Proceedings of IDETC/CIE 2008
ASME 2008 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
August 3-6, 2008, New York City, New York, USA

DETC2008-49372

EFFECTS OF DICHOTOMOUS LEXICAL STIMULI IN CONCEPT GENERATION

I. Chiu and L.H. Shu*
Department of Mechanical & Industrial Engineering
University of Toronto
5 King's College Road
Toronto, ON, M5S 3G8, Canada
*Corresponding author email: shu@mie.utoronto.ca

ABSTRACT

The relationship between language and reasoning motivates us to study the use of language within engineering design. This paper describes our continued investigation of language as stimuli for concept generation. Specifically we investigate dichotomous lexical stimuli that are related to the problem in either a disagreeing, incongruent manner or in an agreeing, congruent manner. This is a follow-up investigation where we extend previous experiments to include both congruent and incongruent stimuli to enable comparison of differences between designer behavior and concepts. A between-subjects think-aloud experiment was performed where participants were presented with a problem and asked to generate concepts to address the problem. Half the participants were provided with incongruent stimuli and the remaining were provided with congruent stimuli.

Participants provided with incongruent stimuli used the stimulus words as verbs more often than the participants provided with congruent stimuli. Verbs possess several properties desirable for use as design stimuli including the increased introduction of new lexicalized concepts to the concept generation process. When two independent raters scored the concepts, there was a positive correlation between the raters that concepts developed with incongruent stimuli were more novel. Understanding the effects of different lexical stimulus types on concept generation contributes to the development of design support tools that exploit the relationship between language and reasoning to increase design novelty.

Keywords: Concept generation, design stimuli, language, dichotomy.

1 INTRODUCTION

Design is challenging because it is ill structured and open-ended (Dym & Little, 2004). To add to the challenge, there is little support for designers in the early stages of design, and designers are often required to rely on experience and intuition (Li & Jin, 2006) rather than on reasoning processes. While there is disagreement on the exact relationship between language and cognitive functions important to design, many agree that there is a connection between language and reasoning (Levinson, 1996; Jackendoff, 1983; Pinker, 2007). While some argue that language influences thought, others maintain that language reflects thought. Regardless of the nature of the relationship, related work suggests that language can be used as a design tool. We have been investigating the use of language in the early stages of design, specifically as related stimuli for concept generation.

Our work focuses primarily on the use of verbs as stimuli, as verbs are the part-of-speech that can often be used to describe functions (Pahl & Beitz, 1996; Stone & Wood, 2000). We use lexical relationships to generate stimulus sets that are related to the problem. In the past, we studied the use of verb taxonomies as stimuli, investigating the use of related verbs that have more specific meaning, i.e., hyponyms, or more general meaning, i.e., hypernyms (Chiu & Shu, 2007b). In this paper, we investigate the effects of dichotomously related stimuli: stimuli that are incongruent with, or disagree with, the problem, and stimuli that are congruent with, or agree with, the given problem. For this investigation, we used think-aloud experiments where participants verbalize their thoughts as they complete the concept generation process.

2 RELATED WORK

2.1 Concept generation

Concept generation is the creative, early stage of the design process, where the goal is to expand the design space by generating as many concepts as possible. Designers are discouraged from fixating on initial ideas and instead encouraged to generate multiple ideas from which the best concept can be selected (Osborne, 1963; Dieter, 2000; Dym & Little, 2003). In concept generation, creative and "wild" ideas are encouraged, and the evaluation of ideas is withheld until later. Techniques that have been developed for generating multiple, creative concepts include "free" brainstorming (Osborne, 1963), synectics (Gordon, 1961), biological analogies (Vakili & Shu, 2001), random stimuli (De Bono, 1992) and TRIZ (Altshuller & Shulyak, 1996).

2.2 Language and design

While language is not usually considered a conventional engineering tool, many researchers have acknowledged the importance of language within design. Language has been applied to requirements gathering (Burg, 1997; Nuseibeh & Easterbrook, 2000), concept generation (Segers, 2004; Chiu & Shu, 2007a,b), design representation (Pahl & Beitz, 1996; Stone & Wood, 2000), and design retrieval and reuse (Stone & Wood, 2000; Yang et al., 2005).

Language is also an important design analysis tool. Analyzing language output from design activities aids in understanding approaches to design (Hey & Agogino, 2007; Grenier & Schmidt, 2007), and provides insight into design team behavior (Mabogunje & Leifer, 1997; Dong et al., 2003; Ji et al., 2007).

Although not specifically a language-based design method, De Bono (1992) suggests that designers relate randomly selected stimuli to their problem to gain new perspectives. The random stimulus could be in the form of a picture from a catalogue or a word from a dictionary. Pictures of an object from a catalogue would likely be represented mentally as a noun. Advantages of incongruent stimuli may be similar to random stimuli, as stimuli in both cases may appear unexpected and non-obvious. However, incongruent stimuli are still related to the problem, and can be systematically generated through lexical sources such as thesauri.

We have previously investigated language as design stimuli. In our biomimetics work, we used language to retrieve relevant biological analogies to be used as design stimuli (Hacco & Shu, 2002; Chiu & Shu, 2007a). Functional keywords expressed as verbs, e.g., "clean" or "remove", were used to retrieve related biological phenomena from natural-language text, e.g., biological phenomena that involve cleaning and removing. In more recent work on the direct use of words as design stimuli, i.e., the previous study that led to the current follow-up investigation, we observed that although participants would often use stimuli as both nouns and verbs, it was verb usage that introduced more new lexicalized concepts into the concept generation process (Chiu & Shu, 2007c).

2.3 Dichotomies in language, reasoning and design

Dichotomy – the division or contrast between two things that are represented as being opposed or entirely different (Oxford, 2003) – appears common in language and reasoning. Antonyms, or opposite words, are universally found in language, and most people demonstrate good intuition in recognizing antonym/synonym pairs (Fellbaum, 1993; Murphy, 2003). For example, while many would agree that the pairs "weighty/weightless" and "heavy/light" convey similar relationships, many will reject "weighty/light" as being an antonym pair. Despite its prevalence in language and wide recognition by language users, antonymy can be difficult to formally categorize and quantify as there are many different types of opposite relations, e.g., contradictory antonyms involving polar opposites such as "succeed/fail" and converse antonyms involving the same action from different perspectives such as "buy/sell" (Fellbaum, 1993). Another obstacle to formal categorization is that antonymy relies on definition by context, such as in a phrase or a sentence (Murphy, 2003). In our case, incongruent stimulus is defined within the context of the problem and by a stimulus word's relationship to other words in the stimulus set.

