials, with the coupling of word and image pairs, and their order of presentation (i.e. whether the curved image was presented on the right or left side, and which vowel configuration was presented above the other) randomized both within and across subjects.

3.2.4- Data Analysis

Matching responses for experimental trials were scored for each participant based on the vowel content of the word chosen for the images in each trial. Because the response to a single image necessarily produced the response to the second image (because it was impossible to choose the same word for both images), data were analyzed only by looking at which vowel configuration was chosen for the first image (i.e. the image on the left). Curve associated vowels *(/oo/* and */oh/)* were assigned a value of 0 and jagged associated vowels *(/ay/* and */ee/)* a value of 1 and response data (both based on vowel choice and response time) was then subjected to a repeated measures analysis of variance with image type (curved or straight) and consonant congruency as within (repeated) factors.

3.3- Results

Participants reliably responded to images in the predicted direction, matching curved images to words with curve associated vowels 72.8% of the time and jagged images to words with jagged-associated vowels 67.9% of the time ($F_{1,570}= 27.19$, P<0.05)

(Figure 3.3). There was no effect of consonant congruency ($F_{2,570}= 0.20$, P=0.82) on vowel choice.

Response time data showed no significant effect for either image type ($F_{1,570}= 0.41$, P=0.53) or consonant congruency ($F_{2,570}=0.37$, P=0.69) but a significant interaction of image type x consonant congruency ($F_{2,570}=3.23$, P<0,05). Participants also responded faster on trials where they answered correctly based on vowel content (4715 ms) than on trials where they responded incorrectly (5571 ms) ($F_{1,658}=16.26$, P<0.05). There was also a noticeable but unstable and non-significant effect towards consonant congruency slowing reaction times.

3.4- Discussion

Response data from experiment 1 were as predicted and paralleled the findings of previous work. In the absence of variation in consonant content, which was previously demonstrated as superordinate in driving previous effects, vowel characteristics of nonsense words were mapped onto images in reliable ways. The curve-associated or rounded vowels /*oh*/ and /*oo*/ were consistently matched with the curved line rounded images, while the jagged-associated or non-rounded vowels /*ee*/ and /*ay*/ were consistently matched with the straight line jagged images. For example, given the pair *Loolooh* and *Leelay* participants consistently matched the former with rounded images and the latter with jagged images. Despite this consistent mapping, reaction time data were suggestive of the possibility that the interaction of vowels and consonants at least influences processing speeds, if not final decisions made by participants. Although no a priori predictions regarding reaction times were made, if pressed it might have been suggested

90

that congruency of consonant class for a given word pair would have increased the time required to respond based on vowel content (insofar as congruency of consonants might distract participants from mapping vowels onto the images). This post hoc prediction was not supported by the data. Although participants were significantly faster on trials where they responded correctly than on trials where they responded incorrectly, consonant congruency had no effect on reaction time. Further, the interaction between consonant congruency and image type revealed that the majority of the interaction was accounted for by the mixed consonant word conditions (i.e., plosive-sonorant and sonorant-plosive words). For jagged images, participants consistently chose mixed consonant words more quickly than either purely plosive or purely sonorant words, while for rounded images the opposite was true. This finding is difficult to interpret directly, but one possibility is that mixed case words (which are more like the words of real English) have their own basins of attraction and biases which should be explored in further experiments. Additionally, as discussed elsewhere, the constraint imposed by the figure generation method used in this experiment produced curved images which were noticeably less "cloud-like" than those of many previous studies- despite the curvature of the lines used to connect their calculus points many still possessed jagged exterior angles and other traits associated with the spikier class of images.

The relationship between the results of this experiment with broader sound-symbolic relationships instantiated in real language raises two potential concerns. The first is the question of external validity and whether the artificial situation presented within this experiment is representative of anything resembling language. The decision to label an object with one of two nonsense words which vary only according to their vowel

content is undoubtedly one which most language users will never have to do. The limitations imposed by the experimental design on flexibility of response further constrains comparison to real language, where speakers performing any similar task would be able to sample from the entire set of phonemes they are capable of producing when creating new linguistic tokens. Second, presenting words to participants visually ignores the channel over which the majority of linguistic content is transmitted (at least historically): the speech signal. Given that the vowel based hypothesis is based at least partially on the shape of lips during articulation, and that the consonant-driven hypothesis focuses on sharp vocal inflection and chaotic spectral patterning, presentation through the visual domain provides at best the activation of associations one full step removed from the level at which they are likely instantiated. Coupled with the fact that visual presentation of graphemes induces possible orthographic confounds (see Koriat & Levy, 1978), an experiment which presents stimuli acoustically is valuable.

In Experiment 3.2, an experimental protocol which allows participants direct access to acoustic speech cues, but which also introduces a novel task which has more direct analogues with object naming is used. Although apart from astronomers and entomologists modern people rarely have the opportunity to name a novel object, objects must be named by someone before the process of cultural transmission (Kirby, Cornish, & Smith, 2008; Pagel, 2009) takes over. Further, although actual naming events are rare, in the real world people constantly choose between synonyms for the same word; a process which can presumably be guided by slight differences in connotation which may themselves be influenced by sound-symbolic biases sorting through acoustic space. In

Experiment 3.2, participants are tasked with creating novel nonsense words from a set of available syllables.

EXPERIMENT 2: WORD CONSTRUCTION EXPERIMENT

3.5- METHODS

3.5.1- Participants

Participants were 22 undergraduate students (14 female, 8 male) who were enrolled in introductory psychology courses at the University of Lethbridge and received course credit for their participation in the experiment.

3.5.2- Image and Syllable Stimuli

The image stimuli used in this experiment were a combination of those used in the above study with a subset of the images used by Westbury (2005) in his priming experiment and images created with a new image generation technique developed in my lab. This new image generation procedure was similar to the previous procedure in that it populated a field with randomly generated but radially constrained calculus points (Fig 3.5-A), but subsequently departed from the creation of figures via the joining of line segments. Instead, for curved images a single number was generated which represented the radius of a circle whose center point was the generated calculus point (Figure 3.5-B). For straight images the circle generated at each point (Figure 3.5-C) for curved images was used as the boundary for the generation of a triangle whose three angles were each generated to fall along the circumference of the circle (Figure 3.5-D). At each of the calculus points generated in step A either a circle or triangle was drawn, and then the

lines which fell inside of the shapes were erased, producing bounded two dimensional shapes which were either jagged (Figure 3.5-E or curved (Figure 3.5-F). Unlike the original figure generation method used in experiment 1, these figures were only equivalent to one another in the simplest sense that their initial calculus points were identical and the circles of the curved figures served as boundaries for generation of triangles for jagged images. This process produced curved images which were on average substantially larger than their jagged counterparts (at least in terms of surface area), and that were additionally much more similar to one another than the jagged images were to each other. Overall, there were 45 image pairs; 1 from Köhler, 4 from Maurer et al, 4 from Westbury, 18 "Old" figures (from the generation method detailed in Experiment 3.1), and 18 "New" figures (which used the generation method presented above).

Syllables for this experiment were created using /t/, /k/, and /p/ as the plosive consonants and /m/, /n/, and /l/ as the sonorant consonants, as in experiment 1 above. In keeping with the findings of Experiment 3.1, the rounded (curve-associated) vowels /oo/ and /oh/ and the non-rounded (jagged-associated) vowels /ee/ and /ay/ were used, adding /ah/ as a curve-associated vowel and /uh/ as a jagged-associated vowel in keeping with the classification of Maurer et al. (2006). All vowels were presented in the roughly phonetic form they are presented here to ensure that differences in assumed pronunciation were minimized. Unlike the previous experiment, syllables were able to vary simultaneously across all dimensions and could thus be of four types: plosive:curve-associated (PCA), plosive:jagged-associated (PJA), sonorant:curve-associated (SCA), and sonorant:jagged-associated (SJA). Within each of these general classes there were 9 possible permutations of consonants and vowels, which participants were exposed to equally.

Auditory versions of the syllable stimuli were prepared using Swifttalker with the computer voice David. All were thus standardized for pronunciation and spoken at the same amplitude and speed, ensuring that no extraneous experimenter supplied prosodic information impacted word choice by participants.

3.5.3- Experimental Design

Experiment 2 was conducted on computer via a graphical interface created using Runtime Revolution v. 2.8.1. On each trial, participants were shown a single image with two columns of syllables below it (see Fig 3.6). Each column of syllables contained four syllables representing all of the possible general consonant-vowel configurations (PCA, JSA, SCA, SJA). When participants clicked on a syllable it was then played to them over headphones and added to the "final word" display (Fig 3.6). Participants were required to choose one syllable from each of the two columns below the image, resulting in standardized two syllable words constructed from six letters. Participants were free to change their answer for either syllable, resulting in a new sound file being played to them and the final word changing. After choosing syllables participants could listen to the 'final word' they had constructed before proceeding to the next trial via a button on screen. There were a total of 90 trials, with each image drawn from the 45 pairs of generated images being seen once. Each of the 9 specific consonant-vowel configurations within the syllable subclasses was seen a total of 10 times, and randomized between trials. Additionally, the order from top to bottom in each column in which the syllable classes were presented was counterbalanced between trials.

3.5.4- Data Analysis

The chosen syllables for each participant were recorded and matching scores of two types were calculated based on the predictions of the consonant and vowel driven interpretations of the Bouba-Kiki effect. Matching scores based on consonant content were calculated based on syllables containing the plosive consonants /t/, /k/, and /p/ being associated with jagged images and the sonorant consonants /m/, /n/, and /l/ being associated with curved images. Matching scores based on vowel content were calculated based on syllables containing the curve-associatied (rounded) vowels /oo/, /oh/, and /ah/ being associated with curved images and the jagged-associated (non-rounded) vowels /uh/, /ee/, and /ay/ beings associated with jagged images.

Response data (both for syllable choice and response time) were then subjected to a repeated measures analysis of variance with image type (curved or straight)as a within (repeated) factor. A final congruency matching score was generated by combining the matching scores of the vowel and consonant categories to determine whether vowels and consonants might not only each have their own effects on sound symbolism for shape, but also whether they might have non-linear combined effects. This congruency matching score was determined for each individual syllable and for the final word itself and analysed via one-sample t-tests against chance.

3.6- Results

On the first syllable, participants responded to images in a fashion which supported both the vowel and consonant driven predictions for the Bouba-Kiki effect (See Figure 3.7). Participants preferentially chose syllables for a given image based on its

consonant class ($F_{1,1936}$=5.45, p<0.05), with curved images being matching with syllables containing sonorant consonants 52.4% of the time and jagged images being matched with syllables containing plosive consonants 54.9% of the time. One sample t-tests against chance reveal that the consonant-jagged image response bias was preferentially driving the effect, as only it was significantly different from chance performance (t_{21}= 2.26, p<0.05). Participants also preferentially chose syllables for images based on their vowel class ($F_{1,1936}$= 10.61, p<0.05), matching curved images with curve-associated vowels 56.9% of the time and jagged images with jagged-associated vowels 51.7% of the time. One sample t-tests reveal that the curve-associated vowel/ curved image response bias was preferentially driving the effect, as only it was significantly different from chance performance (t_{21}=3.84, p<0.05). Finally, the overall syllable congruency analysis revealed that syllables which matched based both on consonant class and vowel class were selected at higher than chance (25%) levels. First syllables which contained both sonorant consonants and curve-associated vowels were selected 29.1% of the time for curved images (t_{21}=2.83, p<0.05) and first syllables containing both plosive consonants and jagged-associated vowels were selected 28% of the time for jagged images (t_{21}=2.09, p<0.05).

On the second syllable, participants did not respond preferentially based on consonant class ($F_{1,1936}$= 0.10, p=0.76), choosing sonorant consonants 60.8% of the time for curved images (t_{21}=6.96, p<0.05) and 59.6% of the time for straight images (t_{21}=6.15, p<0.05). Similarly, for vowels participants did not respond preferentially ($F_{1,1936}$=0.00, p=0.96), choosing jagged-associated vowels 63.7% of the time for curved images (t_{21}=8.99, p<0.01) and 62.6% of the time for jagged images (t_{21}=8.21, p<0.05). The

overall syllable congruency analysis for syllable two revealed that chosen syllables matched on both consonant class and vowel class at greater than chance (25%) levels. A second syllable with both sonorant consonants and curve-associated vowels was selected 31% of the time for the curved images (t_{21}=3.96, $p<0.05$) and a second syllable with both plosive consonants and jagged-associated vowels was selected 32% of the time for jagged images (t_{21}=4.80, $p<0.05$).

Finally, analysis of the final word chosen revealed that participants responded in a fashion which was significantly different from chance (6.25%) for overall syllable congruency. Curved images were paired with two syllables that were contained both sonorant consonants and curve-associated vowels 8.9% of the time (t_{21}=1.95, $p<0.05$). The two syllables chosen for straight images contained both plosive consonants and jagged associated vowels 8.9% of the time (t_{21}=2.41, $p<0.05$).

3.7- Discussion

The results of Experiment 3.2 were more ambiguous than those of Experiment 3.1, although they still clearly demonstrated sound-symbolic biases in word-sound: object-shape matching. Response data for the first syllable of the constructed words were supportive of both a consonant-driven interpretation of the Bouba-Kiki effect and a vowel-driven interpretation. In some ways, this pattern is not unexpected; previous work has suggested a consonant driven effect, while other work, including the results of Experiment 3.1, suggests that vowels carry the Bouba-Kiki effect in the direction predicted by other researchers. Interestingly, the round-mouth articulation hypothesis presented by other authors (in contrast to the less well developed explanation for non-

rounded vowels- see a critique in Nielsen & Rendall, 2011) seems to be supported by the data: participants consistently chose first syllables for words to label curved images that contained the curve-associated (rounded) vowels /oh/ and /oo/. Consistent with the consonant-driven hypothesis, in which inflection patterns and spectral characteristics of sounds drive the effect, participants consistently chose first syllables to label jagged images which contained the plosive consonants /t/, /k/, and /p/. Unfortunately, these data cannot definitely answer why previous work showed a preferential and nearly exclusive ability of consonants to drive the Bouba-Kiki effect. The 2-AFC tasks used in most experiments might provide the most parsimonious answer here, as they are capable of creating the illusion of a two sided (e.g. plosive with jagged images, sonorant with curved images) bias for either interpretation. The inability of participants to choose based on both biases within individual trials might effectively ensure that only the stronger of the two biases is shown in the summary data, despite the fact that concordance rates of 80% based on consonants might suggest that 20% of the time participants are choosing based on vowels.

Response data for syllable 2 were interesting in that they were not supportive of either consonant or vowel based explanations of the Bouba-Kiki effect. Participants demonstrated a consistent bias towards choosing syllables with a sonorant/jagged associated configuration (e.g. *Nee*) regardless of the image type. This finding of an overall syllable preference in the second syllable might be an outcome of participants choosing words that are good fits for reasons other than their sound-symbolic connotations. Ease of pronunciation and adherence to regularities of English language word formation (which

may themselves be based on pronunciation) are two possible explanations for this overall word construction bias.

