The Role of Function, Homogeneity and Syntax in Creative Performance on the Uses of Objects Task

by

Evelyn Forster

A thesis submitted in conformity with the requirements for the degree of Masters
Psychology
University of Toronto

© Copyright by Evelyn Forster 2008
The Role of Function, Homogeneity and Syntax in Creative Performance on the Uses of Objects Task

Evelyn Forster

Masters

Psychology
University of Toronto

2008

Abstract

The Uses of Objects Task is a widely used assessment of creative performance, but it relies on subjective scoring methods for evaluation. A new version of the task was devised using Latent Semantic Analysis (LSA), a computational tool used to measure semantic distance. 135 participants provided as many creative uses for as they could for 20 separate objects. Responses were analyzed for strategy use, category switching, variety, and originality of responses, as well as subjective measure of creativity by independent raters. The LSA originality measure was more reliable than the subjective measure, and values averaged over participants correlated with both subjective evaluations and self-assessment of creativity. The score appeared to successfully isolate the creativity of the people themselves, rather than the potential creativity afforded by a given object.
# Table of Contents

List of Tables .................................................................................................................................................. v

List of Appendices ......................................................................................................................................... vi

The Role of Function, Homogeneity and Syntax in Creative Performance on the Uses of Objects Task .................................................................................................................................................. 1

1.1 The Uses of Objects Task .......................................................................................................................... 3

1.2 Mass and count nouns .................................................................................................................................. 4

1.2.1 Individuation and count-mass syntax ......................................................................................................... 5

1.2.2 Individuation and Feature Salience ........................................................................................................... 6

1.2.3 Individuation and function .......................................................................................................................... 7

1.2.4 Syntax and Feature Salience ........................................................................................................................ 9

1.2.5 Feature Salience in Superordinates ........................................................................................................... 11

1.2.6 Syntax, Function and Creativity ................................................................................................................ 12

2 Purpose .......................................................................................................................................................... 13

3 Experiment 1 ................................................................................................................................................... 15

3.1 Method ........................................................................................................................................................ 15

3.1.1 Participants ............................................................................................................................................... 15

3.1.2 Apparatus ................................................................................................................................................. 15

3.1.3 Procedure ................................................................................................................................................ 15

3.1.4 Calculation of Concept Properties .......................................................................................................... 16

3.2 Results ........................................................................................................................................................ 19

3.2.1 Homogeneity and Functionality Consensus ............................................................................................... 19

3.2.2 Properties associated with Mass-Count Status ......................................................................................... 20

3.2.3 Properties associated with Usenet Ratio ................................................................................................. 21

3.2.4 Distinctiveness ........................................................................................................................................ 22

3.3 Discussion .................................................................................................................................................... 23
3.3.1 Validity of homogeneity and functionality consensus ratings, mass-count status and Usenet ratio .......................................................... 25
3.3.2 Properties defining mass and count nouns ............................................. 25
3.3.3 Distinctiveness Measures ...................................................................... 26

4 Experiment 2 ........................................................................................................ 26

4.1 Method ............................................................................................................... 26

4.1.1 Participants ................................................................................................... 26
4.1.2 Apparatus ..................................................................................................... 27
4.1.3 Stimuli ........................................................................................................... 27
4.1.4 Procedure ..................................................................................................... 27

4.2 Results ................................................................................................................ 28

4.2.1 Scoring Responses ....................................................................................... 28
4.2.2 Reliability of Creativity Scores .................................................................... 29
4.2.3 Analyzing LSA Scores by User ..................................................................... 32
4.2.4 Analyzing LSA Scores by Object .................................................................. 34

4.3 Discussion .......................................................................................................... 38

4.3.1 Reliability and Validity of creativity scores .................................................. 38
4.3.2 Similarity of Group 1 and 2 stimulus properties and user responses ............... 38
4.3.3 Differences between environment and task conditions .................................. 38
4.3.4 What the test says about objects .................................................................. 39
4.3.5 What the test says about people ................................................................... 39
4.3.6 Strategy use ................................................................................................... 40
4.3.7 Future Study ................................................................................................ 40

5 General Discussion ................................................................................................ 41

References .............................................................................................................. 42

Appendices ............................................................................................................ 46
List of Tables

Table 1: Methods of qualitative matching for count and mass nouns according to study

Table 2: Correlations between distinctiveness measures for all objects.

Table 3: Correlations between distinctiveness measures for basic objects.

Table 4: Correlations between LSA scores and fluency, when averaged over objects.

Table 5: Correlations between LSA scores and fluency, when averaged over users.

Table 6: Correlations between LSA scores and object properties.
List of Appendices

Appendix 1: Question posed of homogeneity raters

Appendix 2: Question posed of functionality raters

Appendix 3: Instructions for the Uses of Objects Task

Appendix 4: Think Aloud Instructions

Appendix 5: Think Aloud Cryptarithm

Appendix 6: List of Concepts
The Role of Function, Homogeneity and Syntax in Creative Performance on the Uses of Objects Task

Researchers have long grappled with the problem of creativity assessment: what characteristics are necessary for creativity, and how they can be measured. Studies have been performed to determine what environments best encourage creativity (Amabile, Conti, Coon, Lazenby, & Herron, 1996), what personality traits correlate with it (Helson, 1996; McCrae, 1987) and how it can be squelched (Amabile, 1998), but little can be concluded from these studies if there are still open questions about the validity and reliability of the creativity tests that are implemented.

The two most prominent sides in this testing controversy are the psychometric approach (exemplified by Guilford, 1967), and what Sternberg (1999) calls the confluence approach. The psychometric approach attempts to reduce creativity to its bare bones, and to develop measurement techniques of very specific creative performances that are unambiguous and quantifiable (Plucker & Renzulli, 1999). The confluence approach goes in the opposite direction, assessing activities that combine many skills such as writing a story, drawing a picture or creating an advertisement (Sternberg & Grigorenko, 2002). For years, little progress has been made in determining the most effective method of creativity assessment (especially one that is both objective and quick to perform and score). This study will focus on improving the accuracy and objectivity of the Uses of Objects (UoO) Task, by determining the influence that the count-mass status of a stimulus has on task performance and comparing results to other psychometric and confluence tests.