The concept of disagreement and opposing forces is well established within philosophy. Dialectics, first defined by classical philosophers and then furthered by later philosophers, is based on the use of thesis and antithesis to force synthesis (Lawless, 2005). The concept of opposition is not only found in western philosophy, but in philosophy of other cultures as well.

Design methods using dichotomy include TRIZ, the Russian acronym for TIPS, the Theory of Inventive Problem Solving and argumentative negotiation (Altshuller & Shulyak, 1996; Jin & Geslin, 2007). In TRIZ, the problem is phrased in contradictions to identify parameters to be improved and those degraded as a consequence. Argumentative negotiation, similar to dialectics, involves the verbalization of contradictory demands to move towards agreement and to produce novel solutions in collaborative engineering.

In another one of our recent studies, also partly motivated by the results observed in the previous 2007 experiment, participants were given four problems and associated stimulus sets consisting of words related to the problem both similarly and oppositely in manner. In this pen-and-paper experiment, 42 participants were provided with similar and opposite terms simultaneously. Participants who chose at least one opposite term developed more novel concepts. For one of the problems, the increase in novelty was statistically significant (Chiu & Shu, 2008). Since this most recent previous study was a pen-and-paper experiment where only final results were collected, the follow-up think-aloud experiment reported here will provide more insight into the use of dichotomous stimuli in concept generation.

2.4 Think-aloud experiments in engineering

In a think-aloud experiment, participants verbalize all their thoughts while completing a task. While there are concerns about the validity of think-aloud, e.g., talking about the task will change the task itself, Ericsson & Simon (1993) argue that since verbal on-line reporting draws on short-term memory, i.e., facts and thoughts already present, such verbalization would not alter the thought process. Many precautions were taken to increase the validity of the data. For example, participants were instructed to report only thoughts as they occurred to them and not to plan what they were going to say, nor to judge their thoughts. To discourage conversation, which may affect the participants' thoughts or bias participants, the experimenter sat behind the participant to create the impression that the participant was alone in the room.

Many others have relied on using think-aloud, as it is currently the most feasible and practical way to study mental processes in problem solving and engineering (Rasmussen & Jensen, 1974; Goor & Sommerfeld, 1975; Bhaskar & Simon, 1977; Atman et al., 1997; Benami & Jin, 2002).

The relatively small number of participants in think-aloud studies is generally accepted, provided other methods are used in conjunction (Visser, 2006), which we have and reported elsewhere (Chiu & Shu, 2007b; Chiu & Shu, 2008).

3 MOTIVATION

We are motivated to study the use of lexical stimuli in design because of the established links between language and reasoning, and because language has been recognized as important to design. This specific investigation is motivated by previous results that suggest we continue our study of lexical stimuli by examining dichotomous stimuli, or stimuli that can be classified into opposing categories. Antonyms and synonyms are words that have such a dichotomous relationship and in fact are the only valid verb relationship other than hypernymy/hyponymy. However, there are limitations on antonymy/synonymy such as implication of exact opposition and exclusive binary pairing (Fellbaum, 1993; Murphy, 2003). For example, "hot" and "cold" are considered exact opposites, but while "hot" and "cool" are also incongruent in meaning, it is not quite an antonymous pair. Using only antonym/synonym pairs may result in few stimulus words because of pair restrictions, and many words, especially verbs, lack exact antonyms.

However, we can still investigate dichotomous stimuli by examining words that disagree with the problem, i.e., incongruent stimuli, and words that agree with the problem, i.e., congruent stimuli. The idea of semantic congruency, or sameness in meaning, has been applied to sentence and picture matching tasks where participants compare pictures with sentences to determine if they agree. The sentence may be incongruent with the picture by containing order reversals, logical reversals or negation (Clark & Chase, 1972). Here, we extend the idea of congruency and incongruency to the agreement in meaning between a problem and its stimulus set.

We were motivated to investigate the effects of incongruent stimuli when we could not explain the results of a think-aloud experiment previously reported (Chiu and Shu, 2007c). In the previous experiment, we provided three participants each with three problems and corresponding sets of lexical stimuli related to the keywords of each problem. After conducting think-aloud sessions where participants were told to verbalize all thought processes during concept generation, we categorized how participants used each stimulus word in terms of the part-of-speech, i.e., as noun, verb or noun modifier/adjective. Overall stimulus usage is summarized in Figure 1 below.

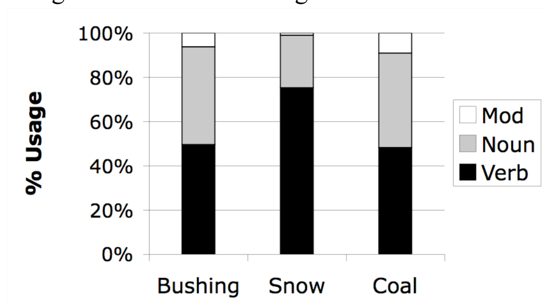

Figure 1: Combined participant POS stimulus use for Participants 1-3 in previous experiment.

We could not explain the increased verb usage of the stimuli for the second, "snow" problem, which was found to be significant (repeated measures ANOVA, $F(2,4) = 17.881$, $p = 0.010 < 0.05$.) Upon re-examining the stimulus set, we noted that while stimulus words were derived from the problem statement for each problem, for the "snow" problem, this led to stimulus words that were incongruent with, or disagreed, with the actual functional goal. Suspecting the difference in verb usage may be related to the difference in stimulus type, we performed a follow-up experiment where a new group of participants were provided with agreeing, or congruent, stimuli for the "snow" problem. Performing the follow-up experiment enables us to compare the effects of incongruent and congruent stimuli for the "snow" problem and to answer the following two questions:

1. Does incongruent meaning in stimuli affect stimulus use in concept generation?
2. Does incongruent meaning in stimuli affect concept novelty?