Congruency data for not only both syllables individually but also the entire selected word were significantly above chance, with participants likely to choose consonant-vowel pairings which were congruent with the predictions of both main interpretations of the Bouba-Kiki effect at levels above chance. For example, given a curved image participants were not only more likely to choose the SCA configuration (e.g. Loo) for both the first and second syllables, but also more likely than chance to choose an overall word form of SCA-SCA (e.g. Loomoh). Similarly, given a jagged image participants were likely to choose a PJA configuration (e.g. Tee) for both the first and second syllables and an overall word form of PJA-PJA (e.g. Teekay). This increased overall congruency is suggestive of super-additive sound-symbolic biases which can be exploited by varying the entire signal within individual phonemes- its presence may effectively explain why previous researchers produced many stimuli whose consonant and vowel content were already perfectly confounded.

3.8- General Discussion

The results of these experiments, which confirm a role for both consonants and vowels, move towards unifying the two main explanations of the Bouba-Kiki effect, for the first time demonstrating that each seems to be responsible for its own portion of the bias. Analysis of responses also reveals that they are most likely shaped by more than just sound-symbolism, as large general response biases in the second syllable suggest a possible influence of overall word form on subject choices. One possible explanation is that this is

reflective of ease of articulation (or co-articulation). Unfortunately, the presence of what some participants may view as taboo syllables (i.e. *'pee'* and *'poo'*) for word construction made an examination of this possibility difficult, although there did seem to be a general trend for participants to prefer symbols in which the vowel sound is articulated in either roughly the same tongue/lip configuration as the preceeding consonant or which occurs farther forward in the mouth/toward the lips. Subjects unwillingness to choose these two syllables despite clustering around their constituent parts in their other responses suggests even further language-constrained input onto sound symbolic processing, a finding which is entirely in line with the predictions of models of language acquisition like those of Monaghan & Christiansen, (2006).

Ramachandran & Hubbard (2001)'s proposal that cross-modal synesthesia-like linkage between visual percepts and motor (or somatosensory) activity are responsible for sound-symbolic biases is potentially amenable to include a link between plosive consonants and jagged images. Although direct auditory-visual synaesthetic linkage is unlikely in this case, both the inflectional patterns of consonants and their acoustic characteristics are potentially amenable to forming cross-modal linkages which yield expectations about relationships between object shapes and word forms. One important question is whether such linkages are a product of the way in which humans and other animals are wired at birth (and thus represent universal sound symbolism) or whether these linkages are learned through experience within a given language (which could be termed incidental sound symbolism). Research in animal communication suggests that certain types of broadband calls which are differentially prevalent in agonistic vs. affiliative contexts are salient to conspecifics in ways that are not learned. Starlings for

example are unable to learn to ignore starling distress calls, although they are capable of learning to ignore other similar sounds (Johnson et al. 1985). Unlike vowel production, even given a common neurological basis for the salience of certain consonant sounds over others, there is no necessary link between the sensory form of certain acoustic signals and the shapes of unfamiliar objects. Thus, while the consonant account does not require a synaesthetic linkage per se, it does point to some sort of communication (or at least redundancy) across sensory domains. Fortunately, recent work has suggested not only that synesthesia is much more prevalent than previously believed (Simner et al. 2006), but also that the sound symbolic biases of synaesthetes are often similar to biases exhibited by normal participants (Ward et al. 2006)

Whatever the proximate explanations for sound-symbolic biases are, they are also intriguing to consider for possible influence on language processing and learning (Imai et al. 2008). It has been demonstrated through modeling and analysis of extant languages that sound symbolic relationships and non-arbitrariness are potentially beneficial, but that as lexicon size increases they begin to hamper language acquisition and transmissibility (Gasser, 2004; Monaghan, Christiansen, & Fitneva, in press). Experiment 3.2's finding that word construction seems to be constrained by more than sound symbolic relationships are thus not only in line with previous work, but might also explain the arbitrariness of language pointed to by conventionalists like Saussure. Sound symbolic relationships can be built up from regularities of speech production and perception and be co-opted to be useful in symbol stabilization, but ultimately the process of lexicalization has more constraints than these sound symbolic biases and it is only natural that given increasing lexicon size and constraints like co-articulation that they

should be washed out somewhat. It is in explaining the interface between arbitrariness

and systematicity that sound symbolism research faces its next major frontier.

Table 3.1. Sample of nonsense words used in previous Bouba-Kiki experiments and their classification based on consonants (S(onorant) or P(losive)) and vowels (JA (jagged associated/unrounded) or CA (curve-associated/rounded).

Kohler		Maurer		Experiment 1	
Maluma (S:CA)	Takete (P:JA)	Baamoo(S*:CA)	Kuhtay (P:JA)	Tookah (P:CA)	Teekay (P:JA)
Davis		Gogaa (S*:CA)	Teetay (P:JA)	Nohmoo (S:CA)	Naymuh (S:JA)
Uloomu (S:CA)	Takete (P:JA)	**Westbury**		**Experiment 2**	
Ramachandran &Hubbard		Loum (S:CA)	Bobe (P:CA)	Keemay	Mootoh
Bouba (S*:CA)	Kiki (P:JA)	Moom (S:CA)	Deeb (P:JA)	(P-JA-S-JA) Meenoo	(S-CA-P-CA) Tohpee
				(S-JA-S-CA)	(P-CA-P-JA)
Legend– P– Plosive Consonants S– Sonorant Consonants					
JA– Jagged-Associated Vowels CA– Curve-Associated Vowels *Denotes which word is *relatively* more plosive given either mixed word construction or pairs with two plosives					
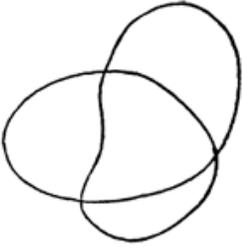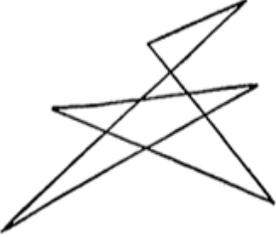
Figure 3.1. Sample of unfamiliar curved (left) and jagged (right) images used in previous

Bouba-Kiki experiments and the experiments presented here
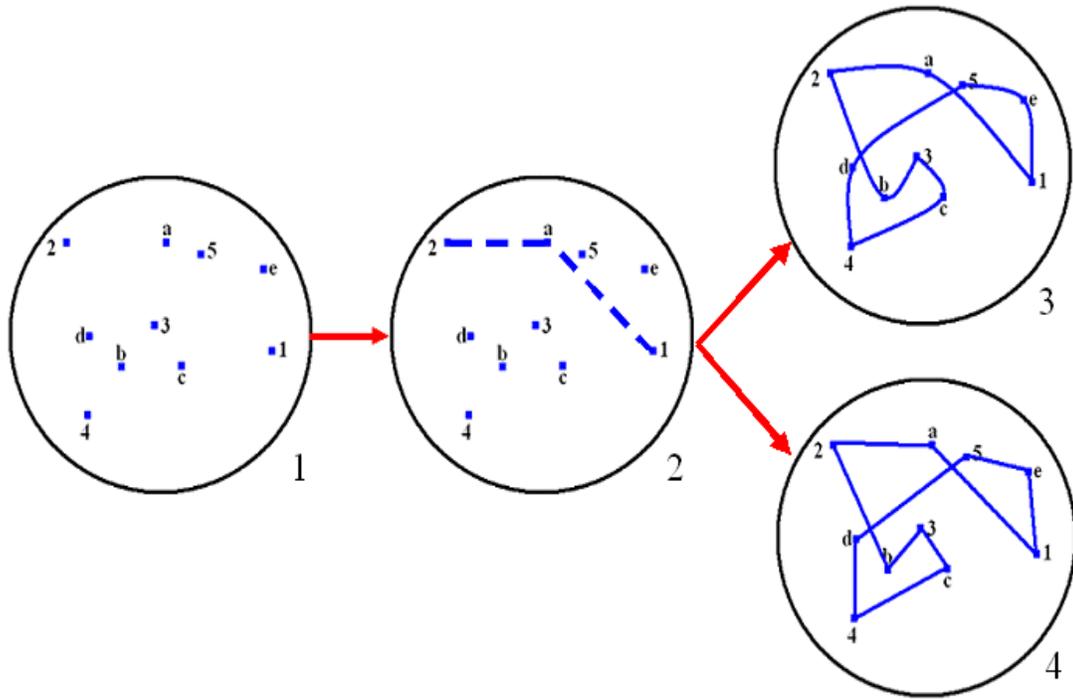
Figure 3.2. Schematic illustrating the process used to create curved and jagged versions of

the same image based on a set of randomly generated calculus points for Experiment 1
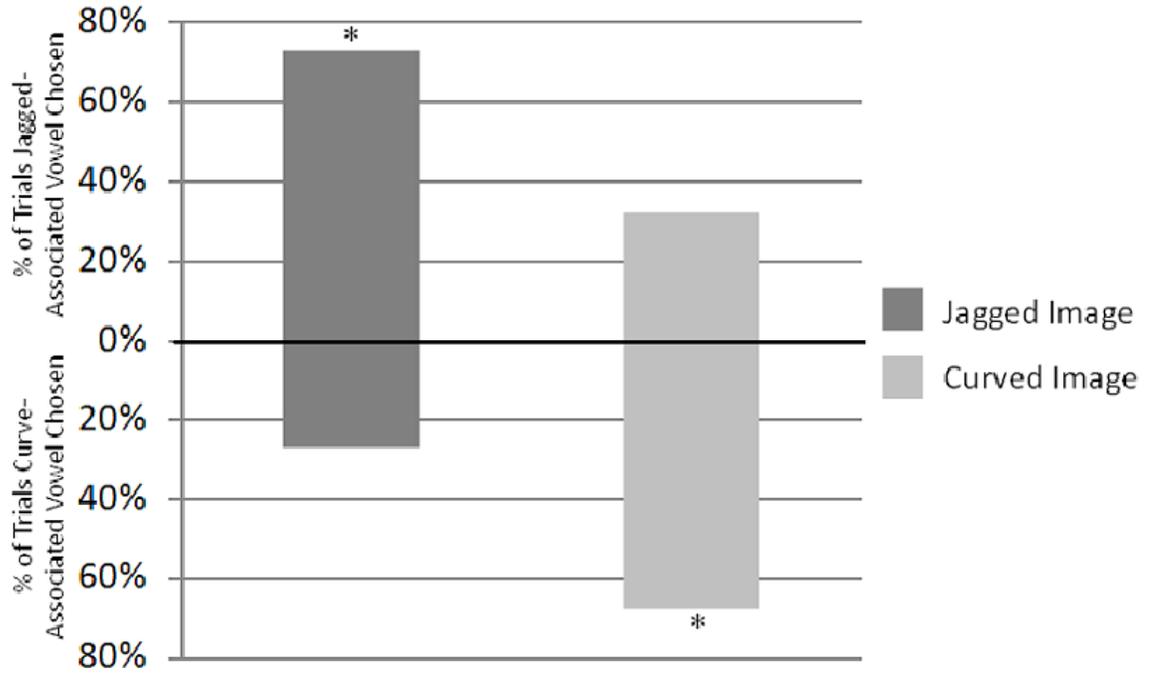
Figure 3.3. Graph of participant responses for Experiment 3.1 demonstrating that participants reliably matched jagged images with words containing jagged-associated vowels and curved images with words containing curve-associated vowels
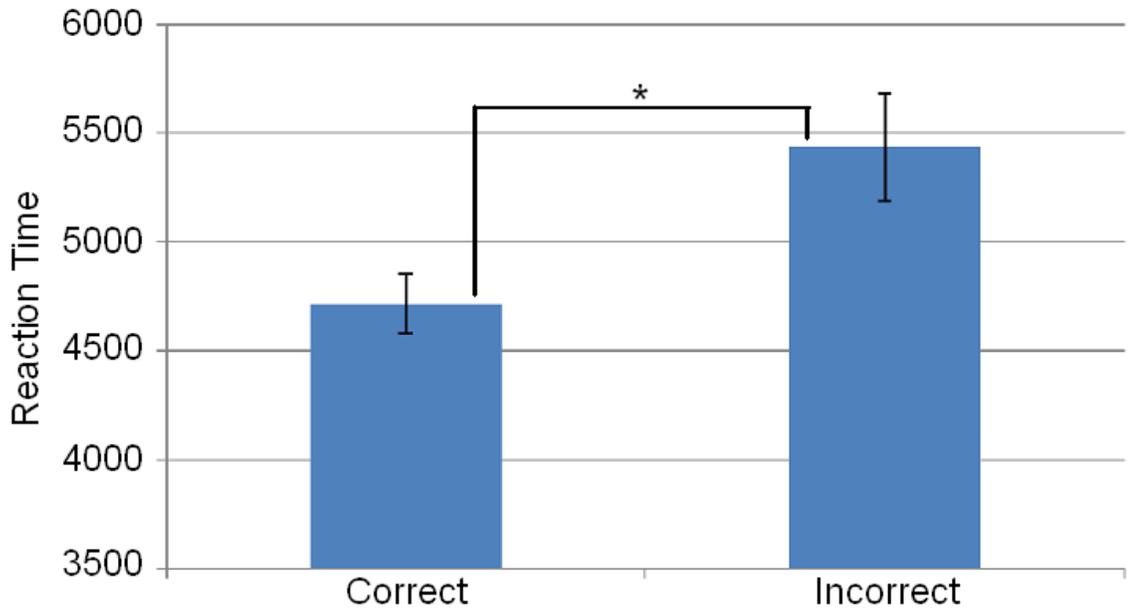
Figure 3.4. Mean (+- SEM) reaction time data from Experiment 3.1 demonstrating

significantly faster reaction times on trials in which participants responded correctly than