Two factors have held back the field for a very long time: a) many factors influence creativity, and b) there are many different ways to be creative. This means that assessments for one person can vary quite dramatically from one day to the next, and results on a scientific or mathematic creativity test will vary greatly from results on an artistic creativity test. Sternberg and Lubart (1992) writes that creativity is a function of six variables: intelligence, knowledge, thinking style, personality, motivation and environmental context. Each of these can fluctuate from day to day, due to changes both in a person’s internal and external environment.

The psychometric approach solves the too-many-factors problem by administering a large battery of short, specific and controlled tests, in the hopes that all aspects of creativity will be
Guilford and Hoepfner (1966) provide 57 psychometric techniques, which ask participants to perform tasks such as grouping and regrouping objects according to common properties and listing the consequences of some sort of unlikely change (such as people no longer needing sleep). These tests are incorporated into Guilford’s (1967) Structure of the Intellect (SOI) tests, which can be broken down into four factors: fluency, flexibility, originality, and elaboration of ideas. Torrance’s (1962, 1974) Tests of Creative Thinking (TTCT) are roughly based on the SOI tests, adding two additional factors (the ability to resist premature closure, and the ability to abstract). Although a large battery of tests cannot necessarily control for factors such as motivation and environmental context, a number of different tests could account for differences in thinking style, personality, and knowledge.

Confluence tests solve the too-many-factors problem by integrating many components of creativity together into two or three long, complex tasks. Whereas original idea generation and flexibility are the primary abilities being measured in psychometric tests, confluence tests can also witness the pruning of ideas, and can follow creativity on a slightly larger scale.

To produce a truly creative work, however (such as a painting, story, film, invention) one must watch it evolve over several months, not twenty minutes. Even Rainer Maria Rilke’s First Elegy, which was written at lightning speed during a burst of creativity, took a day to complete (Freedman, 1998). By using a confluence test one runs the risk of thinking that true creativity is being measured, simply by virtue of it being integrated and natural. But there is not enough time given to develop any sophisticated work, or to see an invention play out over repeated testing and modification.

Another drawback is that confluence tests only test certain domains (humour and storytelling perhaps, or problem solving) and often require specific skills on behalf of the participant. Given that there is the danger with these tests of putting all of one’s eggs into only two or three baskets, the danger is worsened by complicating each test with knowledge or talent requirements. In tests that require specific skills, participants may be limited in their drawing or writing ability, musical knowledge, physical agility or mathematical ability, which could prevent them from performing to their best potential. In other cases, some numerical tests could be mastered simply by brute force and may not actually test creativity. These types of skill requirements are exemplified by two psychometric tests in Guilford and Hoepfner (1966). Two of their mathematical tests have
reliabilities of .80 (Number Rules) and .82 (Number Combinations), and are the first and third most reliable of all 57 tests. The principal components analysis performed in the study showed they both load onto the same group (DSR-Divergent production of symbolic relations), suggesting they measure similar attributes. Those with better intuition about number relations may do consistently better than those with less mathematical ability, leading to the higher reliability reported. Other tests of numerical relations also load onto this group but were predicted to load onto others, suggesting that numerical skill effects may be overwhelming the skills they were intended to test.

Because there is a reliance on domain-specific knowledge, creativity in one domain may not match a person’s creativity in others, requiring that testing incorporate all domains to be a fair assessment. Whereas longer tests can only cover two or three different domains, psychometric batteries can accommodate this by including dozens of different tests.

A third possible solution to the too-many-factors problem is to isolate those factors and create a set of tasks that assess each one. If creativity is a function of problem formulation and idea generation, selection and modification (as suggested by Hayes, 1989) or of fluency, flexibility, originality, and elaboration (as suggested by Guilford, 1967), the solution may be to develop a single, fundamental test for each of these processes and administer them in series, allowing participants to fully develop creative ideas in a carefully controlled environment, and measuring progress continuously. An ideal psychometric test of idea generation would be the UoO Task.

1.1 The Uses of Objects Task

The Uses of Objects Task requires participants to generate as many uses for a given object as possible (allowing for disassembly and reconstruction of objects). Like most psychometric tests, it is quick and easy to administer. It is also quite objective, as scoring is by the number or variety of answers given. The task is widely used, but has the same problem with low reliability as most other tests of creativity.

If a certain use for an object is particularly obvious, it may inhibit the generation of other uses. If the most typical functions of some objects (a hammer is used for hammering) overshadow other possibilities (a hammer could be used for painting), variance in the degree of overshadowing could cause reliability issues. Objects may also vary in composition and homogeneity (many vs.
few unique components), which could also affect idea generation. Not only could object functions have an inhibitory effect on alternate uses, but function may also have an effect on the salience of features. If features such as material are activated over features such as shape, it may be easier to generate structure-independent uses for objects, resulting in a larger number of creative ideas.

Both of these variations are related to count-mass status: count nouns tend to have more obvious functions, and mass nouns tend to be more homogeneous. If people differ in the strategies they use to perform the UoO task, some strategies may be more conducive to count nouns and others to mass nouns, causing an imbalance in testing if the count-mass status of objects is not controlled.

Object familiarity could also play a part in the quality and number of inferences made from object categories. Smith, Shafir & Osherson (1993) found that when assessing the quality of arguments based on category relations (e.g. “Robins and ducks have sesamoid bones; therefore, all birds have sesamoid bones”), judgments were based on similarity between the premise and conclusion categories (robins, ducks and birds). The plausibility of premises and conclusions were only taken into account when the properties (sesamoid bones) were familiar. Thus, if participants are generating uses of objects that they may not understand very well (perhaps involving nuclear chemistry or the construction of musical instruments) or working with unfamiliar objects (such as a theremin or a piece of Boron), their ability to judge the quality of their responses may be diminished.