While there are other metrics associated with evaluating concepts, e.g., quality, quantity and variety (Shah et al, 2000), we feel that novelty – the quality of being new, original or unusual (Oxford, 2003) – is of particular importance to concept generation as it is an important measure of design artifacts (Kan & Gero, 2007).

4 EXPERIMENTAL METHOD

4.1 Participants and raters

Included in this investigation are data from a total of six male participants, all fluent English speakers, recruited from the Faculty of Applied Science and Engineering at the University of Toronto. These participants had varying technical backgrounds in either mechanical or industrial engineering. The three participants of the previous 2007 think-aloud experiment who were given incongruent stimuli for the "snow" problem were fourth-year undergraduate students (Participants 1-3). The three participants of this current follow-up experiment consisted of two first-year Master's students (Participants 4 and 6) and one second-year Ph.D. student (Participant 5). The participants were trained to verbalize prior to the experiment and were paid upon experiment completion.

Two independent raters, a male and a female, were recruited to evaluate the resulting concepts. Both have a graduate level of education in engineering. Rater 1 was in the 2nd year of his Master's program, and Rater 2 had completed her Master's degree. Both raters are familiar with conceptual design. The raters were neither trained prior to concept rating, nor paid.

4.2 Procedure and experimental design

Participants were instructed to "think-aloud" and verbalize all thoughts and reasoning as they worked through the three design problems. They were first given a series of practice arithmetic and word problems to train them in the process of verbalizing while working on a task. The practice problems presented on worksheets allowing for written annotations, e.g., calculations, sketches, etc., were the same between the previous and follow-up experiments.

Three experiment problems were used, the "bushing and pin" problem, the "snow" problem and the "coal" problem. Each experiment problem, also presented on a worksheet, included the problem description and a set of stimulus words generated using keywords related to the problem requirements.

For the "bushing and pin" and "coal" problems, all six participants received congruent, agreeing stimuli. For the "snow" problem, Participants 1-3 received a set of incongruent stimulus words disagreeing with the requirements of the problem, while Participants 4-6 received congruent stimulus words agreeing with the requirements of the problem. Participants were not informed of the relationship between the stimulus set and the problem, and asked to review all stimulus words and develop their concepts based on selected stimuli.

In this paper, only the between-subjects results for the "snow" problem will be compared and described in detail. The results of the "bushing and pin" and "coal" problems are used as a baseline for part-of-speech usage. Here, we are interested in the between-subjects effects of different stimulus types (incongruent and congruent stimuli) tested on two different groups of participants (Participants 1-3 and Participants 4-6) for the "snow" problem. Table 1 below illustrates the experimental design and the gray row highlights the data being compared and discussed in this paper.

Table 1: Experimental design and data.

	Previous Experiment (Participants 1-3)	**Follow-up Experiment (Participants 4-6)**
Problem	***Stimulus Type***	***Stimulus Type***
Bushing	Congruent	Congruent
Snow	Incongruent	Congruent
Coal	Congruent	Congruent

In both experiments, participants were given a total of 15 minutes per problem for concept generation. If they were silent for any length of time, they were prompted to keep talking. The sessions were recorded and transcribed.

The transcriptions were separated into phrases each containing one main verb. We then determined the part-of-speech classes (POS) in which stimulus words were used.

Concepts were first extracted from the transcripts and worksheets, and then provided to the two independent raters. The raters were instructed to score the concepts based on novelty and not to second-guess their initial ratings. The raters were not informed of the identities of the participants, nor were they informed that any stimuli were provided at all. The raters were asked to score the concepts between 0 and 10, with 0 indicating a concept that is not novel at all, 5 indicating a concept that is of average novelty, and 10 to indicate that the concept is extremely novel. Numerical scores were used to facilitate statistical analyses.

4.3 Experiment Problems

Three different problems were used to collect verbalizations and data from the six participants. Problem 2, the "snow" problem, is the problem of interest for this current investigation and thus will be described in detail. The "snow" problem and the associated incongruent and congruent stimulus sets are presented below. The other two problems presented to the participants were the "bushing and pin" alignment problem adapted from Kosse (2004) and the "coal" storage problem, adapted from Dieter (2000). Again, data collected from the "bushing and pin" and "coal" problems served as baseline comparisons for between-subject POS usage.

<u>Snow Problem Statement</u>: In Canada, snow is readily available in the winters and has good insulation qualities due to the amount of air in it. However, if the snow is packed to the point it becomes ice, it is less insulative due to the loss of air. Come up with a concept to enable snow to be used as an additional layer of insulation for houses in the winter.

4.4 Word stimulus sets

WordNet was used as a language framework to generate the related stimulus sets. WordNet is an online lexical database that is organized according to psycholinguistic theories of human lexical memory, where words are stored in hierarchies according to their semantic relatedness to other words (Miller et al., 1993). This is unlike a dictionary where words are organized alphabetically regardless of semantic relationships to adjacent entries. Verbs are stored in hierarchies according to the verb specificity. More general verbs are known as hypernyms, or superordinate verbs, while more specific verbs are known as hyponyms, or subordinate verbs. For example, the verb "to move" is a hypernym of the more specific verb "to walk", and the verb "to amble" is a hyponym of the verb "to walk" because it denotes a more specific manner of walking.

The incongruent stimulus set was generated using the keywords "pack" or "compact", words taken directly from the problem statement. These keywords are incongruent with the problem because the problem statement states that packing the snow removes air and thus the desired insulating qualities of snow. All words in the stimulus set are either hypernyms or hyponyms of these incongruent keywords. The stimulus words are: arrange, bundle, change, compress, constrict, contract, force, impact, move, push, squeeze, tighten and wad. The stimulus words were displayed randomly in a grid for the experiment, but listed above in alphabetical order for brevity.

The second, congruent stimulus set was generated based on the agreeing keyword "insulate". There were few hypernyms/hyponyms in WordNet for "insulate". Therefore, the keywords "surround" and "protect", taken from the WordNet gloss, or definition, of "insulate", were also used to generate a comparably sized stimulus set and one with a similar distribution of words between the different levels of word specificity (WordNet, 3.0). The words from the congruent stimulus set are: blanket, control, cover, defend, enclose, immerse, pack, preserve, prevent, restrain, restrict, submerge, touch. It is interesting to note that "pack" was found in the congruent stimulus set. Having an antonym in the hypernym/hyponym hierarchy is not uncommon, e.g., "rise" and "fall" are co-hyponyms of "move", (Fellbaum, 1993) and agreement and disagreement, like antonymy, is defined within the context of the problem and other stimuli.