on trials in which they responded incorrectly
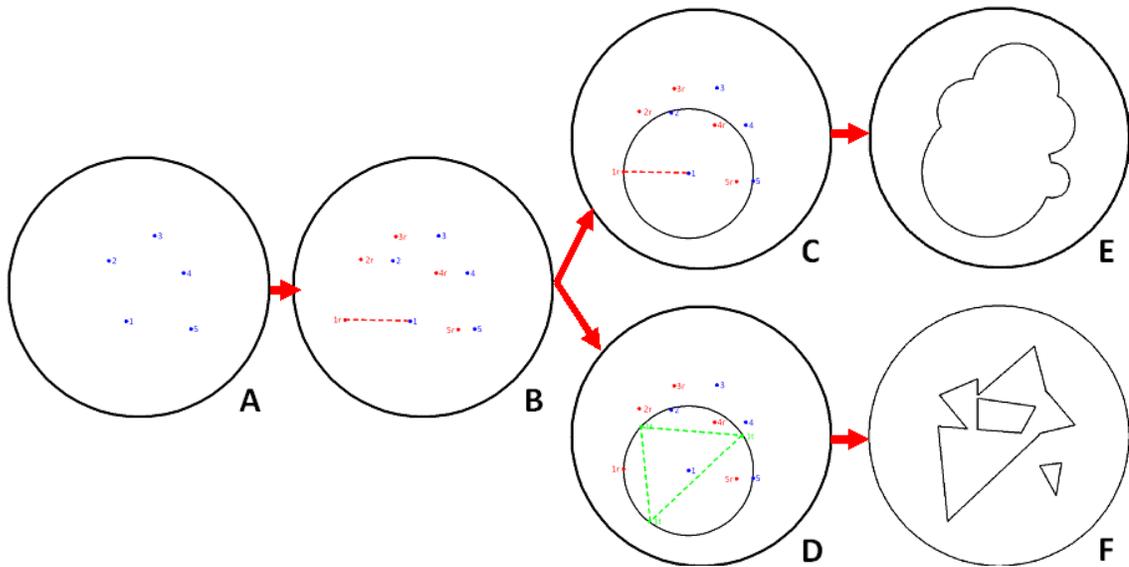
Figure 3.5. Schematic illustrating the process used to create curved and jagged images based on randomly generated calculus points for Experiment 2 (Figure generation method #2)- note: actual figure generation method used a larger number of generated triangles and circles and thus produced slightly more complex figures)
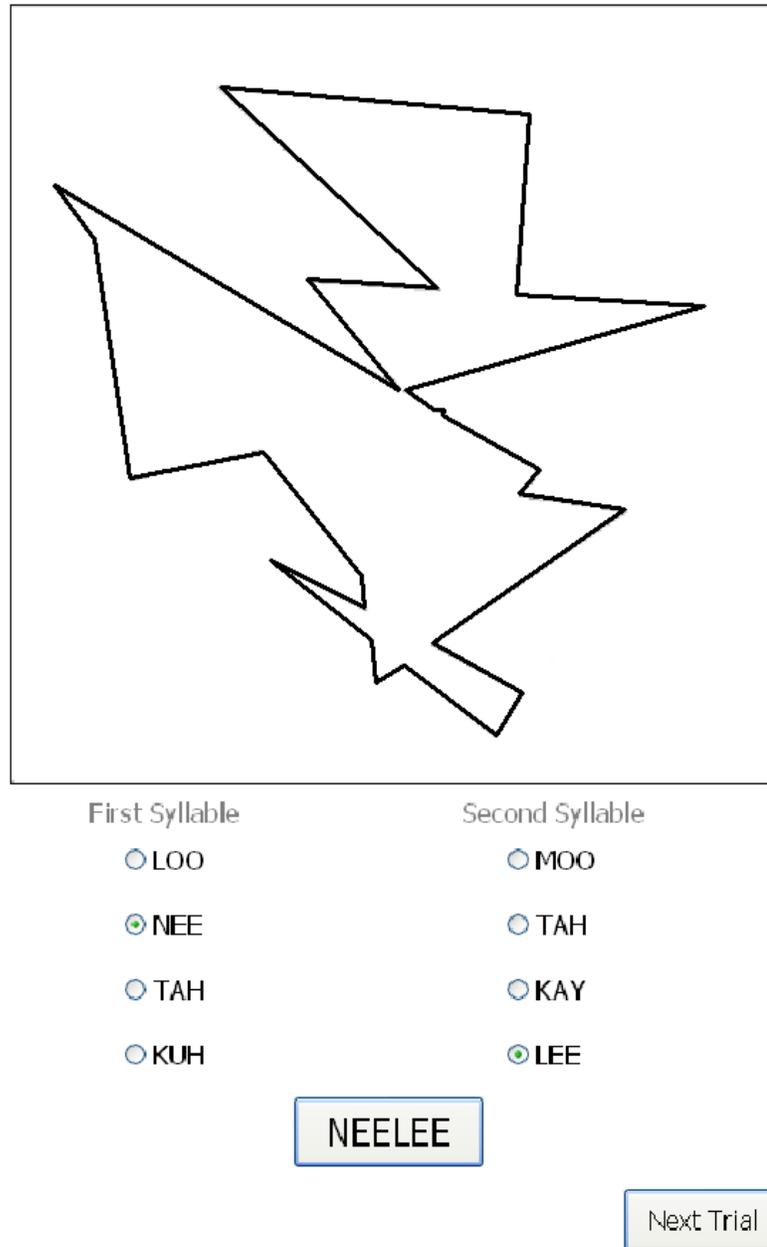
Figure 3.6. Schematic of the typical interface for a trial seen by a participant in

Experiment #2
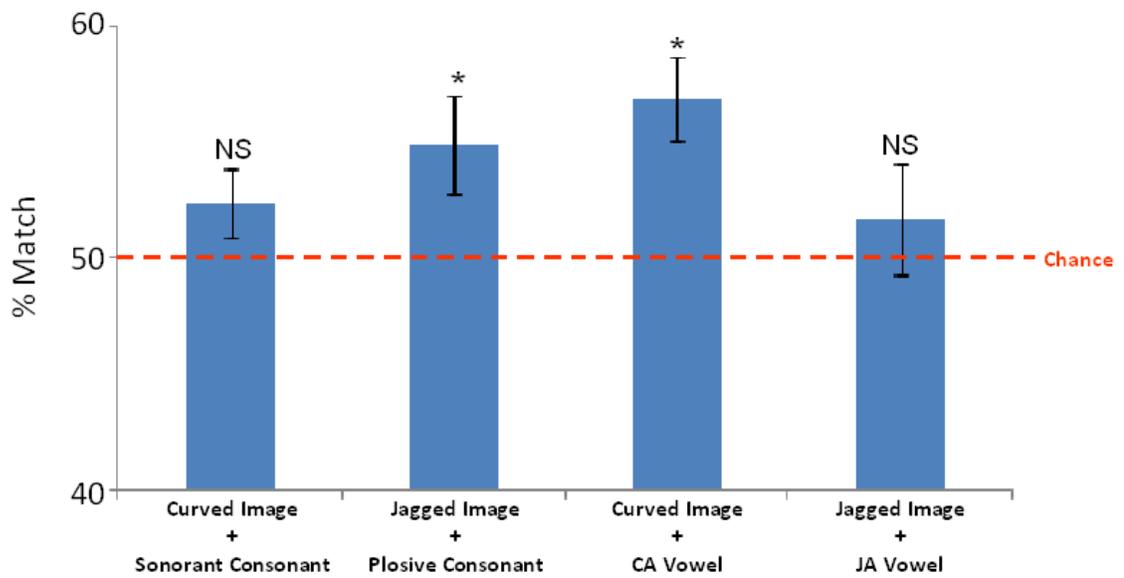
Figure 3.7. Mean (+- SEM) matching rates for syllable 1 of Experiment 3.2. Demonstrates

that jagged images were paired with plosive consonants at rates significantly above

chance and curved images were paired with curve-associated vowels significantly above
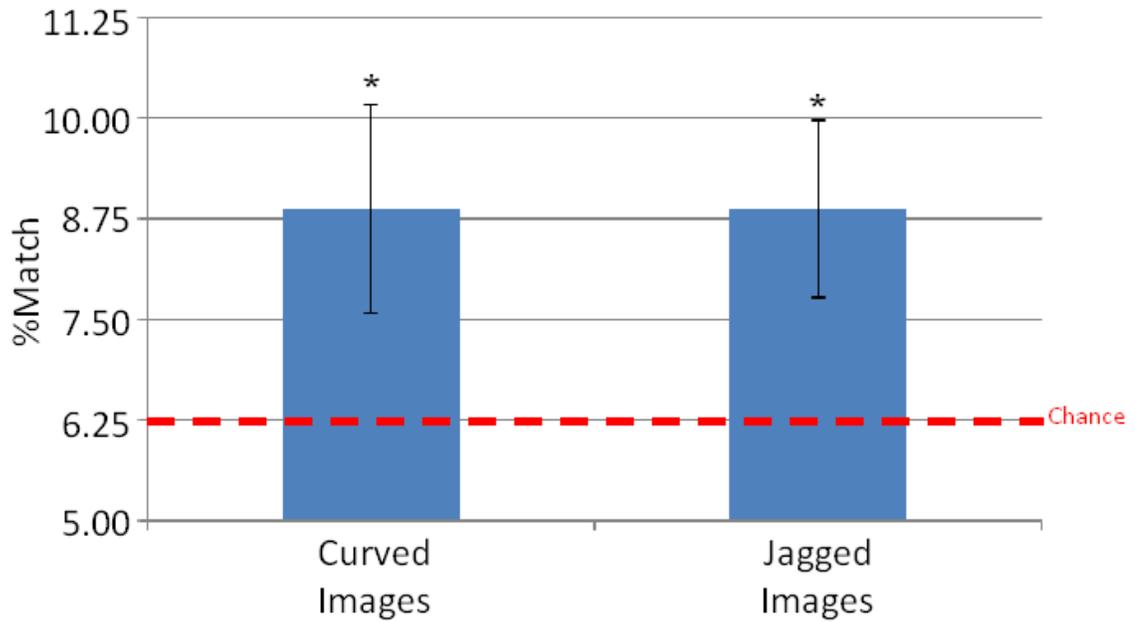
chance.

Figure 3.8. Mean (+- SEM) overall matching rates based on the congruent matching of figures and words for both their consonant and vowel characteristics. Demonstrates that curved images were matched at rates greater than chance (6.25%) with words containing both sonorant consonants and curve-associated vowels. Jagged images were matched at rates greater than chance (6.25%) with words containing both plosive consonants and jagged-associated vowels.

CHAPTER 4

THE MAGNITUDE AND ORIGIN OF SOUND-SYMBOLIC BIASES IN PROCESSING ARTIFICIAL WORD MATERIAL AND THEIR IMPLICATIONS FOR LANGUAGE LEARNING AND TRANSMISSION

4.1- Introduction

A fundamental tenet of contemporary linguistics, codified by de Saussure (1983), is that word meaning arises by convention because there are no clues to meaning in the structure of words themselves. In short, word form is arbitrary. This dictum survives to the present despite sporadic reports to the contrary in the form either of studies showing structural consistencies in the form of certain word classes within and between languages (e.g., Berlin, 1994; Brown, Black, & Horowitz, 1955), or of experiments showing that, for artificial word material, people have strong and consistent expectations about appropriate word structure (Köhler, 1947). Such findings are often lumped under the banner of sound symbolism, a term that flags the possibility that language structure is not entirely arbitrary. However, they have done little to challenge the classic Saussurean dictum. Instead, such findings have generally been discounted and marginalized, at best treated as curiosities vastly outweighed by the overwhelming arbitrariness of language (Newman, 1933; Newmeyer, 1993).

However, there has been a recent resurgence of interest in sound symbolism, one that appears to pose much greater challenges to the Saussurrean dictum. This resurgence reflects two complementary developments. First, there has been considerable systematic

research into the phenomenon of synaesthesia. While initially deemed peculiar or aberrant, contemporary research demonstrates that cross-modal neurosensory associations of the sort first described for synaethesia are about 100 times more common than previously expected (Simner et al., 2006). It has also shown that normal individuals often manifest many of the same cross-modal activity biases as synaesthetes (Ward et al., 2006; cf. Brang et al., 2010). Such findings have led some to propose that cross-modal connectivity is, in fact, a basic feature of neural organization and that synaesthetes simply manifest an exaggerated form of it (Spector & Maurer, 2009).

Complementing this development has been a renewed and more sophisticated focus on the bio-cultural origins of language. This research program has specifically included both experimental and modeling studies of the role of sound-symbolism in facilitating language acquisition and affecting the differential survival and learnability of word forms (e.g. Imai, Kita, Nagumo, & Okada, 2008; Kirby,Cornish, & Smith, 2008; Monaghan, Christiansen, & Fitneva, in press).

Taken together, these two research developments have helped to illuminate a plausible mechanistic basis for sound-symbolism in the cross-modal neural connectivity that appears to be a common organizational feature of the nervous system (Rouw & Scholte, 2007), and they have helped to establish a potentially important functional role of sound symbolism in language acquisition and transmission. However, this work still must deal with an outstanding paradox. On the one hand, the mechanistic and functional accounts help to explain the strong biases observed in past experiments testing people's naïve expectations about the appropriate structure of artificial language material. On the

other hand, the structural patterns implicated by people's naïve expectations do not appear to be manifest in the words of real languages. More succinctly, if the biases are so strong, why are they not manifest in the structure of real language?

A paradigm example of this paradox lies in the classic Bouba-Kiki effect first described by Köhler (1947). This effect involves a bias in which people consistently map unfamiliar objects that are either jagged or rounded in form to nonsense words containing specific types of consonants or vowels (Maurer et al., 2006; Nielsen & Rendall, 2011; Tarte & Barritt, 1971; Tarte, 1974; Westbury, 2005). Across multiple studies, the matching biases are consistent and large (upwards of 90% consistent matching). Given the consistency and magnitude of these effects, it is odd that there is so little tangible evidence for this effect in real languages.

I have proposed that a potential solution lies in the transparency of the task administered in most experiments. For example, most experimental designs have used a simultaneous presentation procedure in which two unfamiliar object images and two nonsense words are displayed at the same time. This procedure allows participants to make direct comparisons between the different object shapes and the different word types which might make the associations between them obvious. As a result, such experiments produce very high concordance rates that overstate the magnitude of the bias and thus the expectations about its prevalence in real language. In support of this proposal, I found that, when the experiment is modified only slightly to involve sequential presentation of single images, the same matching biases are observed but the magnitude of the effects are much smaller (~60%; Nielsen & Rendall, 2011). This finding helps to

resolve the paradox by reconciling previous apparent discrepancies between the magnitude of the matching bias and the dearth of evidence for it in real language. What remains unclear is whether the bias that this work confirms to be real but weak is something that participants have inherently or that they construct in the task itself.

To test this possibility, I designed an experiment to teach participants matching rules that were either congruent or incongruent with the consistent matching patterns documented previously. If participants have no pre-existing matching bias, then both sets of participants should perform equally well. However, if participants bring a natural bias to the task, then it should affect their performance, facilitating performance by participants in the congruent rule condition and interfering with performance by participants in the incongruent rule condition.

4.2- METHODS

4.2.1- Participants

Participants were 48 undergraduate students (36 female, 12 male) who were enrolled in introductory psychology courses at the University of Lethbridge and received partial course credit for their participation.

4.2.2- Word and Image Stimuli

The image stimuli used in this experiment were created using a radially constrained mathematical formula which created curved and jagged images which were equivalent to one another in all but the curvature of their lines (see Figure 4.1, also see Nielsen & Rendall, 2011 for full details of figure generation). This figure generation

procedure allowed for the wholly objective production of a large number of image sets that were thereby not influenced by implicit experimenter biases in image selection which has been a criticism of past work. A total of 46 pairs of images were created, 40 of which (80 total images) were used in the experimental trials and 6 of which (12 total images) were used in the demonstration trials.

Word stimuli were created for this experiment using the plosive consonants /t/, /k/, and /p/ and the sonorant consonants /m/, /n/, and /l/ and the vowels /a/, /e/, and /u/. Words were created by combining consonants and vowels in c-v-c-v order, producing a set of four letter words that were two syllables long. To ensure that all words were of the same syllabic length the vowel /e/ was excluded as a possible terminal letter (to avoid words such as *pake*, which are four letters long and can be produced as a single syllable, *payk*). Words were also constructed of either entirely plosive consonants or entirely sonorant ones, with no mixing of the two consonant classes within words. This produced 54 possible permutations of each overall word form.