1.2 Mass and count nouns

Relationships between objects and their linguistic and cognitive representations are clearest among basic-level mass and count nouns, such as cement (a mass noun) and a dog (a count noun). The behaviour of other types of mass and count nouns is less predictable.

No studies have been found that link feature salience or object function to performance in a task such as the UoO task, but it is predicted that high performance on the task is associated with the use of structure-independent functions instead of structure-dependent functions, and the activation of a variety of features such as material, weight, and colour (which are activated by substances) rather than a single property such as shape (which is activated by individual objects).
Many studies have linked feature salience and object function to individuation and structural complexity. Individuation is, in turn, related to count-mass syntax.

### 1.2.1 Individuation and count-mass syntax

Whereas the distinction between mass and count nouns was originally considered rather arbitrary (Bloomfield, 1933; Gleason, 1969; Markman, 1985; McCawley, 1975; Palmer, 1971; Quine, 1960; Ware, 1979; Whorf, 1962), the prevailing view recently has been that count nouns and mass nouns are distinguished by their individuality (Bloom, 1999; Bloom & Kelemen, 1995; Bunt, 1985; Wisniewski, Imai, & Casey, 1996; Wisniewski, Lamb, & Middleton, 2003). This is referred to by Wisniewski, Lamb, & Middleton as the Cognitive Individuation Hypothesis.

Count nouns are discrete and bounded (Allan, 1980; Barner, 2003, 2005; Gordon, 1985; Joosten, 2003; Wisniewski, Lamb, & Middleton, 2003), and can therefore be quantified as individual, countable things. This is why count nouns can be preceded by numbers and words such as few or many, whereas mass nouns are preceded by words like much and some (“I have five candy bars”; “I have some water”; “I have many yachts”). Only count nouns can be used in both singular and plural form, and so it is inappropriate to say “I have gravels in my shoe,” because gravel is a mass noun.

Mass nouns tend to describe homogeneous and unbounded substances. They are arbitrarily divisible in that dividing clay (for example) in half will still result in clay (Gillon, Kehayia & Taler, 1999; Middleton et al., 2004; Wisniewski, Lamb, & Middleton, 2003). Comparatively, dividing a dog in half will result in the possession of half a dog. In a slight modification of this notion, Barner and Snedeker (2006) argue that mass nouns are simply nouns that lack enumerating features. (This is why people are forced to use “three pieces of cake” or “three grams of millet” instead of “three pencils” when describing the quantity of a mass noun.) Mass nouns may label objects or groups of objects because they are less individuated than count nouns, but not because they are completely non-individuated. A prime example is the mass superordinate (such as furniture), which is composed of distinct objects but treated as a mass noun.
1.2.2 Individuation and Feature Salience

To understand the set of salient qualitative features being summoned by an object, a matching paradigm is often used. Some studies offer a standard object and the choice of two alternatives, which correspond to two different properties. The selection of a particular type of match over another offers an idea of the relative salience of the two properties. The properties examined are usually those associated with shape or material. By varying the individuation of different types of objects, substances and categories, relationships can be determined between individuation and feature use. Cross-cultural studies suggest that English speakers pay closest attention to shape when dealing with individuated objects, and that speakers of numeral classifier languages (those that do not distinguish between count and mass nouns, such as Japanese or Yucatec Mayan) have a lower tendency to individuate objects and pay closer attention to material (Imai & Gentner, 1997; Lucy, 1992; Mufwene, 1984). Despite a lower tendency to individuate, speakers of numeral classifier languages can still agree on the mass/count status of an object (Markman, 1985). They are still able to differentiate individuated and non-individuated objects, but for cultural or linguistic reasons it is less salient to them.

Imai and Gentner (1997) studied English and Japanese speakers, showing them complex objects (which had distinct parts and distinct functions associated with them, such as a lemon juicer), simple objects (which were solid objects with distinct shapes, such as wax in the shape of a kidney), and substances in the shape of complex objects (such as sand in an S-shape). Participants were shown one of the three, as well as two alternatives: one that matched in shape but not material, and one that matched in material but not shape. They were then asked which of the second and third objects best matched the first. Both Japanese and English speakers chose to match by shape for complex objects. For simple objects, English speakers responded most often with the shape alternative, whereas Japanese speakers paid more attention to material. Japanese speakers matched substances by material, whereas English speakers responded at chance.

Barner & Snedeker (2006) found that complex objects are viewed as more individuated than simple objects and substances. According to Prasada (1999), complex objects are attributed with a specific, constructive process that individuates them. This could mean that shape is more salient when considering individual objects, even among Japanese speakers who pay less attention to individuation. If Japanese speakers do indeed individuate less and are therefore more
likely to view objects as un-individuated, their matching by material could mean that low individuation encourages awareness of attributes associated with material. In none of the three conditions did English speakers respond above chance with matching of the substances, but this could be due to a tendency to view the substances as individuated because they were formed into complex shapes.

Lucy (1992) performed a similar experiment to Imai and Gentner’s with American and Mayan men. Inanimate objects were used and there was no distinction between complex objects, simple objects, or substances. The result was consistent with Imai and Gentner: American men chose to match by shape, and Mayan men chose material. Lucy used 8 objects in his study, and all but one of these (kernels of corn) was an individuated object. English speakers matching by shape confirms the suggestion that individuation emphasizes shape over material.

The above studies show that people have a greater tendency to define individuated or complex objects by their shape than by material. It is also likely that the converse is true; non-individuated substances encourage people to consider material over shape. A number of features could be salient to material matching, such as colour, texture and function, but these features have yet to be tested in such matching experiments. Attentional biases towards shape for complex objects and material for substances are seen in children as young as two years old (Imai & Gentner, 1997; Soja, Carey & Spelke, 1993; Jones & Smith, 2002).