All stimulus sets were generated from WordNet verb hierarchies and consist of verbs residing in levels 1 through 4 of the hierarchy. Figures 2 and 3 show partial hierarchies of the incongruent and congruent stimulus sets, respectively. Each stimulus set contained 13 verbs with different levels of specificity and degrees of intransitivity. While every effort was made to balance levels, intransitivity as well as the familiarity versus obscurity of the words selected, the stimulus sets were restricted by the available words with desired properties within WordNet. All stimulus words were presented in root verb format. No definitions or example uses of words were given.

Solid boxes indicate original keywords, and dashed boxes indicate words in the hierarchy used to generate other words. Words in neither dashed nor solid boxes were given as stimuli. The number after the word indicates which meaning, or sense, enumerated in WordNet, was used to generate the hierarchy. A smaller sense number indicates greater familiarity. Figures 2 and 3 show that different, but familiar, senses of "pack" appeared in the two stimulus hierarchies. Sense 1 appears in the congruent stimulus set, i.e., "arrange in a container", and sense 3 appears in the incongruent set, i.e., "compress into a wad".

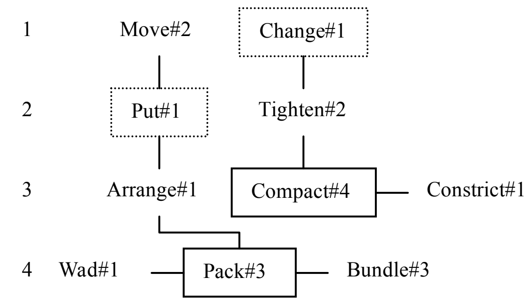

Figure 2: Partial stimulus hierarchy of the incongruent stimulus set.

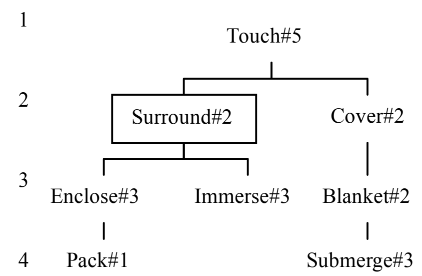

Figure 3: Partial stimulus hierarchy of the congruent stimulus set.

5 OBSERVATIONS AND RESULTS

After transcribing the concept generation sessions, part-of-speech (POS) usage was determined and concepts were summarized and provided to the independent raters. We describe the results below.

5.1 Stimulus part-of-speech use

Stimuli parts-of-speech were determined by generating keyword-in-context (KWIC) lists (Weber, 1990), so that stimulus use could be examined in context of the sentence or phrase. A typical KWIC list is given below for the keyword "constrict" from the incongruent stimulus set:

1. same as CONSTRICT,
2. tighten and CONSTRICT.
3. CONSTRICTING is all about not letting the heat leave.
4. so you want to CONSTRICT the motion of the heat.
5. CONSTRICT, yeah, from compress,

Instances of stimulus use were categorized as noun, verb or modifier. Unspecified use was categorized as "unknown", to describe instances when participants listed words or uttered words without further context. Note that POS usage was only examined for stimulus words provided to participants. A study of overall POS usage may also yield interesting results regarding how designers use language in design. Figure 4 summarizes the part-of-speech stimulus use for each problem in the follow-up experiment. These participants were provided with congruent stimuli for all three problems. A repeated-measures ANOVA comparing stimulus verb usage by the three participants of the follow-up experiment shows no difference in stimulus verb use between the three problems ($F(2,4) = 1.915$, $p = 0.261 > 0.05$). In the previous, original experiment, there was a significant increase in verb use for the "snow" problem where participants were given incongruent stimuli, as shown in Figure 1. For comparison, Figure 4 shows the stimulus POS use for the follow-up experiment.

Figure 5 shows the difference in verb usage between the previous experiment and the follow-up experiment for the "snow" problem.

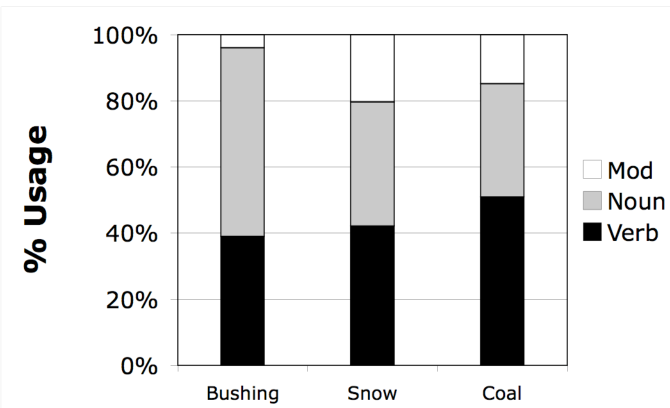

Figure 4: Combined participant POS stimulus use for Participants 4-6 in the follow-up experiment (congruent stimuli.)

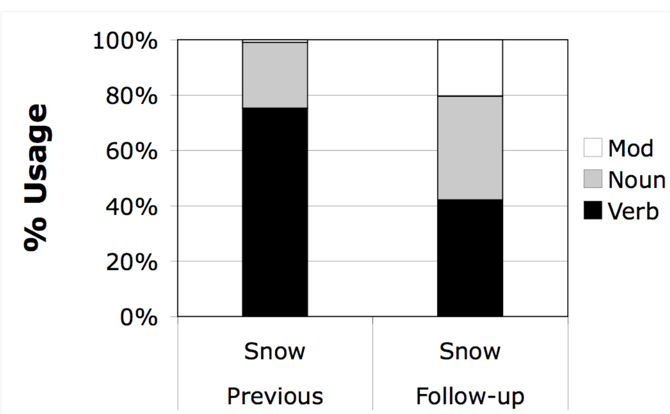

Figure 5: Difference in stimulus POS use for the "snow" problem. Incongruent stimuli from previous experiment on left; congruent stimuli from follow-up experiment on right.