4.2.3- Experimental Design

This experiment was conducted on computer via a graphical interface created using Runtime Revolution v. 2.8.1. Participants were split into two conditions, and then shown a series of demonstration trials in which they were tasked with learning a rule for the association between unfamiliar images and non-sense words. In each demonstration trial, participants were presented with a single image with a nonsense word beneath it. After a brief delay, they were prompted with either a large green check-mark or a large red "X" below the word to inform them of whether the word:object pairing was correct or not.

117

Participants were given no instructions on what characteristics of images and words were to be considered salient for matching. Each participant was shown 12 of these demonstration trials (six correct, six incorrect), three from each of the four possible pairings of images and word classes (3 plosive words with jagged images, 3 plosive words with curved images, 3 sonorant words with jagged images, 3 sonorant words with curved images).

Participants in Condition 1 were presented with demonstration trials which were congruent with sound-symbolic biases observed in past studies: they were shown green check-marks for demonstration trials where curved images were paired with sonorant words and where jagged images were paired with plosive words. Participants in Condition 2 were presented with demonstration trials which were incongruent and ran counter to previous observed response patterns; they were shown green check-marks for demonstration trials where curved images were paired with plosive words and where jagged images were paired with sonorant words. To ensure that the learnability of the stimulus sets was not systematically biased between participants, all participants were exposed to the same stimulus sets in demonstration trials and in the same order, the only difference being the congruency of the rule they were taught.

Following the demonstration phase, participants received 80 experimental trials in which 80 of the remaining 92 possible word permutations were used (40 of each plosive and sonorant words sampled randomly for each participant). In each trial, a single image was presented for two seconds with one of the nonsense words. Participants were then able to respond by using the keyboard to indicate whether or not the presented pairing

matched the rule they had learned during the demonstration phase of the experiment. They responded by pressing the "Z" key on the keyboard for a correct match and the "/" key for an incorrect match. Order of presentation for images was randomized within and across subjects.

4.2.4- Data Analysis

D` values were calculated for each participant based on the rule that they had learned in the demonstration phase of the experiment. D` values for each condition were compared to chance using a one sample t-test and to each other using a two sample t-test. A repeated measures Analysis of Variance (rmANOVA) was used to test the effects of experimental condition on reaction time. I predicted that participants in the congruent condition would perform better and faster than participants in the incongruent condition.

4.3- Results

Participants in the congruent condition of the experiment responded correctly to the rule they had learned at rates above chance (t_{23}=3.07, p<0.01), with an average d` score of 0.17 corresponding to 53.3% correct. Participants in this condition also performed better than those in the incongruent condition (t_{46}= 2.4, p<0.01). Participants in the incongruent condition performed at rates that were not different from chance (t_{23}=0.18, p=0.43), with an average d` score of 0.006, corresponding to 50.4% correct. There was no effect of experimental condition on reaction time (F=0.03, p=0.86).

4.4- Discussion

Discrimination performance (d`) was low in both experimental conditions which supports my previous findings (Nielsen & Rendall, 2011) that the relative transparency of the experimental task in previous studies has yielded inflated concordance rates that have overestimated the magnitude of the matching bias and thereby inflated expectations about its prevalence in the structure of real languages. At the same time, d` scores were influenced by experimental condition, with participants in the congruent condition performing modestly but significantly better than chance and better than participants in the incongruent condition who performed at chance levels. This pattern indicates a real but weak bias toward the rule embodied in the congruent condition that facilitated performance in this condition and interfered with performance with the opposite rule in the incongruent condition.

Together, these findings help to reconcile the apparent paradox arising from past work which pointed to a very robust and strong matching bias for artificial words with little or no evidence of it in the structure of real languages. The solution is probably two-pronged. First, the magnitude of the bias is much smaller than previously believed. Second, there are other factors that affect the likelihood that the bias will be manifest in fully developed lexicons. For example, Gasser (2004) has shown that as lexicon size increases arbitrariness in word forms becomes critical to avoid confusion among related words. This finding is buttressed by more direct attempts to model the process of lexicalization. Monaghan et al. (in press) have shown that structural regularities (non-arbitrariness) can be important to define word class membership. However, within word

classes, arbitrary distinctions among items are more stable and easier to learn. Together, these findings suggest that the greatest functional value for sound symbolic regularities probably comes during the period of language acquisition prior to lexicon crystallization. Consistent with this proposal, Imai et al., (2008) have shown that sound symbolic relationships in words facilitate noun learning for both English and Japanese children.

Any interpretation of the functional value for language acquisition of the sound-symbolic matching bias demonstrated here and in previous experiments hinges on what one believes to be the source of the demonstrated bias. Here, the distinction is between a learned bias (that results from language exposure) and a learning bias (that predates and influences language exposure). In defense of the former, it might be argued that the weak bias observed in the congruent condition of my experiment, and that also matches the bias pattern reported in previous experiments, is merely incidental. It stems from the fact that adult speakers have considerable experience with their own language which embodies the structural regularities tested. However, this position runs afoul of several important points, namely that the bias observed is specifically reported to be lacking in real languages (e.g. Newman, 1933; but see Johnson, 1967) and yet the bias has nevertheless been demonstrated not only for speakers from very disparate language groups (Huang et al., 1969 ; Nygaard, Cook, & Namy, 2009) but also for small children with minimal language exposure (Maurer et al., 2006).

Hence, my results provide more support for the possibility of a natural learning bias that facilitated performance in the congruent condition and interfered with it in the incongruent condition. From a functional standpoint, the bias might reflect broader

121

affective-semantic relationships embodied in the physical structure of vocal signals. As reviewed elsewhere, these relationships include the way different voiced sounds (e.g., harsh, noisy and punctuate sounds versus smoother and more harmonic ones) have naturally different hedonic value because they are habitually associated with situations that have very different social and behavioural salience and consequences for listeners (Morton, 1977; Owren & Rendall, 1997, 2001). Such relationships are common in humans, both in the prosodic and paralinguistic dimensions of speech (Bachorowski & Owren, 1995; Murray & Arnot, 1993; Owren et al., 2005; reviewed in Rendall 2003) and also in a variety of non-linguistic vocal signals (e.g., laughter, crying; Bachorowski, Smoski, & Owren, 2001; Protopapas & Eimas, 1997). They are also manifest far more widely in the vocal signals of nonhuman primates and many other animal species (Rendall & Owren, 2009). As a result, there are some very broad relationships between the physical structure of voiced sounds and the natural affective salience they have for listeners. These vocal affective relationships might include, at least in a limited way, the various phonemes of language that thereby inherit some natural semantic potential.

This functional account dovetails with recent research on cross-modal neural processing that provides a mechanistic account of the integration of cross-modal perceptions and explains why some forms are more salient than others. Concurrently, language modeling research has demonstrated that a combination of arbitrariness and sound-symbolic relationships is ideal for language stabilization and transmission. Hence, by integrating such mechanistic and functional insights it might be possible to determine the kinds of non-arbitrary, sound-symbolic relationships that should arise in languages and why.

122
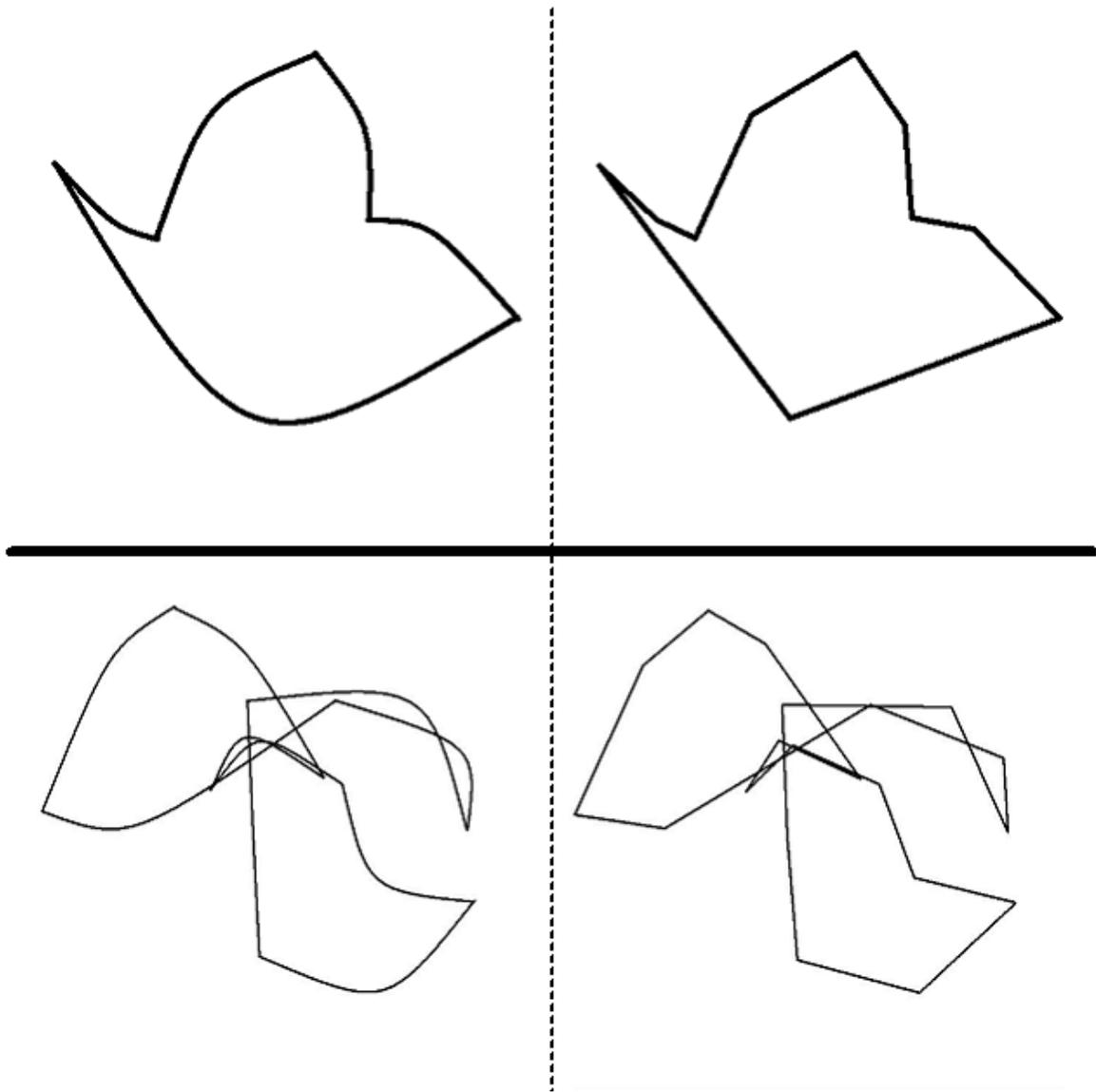
Figure 4.1. Side-by-side comparisons of sample curved and jagged images generated with my randomized image construction method. Note that the curved and jagged images are only subtly different.
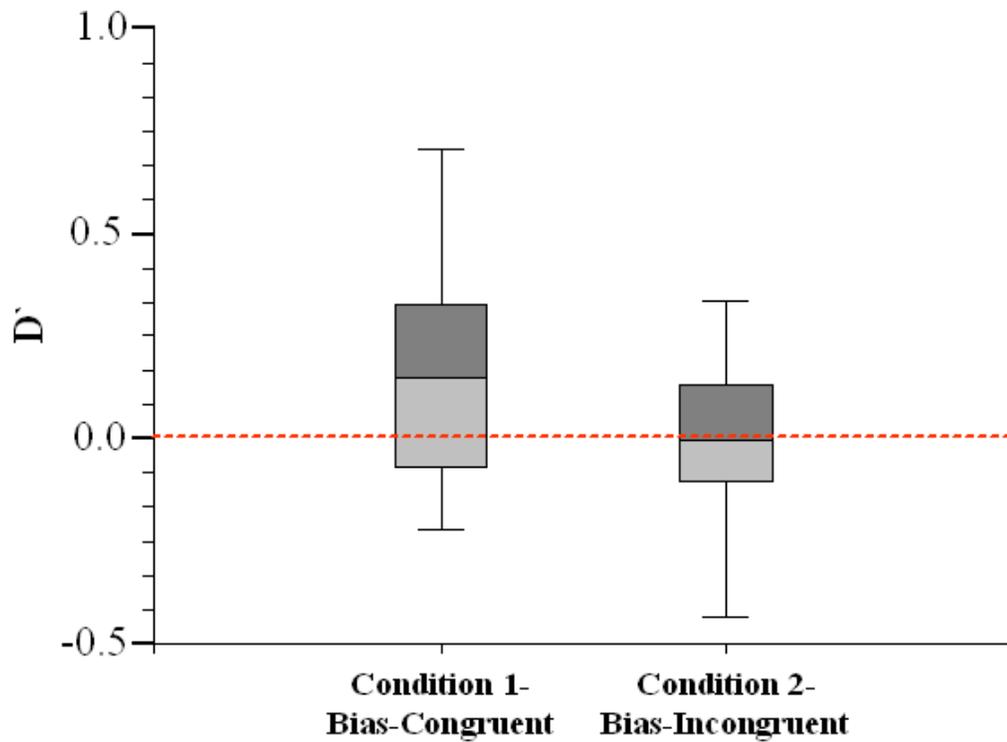
Figure 4.2. D` performance by participants in Conditions 1 and 2 scored by their ability
to follow the rule they learned in the demonstration phase of the experiment.
Performance in Condition 1 was significantly better than chance (P<0.01) and
significantly different from performance in Condition 2, while performance in Condition 2
was not different from chance (see text for statistics).

CHAPTER FIVE: SUMMARY AND CONCLUSIONS

5.1- The Bouba-Kiki Effect Revised and Revisited

The Bouba-Kiki effect, originally described by Köhler in his 1929 book *Gestalt Psychology* describes a reliable pairing of nonsense words based on their acoustic characteristics to unfamiliar two dimensional objects based on their visual characteristics. For nearly a century the effect has periodically been re-examined with very little additional information learned. Experiments designed to better explore the effect have had marginal success, falling prey to the same pitfalls that contributed to the marginalization of Köhler's original work. Typically, the number of stimuli created for any given experiment has been fewer than ten pairs, with the precise differences between curved and jagged images never clearly delineated, and the content of constructed nonsense words confounded in such a way that making claims about what word characteristics were driving associations was impossible. Furthermore, the experimental protocols themselves were often poorly designed and balanced, raising concerns about what relationship, if any, large experimental effects might have to any real world phenomena. The experiments presented in this thesis have gone some way towards addressing these concerns, using improved methodology for stimulus construction, but also tweaking experimental protocols to more thoroughly dissect what appears to be a complex bias. By more carefully exploring the Bouba-Kiki effect I have been able to link the effect to sound symbolism more generally, and with an eye towards both functional and mechanistic explanations of cross modality to steer towards future integration of sound symbolism with models of both language acquisition and transmission.

Collectively, by improving upon stimulus generation, experimental design, and dealing with small portions of the effect in isolation, I have been able to provide evidence that the Bouba-Kiki effect is a *bona fide* example of universal sound symbolism.