### 1.2.3 Individuation and function

A theory that is currently enjoying some popularity is that individuation and/or count-mass syntax is reliant on an object’s function. When it is manipulated alongside other properties of objects, function appears to be a primary property in determining count-mass status (Gathercole, Cramer, Somerville & Jansen op de Haar, 1995; Gathercole & Whitfield, 2001; Keil, 1989; Krnel, Watson & Glazar, 1998; Nelson, 1996).

Assessment of function appears to be a relatively automatic process when viewing objects. When objects are graspable, activation has been observed in parietal and prefrontal motor areas when the objects are viewed (Chao & Martin, 2000; Grafton, Fadiga, Arbib, & Rizzolatti, 1997; Handy, Tipper, Borg, Grafton, & Gazzaniga, 2006; Martin, Wiggs, Ungerleider & Haxby, 1996). This activation is lower when participants have extensive experience manipulating the object,
and high when the object is less familiar (Handy et al., 2006). Viewing a door knob, for example, does not activate motor regions but instead is associated with response in a small bilateral area of lateral occipital cortex. If a novice views a climbing hold, a much larger area of occipital and temporal cortex is activated. Expert climbers, however, show the same activation as for the door knob. This suggests that hands-on, extensive experience grasping an object reduces the level of analysis required of—and performed by—the motor cortex, when that same object is viewed.

Prasada (1999) determined that objects with regular shapes were perceived as being products of a directed process (presumably because it is intended for a specific function), whereas irregular shapes were not. “Regular” shapes are described in Prasada et al. (2002) as having “straight edges and curves with constant or smoothly changing curvature.” When regular and irregular objects were shown to subjects, regular shapes were given a count noun and irregular shapes were given a mass noun, suggesting that individuation (denoted by use of a count noun) could be associated with structure and therefore with an intended function.

In order to see whether irregular shapes could also be associated with a directed process, irregular shapes were shown being used for structure-dependent functions (such as a key being used in a lock), and structure-independent functions (such as a key being used to ring a bell). When the structure of the object was necessary for the function, it was given a count noun. When the structure was incidental (a key shape is not required to ring a bell), a mass noun was considered more appropriate. In realistic terms, this would suggest that people see a key as “a key” when it is used to turn a lock, but as “metal” when used to ring a bell. Structure-dependent functions therefore suggest individuation more than structure-independent functions.

Gathercole et al. (1995) displayed substances in particular shapes to participants in three conditions: one with shape-specific functions illustrated, one showing substance-specific functions and one with no associated function. Shape-specific function predisposed people to consider shape, and substance-specific function predisposed them to consider material. When no function was associated with the object, selection followed syntax. They matched by shape in count syntax, whereas for mass syntax they matched by substance. In the case where no function or syntax information was provided, objects were matched by shape. Therefore, the prioritization for these participants appeared to be function > syntax > shape. Results of qualitative matching experiments have been summarized in Table 1.
Table 1

Methods of qualitative matching for count and mass nouns according to study

<table>
<thead>
<tr>
<th>Study</th>
<th>Basic-level</th>
<th>Superordinates, Flexible nouns, aggregates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Mass</td>
</tr>
<tr>
<td>Barner &amp; Snedeker (2006) a</td>
<td>SHA</td>
<td>MAT</td>
</tr>
<tr>
<td>Gathercole et al. (1995) b</td>
<td>SHA</td>
<td>MAT</td>
</tr>
<tr>
<td>Imai &amp; Gentner (1997) a</td>
<td>SHA</td>
<td>MAT</td>
</tr>
<tr>
<td>Lucy (1992)</td>
<td>SHA</td>
<td>--</td>
</tr>
</tbody>
</table>

Note. SHA = shape, MAT = material. Shape and material were the only options given.

a Complex objects are treated as count nouns, substances are treated as mass nouns.
b In the no-demonstration-of-function condition, participants matched by syntax.

### 1.2.4 Syntax and Feature Salience

Feature salience may be linked to count-mass syntax indirectly through individuation, but there is a direct link as well. Interpretation of count-mass flexible nouns (those that can have both mass or count status, like *stone(s)* or *string(s)*) depends on the syntax used to refer to them. Using a particular word or syntax suggests a certain interpretation. If a speaker denotes something as “a table,” they are making its table-ness salient. If they say “some wood,” they are indicating their interest in its material properties as opposed to its structural properties. For ease of communication, the receiver adopts a similar interpretation. Wisniewski, Lamb and Middleton (2003) give an example using the phrases “a bench” and “move over and give me some bench.” Hearing the first causes a listener to interpret the bench as a discrete object. Hearing the second brings to mind an area of space instead. Syntax is responsible for much of how people perceive these flexible syntax nouns, and by extension the salient properties of those objects.

These differences in syntax and feature salience are also present for aggregates and superordinates. Count mass syntax goes hand-in-hand with individuality, structure and function,
which each have a great influence on feature salience. Aggregates are groups of small, homogeneous components, and can be either plural count nouns (like toothpicks and grapes) or mass nouns (like rice and grass). Superordinates have a similar distinction. There are plural count superordinates like “vehicles” and mass superordinates like “ski gear.” Aggregates and superordinates behave in slightly unusual ways and require additional rules to be classified. Wierzbicka (1988) provided four dimensions on which to base the distinction for these forms: perceptual conspicuousness, arbitrary divisibility, heterogeneity, and quality of interaction.

_Perceptual conspicuousness_ refers to how easy it is to perceive individual entities within the object. It can be used to differentiate aggregates like “noodles” and “rice.” In Middleton et al. (2004), whether a mass noun or plural count noun was used to describe a group was related to the ease with which a member of an aggregate was distinguishable. This phenomenon is described in depth in Dehaene (1997). Individual objects in groups of 1, 2 or 3 can be easily enumerated without requiring conscious counting, whereas objects in groups of 23 or 27 can only be approximated in number. The length of time to enumerate objects past four (and the number of errors in counting them) increases almost exponentially with the number of objects. In a pot of 20,000 grains of rice, each grain becomes quite indistinguishable.