Comparing the percentage of stimuli used as verbs with an independent T-test shows that there is a significant increase in verb use by participants who were given incongruent stimuli in the previous experiment, $t(4) = 5.609$, $p = 0.005 < 0.05$.

5.2 Concept novelty

Concepts were compiled from the transcripts and worksheets, and then provided to the raters. The raters were not given information regarding stimulus type nor associated stimulus words for each concept. The concepts were randomized and sent to the raters via email along with instructions to rate each concept from 0 – not novel, to 10 – very novel. The participant concepts, stimulus type, associated stimulus words and rater scores are given in Table 3. The descriptions of the concepts in this table are the same as the descriptions given to the raters.

Spearman's correlation factor, r, shows a marginally significant correlation between the two raters, $r = 0.51$, $p = 0.054 > 0.05$. Spearman's correlation factor ranges from 0, indicating no correlation, to +/-1, indicating either perfect positive correlation or perfect negative correlation.

Another measure of agreement between two raters is Cohen's kappa, κ, which subtracts the probability that the same rating occurred by chance. The novelty scale of 0-10 was aggregated into the following categories: 0-3 rating – not novel, 4-6 rating - average novelty and 7-10 rating – novel. Cohen's kappa was calculated to be $\kappa = 0.24$, which can be interpreted as "fair agreement" between raters (Landis & Koch, 1977).

The raters scored the concepts developed with incongruent stimuli as more novel than the concepts developed with congruent stimuli. However, a dependent T-test for each rater shows there is no significant difference in concept novelty between the stimulus types. Table 2 summarizes the mean novelty from each rater for each stimulus type.

Table 2: Rater 1 and Rater 2 mean concept novelty.

Rater	Mean Novelty: Incongruent (N=5)	Mean Novelty: Congruent (N=6)	t(9)	p
Rater 1	6.4	4.6	1.09	0.15
Rater 2	4.7	3.7	1.02	0.16

Table 3: Participant concepts sorted by stimulus type

	Concepts and Associated Stimuli	Participant #	Rater 1	Rater 2
Incongruent stimuli	**Concept #3**: Create a snow blanket that can be applied over windows at night. **Associated Stimuli:** Force, bundle, constrict, squeeze	3	3	2
Incongruent stimuli	**Concept #6:** Collect all the snow in one place, then use something like a leaf blower to throw it back out on the parts of the house you want covered. **Associated Stimuli:** Arrange, move	2	10	6
Incongruent stimuli	**Concept #8:** Use a funnel to direct snow down the side of the house walls and pump to re-circulate the snow to prevent compaction. Use sensors to detect fluffiness of snow and something like a blender to chop up ice. Dehumidify the snow so it doesn't get too wet and form into ice. **Associated Stimuli:** Constrict, move, force	1	7	3
Incongruent stimuli	**Concept #9:** Somehow keep layers of snow on the roof, and also force snow around the house by having some sort of outside barrier. The barrier will have a release mechanism to let snow out to prevent too much packing. **Associated Stimuli:** Force, constrict	3	4	5
Incongruent stimuli	**Concept #11**: Make snow bricks and stack them next to the house. Make a snow bricks attachment for your snow blower. **Associated Stimuli:** Compress, compact, constrict, arrange, change	2	8	7
Incongruent stimuli	Average concept novelty for incongruent stimuli		6.4	4.6
Congruent stimuli	**Concept #1**: Have a control system that detects when it rains and then notifies you. Then you can either manually or automatically cover or envelop your house with a plastic sheet so that the snow doesn't get wet. Also detect when there is too much snow, which will cause too much pressure and pack the snow. When this happens, activate the blanket to cover the house to prevent more snow from landing on the snow. **Associated Stimuli:** Control, cover, blanket	4	5	3
Congruent stimuli	**Concept #2:** Create a snow bed to enclose the area you want to insulate. This is a box or barrier to prevent loss of snow in the system. **Associated Stimuli:** Enclose, prevent, restrain, restrict	5	3	3
Congruent stimuli	**Concept #4:** Use artificial snow as an insulator, something that doesn't lose air. **Associated Stimuli:** none	4	8	4
Congruent stimuli	**Concept #5:** Wall spacer – surround the house with another wall with a cavity to collect the snow and have openings at the bottom to allow old snow to be removed. Add a hydrophobic cover on the roof to deflect the snow down the sides into the spacer. Use a control system/device to lift the middle of the cover or blanket on the roof to allow snow to slide down the sides to the space/separator. Cover the spacer to prevent rain and wind from getting into the spacer to preserve the snow from the sun, wind and rain. **Associated Stimuli:** Blanket, preserve, enclose, cover	6	7	4
Congruent stimuli	**Concept #7**: Have a device to retain snow and to remove ice and packed snow and to refresh the snow. This device can monitor condition of the snow and control amount of ice or water and removes that and refreshes the snow. **Associated Stimuli:** Submerge, prevent, preserve, control	5	3	3
Congruent stimuli	**Concept #10:** Use a blanket to cover the snow, to insulate the snow layer. **Associated Stimuli:** Cover, blanket	5	2	5
Congruent stimuli	Average concept novelty for congruent stimuli		4.7	3.7

6 DISCUSSION

In this section, we first discuss POS usage and then concept novelty. We highlight the increased verb usage observed in association with incongruent stimuli, and provide evidence from our own and others' previous work suggesting that verbs play an important role in concept generation. Then we consider concept novelty while examining specific instances of agreement between raters.

6.1 Stimulus POS

After completing the previous experiment with Participants 1-3, who were inadvertently provided with incongruent stimuli for the "snow" problem, we noticed that the verb use was higher for that specific problem. Possible contributions to this difference composed prior to the follow-up experiment were:

1. The stimulus set was incongruent with, or disagreed with the problem.
2. The problem was more novel than the other two problems, which were more likely to have been encountered in engineering curricula.

Upon analysis of the follow-up experiment data, we found that the subsequent subjects, Participants 4-6, who were provided with congruent stimuli, did not demonstrate the same increase in verb use for the "snow" problem. Therefore, it appears that increased verb usage may be attributed to the difference in stimulus type rather than to differences in problem novelty or type.