5.1.1- Stimulus Construction

In both Chapters Two and Three I presented new methods of figure generation that standardized techniques for generating images that varied systematically in their structure. In the first figure generation method, the generated images differed only in the curvature of their lines and were thus very subtly different from one another, allowing us to confirm that differences between stimuli need not be very large for significant differences in response tendencies to emerge. In the second figure generation method the figures were, returning to a more classic experimental set-up, far more stylized and differentiated from one another. Surprisingly, I found no consistent differences in response bias between the two image sources, nor any consistent significant differences between my images and those generated by previous researchers into the Bouba-Kiki effect. Again, this is promising in that it deals with concerns that have previously been raised about whether the biases displayed in the Bouba-Kiki effect could have any bearing on word use and language evolution outside of the lab. Even given very small differences between images, and even when two nearly identical images were presented side-by-side participants still demonstrated nearly identical matching biases. In both cases, the use of a standardized procedure for the generation of new images allowed me to remove experimenter bias as a possible source of subsequent participant bias; images I created were not chosen to be specifically effective at eliciting response biases, but merely

to run alongside predictions regarding which visual characteristics might be most salient to participants.

In addition to developing new figure generation methods that controlled for experimenter bias, in each experiment I created a large set of newly generated nonsense words, varying the characteristics depending on the way they were presented and which specific hypothesis was being addressed. In total, I presented several hundred word form polymorphisms, encompassing the entire acoustic space of the letters that were used in my experiments and allowing me to test the effects of both vowels and consonants individually as well as congruently. In the case of both word and figure generation, the creation of standardized procedures allowed for the use of large corpuses of stimuli, and thus a far greater number of trials, addressing a problem inherent in the majority of research into the Bouba-Kiki effect.

Finally, by presenting stimuli in a standardized auditory form in two of my experiments I attempted to reduce possible influences from the orthographic form of letters which might have otherwise skewed participant responses; although as mentioned previously, insofar as these orthographic confounds exist they may serve as evidence of the pervasiveness of sound symbolic biases rather than representing a confound of some kind.

5.1.2- Experimental Design and Transparency

Previous research into the Bouba-Kiki effect was heavily critiqued for the way experiments were designed and implemented; in many cases being delivered to large

groups of participants simultaneously(e.g. Davis, 1961), consisting of only a few trials (Maurer et al., 2006; in some cases a single trial e.g. Köhler, 1947), and using only a limited set of stimuli. In each of these cases one major concern was that experimental manipulations might be transparent to participants, ensuring that their responses would fall in line with predictions made by experimenters due to demand characteristics of the experiments. In Chapter 4 I presented an implicit experiment designed to specifically address problems of experimental transparency and demonstrated that participants were more capable of learning response rules which were in line with biases demonstrated by previous researchers, and that attempting to learn a response rule incongruent with their existing response bias produced interference, causing them to answer at chance and display no response bias at all.

Across all of my experiments, concordance rates dropped steadily as stimuli became less stylized and more similar to one another and also as individual experimental trials became less transparent. Rather than being discouraging, this finding suggests that even very small differences are amenable to influence from sound symbolic biases, as even when only single images were presented with two similarly constructed nonsense words biases were still observed.

5.1.3- The Differential and Combinatorial Roles of Vowels and Consonants

The majority of previous research on the Bouba-Kiki effect has focused almost exclusively on the role of differentially salient vowels for driving the response biases observed in participants. For a number of reasons, I considered this position somewhat suspect: first, the acoustic space for consonants is much larger than the acoustic space for

vowels, suggesting that of the two a bias might be more effective in the case where contrasts between classes of sounds could potentially be largest; second, although Ramachandran and Hubbard's (2001) suggestion that cross-modal connections that associated the round-mouthed articulation of rounded vowels like the */oo/* in *'bouba'* was very plausible, the connection of jagged shapes to so called 'non-rounded' vowels seemed questionable with no specific justification; and third, there seemed to be a confounded coupling of sonorant consonants with rounded vowels and plosive consonants with non-rounded vowels in nearly all previous stimuli pairs, making it impossible to parse the relative effects of different types of consonants and vowels.

In Chapter One I removed the confound produced by the coupling of sonorant consonants with rounded vowels and plosive consonants with non-rounded vowels by swapping consonants across word pairs and demonstrated that the effects of previous researchers, which had previously been ascribed to the vowel content of the words, were better explained in terms of their contrasting consonant content. In the Experiment 2.1I provided further evidence of consonant driven biases using stimuli presented in the auditory domain.

Despite my findings regarding the effect of consonants, it seemed likely that vowels might still be responsible for some portion of the effect, even if their relative influence had been washed out by consonant-driven associations. After all, as Westbury (2005) had pointed out, the split between consonants and vowels is somewhat artificial and a split between stops and continuants might actually be more valuable. Thus, in Experiment 2.1 presented an experiment where the consonant content of word pairs was

held constant while the vowel content varied systematically. The results of that experiment confirmed the findings of previous research done by Tarte (1974; Tarte & Barritt, 1971) which similarly held consonants roughly constant. Given that consonants and vowels could both be shown to produce biased response patterns in pairing words to images, an experiment was designed to determine what their cumulative effects would be. So, in Experiment 3.2 participants were given a complete sample of all of the consonant and vowel configurations possible for each individual syllable and were tasked with constructing words from scratch. The results of this experiment demonstrated that, when given the ability to choose freely, participants displayed both consonant and vowel driven biases, matching plosive consonants to jagged figures and rounded vowels to curved images at rates above chance. Additionally, there seemed to be a small effect of consonant and vowel congruency, with participants choosing vowels and consonants jointly in their predicted direction at levels above chance. Collectively, the results of these four experiments demonstrated that far from being a simple effect, the Bouba-Kiki effect is driven by at least two perceptual biases, although the consonant driven bias seems stronger given its ability to completely wash out the effect of vowels in confounded word pairs.

5.1.4- Unresolved Issues

The research presented in this thesis has come a long way in clearly demonstrating some of the important factors that drive the Bouba-Kiki effect and in what ways those factors are related to one another, but there is still a great deal of work to do, both in examining further stimulus characteristics which are important for the effect, and also in

130

determining what mechanisms are responsible for the observed response biases (Section

5.3). Additionally, research into the Bouba-Kiki effect and sound symbolism more

generally, needs to be married with functional explanations of language structure, both

through computer modeling and through experiments and analyses focused on lateral

cultural transmission (Section 5.4).

Although the sample of phonemes that made up the experimental stimuli in the

preceding chapters was far larger than has been used in any previous research, it still

accounted for a very small percentage of the total number of phonemes available in the

English language, instead focusing in the case of consonants on exemplars of sonority vs.

plosivity and in the case of vowels on those vowels which have been demonstrated by

previous researchers to be responsible for the majority of sound symbolic biases. In the

case of consonants, even in the best case my research only examined a total of 8

phonemes, which is less than half of the total number of legal English phonemes, and

samples even less of the overall acoustic space and total number of producible consonant

sounds. Other sound symbolic work, especially the work analyzing the phonemic content

of swear words (Lancker, 1999) has shown that along the sonority hierarchy (Clements,

1990; Parker, 2002; Parker, 2008) swear words tend to cluster near the fricatives, and

Yardy (2011) found that generally the higher the plosivity of a letter the more likely it

was to be found within a swear word. Given an explanation of the Bouba-Kiki effect as

being moderated by the inherent affective influence of different speech sounds, I would

predict a similar finding for any extension of the effect to analyze a greater corpus of

consonant sounds- in the case of the Bouba-Kiki effect it is unlikely to be the specific

sounds that are important, but only their relative acoustic characteristics.

My experiments were designed to make principled distinctions only between various consonants across broad categories (sonorant vs. plosive) but never specifically within those categories to test the relative strengths of individual consonants for driving the effect. Much as researchers have relied only on vague notions of "curviness" vs. "jaggedness" in figure construction I have never specifically analyzed which acoustic characteristics of sounds were most important, instead relying on the chaotic spectral patterning of plosive consonants generally as an explanation of their link to more chaotic images without analyzing in any further detail what specific aspects of the structure of the consonant sounds themselves are responsible for the effects.

As acknowledged above, despite standardizing figure construction to remove bias and allow for easy creation of a large corpus of stimuli, which specific characteristics of figures participants are focused on in making their matching choices is an open question. Future work may require the use of new technology like eye-tracking software to determine which specific aspects of stimuli are attended to preferentially during matching tasks, whether it be the curvature of lines, the number of acute angles, or some more holistic aspect of the images (their 'gestalt'). Use of eye-tracking software would also ideally allow for the analysis of what characteristics of printed words and letters participants were preferentially attending to (if any) to attempt to determine whether orthographic associations were present (as they seem to be in some synaesthetes, with color concurrents being similar for similarly shaped letters (Brang et al., 2011).

5.2- Arbitrariness and Non-Arbitrariness: A Paradox Explained

Overall, sound symbolic research has a history of being marginalized because evidence for its presence in fully robust human languages has been sparse. In some cases (e.g. Newman, 1933) this has almost certainly been due to unfair analysis of large word corpuses, but even when clear evidence of sound symbolism has been found it has been present mostly in certain word types (e.g. those denoting size; Johnson, 1967). As discussed in Chapter four, this presents a paradox: how are observed experimental biases so large without evidence for sound symbolism pervading languages in a way that should make them immediately obvious? One possibility is that experimentally observed sound symbolic biases are artificially inflated by task transparency, which was clearly demonstrated throughout the experiments presented here: as task transparency fell off, so too did the performance of participants. Unfortunately this finding can only be part of the answer to the paradox of experimental ubiquity and lexical rarity, as at some level we should expect sound symbolic biases to appear in language.

Based on modeling work (Gasser, 2004) it seems possible that as lexicon size increases it becomes necessary for arbitrary language structures to emerge. Strictly speaking, even if sound-symbolic biases provided a large benefit for language learning or transmission, that benefit would eventually disappear as more and more unique symbolic tokens were required: there are diminishing returns on sound symbolic biases. Thus, it might be during the process of language acquisition that sound symbolic relationships would be most prevalent, later being washed out as communication with increasing numbers of arbitrarily structured linguistic tokens become necessary. Not only does

research suggest that mothers use sound symbolic language preferentially when teaching their children (Imai et al., 2006), but Monaghan et al. (2006) reported that the optimal language would be one in which category membership is denoted by sound symbolic relationships while within categories sound symbolism is actually problematic. Collectively, these divergent fields of research suggest an answer to the paradox: sound symbolic relationships can be beneficial for learning and thus might be found in language, especially in categories which are learned early (such as for example, words for size), but when lexicons are fully crystallized there should be little evidence of sound symbolic relationships. Once crystallization has occurred adults are still prone to leaning on sound-symbolic associations when labeling new objects and are still differentially affectively impacted by certain phonemes but the information bearing portions of language are mostly unaffected by those biases.

5.3- Synaesthesia and Cross-Modal Communication

Synaesthetes might serve as a good model system to assess the differential affective salience of phonemes and their associations to other modalities. Rather than being peculiar, synaesthetes may reflect exaggerated or undifferentiated neural connections which exist in all people, effectively maintaining a hypertrophied form of the naïve perceptual biases underlying sound-symbolic associations like the Bouba-Kiki effect. Cross-modal associations might be strong enough to influence learning, allowing for an increased ability to learn structural regularities in concurrently presented stimuli to multiple senses (e.g. Frings & Spence, 2010) which would only be strengthened in synaesthetes. For normal individuals, at some point perceptual biases must be ignored for

134

stabilization of fully formed lexicons, but it may be the case that for synaesthetes this stage is impossible, perhaps serving to explain why autism and synesthesia co-occur at rates well above chance (Asher et al., 2009) with autistic individuals unable to prune or otherwise modulate (through higher-order downstream effects) cross-modal associations. Hence, they retain somewhat undifferentiated sensory experiences. Asher et al.'s (2009) study also provides circumstantial evidence suggesting that synesthesia is directly caused by an increased number of cortical connections, as it also co-occurs with epilepsy at rates well above chance.

Given a large body of research, it seems unlikely that the shift from undifferentiated sensory processing to a differentiated one is reflective of a complete ablation of connections between newly specialized cortical areas. It has been shown that in normal members of the population synaesthetic experiences can be induced (Saenz & Koch, 2008). Jacome (2011) for example, demonstrated that sound could induce perceptual flashes of light in subjects with herniated optic chiasma (who by definition could not possibly see), suggesting that connections between the visual and auditory cortex are pervasive even in non-synesthetes. Further, when normal individuals are blindfolded for a sufficient length of time (less than a week) their visual cortex can be observed via brain scans to process auditory information. One explanation could certainly be the creation of new neural connections that appropriate now unused tissues, but the amount of time required for the striate cortex to process auditory information is too short for such connections to grow, suggesting that they must already exist (Doidge, 2007). Finally, in blind subjects with experience reading Braille, the somatosensory homunculus of the reading fingertip does not increase in size, but rather increased efficiency at reading

Braille is mediated by activation of the visual striate cortex (Sadato et al., 1996)

Collectively, these findings suggest a view of the brain that is massively redundant, with

structural modularity emerging as a function of which types of tissue are optimal for a

given task, with other areas of the brain being more or less capable of performing similar

tasks (Szathmary, 2010).

Redundancy in processing may be valuable for a number of reasons, least of all

the plasticity it allows when recovering from complications in development or other

events which destroy or entirely ablate important brain areas. Whether the potential

benefits of cross-modality for learning statistical regularities were themselves selected for

or are a spandrel (Gould & Lewontin, 1979) of selection for redundancy is unknown, but

the potential implications of such cross modal connections are far reaching, not only in

explaining language and the remarkable pattern-recognition abilities of humans (Fahle,

1994), but also potentially in explaining abnormal phenomena like synesthesia, autism

(Asher, 2009), epilepsy, and perhaps even more peculiar phenomena like Anton's

syndrome, where despite being cortically blind those afflicted believe they are still

completely capable of seeing.

Attention to plausible mechanistic instantiations of cross-modal connectivity

raises a large number of questions pertaining to the way in which the brain functions

more generally and the outcomes of its organization. By using mechanism as a foundation

by which to study the Bouba-Kiki effect and other examples of sound symbolism we are

able to augur not only what types of sound-symbolic linkages might be present, but also

to determine what similarities exist between normal members of the population and those

with aberrant levels of connectivity that seem reflective of an earlier developmental stage. Given the broad way in which the neural architecture underpinning sound-symbolic biases might influence other types of processing, there are important potential implications for language, learning, and the manipulation of the behaviour of others (and with proper feedback mechanisms, manipulation of one's own behaviour).

5.4- Implications and Future Research

In Section 5.1. I outlined a number of unresolved issues that pertain directly to the Bouba-Kiki effect as a psychological phenomenon. To move forward with research, a number of these unresolved issues must be dealt with in a systematic way so that future researchers are better able to interpret their data. I clearly demonstrated through the course of my experiments that how stimuli and tasks are structured can greatly influence interpretation; what seemed like an effect driven by vowels was revealed to be an effect driven by consonants, only to be revealed upon more careful examination to reflect both consonant and vowel driven biases operating independently and in conjunction with one another. Although I have principled reasons to believe that some overall perceptual gestalt resembling sonority is responsible for the word sound portion of the observed sound/shape biases, the degree to which variation in phonemic content carries over to variation in response patterns remains unstudied. Similarly, although my figure generation techniques improved upon previous work, I still do not know what specific characteristics of figures participants are responding to, or how those characteristics, presented as two dimensional representations, correspond to real world three-dimensional objects (although I have reason to believe the correspondences are similar; Liu, 1997).