_Arbitrary divisibility_ allows one to differentiate between aggregates and collections like _a convoy_ and _rice_. Rice with 2.5 rice grains removed is still _rice_, but _a convoy_ can only have two whole trucks removed—not halves—to remain a convoy. It can be argued that a single chair is still _furniture_, but half a chair is certainly not. If something can be divided arbitrarily and still serve the same function, it is usually labeled with a mass noun. Division in mass nouns is at the level of atoms, whereas division in count nouns is (at best) at the level of objects.

_Heterogeneity_ can be used to differentiate mass superordinates (_furniture_) and count superordinates (_weapons_), but there is still debate over how it can be done. According to Wierzbicka, groups that are made up of different kinds of objects cannot be counted in the same way as homogeneous groups can, and are therefore treated as mass nouns. Bloom (1990) argues the opposite: that members of mass superordinates co-occur (this is confirmed by Wisniewski, Imai, & Casey, 1996) and are therefore treated as more similar than members of count superordinates. Item similarity deemphasizes individuality, causing the group to be represented as a mass noun. The problem with both of these arguments is that some studies have found

*Quality of interaction* is the last dimension. Middleton et al. (2004) found that people tend to interact with only one or a few members of plural count nouns, but with multiple members of mass noun aggregates. When participants were forced to handle individual grains of an unknown aggregate, they were more likely to choose a plural count noun than those who viewed the mound but did not interact with it.

The count/mass distinction appears to be related to individuality, structure and function. For aggregates and superordinates, individuation depends on a number of factors including perceptual discrimination, heterogeneity (of substance or function) and interaction. For mass superordinates, it is primarily function.

### 1.2.5 Feature Salience in Superordinates

Count noun objects can be used on their own to fulfill a single function through a single action. Mass superordinates have a general purpose that can only be fulfilled by performing several different functions and therefore by using a combination of different objects.

Prasada (1999) confirmed that members of mass superordinates were more likely to be used together than members of count superordinates. A member of a mass superordinate may be physically individuated, but in terms of functionality it is reliant on the proximity of other members in the category. This is also likely to be true for most mass aggregate functions. People do not throw a single piece of confetti in the air or boil a single grain of rice.

Although previous research has suggested that mass superordinates do not quantify over individuals (Bloom, 1990; Wierzbicka, 1988; Wiesniewski et al., 1996), Barner and Snedeker (2003, 2005) have provided possible evidence that they do, showing that participants can quantify mass superordinates by number. However, only number and size were provided as differentiating characteristics in these studies, and item variation and other salient characteristics of mass superordinates were not varied. In the absence of these properties, differentiation might have defaulted to number.
McCawley (1975) has suggested that superordinate mass nouns are quantified by the extent to which their function is fulfilled. This does not disallow quantification by number if necessary (for example, if one needs to compare two groups that have equal ability to fulfill a function but one contains twice as many items), but overall fulfillment of function is the preferred measurement. McCawley found that when asked to quantify by the level of variation of objects or by number, people quantified by object variation. This finding was confirmed in Gordon and Rodman (2006).

Wisniewski, Imai, & Casey (1996) found that information pertaining to individual objects was de-emphasized in members of mass superordinates. They found that their interactions with members of mass superordinates were more likely to involve multiple category items (in spatial and temporal proximity), whereas interactions with members of count superordinates involved items in isolation. Participants were more likely to mention membership in a count superordinate (“is a weapon”) than a mass superordinate category (“is furniture”) as an important characteristic. When asked to confirm the statement “a bomb is a kind of weapon,” people were quicker to respond than to the statement “a chair is a kind of furniture.” However, responding to “a chair and a table are furniture” was faster than responding to “a gun and a bomb are weapons,” reinforcing the notion that members of mass superordinates are considered in combination.

1.2.6 Syntax, Function and Creativity

When people interact with an object, the most important task to complete is usually not simply remembering how many there is or what colour it is. The most important task is to determine how they can manipulate it to suit their needs. How they think to manipulate it will depend on the individuation, structure, and function of the object. Because these properties appear to be determiners of count-mass syntax (or at the very least are closely tied to it), syntax could be a predictor of feature salience and object use, and therefore success on the UoO Task.

Count nouns have specific dimensions over which they are specified quantitatively and qualitatively. Mass nouns are underspecified and have more flexible associations, which is why a variety of dimensions become salient. By sorting objects and concepts into a strongly-defined group (count nouns) and one that is more loosely defined, people can easily identify count nouns
that require less processing and focus on the mass nouns whose most appropriate current function may require more attention and ingenuity to ascertain.

When we are interacting with a complex, individuated object like a car or an apple, there is a high chance that there is already a familiar use associated with the object. A car is used for transportation, and an apple is used for nourishment. A car tire can be cut up and sewn into a shirt and an apple can be used to prop a window open, but these uses are less common and other objects are more appropriate for those uses.

Sand, on the other hand, can be melted down and turned into glass or computer chips. Water can be drunk, or it can be used to harvest nuclear energy. Because substances are less complex, they have more flexible uses and people are required to think about them in a more flexible manner.

Substances, mass aggregates, mass superordinates and flexible nouns in mass form have a variety of different possible uses, and should therefore trigger a set of basic material properties (such as size, weight, colour, texture, etc.) that can be applied in a wide variety of ways. Individuated objects, count aggregates, count superordinates and flexible nouns in count form are more likely to be already associated with a particular use specific to its material properties, and so consideration of those properties would be redundant. Instead, shape and number are more salient.

Mass nouns may induce deeper processing and encourage the retrieval of a greater number of features, as well as a higher proportion of features that permit creativity and idea generation. This means that it may be harder to generate ideas based on count nouns, because either the wrong concepts are activated (shape instead of material) or because their associated functions are too strong to ignore.