Our previous work showed that verb use in concept generation is key to introducing new arguments (words or phases) into the concept generation process. We found that verbs introduced 65.5% of new arguments, while nouns and noun modifiers only introduced 18.9% and 15.5% of new arguments respectively. These new arguments are words and phrases that represent new lexicalized concepts, or ideas. The following examples identify the stimuli used and the resulting new arguments. The stimuli used are given in italics and the new arguments are underlined.

"*Constrict* the <u>motion of heat</u> [from leaving the house]"

"[using a] <u>funnel</u> to *constrict* snow"

"*arrange* snow <u>particles</u>"

These new arguments (words or phrases) are not given in the problem statement, nor are they only related to the general process of reasoning about the problem, as clarified below.

Stimulus words used as nouns were often done so in the context of reasoning about the word itself or the general problem solving process. The following examples demonstrate reasoning about the use of stimulus words "touch" and "prevent":

"…the word *touch"*

"so the word *prevent* is telling me that this concept needs to *prevent* two things…"

In their work on evaluating design group success in early phases, Mabogunje and Leifer (1997) found that an increased number of distinct noun phrases is a good indicator of success. They found that noun phrases changed throughout the design process whereas verb phrases associated with the design remained relatively fixed. Even though the verbs, or the functions, of a design can be fixed early on, the syntax of sentences shows that it is the verb that introduces new noun phrases and relates noun phrases to other noun phrases. This is shown in the language equations below:

$$(1)\ S \rightarrow NP + VP$$
$$(2)\ NP \rightarrow N + PP$$
$$(3)\ PP \rightarrow P + NP$$
$$(4)\ VP \rightarrow V + NP$$

Where S = sentence, NP = Noun phrase, N = Noun, VP = verb phrase, V= Verb, PP = Prepositional phrase, P = Preposition. Substituting (4) into (1) gives (5):

$$(5)\ S \rightarrow NP + (V + NP)$$

Related work in linguistics also shows that verbs have properties that may be advantageous for concept generation. The first such property is flexibility and mutability, where verbs can take on slightly different meanings depending on their arguments (Gentner and France, 1988). Thus, verbs are more flexible in meaning and can be used in a wider range of phrases while still "making sense". It was seen above that participants were able to apply the word "constrict" to either heat or snow in phrases that made sense grammatically and semantically.

Another property of verbs is that they invoke the inclusion of semantic filler roles (McRae, et al., 2005). These semantic filler roles include agent roles, patient roles, instrument roles and location roles. Table 5 gives typical examples for each filler role.

Table 5: Examples of verbs and semantic filler roles.

Verb	Filler role
Hammer	Agent – carpenter
Throw	Patient – ball
Chop	Instrument – axe
Eat	Location – cafeteria

This suggests that once participants use stimuli as verbs, semantic filler roles associated with the verb can be easily introduced into the concept generation process, as the participants are primed to provide the filler roles.

In general, verbs can evoke more relational information, whether syntactically, e.g., through the ability to introduce and relate new noun phrases, or semantically, e.g., through introduction of new filler roles. Our results suggest that providing incongruent stimuli increases their use as verbs, which in turn increases the introduction of new words and phrases to the concept generation process. Because incongruent stimuli may appear unexpected and puzzling to the designer, it may force designers to reason more to reconcile the differences between the stimuli and problem statement. Since verbs are relational, more verb use may be required to relate the disparate stimuli and problem statement.

Comparing Figure 1 and Figure 4 also reveals that there was an increase in modifier use in the follow-up experiment where participants were provided with congruent stimuli. Previously, we found that while use of stimulus words as modifiers was rare, modifiers introduced a high proportion of new arguments. However, in the current follow-up experiment, many instances of modifier use were attributed to the use of "pack" as a modifier for "snow", as either a pre-modifying adjective e.g., "the packed snow", or as a post-modifying adjective e.g., "the snow is packed". Since "pack" was used consistently with the problem statement, the use of "pack" as a modifier did not introduce new arguments in the process. This supports that context from the rest of the stimulus set is important to defining the incongruency or congruency of individual stimulus words.

The participants' use of "pack" may reflect the priming effect of the modifier use of "pack" in the problem statement. Other studies show that when English speakers are provided with an adjective and asked to respond with *another word*, 65% of the responses were also in the adjective class. When provided with nouns, 79% of the response words were nouns, and when provided with verbs, 43% of the response words were verbs (Fillenbaum & Jones, 1965).

6.2 Concept novelty

Using two or more raters and determining inter-rater agreeability is common practice in medical diagnostics and psychological studies (Landis & Koch, 1977; Shrout & Fleiss, 1979). Others have used this practice in engineering design studies (Perttulla & Liikkanen, 2006). A generally accepted "good" level of inter-rater reliability is in the range of $\kappa > 0.75$, while Altman (1991) suggests that $\kappa < 0.40$ is poor agreement. However Landis and Koch (1977) suggest that the reliability of $\kappa = 0.24$ found between our two raters is "fair agreement". Landis and Koch's scale is as follows:

- Poor agreement = Less than 0.20
- Fair agreement = 0.20 to 0.40
- Moderate agreement = 0.40 to 0.60
- Good agreement = 0.60 to 0.80
- Very good agreement = 0.80 to 1.00

In many applications, such as in medical diagnostics, raters are first trained, which influences and increases inter-rater reliability. The raters in this experiment were not trained to prevent biasing them towards what the experimenters consider "novel" and they were instructed to not second-guess their ratings. Concept evaluation is a difficult task, and the concepts presented here have different levels of detail, making it challenging to compare concept to concept. Therefore, a fair level of agreement on a difficult rating task seems acceptable at this point of the study. Furthermore, the difference between the raters was more related to the magnitude of the score they assigned concepts, rather than their relative ranking of these concepts. In general, Rater 1 had a larger spread in his overall novelty scores (SD=2.66), while Rater 2 had a smaller spread in her overall novelty scores (SD=1.51).