Development in each of these areas is needed to move forward with research into the Bouba-Kiki effect and sound symbolism more generally, especially in connecting sound symbolic biases to important research topics like the evolutionary roots of language, the process of language acquisition within the lifetime of an individual, transmission of language and its change across generations, non-propositional uses of language in humans, and finally the ways in which each of these fields of research might interact with research into animal communication. A dissection of how research into the Bouba-Kiki effect and sound symbolism more broadly can be married to research in each of the above fields is the focus of the following, final sections.

5.4.1- The Emergence of Language

Language doesn't fossilize, and so the evolution of language has been labeled as one of the hardest remaining questions in all of the biological sciences (Christiansen & Kirby, 2003). Simply put, even animals who are trained from birth to be proficient in language never reach the level of language proficiency of even toddlers, developing overall vocabularies of less than five hundred words. Human beings are alone amongst all of the world's other organisms as the sole user of language, and because physical evidence for language was impossible to acquire prior to the invention of writing, we have no grounds on which to base claims about which stage of human evolution language first emerged, other than that it emerged after the split with chimpanzees. Pinker and Bloom (1990) have claimed that all complex behaviours must be a product of direct natural selection acting on their emergence, and have thus claimed that specific modules must be necessary for my uniquely human abilities. In support of this claim, Pinker has subsequently

138

offered up evidence that the FOXP2 gene (Pinker, 2007; see Lai et al., 2001) is implicated in language, due to the fact that a single British family in whom the gene was mutated displayed broad language deficits which were reliably passed down to subsequent generations. Humans also have another profound difference with other animal species: my cortex is much larger than almost every other animal relative to my body size (Jerison, 1973; cf Roth & Dicke, 2005), which has led to the conclusion that brain size and language must be correlated, and thus that directional selection pushed us towards language use, which required new neural tissues and cognitive modules.

Other researchers have taken the large brains of humans and related it to language in an entirely different way, claiming that the brain reflects selection for a large domain general processor that is subsequently used for language, rather than direct selection for language producing a large brain. This approach roughly claims that language is a spandrel of selection for other traits, like a larger memory and pattern recognition abilities that would be functional in a variety of contexts.

Uniting these two approaches seems impossible, with the first assuming modularity that doesn't seem to be present in the brain in much more than a functional sense (Fiez et al., 1995; Poldrack et al., 2001), and the second falling prey to the same problem that all explanations invoking exaptations do; claiming that language emerges as a reliable functional outcome of selection for a number of other abilities completely ignores the functional pressure that language must exert, if not at the time of its acquisition, then at least subsequently.

Gould's idea of spandrels itself is based on surfaces which emerge from the process of creating domed ceilings and which were often subsequently painted by artists. Spandrels are seen as beautiful, but not because of any design of the architect who sought only to produce functional arches; their function is emergent. Unfortunately, this argument ignores the fact that spandrels themselves are not beautiful at all; they merely create an additional surface on which selection can operate; in the case of cathedrals this gave artists a place to paint or engrave, while in the case of evolution this allows for the emergence of new traits that can be selected for. In this sense, the idea of a spandrel or exaptation is interesting, but entirely uninformative in the context of evolution. No trait can be selected for *de novo*, and thus all traits can be seen as exaptations of traits which were originally produced for some other purpose (or in some cases for no purpose at all). Language could only truly be considered a spandrel if it was an outcome which itself required absolutely no selective pressure; roughly if culture and interactions with other organisms assumed the role of a cathedral painter and made the otherwise useless spandrel both functional and beautiful. Certainly, cultural processes and interaction with conspecifics are important mediators of language, but the claim that language can emerge without any selection is one for which there is no real evidence, and additionally one which ignores continuity with animal communication systems. Furthermore, claiming that language is merely an outcome of selection for other traits obscures the search for the source of language: if the co-occurrence of a number of traits is a necessary pre-condition for the emergence of language then it is those traits I must in turn elucidate.

So I arrive at the crux of the problem of tracing the history of language evolution: language appears to have a genetic component, suggesting that at some level of

organization it has been selected for directly, but evidence for distinct neural modules is lacking or confusing; and language appears to be based at least in part upon cognitive processes which are shared with other types of information processing, but cannot be placed entirely on the shoulders of those cognitive processes. At this stage, it seems that the answer to how language evolved likely lies somewhere between the two: natural selection need not produce modules in the strictest sense, but can act upon existing processes, strengthening them to set the boundary conditions for the emergence of language. Subsequently, cultural evolution and individual learning can act upon these boundary conditions and shape the specific instantiations of language, all the while being constrained by those boundary conditions for the ability to use language productively and also by perceptual and cognitive biases which seem otherwise unrelated.

5.4.2- Language Acquisition and Learning

Studied in isolation, the question of language evolution seems insoluble; because language leaves no fossils I cannot determine where in our history language emerged, what cognitive abilities were necessary for its emergence (and whether any were unique to language), or what role cultural transmission and individual learning played in its development. Fortunately, the question of the evolution of language is one which can be broken down into its constituent parts; by carefully examining the language capabilities of children and the way in which their mastery of language emerges through development I can determine how much of the uniqueness of language is taken care of by the learning process. In addition to examinations of the ways in which children learn language, the errors that are commonly made both in speech production and in grammatical structure

can also be illuminating; clearly showing signs of the developmental pathways down which language proceeds in the course of learning. Computer modeling or the use of artificial agents has also become a productive field, especially in delineating what sorts of cognitive processes are required for the emergence of language. Given similar inputs in the course of language learning, the differences which emerge between artificial agents and children can be illustrative of what sorts of learning processes or biases children might have to make up the difference.

Sound symbolism is one such bias which might be important for learning, but which is typically ignored in modeling of language, largely because the information metaphor is so prevalent. One especially difficult part of language learning is the learning of verbs. In the case of nouns, it is generally quite simple, through ostension (in which indexical pointing is associated with linguistic tokens; Gliga & Csibra, 2009), to teach children what one is referring to when applying a new linguistic token. Ostension, however, is impoverished in the case of verb learning (Hirsh-Pasek & Golinkoff, 2006), because in pointing a verb there are many single things that a person might be pointing to; for example saying the word "trot" and pointing at a horse might leave a naïve listener questioning whether one is pointing at the horse, the action of running, the direction the horse is running in, or some gestalt of the entire phenomenon. In a series of studies, Imai et al. (2008; Nagumo et al., 2006) demonstrated that Japanese mothers use sound symbolic words more often when describing actions to their infants and that this use of sound symbolic language allowed Japanese speaking children to learn verbs at a faster rate than English speaking children, whose mothers do not typically use sound symbolic language. Subsequently, Imai et al. (2008) also showed that English speaking

children were able to learn Japanese sound-symbolic words for actions at rates similar to Japanese children and greater than normal english speaking children, suggesting that the learning bias cannot be accounted for by an overall presence of more sound symbolism (and thus more experience with sound symbolism) in Japanese children compared to English ones. Nygaard et al. (2009) extended this work, demonstrating that English children who were otherwise inexperienced were more accurate in learning Japanese words when those words were antonyms and responded faster than when learning randomly paired word meanings, suggesting that the phonetic structure of the words was facilitating learning. Finally, Brandone, Golinkoff, Pence, and Hirsh-Pasek (2007) demonstrated the importance of context, showing that children are only able to learn verbs when the context of the verb's presentation and the action are perceptually matched, suggesting not only that contextual information is important for language learning, but also that children are unable to ignore contextual information and learn non-matching actions, only gaining an ability to do so at approximately three years of age. Parault (2006; Parault & Schwanenflugel, 2006; Parault & Parkinson, 2008) has provided additional evidence that sound symbolic words are more easily learned than words that do not contain sound symbolism, demonstrating that 5th and 6th graders were able to learn sound symbolic words more easily, but also that adults when exposed to obsolete words were able to provide better definitions for those words with sound symbolic forms.

Not only do children rely on sound symbolic correspondences when learning new words, but they also use sound symbolism when creating novel names for unfamiliar objects. Samuelson and Smith (2005) found that when children were tasked with the

spontaneous naming of unnamed objects they attended preferentially to shape over texture or color, naming objects in line with the predictions of sound symbolic relationships like the Bouba-Kiki effect.

Computer modeling of language acquisition has a large advantage over experiments with children, but that advantage is also its potential downfall. Researchers who program their own computer models or artificial agents are able to explicitly set up the type of learning algorithms, perceptual systems, and other parameters which characterize their systems. Ostensibly, this allows researchers to tightly control the assumptions that their agents make, and thus to model language using the most minimalist approaches possible. Unfortunately, this is not always the case, and some researchers (e.g. Steels, 2003) have specifically addressed the need for researchers to be explicit about what abilities their agents have that allow them to engage in iterated learning and language acquisition.

Nearly all computer models that attempt to address language learning do so through processes of iterated learning and language games between agents, who have shared reference and thus the ability to engage with one another. Steels (2003) has pointed out that shared reference is perhaps the most important of all traits in language learning, and has demonstrated that postural elements of body language are important for the learning of language. Robots with shared body plans are able to simulate their own movement and as a community arrive at a common symbol system that allows them to ask other robots to move in specific ways. Quinn (2001) has gone so far as to demonstrate the embodied agents need not even have dedicated communication channels

for communication to emerge; functional behaviours based simply on movement can emerge between similarly structured embodied agents, and these early functional behaviours can form the basis for later communication systems.

The importance of body plans and physical constraints is clearly demonstrated in the work of Oudeyer (2002), who has demonstrated that sensory-motor coupling dynamics are likely responsible for the forms of phonemic coding that commonly arise in human languages. Specifically, the physical instantiation of the vocal tract and its innervations leads naturally to the emergence of certain phonemic codings. Oudeyer's model system- a vowel synthesizer with constraints similar to the human vocal tract- has no built in assumptions other than physical constraints, and still arrives at a similar system to human languages, suggesting that in some cases the physical structure of a trait alone is enough to determine its developmental outcomes without the influence of any functional inputs.

Other models have been able to parallel the emergence of symbol-based communication systems in existing animals. Loula, Gudwin, El-Hani, and Queiroz (2010) have for example demonstrated that in a population of autonomous agents a self-organized referential system for predator specific alarm calls emerged, with a distinct sound for each predator type, similar to the purportedly referential alarm calls of vervets. Loula's work is interesting, but is an important example of how computer modeling can overreach when claiming to parallel human or animal communication systems. First, the referential system of Loula's artificial agents stabilizes first after many thousands of iterations of calling, before which symbolic tokens have no inherent valence and do not

elicit predator-avoidance responses. The obvious implication is that generations of monkey callers cannot survive long enough to arrive at this kind of calling system. Similarly, because the system is entirely information-based there is no such thing as "similarity" between two signals; signal number 49 and 50 are no more similar to each other than are symbol 1 and 100. Finally, we know from animal research more generally that signals themselves can have salience before they stabilize (if they ever do) as bona fide symbolic tokens.

Computer modeling is thus a mixed bag when it comes to research into language acquisition; a great deal of convincing work has been undertaken that highlights the roles of embodiment and shared reference in establishing symbolic associations, further flagging the importance of establishing not only what abilities a given system has, but also the possibility that the ability to play language games themselves might itself be an important trait to consider. On the other hand, most research still relies on unconstrained modeling which is not biologically plausible and which assumes from the start that language and communication are arbitrary, rendering conclusions about how well computer language parallels human language impossible. By marrying sound symbolism and other perceptual biases to computer modeling it should become increasingly possible to determine what abilities humans might have that are peculiar, and the inclusion of constraints which are themselves functional might actually lead to computer models of language acquisition that are less complex than systems which work from truly arbitrary symbols to obtain a lexicon and syntax (Nowak, Plotkin, & Jansen, 2000).

5.4.3- Language Transmission

The study of how languages change over time has been implicated as a possible source of information regarding what the real purpose of language is, with authors like Pagel (2009) pointing out that certain words of increased salience and use are unlikely to change over time, while other words which are less commonly used or relatively unimportant are liable to change over time. Similarly, Kirby et al. (2008) have demonstrated that cultural transmission of language proceeds in a fashion in which languages become increasingly easier to learn and more structured over time. Although each of these approaches is interesting, they largely ignore what characteristics of words make them more or less effective carriers of meaning than other words, and thus potentially miss out on examining sound-symbolic biases which can be selected for as multiple word-form polymorphisms are selected for their transmissibility.

Despite this shortcoming, models of cultural transmission of language and its evolution within linguistic communities might explain a great deal about language. The process of creolization is one fascinating case in point, as language impoverished children move from simplistic pidgin languages to fully lexicalized and grammar-rich creoles within single generations (e.g. Hymes, 1971). Similarly, American Sign Language was a fairly impoverished and simplistic language before effectively being put into the hands of young deaf/mute children and undergoing processes of cultural transformation which led to greater syntactic structure and productivity. Because of the power of culture in shaping and strengthening language, some have claimed that the process of language emergence is mostly a social one (e.g. Pagel, 2009), but this claim does not align well with

the absence of language in other animals with some evidence of culture and cultural transmission (Nakamichi, Kato, Kojima, and Itoigawa, 1998), or with the inability of even human-acculturated primates to learn language from humans (e.g. Gardner, Gardner, & van Cantfort, 1979).

Sound symbolism might thus be profitably integrated not only into models of language learning and explanation of *in vivo* language learning in children, but also into models of language evolution that focus on its lateral transmission. Incorporation of sound symbolic biases into these types of modeling, rather than making things more complex, might ultimately serve to demonstrate that language is a complex process which emerges from a suite of much simpler adaptations and learning processes, the confluence of which is unique to humans, even if no single trait can be said to be selected directly for language. Alternatively, the delineation of traits that are important for language emergence and that are common to computer models or other animal communication systems might illuminate what traits are, by contrast, uniquely human and thus likely to have been targets of selection.

5.4.4- Affect Modulation and the Everyday Use of Language

Despite the obvious benefits which humans have accrued through the use of symbolic communication, a great deal of my language use still involves social interaction that is not entirely reliant upon the transmission of propositional content. As often as I engage in interactions with other people in which the transmission of information is important, I engage in interactions that closely parallel social grooming in animals; forming friendships, relationships, business partnerships, and even temporary alliances

148

which are an important part of being a human being. In fact, more often than not it seems that it is not **what** we say that is important in the context of social interactions, but **how** we say it. Subtle changes in intonation can convey widely different meanings without any of the supposedly arbitrary symbols of language changing. Sarcasm (Pexman & Olineck, 2002) and verbal irony are two examples of communication that are entirely mediated by prosody and context, which computers, for all their current sophistication, would never be able to understand, but which nevertheless have potentially important implications for human communication and social interaction.