2 Purpose

This study had two primary purposes. The first was to refine the UoO task by taking into account the relationship between function and count-mass status. If count-mass status (or some related property of count and mass nouns) influences performance when a given strategy is used, this must be controlled. In addition, this study explored the efficacy of using semantic analysis tools to automate the scoring process for the UoO task. This could remove some human error as well.
as speeding up analysis. Responses to the UoO Task were analyzed using Latent Semantic Analysis (LSA; Landauer & Dumais, 1997), in order to measure the semantic similarity of responses.

The secondary purpose was to isolate all the features commonly activated by each type of noun (basic-level, aggregate, flexible and superordinate). Choices for qualitative differentiation have usually been limited to shape and material. Studies of qualitative differentiation have been largely driven by researchers’ intuitions of what two properties might be most salient in particular situations. Analysis of the uses suggested for each object will allow more features to be determined, so that future matching tasks can more accurately determine how feature salience and syntax/individuation interact.

Normally, if participants place a high priority on their own creativity, some level of anxiety may arise if they are incapable of thinking up many uses for the objects provided. If they are unsure of themselves, they may not put in as much effort or give up entirely (especially if the test appears to require knowledge they do not have). In the case of this study, however, anxiety was minimized by the use of an anonymous World Wide Web interface, which has been shown to improve performance on the UoO Task because participants feel more at ease and are not worried they will be personally judged (Joinson, 1999). They can access the interface from wherever they are comfortable as long as they have an Internet connection.

These past hundred years have seen the emergence of creativity research as a distinct field, with hundreds of psychological, managerial and educational research studies on how to improve creativity. Without standardized, reliable creativity tests, however, the field has little chance of moving forward.

The next step in improving psychometric creativity tests is to understand how the visual, conceptual and syntactic properties of stimuli influence how participants represent the objects, how they manipulate them and how they reason with them. Many implementations of the UoO task may not consider the larger network of knowledge people are accessing when they are dealing with each object. By determining the features that are activated when people generate uses of objects, we can improve our understanding of creativity as well as the construction of the task itself.
3 Experiment 1

3.1 Method

3.1.1 Participants

28 independent raters (23 female, 3 male, 2 declining to answer) provided common uses for the objects in the study. All participated online, and all but one were currently living in either the United States of America or Canada. All spoke English as their first language. The mean age of the participants was 35.25 (SD = 11.44).

15 additional participants (12 female, 2 male, 1 declining to answer) gave ratings of homogeneity for each of the objects. All participated online, and all were currently living in either the United States of America or Canada. Thirteen spoke English as their first language. The mean age of the participants was 32.43 (SD = 8.86).

3.1.2 Apparatus

All data collection was performed online, through the personal computers and web browsers of distant participants. Information was delivered to the web browser by an Apache server running PHP and MySQL, and timing and sound was interpreted by the browser through JavaScript.

3.1.3 Procedure

Participants were shown 40 objects via the Internet (the full list is included in the Appendix). 19 of the objects were displayed in mass syntax (e.g. “stone”) and 21 were displayed in count syntax (e.g. “a stone”). Images were shown above the words, and were chosen to reflect the object’s mass-count status; if count status was used, objects were displayed as individuated and when mass status was used, objects were shown in as non-individuated a manner as possible. In some cases, objects had to be depicted as a pile of “stuff” or pieces of stuff, but containers were avoided.\(^1\) Superordinate nouns were depicted using several pictures, each a possible object within the superordinate group.

\(^1\) Metal, for example, was pictured as a pile of metal bricks. Rice and pepper were shown in large piles.
Participants were divided into two groups. One group was directed to state the most common way they would use the object, which was used to determine the functional consensus about the objects. The other group was asked to rate the objects on their homogeneity. Examples of homogeneous and non-homogeneous objects were given at the beginning of the set to clarify the meaning of the word.

3.1.4 Calculation of Concept Properties

3.1.4.1 Functionality Consensus

The functions provided by participants were coded into broad general categories by the author. For example, “play with my daughter”, “to play in”, and “play in it” were all coded as “play.” The frequency of the most common response category was divided by the total number of responses for a measure of agreement on the most common use for an object. This property was labeled as functionality consensus.

3.1.4.2 Homogeneity

Homogeneity was rated by participants on a scale from 1 (not homogeneous) to 7 (very homogeneous) and then averaged. The value was used as-is and no modification was necessary. A second method was used as a verification of validity. The number of different materials in the object was counted by the author and capped at ten.

3.1.4.3 Familiarity and Word Frequency

Two measures from McRae et al. (2005) were used: the familiarity of the concept (as reported on a scale from 1 to 9 by 20 people), and the word frequency of each concept according to the British National Corpus (BNC).\(^2\) Information from McRae et al. was available for only 21 of the 40 objects, so the BNC frequencies were updated with information from the most recent version of the corpus in order to include all 40.

\(^2\) An additional measure of word frequency was included from Kucera and Francis (1967). For basic-level concepts, the KF to BNC correlation was ___ and for all objects it was \(r = 0.963 (p < .001)\). For simplicity—and because the LSA system separates singular from plural words unlike the KF measure, which sums over singular and plural uses—the BNC was used.
3.1.4.4 Mass-Count Status

Mass-Count status was determined a priori as the most common syntax used for each object. When a particular syntax was very ambiguous (like for stone or metal), mass or count status was imposed on the object according to the author’s intuitions so half were mass syntax and half were count. Mass-Count status was therefore not used as a measure of syntactic ambiguity or even the probability of one syntax over the other; only the specific syntax used to display an object to participants. No attempt was made to randomize syntax assignment.

3.1.4.5 USENET ratio and Ambiguity

USENET ratio, on the other hand, was developed as a measure of the probability of one syntax over another. Nouns were coded as belonging to three levels: basic, ambiguous and superordinate, using the definitions provided earlier. While superordinates were relatively easy to discriminate, it was more difficult to draw a line separating basic from ambiguous nouns. To serve as a rough guide for such a line, the frequency of phrases using a word in count syntax was compared to the frequency of word use in mass syntax, to give a measure of the relative prevalence of mass or count syntax for the word. The frequencies of six phrases were retrieved for each word, as below.