Both raters had a higher mean novelty for concepts developed from incongruent stimuli, although the difference is not statistically significant. Both raters scored Concepts 11 and 6, the "snow bricks" and "snow throwing" concepts, as being amongst the top three most novel concepts. Rater 1 ranked the above concepts as his second (tied) and first most novel concepts, and Rater 2 ranked the above concepts as her first and second most novel concepts respectively. Participant 2, who was provided with incongruent stimuli, generated both concepts. It is possible that Participant 2 is overall a better designer and more creative. However, all participants were volunteers, and were informed that this was a design study prior to the start of the experiment, and thus we assume were suitably competent and motivated.

Both raters also ranked the same concepts as being amongst the least novel. Rater 1 scored Concepts 2, 3, 7, concepts that describe retaining fallen snow on specific areas of the house, such that these concepts are in a three-way tie for second least novel. Rater 2 rated the same three concepts as amongst her least novel as well, with Concepts 2 and 7 tied for second least novel and Concept 3 as least novel. Concepts 2 and 7 were developed from congruent stimuli, while Concept 3 was developed from incongruent stimuli.

Incongruent stimuli appear to increase concept novelty. While the difference is not significant, the independent raters agreed on their rankings of the most and least novel concepts. We hypothesize that the increased verb use observed with the use of incongruent stimuli promotes the introduction of new lexicalized concepts into the concept generation process. This in turn contributes to increased concept novelty.

7 SUMMARY AND FUTURE WORK

We are motivated to investigate language as design stimuli because of the relationship between language and reasoning. We investigate incongruently and congruently related stimuli because of the importance and prevalence of dichotomous relations in language and reasoning, e.g., antonyms and dialectics. Incongruent stimuli may also offer advantages of random stimuli such as being non-obvious and unexpected. However, incongruent stimuli are related to the problem and thus can be systematically generated from lexical resources. We decided to complete a follow-up study in an attempt to explain increased verb use for a problem where we inadvertently provided incongruently related stimuli. Completion of this follow-up experiment allowed us to compare participant behavior and concepts resulting from dichotomous stimuli to answer the following questions:

1. Does incongruent meaning in stimuli affect stimulus use in concept generation?
2. Does incongruent meaning in stimuli affect concept novelty?

Our answers are as follows:

1. POS analysis shows that participants provided with incongruent stimuli used the stimuli more often as verbs.
2. Raters scored concepts developed from incongruent stimuli as more novel.

Many agree that verbs are the best part-of-speech for design description because verbs can convey the actions of functions. Our previous work showed that stimuli used as verbs introduced more new words or phrases into the concept generation process than nouns. Additionally, related work in linguistics show that verb meaning is more mutable and flexible, and verbs can stimulate the inclusion of semantic filler roles. The relational nature of verbs allows different words and phrases to be related to each other. This combined work suggests that increased verb use of stimuli is desirable for concept generation. Participants provided with incongruent stimuli used the stimuli as verbs more often, perhaps because verbs enable the forming of relationships between the incongruent stimuli and the problem. The unexpected nature of the incongruent stimuli may have forced participants to make new connections to solve the problem, introducing novelty in the process.

While the stimulus sets were incongruently and congruently related within the context of the problem, they were not dichotomously related to each other, i.e., “to pack” is not the direct opposite of “to insulate”. Future work could include using stimulus sets that are in direct opposition to each other. Future experiments may also benefit from having fewer words in the stimulus set as we try to determine which words are best as stimuli.

Regarding concept evaluation, future experiments could include training for the raters. The challenge is to train the raters without biasing them. In experiments where a large number of concepts are collected, it is possible to have novelty defined based on the number of distinct concepts per concept category, as done in our recent work (Chiu and Shu, 2008). However, in think-aloud experiments where there is typically a smaller sample size, (time and resource limitations prevent the use of a large number of participants), novelty as perceived by independent raters may be the best approach. It may be useful to have more than two raters. However, related work used only two independent raters for many tasks.

Overall, incongruently related lexical stimuli, or stimuli that disagree with the problem statement, encouraged increased verb use of stimuli, and the subsequent concepts demonstrated increased novelty. Continued work will seek to understand the relationship between language and design, and specifically how this relationship can be used to facilitate conceptual design and to generate more novel concepts.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC), the participants and raters of this experiment.

REFERENCES

Atman, Cynthia J., Chimka, Justin R. and Karen M. Bursic, 1997, Results from a verbal protocol study of the design process, Proc. of Frontiers in Education Conf., v.3, p.1283.

Altman, D.G., 1991, Practical Statistics for Medical Research. London England: Chapman and Hall.

Altshuller, G.S., Shulyak, L., 1996, And Suddenly the Inventor Appeared: TRIZ, the Theory of Inventive Problem Solving, 2/e, Technical Innovation Center, Worcester, MA.

Benami, O., Jin, Y., 2002, Creative stimulation in conceptual design, ASME DETC/CIE, Montreal, Canada, DETC2002/DTM-34023.

Bhaskar, R., Simon, H. A., 1977, Problem solving in semantically rich domains: An example from engineering thermodynamics, Cognitive Science, 1:193-215.

Burg, J.F.M., 1997, Linguistic Instruments In Requirements Engineering, Vrije Universiteit, Amsterdam, PhD thesis.

Chiu, I., Shu, L.H., 2007a, Biomimetic design through natural language analysis to facilitate cross-domain information retrieval, Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 21/1:45-59.

Chiu, I., Shu, L.H., 2007b, Using language as related stimuli for concept generation, Artificial Intelligence for Engineering Design Analysis & Manufacturing, 21/2:103-121.

Chiu, I., Shu, L.H., 2007c, Understanding the use of language stimuli in concept generation, ASME IDETC/CIE, Las Vegas, NV, IDETC2007-35772.

Chiu, I., Shu, L.H., 2008, Use of opposite-relation lexical stimuli in concept generation, Annals of CIRP, 57/1, in press.

Clark, H.H., Chase, W.G., 1972, On the process of comparing sentences against pictures, Cognitive Psychology, 3:472-517

De Bono, E., 1992, Serious Creativity, HarperCollins, NY.

Dieter, G.E., 2000, Engineering Design: A Materials and Processing Approach, *3rd Edition*, McGraw-Hill, NY.

Dym, C.L., Little, P., 2004, Engineering Design, A Project-Based Introduction, 2nd Edition, John Wiley & Sons, Inc., NJ.