Insofar as the theory of *affective semantics* is correct, perceivers of communication, both in humans and in other animals are largely helpless when it comes to modulating the way in which they are affected by the verbal behaviour of others, suggesting that manipulation through auditory conditioning is an important part of language that cannot be overlooked, especially when considering the evolutionary roots of language insimpler forms of communication. Examinations of language learning in children, simulated language acquisition in artificial agents, and lateral transmission of language through cultural influences are important parts of the puzzle in understanding the evolution of language. The fingerprints of sound symbolism can already be seen in the language learning abilities of children, and more biologically plausible models of language acquisition are on the horizon that will take into account perceptual biases, delineating how much of the roots of the phenomenon of language might be shared with the communication systems of other animals (and in turn, how much is uniquely human).

5.4.5- Implications for and from Animal Communication

Perceptual biases like those described by Morton (1994) are ubiquitous in the animal kingdom in animals that use auditory means of communication. Recognizing that prior to any process of lexicalization these auditory signals have inherent meaning which is not arbitrary has been an important influence on research into animal behaviour, although it has still largely been ignored in favor of anthropocentric models of animal behaviour that look for similarities to human behaviour rather than looking at humans for the influence of my animal ancestry. My anthropocentric views of animals are largely mediated by my own perceptual and cognitive biases; simply put, we do not share the *umwelt* of other species and can in most cases only guess at what their perceptual experiences are like, assuming they have perceptual experiences that we could understand at all (Nagel, 1974). Despite this fact, we do share a great deal of our perceptual equipment with other mammals, especially those with whom we are closely related, and so there should be a great number of non-linguistic parallels between us and my closest relatives that could be necessary, but not sufficient for the emergence of language.

The ubiquity of cross-modal neural processing in humans is also a potentially valuable area of research in animals, as it seems likely that connections between cortical areas in which events co-occur temporally can have a number of functional outcomes for learning that might be present in other animals and that might underlie all sorts of social behaviour. In researching animal communication further, it will be important not to consider animals as deficient or degraded versions of humans for whom parts of language are missing. Given the continuity between us and other animal species the difference

between communication systems is certainly striking; humans have millennia old oral and written traditions and now have as a species produced far more written information than there is in an entire human genome (Sagan, 1977). At the same time, the pervasiveness of information in human culture does not fully explain language; a great deal of our everyday interactions are similar to those of other animals and deal with the establishment and maintenance of social relationships. Taking language to be a completely unique phenomenon and elevating it above all other types of communication only frustrates the process of explaining its origins.

5.5- Conclusion

Fire trucks are loud and obnoxious; ice cream trucks are melodious and inviting. *Fuck* grabs attention and increases arousal more effectively than does *Snuggle*. When asked whether the word *nosu*-nosuing is a verb to describe a hopping target or a walking one, English speakers are more likely to answer that it is the word for hopping, despite lack of familiarity with Japanese. Words for large objects like *huge, chunk, enormous,* and *humongous* contain low back vowels at higher rates than chance. When asked to pair unfamiliar line drawings with nonsense words, participants are more likely to pair the word *Bouba* with a rounded or curved figure and the word *Kiki* with a jagged figure.

Each of the above is an example of sound symbolism as presented in this thesis. In these cases rather than sounds being arbitrary carriers of symbolic and propositional content the inherent salience of the sounds themselves has a functional outcome in the absence of any symbolic reference. For the majority of the 20th century these examples of sound symbolism have been marginalized in the face of the informational model of

151

language transmission, which has systematically ignored the importance of the communication channel over which language is typically transmitted and has thus alienated human language from the communication systems of other animals with whom we share a clear developmental and evolutionary history.

In the last decade there has been a resurgence of interest from a number of fields in explaining not only the presence of sound symbolic biases in neurological terms, but also in accounting for functional benefits that have led to their emergence. By acknowledging not only proximate mechanistic causes of sound symbolic biases but also their functional roots the connection between human and animal communication systems has been reestablished.

In the course of this thesis I have presented a series of experiments on a single sound symbolic bias known as the Bouba-Kiki effect, while linking my work to sound symbolism and its broader implications. My experiments have improved upon research methodologies employed by previous researchers and have in the process pointed to a number of areas for future research that might productively examine the question of the emergence of language more generally; a problem that has been described as the most difficult question remaining in psychology.

The implications of the research presented in this thesis have at times been far reaching and speculative, but the potential importance of small biases to much more complex processes cannot be overstated, especially in the case of complex systems which are self-referential (Hofstadter, 2007), where the boundary conditions can generally account for a great deal of the subsequent behaviour of the system.

The Bouba-Kiki effect, far from being marginalia, is a *bona fide* example of universal sound symbolism that is shared not only by all humans, but that is also likely conserved in some ways across a number of animal species. Humans do however appear unique in our ability to manipulate symbols productively and transmit information faithfully in conventionally established fashions, but information transmission alone cannot account for the constant use of language in the establishment and maintenance of social relationships, which is based in no small part upon manipulation the behaviour of other humans. Furthermore, communication, when not framed in the light of information transmission, has a plethora of clear homologues in the communication systems of other animals, suggesting that our own communication system is likely more similar to theirs than we typically appreciate.

As a language universal, the Bouba-Kiki effect is likely more important for understanding language than has generally been assumed, as not only do these effects appear universal in humans, they also have a rich history in the communicative systems of other animals, which must have in some way prefaced our own linguistic abilities. In the long term, the answers to the hardest question remaining in psychology might actually be quite simple, and sound symbolism is likely to be an important part of the confluence of traits that allow for the emergence and stabilization of language, both in evolutionary time and in the course of the lifespans of individual humans.

REFERENCES

Aleman, A., Rutten, G. J., Sitskoorn, M. M., Dautzenberg, G., & Ramsey, N. F. (2001). Activation of striate cortex in the absence of visual stimulation: An fMRI study of synesthesia. *Neuroreport, 12,* 2827–2830.

Aronoff, J. (2006). How we recognize angry and happy emotions in people, places and things. *Cross-Cultural Research, 40,* 83-105.

Asher, J.E., Lamb, J.A., Brocklebank, D., Cazier, J.B., Maestrini, E., Addis, L., Sen, M., Baron-Cohen, S., & Monaco, A.P. (2009). A whole-genome scan and fine-mapping linkage study of auditory-visual synesthesia reveals evidence of linkage to chromosomes 2q24, 5q33, 6p12, and 12p12. *The American Journal of Human Genetics, 84(2*), 1–7.

Bachorowski, J.-A., & Owren, M. J. (2008). Vocal expression of emotion. In M. Lewis, J. M. Haviland-Jones, & L. Feldman Barrett (Eds.), *Handbook of Emotions* (pp. 196-210). New York: Guilford Press.

Bachorowski, J.-A., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science, 6,* 219-224.

Bachorowski, J.-A., Smoski, M.J., & Owren, M.J. (2001). The acoustic features of human laughter. *Journal of the Acoustical Society of America, 10,* 1581-1597.

Bar, M. & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science, 17,* 645-648.

Bentley, M., & Varon, E. J. (1933). An accessory study of phonetic symbolism. *American Journal of Psychology, 45,* 76–86.

Berlin, B. (1994). Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In Hinton L., Nichols, J. & Ohala, J.J. (Eds.), *Sound symbolism* (pp. 76-93). Cambridge: Cambridge University Press.

Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology, 37,* 204-215.

Bolinger, D. (1978). Intonation across languages. In J.H. Greenberge (Ed.), *Universals of Human Language. Vol. 2:Phonology* (pp. 471-524). Stanford, CA: Stanford University Press.

Brandone, A. C., Golinkoff, R. M., Pence, K. H., & Hirsh-Pasek, K. (2007). Action speaks louder than words: Young children differentially weight perceptual, social, and linguistic cues to learn verbs. *Child Development, 78,* 1322 - 1342.

Brang, D., Hubbard, E.M., Coulson, S., Huang, M., & Ramachandran, V.S. (2010). Magnetoencephalography reveals early activation of V4 in grapheme-color synesthesia. *NeuroImage, 53*, 268-274.

Brang, D., Rouw, R., Ramachandran, V.S., & Coulson, S. (2011). Similarly shaped letters evoke similar colors in grapheme-color synesthesia. *Neuropsychologia, 49,* 1355-1358.

Brown R.W., Black A.H., & Horowitz A.E. (1955). Phonetic symbolism in natural languages. *Journal of Abnormal Social Psychology, 50,* 388-93.

Burton, H., Snyder, A.Z., Diamond, J.B. & Raichle, M.E. (2002). Adaptive changes in early and late blind: A fMRI study of verb generation to heard nouns. *Journal of Neurophysiology, 88,* 3359 –3371.

Callejas, A., Acosta, A., & Lupianez, J. (2007). Green love is ugly: Emotions elicited by synaesthetic grapheme-color perceptions. *Brain Research, 1127,* 99-107.

Calvert, G.A., Hansen, P.C., Iversen, S.D., & Brammer, M.J. (2001). Detection of Audio-Visual integration sites in humans by application of electrophysiological criteria to the bold effect. *NeuroImage, 14,* 427-438.

Cheney, D. L., & Seyfarth, R. M. (1996). Function and intention in the calls of nonhuman primates. *Proceedings of the British Academy, 88,* 59–76.

Cheney, D. L., & Seyfarth, R. M. (1998). Why animals don't have language. In G.B. Pearson (Ed.), *The tanner lectures on human values* (pp. 174-209). Salt Lake City, UT: University of Utah Press.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Christiansen, M., & Kirby, S. (2007). *Language evolution.* Oxford, England: Oxford University Press.

Clements, G. N. (1990). The role of sonority in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I. Between the grammar and physics of speech* (pp. 283–333). Cambridge, MA: Cambridge University Press.

Curtiss, S. (1977). *Genie: a psycholinguistic study of a modern-day "wild child".* Boston, MA: Academic Press.

Darwin, C. (1871). *The descent of man, and selection in relation to sex* (1st ed.). London, England: John Murray.

Davis, R. (1961). The fitness of names to drawings. A cross-cultural study in Tanganyika. British *Journal of Psychology, 52,* 259-268.

Dawkins, R., & Krebs, J.R. (1978). Animal signals: information or manipulation? In J.R. Krebs & N.B. Davies (Eds.), *Behavioural Ecology: An Evolutionary Approach* (pp. 282-309). Oxford: Blackwell Scientific Publications.

Day, S. (2004). Trends in synesthetically colored graphemes and phonemes—2004 revision. Available from http://home.comcast.net/~sean.day/Trends2004.htm

De Saussure, F. (1983). *Course in General Linguistics*, Eds. Charles Bally and Albert Sechehaye. Translated by Roy Harris. La Salle, Illinois: Open Court.

Diffloth, G. (1994) i:big, a:small. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 107-114). Cambridge, MA: Cambridge University Press.

Doidge, N. (2007). *The Brain the Changes Itself*. New York, NY: Penguin Paperbacks.

Dravnieks, A. (1985). Atlas of odor character profiles. PA: American Society for Testing and Materials.

Dunbar, R. (2003). The Origin and Subsequent Evolution of Language. In M. Christiansen & S. Kirby (Eds.), *Language Evolution* (pp. 219-234). Oxford: Oxford University Press.

Fahle, M. (1994). Human pattern recognition; parallel processing and perceptual learning. *Perception, 23(4),* 411-427.

Fenton, M.B., & Licht, L.E. (1990). Why rattle snake? *Society for the Study of Amphibians, 24,* 274-279.

Fiez, J.A., Raichle, M.E., Miezen, F.M., Patersen, S.E., Tallal, P., & Katz, W.F. (1995). PET studies of auditory and phonological processing: Effects of stimulus characteristics and task design. *Journal of Cognitive Neuroscience, 7,* 357-375.

Fitch, W.T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Science, 4,* 258–267.

Fitch, W.T., & Hauser, M.D. (1995). Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on "honest" advertisement. *American Journal of Primatology, 37,* 179-190.

Gardner, R.A., Gardner, B.T., & van Cantfort, T.E. (1989). *Teaching sign language to chimpanzees.* New York: SUNY Press.

Frings, C., & Spence, C. (2010). Crossmodal congruency effects based on stimulus identity. *Brain Research, 1354,* 113-122.

Gallace, A., Boschin, E., & Spence, C. (2011). On the taste of "Bouba" and "Kiki": An exploration of word-food associations in neurologically normal participants. *Cognitive Neuroscience, 2,* 34-46.

Gasser, M. (2004). The origins of arbitrariness of language. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society,* 434-439. Mahwah, NJ: Lawrence Erlbaum.

Gebels, G. (1969). An investigation of phonetic symbolism in different cultures. *Journal of Verbal Learning and Verbal Behaviour, 8,* 310-312.

Ghazanfar, A.A., & Hauser, M.D. (1999). The neuroethology of primate communication: Substrates for the evolution of speech. *Trends in Cognitive Science, 3,* 377-384.

Ghanzanfar, A.A., & Rendall, D. (2008). The evolution of human vocal production. *Current Biology, 18(11),* 457-460.

Ghazanfar, A.A., Turesson, H.K., Maier, J.X., van Dinther, R., Patterson, R.D., & Logthetis, N.K. (2007). Vocal-tract resonances as indexical cues in rhesus monkeys. *Current Biology, 17,* 425-430.

Gil-da-Costa, R., Braun, A., Lopes, M., Hauser, M.D., Carson, R.E., Herscovitch, P., & Martin, A. (2004). Toward an evolutionary perspective on conceptual representation: Species-specific calls activate visual and affective processing systems in the macaque. *Proceedings of the National Academy of Sciences, 101,* 17516-1752.

Gliga, T., & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words, *Psychological Science, 20,* 347-353.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review, 95(1),* 384–394.

Gould, S. J., & Lewontin, R. C. (1979) The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings Of The Royal Society of London, 205,* 581-598.

Gray, J. A., Parslow, D. M., Brammer, M. J., Chopping, S., Vythelingum,G. N., & Ffytche, D. H. (2006). Evidence against functionalism from neuroimaging of the alien color effect in synesthesia. *Cortex, 42,* 309-318.

Lane, H. (1976). *The wild boy of Aveyron.* Cambridge, MA: Harvard University Press.

Herzog, M., & Hopf S. (1983). Effects of species-specific vocalizations on behaviour of surrogate-reared Squirrel monkeys. *Behaviour, 86,* 197-214.

Hinton, L., Nichols, J., & Ohala, J. (1994). *Sound symbolism*. Cambridge, MA: Cambridge University Press.

Hirst, D.J., & Di Cristo, A. (1998). *Intonation systems: A survey of twenty languages*. Cambridge, MA: Cambridge University Press.

Hirsh-Pasek, K., & Golinkoff, R. M. (2006). *Action meets word: How children learn verbs*. Oxford: Oxford University Press.

Hockett, C.F. (1960). The origin of speech. *Scientific American, 203*, 88-96.

Hofstadter, D. (2007). *I am a strange loop*. New York: Basic Books.

Huang, Y.H. (1969). Universal magnitude symbolism. *Journal of Verbal Learning and Verbal Behaviour, 8*, 155-156.

Hymes, D.H. (1971). *Pidginization and creolization of languages*. Cambridge, MA: Cambridge University Press.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*, 54-65.