Three phrases were used to assess the frequency of count syntax: “if a stone is” and “as a stone is.” Three phrases were used to assess the same for mass syntax: “if stone is” and “as stone is.” “Is” was used as the final verb because of its high frequency; the same reasoning was used to choose the three beginning subordinators. The USENET corpus (2005-2007), provided by the Westbury Lab and containing over 13 billion words, was chosen because of its size\(^3\). The ratio was calculated with the following equation:

\[
UR = \log_{10}\left(\frac{\#\text{count phrases}}{\#\text{mass phrases}}\right)
\]

---

\(^3\) The British National Corpus (BNC) and The Corpus of Contemporary American English (COCA) did not have a sufficient word count to support the statistics. Even phrases for common words such as “a ball” were low in frequency, suggesting that searches through the corpora for references to colanders would be unsuccessful for both count and mass status.
where UR is the Usenet ratio. Ratios could not be calculated for five objects because no phrases were found for both mass and count syntax.

A high Usenet ratio would suggest the word is more often used as a count noun, and a low ratio would suggest greater use as a mass noun. The ratios of count and mass nouns sometimes overlap, leaving a murky middle ground where use is ambiguous. The middle ten basic-level nouns (*a radio, a radio, pepper, chain, a computer mouse, a telephone, an orange, a door, a stone, metal*, and *a pencil*) were labeled as ambiguous, making the range of ambiguity from −.71 to .60. The number of words in search phrases were not equal, and frequencies of mass usage may have differed from count usage. The mean count phrase frequency was 78.48 (SD = 166.82) and mean mass frequency was 68.23 (SD = 173.36). The mean Usenet ratio of all 40 objects was −.19 (SD = 1.19).

Some count nouns (*a bottle cap, a radio, a computer mouse, a telephone, an orange*, etc.) had unexpectedly low Usenet ratios, despite having all the typical characteristics of unambiguous count nouns. Each of these nouns were individuated and had high functionality consensus (between .50 and .96). Their homogeneity values were not significantly different from other count nouns (t = .86, p = .40), and they did not appear to be arbitrarily divisible. While two of these objects cannot be referred to with mass syntax such as “some bottle cap” or “a piece of telephone,” it is possible to play “some radio” and eat “some orange.” These phrases make slightly different use of the words, referring mostly to an arbitrarily divisible segment of time in which a radio is played, or a piece of prepared food, and not to the objects themselves. They do have enough syntactic ambiguity, however, to be used in such contexts frequently enough to have low Usenet ratios.

### 3.1.4.6 Distinctiveness

Several measures of concept and feature distinctiveness were provided by McRae et al. (2005) and used in this analysis. Distinguishing features were considered to be those that occurred in fewer than three concepts analyzed. For example, *<moos>* was reported for only one concept, “cow”, and was therefore a distinguishing feature of “cow-ness.” Distinguishing features were used in two measures in this study: the percentage of features that were distinguishing, and whether the most common feature reported was a distinguishing feature.
The distinctiveness of a feature was a similar measure, equal to the inverse of the number of concepts in which the feature occurred. The mean distinctiveness of all of a concept’s features and the distinctiveness of the most common feature reported were recorded for each object.

Cue validity of a concept’s feature was given as the probability of the concept given the occurrence of that particular feature. Because it appeared in only one concept, for example, \textit{moos} would be given a cue validity of 1.0. The mean cue validity for all of the features of each concept was used in this study, as well as the cue validity of the most common feature reported.

3.2 Results

Equality of variances was measured using Levene’s Test, and if variances were significantly different (p < .05) a t-test assuming unequal variances was used.

3.2.1 Homogeneity and Functionality Consensus

Nearly all properties of the objects were collected (in McRae et al. (2005), or this study) using no syntax (each concept labeled as “stove” rather than “a stove” or “some stove,” for example). Although this could have been interpreted as a form of mass syntax, interpretation would have been relatively constant because all words were presented in this pseudo-mass form\(^4\). However, there was a risk that the framing of the objects in either mass or count syntax in the current study (or the use of individuated or non-individuated objects for their respective images) could affect ratings of homogeneity and functionality consensus, especially those objects with a more ambiguous count or mass tendency.

A 2x2 (ambiguity x mass-count status) ANOVA was calculated for functionality consensus and no significant main effects or interactions were found. A strong main effect of mass-count status was found in a 2x2 (ambiguity x mass-count status) ANOVA for homogeneity, however (F(1,26) = 12.74, p < .005). As was predicted, mass nouns were seen as being more homogeneous. This was true for both ambiguous and non-ambiguous nouns, suggesting that

\(^4\) The only possible issue this might present would be in the interpretation of ambiguous nouns, which may be pushed in the direction of mass status by the lack of count syntax.
there was an effect of presentation syntax. Either the phrasing used for the object or the photograph depicting the object was influencing participants.

A t-test did show an effect of mass-count status on homogeneity \((t = 2.64, p < .05)\) especially when only ambiguous objects were included \((t = 3.20, p < .05)\). Of the ambiguous objects, a computer mouse, telephone, door, radio and pencil were rated as having homogeneity below the mean \((M = 4.24)\); the rest were homogeneous objects like chain and a stone, and were given moderately high homogeneity values. In other words, the syntax used to display the objects did not seem to push homogeneity ratings in one direction or another when the Usenet ratio suggested an ambiguous syntax. Participants appeared to be paying little attention to syntax, picture individuation, or both. The main effect was caused by the nature of the objects themselves, not the presentation syntax or imagery. Underscoring this, the correlation between homogeneity ratings and a count of the number of materials combined to make the object was significant \((r = .46, r < .05)\).