Dong, A., Hill, A.W., Agogino, A.M., 2003, A document analysis method for characterizing design team performance, Journal of Mechanical Design, 126/3:378-385.

Ericsson, K.A., Simon, H.A., 1993, Protocol Analysis: Verbal Reports as Data, MIT Press, Cambridge, MA.

Fellbaum, C., 1993, English Verbs as a Semantic Net, Five Papers on WordNet, pp. 40-61. ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps.

Fillenbaum, S., Jones, L. V., 1965, Grammatical contingencies in word association, Journal of Verbal Learning and Verbal Behavior, 4: 248-255.

Gentner, D., France, I., 1988, The verb mutability effect: studies of the combinatorial semantics of nouns and verbs, in *Lexical Ambiguity Resolution* (Smal, S., Cottrell, G., Tanenhaus, M., Eds). Morgan Kaufmann, Los Altos, CA.

Goor, A., Sommerfeld, R. E., 1975, A comparison of problem-solving processes of creative students and noncreative students, Journal of Educational Psychology, 67(4), 495-505.

Gordon, W. J. J., 1961, Synectics, The Development of Creative Capacity, Harper & Row, New York, NY.

Grenier, A.L., Schmidt, L.C., 2007, Analysis of engineering design journal sketches and notations, ASME DETC/CIE, Las Vegas, NV, Sept 4-7, 2007, DETC2007-35360.

Hacco, E., Shu, L.H., 2002, Biomimetic concept generation applied to design for remanufacture, ASME DETC/CIE, Montreal, Canada, DETC2002/DFM-34177.

Hey, J.H.G., Agogino, A.M., 2007, Metaphors in conceptual design, ASME DETC/CIE, Las Vegas, NV, Sept 4-7, 2007, DETC2007-34874.

Jackendoff, R., 1983, Semantics and Cognition, MIT Press Cambridge, MA.

Ji, H., Yang, M.C., Honda, T., 2007, A probabilistic approach for extracting design preferences from design team discussion, ASME DETC/CIE, Las Vegas, NV, Sept 4-7, 2007, DETC2007-35455.

Jin, Y., Geslin, M., 2007, Study of argumentative negotiation in collaborative engineering, ASME DETC/CIE, Las Vegas, NV, DETC2007-35710.

Kosse, V., 2004, Solving Problems With TRIZ: An Exercise Handbook, 2nd Edition. Ideation Int'l Inc., Southfield, MI.

Landis J.R., Koch G.G., 1977, The measurement of observer agreement for categorical data, Biometrics, 33:159-174.

Lawless, A., 2005, Plato's Sun, An Introduction to Philosophy. University of Toronto Press, Toronto, ON.

Levinson, S., 1996, Language and Space, Annual Review of Anthropology, 25:353-382.

Li, W., Jin, Y., 2006. Fuzzy preference evaluation for hierarchical co-evolutionary design concept generation, ASME IDECT/CIE, Philadelphia, PA, DETC2006-99312.

Mabogunje, A., Leifer, L., 1997, Noun phrases as surrogates for measuring early phases of the mechanical design process, ASME DETC/CIE 1997, Sacramento, CA.

McRae, K., Hare, M., Elman, J.L., Ferretti, T., 2005, A basis for generating expectancies for verbs from nouns, Memory & Cognition, 33:7/1174-1184.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K., 1993, Introduction to WordNet: an on-line lexical database, in Five papers on WordNet, pp. 1-25, ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps.

Murphy, M.L., 2003, Semantic Relations and the Lexicon, Antonymy, Synonymy, and Other Paradigms, Cambridge University Press, Cambridge, UK.

Nuseibeh, B., Easterbrook, S., Requirements engineering: A roadmap, in A.C.W. Finkelstein (ed), The Future of Software Engineering, IEEE Computer Society Press.

Osborne, A. F., 1963, Applied Imagination: Principles and Procedures of Creative Problem Solving, Charles Scribner's Sons, New York, NY.

Oxford, 2003, The Oxford American Dictionary and Thesaurus, Oxford University Press, New York.

Pahl, G., Beitz, W., 1996, Engineering Design, A Systematic Approach, K. Wallace, L. Blessing, and F. Bauert, Trans., K. Wallace, Ed., 2/e., Springer-Verlag London Ltd., London, UK.

Perttula, M.K., Liikkanen, L.A., 2006 Structural tendencies and exposure effects in design idea generation, ASME DETC/CIE 2006, Philadelphia, PA, DETC2006-99123.

Pinker, S., 2007, The Stuff of Thought, Language as a Window into Human Nature, Viking, New York, NY.

Rasmussen, J., Jensen, A., 1974, Mental procedures in real life tasks: A case study in electronics trouble shooting, Ergonomics, 17:293-30.

Segers, N., 2004, Computational representations of words & representations of words & associations in architectural design: Development of a system to support creative design, Bouwstenen 78, Technische Universiteit Eindhoven, Ph.D. Thesis.

Shah, J.J., Kulkarni, S.V., Vargas-Hernandez, N., 2000, Evaluation of idea generation methods for conceptual design: effectiveness metrics and design of experiments, Journal of Mechanical Design, 122:377-384.

Shrout, P.E., Fleiss, J.L., 1979, Intraclass correlations: Uses in assessing rater reliability, Psychological Bulletin, 86/2: 420-428.

Stone, R.B., Wood, K.L., 2000, Development of a Functional Basis for design, Journal of Mechanical Design, Transactions of the ASME, 122:359-369.

Ullman, D., 2003, The Mechanical Design Process, Third Edition, McGraw-Hill, New York, NY.

Vakili, V., Shu, L.H., 2001, Towards biomimetic concept generation, ASME DETC/CIE, Pittsburg, PA. DETC2001/DTM-21715.

Visser, W., 2006, The Cognitive Artifacts of Designing, Lawrence Erlbaum Associates, Publishers, Mahwah, NJ.

Yang, M.C., Wood, W., Cutkosky, M., 2005, Design information retrieval: A thesauri-based approach for reuse of information design information, Engineering with Computers, 21:177-192.

Weber, R. P., 1990. Basic Content Analysis, 2nd Edition. Sage Publications, Newbury Park, CA.

WordNet, 3.0, http://www.cogsci.princeton.edu/~wn.