Jacome, D. (2011). Sound induced photisms in pontine and extra-pontine myelinolysis. Available from: http://www.webmedcentral.com/wmcpdf/Article_WMC001296.pdf

Jerison, H.J. (1973). *Evolution of the brain and intelligence*. New York: Academic Press.

Johnson, R.C. (1967). Magnitude symbolism of English words. *Journal of Verbal Learning and Behaviour, 6*, 508-511.

Johnson, R., Cole, P., & Stroup, W. (1985). Starling response to three auditory stimuli. *Journal of Wildlife Management, 49,* 620-625.

Kim K.-O. (1977). Sound symbolism in Korean. *Journal of Linguistics, 13*, 67-75.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Science, 105,* 10681-10686.

Köhler, W. (1947). *Gestalt Psychology, 2nd Ed*. New York, NY: Liveright.

Koriat, A., & Levy, I. (1977). The symbolic implications of vowels and of their orthographic representations in two natural languages. *Journal of Psycholinguistic Research, 6,* 93-104.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition, 114,* 19-28.

Lai, C.S.L., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., & Monaco, A.P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature, 413,* 519-523.

Lancker, D.V., & Cummings, J.L. (1999). Expletives: Neurolinguistics and neurobehavioural perspectives on swearing. *Brain Research Reviews, 31,* 83-104.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211-240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes, 25,* 259-284.

Lester, B.M., & Boukydis, C.F.Z. (1985). *Infant crying: theoretical and research perspectives.* New York: Plenum Press.

Levitin, D. J. (2007). *This is your brain on music.* New York, NY: Plume.

Lieberman, P. (1968). Primate vocalizations and human linguistic ability. *Journal of the Acoustical Society of America, 44,* 1574-1584.

Lindauer, M. (1990). The effects of the physiognomic stimuli taketa and maluma on the meanings of neutral stimuli. *Bulletin of the Psychonomic Society, 28,* 151-154.

Liu, C.H. (1997). Symbols: Circles and Spheres Represent the Same Referents. *Metaphor and Symbol, 12,* 135-147.

Loula, A., Gudwin, R., El-Hani, C.N., & Queiroz, J. (2010). Emergence of self-organized symbol-based communication in artificial creatures. *Cognitive Systems Research, 11,* 131-147.

MacNeilage, P., & Davis, B.L. (2000). On the origin of internal structure of word forms. *Science, 288,* 527-531.

Malmberg, B. (1964). Couches primitives de structure phonologique. *Phonetica, 1,* 221-227

Marks, L.E. (1996). On perceptual metaphors. *Methaphor and Symbolic Activity, 11,* 39-66.

Marler, P. (1955). Characteristics of some animal calls. *Nature, 176,* 6-8.

Matychuk, P. (2005). The role of child-directed speech in language acquisition: a case study. *Language Sciences, 27,* 301-379.

Maurer, D., & Mondloch, C. (2004) Neonatal synesthesia: A re-evaluation. In L. Robertson & N. Sagiv (Eds.), *Synesthesia: Perspectives from cognitive neuroscience* (pp. 193-213). Oxford: Oxford University Press.

Maurer, D., Pathman, T., & Mondloch, C.J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science, 9,* 316-322.

Maynard-Smith, J. and Szathmary, E. (1997). The major transitions in Evolution. Oxford: Oxford University Press.

Monaghan, P., & Christiansen, M.H. (2006). Why form-meaning mappings are not entirely arbitrary in language. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

Monaghan, P., Christiansen, M.H., & Fitneva, S.A. (In Press). The Arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General.*

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

Morton, E.S. (1977). On the occurrence and significance of motivation-structural rules in some birds and mammal sounds. *American Naturalist, 111,* 855–869.

Morton, E.S. (1994). Sound symbolism and its role in non-human vertebrate communication. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 348-365). Cambridge, MA: Cambridge University Press.

Muggleton, N., Tsakanikos, E., Walsh, V., & Ward, J. (2007). Disruption of synaesthesia following TMS of the right posterior parietal cortex. *Neuropsychologia, 45,* 1582-1585.

Murray, I.R., & Arnott, J.L. (1993). Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *Journal of the Acoustical Society of America, 93,* 1097-1008.

Nagel, T. (1974). What is it like to be a bat?. *Philosophical Review, 83(4),* 435-450.

Nagumo, M., Imai, M., Kita, S., Haryu, E., & Kajikawa, S. (2006). *Sound iconicity bootstraps verb meaning acquisition.* Paper presented at the XVth International Conference of Infant Studies, Kyoto, Japan.

Nakamichi, M., Kato, E., Kojima, Y., & Itoigawa, N. (1998). Carrying and washing of grass roots by free-ranging Japanese macaques at Katsuyama. *Folia Primatologica: International Journal of Primatology, 69,* 35-40.

Newman, S.S. (1933). Further experiments in phonetic symbolism. *American Journal of Psychology, 45,* 53-75.

Newmeyer, F.J. (1993). Iconicity and generative grammar. *Language, 68,* 756-796.

Ngo, M.K., Mishra, R., & Spence, C. (2011). Assessing the shapes and speech sounds that people associate with chocolate samples varying in cocoa content. *Food Quality and Preference, 6,* 567-572.

Nielsen, A., & Rendall, D. (2011). The sound of round: Evaluating the role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology, 65(2),* 115-124.

Nikolie, D., Lichti, P., & Singer, W. (2007). Color opponency in synesthetic experiences. *Psychological Science, 18,* 481– 486.

Nowak, M.A., Plotkin, J.B., & Jansen V.A.A. (2000). The evolution of syntactic communication. *Nature, 404,* 495-498.

Nuckolls, J.B. (1999). The case for sound symbolism. *Annual Review of Anthropology, 18,* 225-252.

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition, 112,* 181–186.

Ohala, J. (1994). The frequency code underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 325-347). Cambridge, MA: Cambridge University Press.

Oudeyer, P. (2002). Phonemic coding might result from sensory-motor coupling dynamics. In B. Hallam, D. Floreano, J. Hallam, G. Hayes, & J.-A. Meyer (Eds.), *From animals to animats 7: Proceedings of the seventh international conference on simulation of adaptive behaviour* (pp. 406-416). Cambridge, MA: MIT Press.

Owings, D.H., & Morton, E.S. (1998). *Animal vocal communication: A new approach.* Cambridge: Cambridge University Press.

Owings, D., & Zeifman, D.M. (2004). Human infant crying as an animal communication system: Insights from an assessment-management approach. In D.K. Oller & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach* (pp. 151-170). Cambridge, MA: MIT Press.

Owren, M.J., & Rendall, D. (1997). An affect-conditioning model of nonhuman primate vocalizations. In D.W. Owings, M.D. Beecher, & N.S. Thompson (Eds.), *Perspectives in Ethology, Vol. 12: Communication* (pp. 299-346). New York: Plenum Press.

Owren, M.J., & Rendall, D. (2001). Sound on the rebound: Returning form and function to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology, 10(2)*, 58-71.

Owren, M.J., Amoss, R.T., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology, 73*, 530-544.

Owren, M. J., Dieter, J. A., Seyfarth, R. M., & Cheney, D. L. (1993). Vocalizations of rhesus (Macaca mulatta) and Japanese (*M. fuscata*) macaques cross-fostered between species show evidence of only limited modification. *Developmental Psychobiology, 26*, 389-406.

Owren, M.J., Rendall, D. & Ryan, M.J. (2010). Redefining animal signaling: Influence versus information in communication. *Biology & Philosophy 25*, 755-780.

Pagel, M. (2009). Human language as a culturally transmitted replicator. *Nature Reviews Genetics, 10*, 405-415.

Parault, S. J. (2006). Sound symbolic word learning in written context. *Contemporary Educational Psychology, 31*, 228-252.

Parault, S. J., & Parkinson, M. (2008). Sound symbolic word learning in the middle grades. *Contemporary Educational Psychology, 33*, 647-671.

Parault, S. J., & Schwanenflugel, P. J. (2006). Sound symbolism: A possible piece in the puzzle of word learning. *Journal of Psycholinguistic Research, 35*, 329-351.

Parker, S. (2002). *Quantifying the sonority hierarchy*. Ph.D. thesis: University of Massachusetts, Amherst, MA.

Parker, S. (2008). Sound level protrusions as physical correlates of sonority. *Journal of Phonetics, 36*, 55-90.

Patterson, F.G.P., & Matevia, M.L. (2001). Twenty-seven years of project Koko and Michael. In B.M.F. Galdikas, N.E. Briggs, L.K. Sheeran, G.L. Shapiro, & J. Goodall (Eds.), *All apes great and small: African Apes* (pp. 165-176). Boston: Kluwer Press.

Pexman, P. M., & Olineck, K. M. (2002). Does sarcasm always sting? Investigating the impact of ironic insults and ironic compliments. *Discourse Processes, 33*, 199-217.

Pinker, S. (2007). *The stuff of thought*. Cambridge, MA: Harvard University Press.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioural and Brain Sciences, 13(4)*, 707-784.

Poldrack, R.A., Temple, E., Protopapas, A., Nagrajan, S., Tailal, P., Merzenich, M., & Gabrieli, J.D. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. *Journal of Cognitive Neuroscience, 13,* 687-697.

Protopapas, A., & Eimas, P.D. (1997). Perceptual differences in infant cries revealed by modifications of acoustic features. *Journal of the Acoustical Society of America, 102,* 3723-3733.

Quinn, M. (2001). Evolving communication without dedicated communication channels. *Advances in Artificial Life: Sixth European Conference on Artificial Life, 2159,* 357-366.

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

Rendall, D. (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *Journal of the Acoustical Society of America, 113,* 3390-3402.

Rendall, D. (2009). The evolution of language and the (ir)relevance of primate communication. *Evolutionary Anthropology, 18,* 230-233.

Rendall, D., & Owren, M.J. (2010). Vocalizations as tools for influencing the affect and behaviour of others. In S. Brudzynski (Ed.), *Handbook of Mammalian Vocalizations* (pp. 177-186). Oxford: Academic Press.

Rendall, D., Kollias, S, Ney, C., & Lloyd, P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry. *Journal of the Acoustical Society of America, 117,* 944-955.

Rendall, D., Owren, M.J., & Rodman, P.S. (1998). The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *Journal of the Acoustical Society of America, 103,* 602-614.

Rendall, D., Owren, M.J. & Ryan, M.J. (2009). What do animal signals mean? *Animal Behaviour, 78,* 233-240.

Rich, A.N., & Mattingley, J.B. (2010). Out of sight, out of mind; Suppression of synaesthetic colours during the attentional blink. *Cognition, 114,* 320-328.

Richer, F., Beaufils, G.A., & Poirier, S. (In press). Bidirectional lexical-gustatory synesthesia. *Consciousness and Cognition.*

Roblee, L., & Washburn, M.F. (1912). The affective values of articulate sounds. *The American Journal of Psychology, 23,* 579-583.

Roth, G., & Dicke, U. (2005). Evolution of the Brain and Intelligence. *Trends in Cognitive Sciences, 9(5)*, 250-257.

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

Sadato, N., Pascual-Leone, A., Grafman, J., Ibanez, V., Deiber, M.P., Dold, G., & Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature, 380*, 526-528.

Saenz, M., & Koch, C. (2008). The Sound of change: Visually-induced auditory synesthesia. *Current Biology, 18*, 650-651.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110-114.

Sagan, C. (1977). *The dragons of Eden: Speculations on the evolution of human intelligence*. New York: Ballantine Books.

Samuelson, L.K., & Smith, L.B. (2005). They call it like they see it: spontaneous naming and attention to shape. *Developmental Science, 8*, 182-198.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12*, 225-239.

Savage-Rumbaugh, S., & Lewin, R. (1994). *Kanzi: The ape at the brink of the human mind*. New York: Wiley.

Seyfarth, R.M., & Cheney, D.L. (1986). Vocal development in vervet monkeys. *Animal Behaviour, 34*, 1640-1658.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: evidence for predator classification and semantic communication. *Science, 210*, 801–803.

Simner, J. (2006). Beyond perception: Synaesthesia as a psycholinguistic phenomenon. *Trends in Cognitive Science, 11*, 23-29.

Simner, J., & Hubbard, E.M. (2006). Variants of synesthesia interact in cognitive tasks: Evidence for implicit associations and late connectivity in cross-talk theories. *Neuroscience, 143*, 805-814.

Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S.A., Fraser, C., Scott, K., & Ward, J. (2006). Synesthesia: The prevalence of atypical cross-modal experiences. *Perception, 35*, 1024 –1033.

Skinner, B.F. (1957). *Verbal behaviour*. Acton, MA: Copley Publishing Group.

Smith, E.A. (2010). Communication and collective action: language and the evolution of human cooperation. *Evolution and Human Behaviour, 31*, 231-245.

Spector, F., & Maurer, D. (2009). Synesthesia: A new approach to understanding the development of perception. *Developmental Psychology, 45*, 175-189.

Spence, C., & Gallace, A. (2010). Tasting shapes and words. *Food Quality and Preference, 22*, 290-295.

Steels, L. (2003). Evolving grounded communication for robots. Trends *in Cognitive Science. 7(7)*, 308-312.

Steels, L., & Spranger, M. (2008). Can body language shape body image? *Artificial Life, 11*, 577-584.

Stevenson, R.J., & Oaten, M.J. (2010). Sweet odours and sweet tastes are conflated in memory. *Acta Psychologica, 134*, 105-109.

Szathmary, E. (2010). Evolution of language as one of the major evolutionary transitions. In S. Nolfi & M. Mirolli (Eds.), *Evolution of communication and language in embodied agents* (pp. 37-53). Springer Verlag.

Tarte, R.D. (1974). Phonetic symbolism in adult native speakers of Czech. *Langauge and Speech, 17*, 87-94.

Tarte, R.D. (1981). The relationship between monosyllables and pure tones: An investigation of phonetic symbolism. *Journal of Verbal Learning and Verbal Behaviour, 21*, 352-360.

Tarte, R.D., & Barritt, L.S. (1971). Phonetic symbolism in adult native speakers of English: three studies. *Language and Speech, 14*, 158-168.

Terhune, D.B., Cardena, E., & Lindgren, M. (2010). Disruption of synaesthesia by posthypnotic suggestion: an ERP study. *Neuropsychologia, 48*, 3360-3364.

Terrace, H.S., Petitto, L.A., Sanders, R.J., & Bever, T.G. (1979). Can an ape create a sentence? *Science, 206(4421)*, 891-902.

Ultan, R. (1978). Size-sound symbolism. In J.H. Greenberg, C.R. Ferguson, & E.A. Moravcsik (Eds.), *Universals of Human Language, Volume 2* (pp. 527-568). Stanford, CA: Stanford University Press.

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.

Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain & Language, 93*, 10-19.

Witthoft, N., & Winawer, J. (2006). Synesthetic colors determined by having colored refrigerator magnets in childhood. *Cortex, 42,* 175-183.

Yardy, B.J. (2010). *Sound symbolism, sonority, and swearing: An affect induction perspective.* Master's Thesis: University of Lethbridge, Lethbridge, Alberta.