An effect of mass-count status on functionality consensus was observed when superordinates were eliminated \((t = 2.58, p < .05)\). Among basic objects, rice, paint and newspaper were the only mass nouns with high functionality consensus \((\text{above the mean of } .63)\). An additional ambiguous mass noun had the same high consensus \((\text{pepper})\). The basic count nouns with low functionality consensus were a bottle cap, a ball and a wheel. Ambiguous basic count nouns below a consensus of \(.63\) were a computer mouse, a door, and a stone. It is debatable whether the innate properties of these objects caused an accidental effect of mass-count status on functionality consensus, or whether the presentation of the objects (via images or syntax) modified the responses of participants. A third reason may have been that certain properties of the objects affected the selection of presentation syntax by the author, which will be discussed in the next section.

### 3.2.2 Properties associated with Mass-Count Status

Independent samples t-tests were performed to assess what properties could be associated with mass-count status. All levels of nouns were included in the analyses. There were absolutely no differences in the percentage of distinguishing features, the mean distinctiveness or mean cue validity of the features of mass and count concepts. \(p\)-values were above \( .92\) for all three properties. However, significant differences were observed for the distinctiveness \((t = 3.67, \)
$p < .005$) and cue validity ($t = 3.61, p < .005$) of the most reported feature, and whether that feature was distinguishing ($t = 4.39, p < .001$), when divided by mass-count status. Count nouns tended to have more distinct and distinguishing popular features, with higher cue validity.

This is most pronounced in basic nouns; when ambiguous and superordinate nouns were removed, differences were even more significant for primary feature distinguishability ($t = 4.58, p < .001$), distinctiveness ($t = 4.07, p < .001$) and cue validity ($t = 4.24, p < .001$).

When divided into high and low Usenet ratio, primary distinctiveness was almost significant ($t = 1.83, p < .10$), but none of the other distinctiveness measures reached even this level. No correlations were observed, either. In other words, distinctiveness of primary features may have been unconsciously influential on the assignment of mass and count nouns (especially basic ones), but this did not appear to be an obvious natural characteristic of mass and count status. There also appeared to be no difference in word frequency (in the BNC) by mass-count status, nor with familiarity. No other characteristics appeared to be associated with mass-count selection.

### 3.2.3 Properties associated with Usenet Ratio

A t-test showed a strong effect of mass-count status on Usenet Ratio ($t = 4.44, p < .0001$). While the syntax used to present the objects was very important, it was Usenet ratio that would be relied upon for an unbiased representation of mass or count status tendency. In an attempt to answer the question of what determined mass or count status, this ratio was modeled against the other measures available for each object. Usenet ratio was strongly related to BNC frequency, because rare words had a lower proportion of both mass and count phrases, making them appear more ambiguous. BNC was included in the second model in order to partial out the effect of word frequency.

For basic-only objects, automatic backward stepwise regression revealed that homogeneity and familiarity were the two strongest predictors of Usenet Ratio, accounting for 68.4% of the variance and generating a highly significant model ($F(2,7) = 7.57, p < .05$). Both homogeneity ($b = -.52, p < .05$) and familiarity ($b = .57, p < .05$) showed significant effects. A model with BNC included accounted for 74.7% of the variance, but was less significant ($F(3,6) = 5.92$,
p < .05). A t-test also indicated that functionality consensus was also slightly different between the top and bottom Usenet ratios (t = 1.95, p = .067).

When all objects were included, familiarity, homogeneity, BNC frequency, primary distinguishing, primary cue validity and primary distinctiveness were all strong predictors. The model accounted for 86.9% of the variance in Usenet Ratio, and was also very significant (F(6,12) = 13.32, p < .001). Familiarity (b = .50, p < .002), homogeneity (b = −.70, p < .001), BNC frequency (b = −.58, p < .002), primary distinguishing (b = −3.79, p < .001), primary cue validity (b = −5.84, p < .001) and primary distinctiveness (b = 2.22, p < .002) were all significant to the model.

3.2.4 Distinctiveness

Distinctiveness measures were highly inter-correlated for the objects in question. Correlations are reported in Tables 2 and 3. While mean values correlated significantly with one another and so did primary values, correlations between primary and mean values were much lower.
Table 2
Correlations between distinctiveness measures for all objects.

<table>
<thead>
<tr>
<th></th>
<th>Primary Distinctiveness</th>
<th>Primary Cue Validity</th>
<th>Percentage Distinguishing Features</th>
<th>Mean Distinctiveness</th>
<th>Mean Cue Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>-0.900**</td>
<td>-0.968**</td>
<td>-0.430</td>
<td>-0.443*</td>
<td>-0.454*</td>
</tr>
<tr>
<td>Distinguishing</td>
<td></td>
<td></td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Primary</td>
<td></td>
<td></td>
<td>.966**</td>
<td>.270</td>
<td>.348</td>
</tr>
<tr>
<td>Distinctiveness</td>
<td></td>
<td></td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Primary Cue Validity</td>
<td>.374</td>
<td>.425</td>
<td>.433*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distinguishing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage Distinguishing Features</td>
<td>.000</td>
<td>.095</td>
<td>.000</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Mean Distinctiveness</td>
<td>.975**</td>
<td>.975**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Cue Validity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).
Table 3

Correlations between distinctiveness measures for basic objects.

<table>
<thead>
<tr>
<th></th>
<th>Primary Distinctiveness</th>
<th>Primary Cue Validity</th>
<th>Percentage Distinguishing Features</th>
<th>Mean Distinctiveness</th>
<th>Mean Cue Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Distinctiveness</td>
<td>-.885**</td>
<td>-.964**</td>
<td>-.184</td>
<td>-.224</td>
<td>-.254</td>
</tr>
<tr>
<td>Distinguishing</td>
<td>.000</td>
<td>.000</td>
<td>.548</td>
<td>.461</td>
<td>.402</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Primary Distinctiveness</td>
<td>.968**</td>
<td>.080</td>
<td>.191</td>
<td>.205</td>
<td></td>
</tr>
<tr>
<td>Distinguishing</td>
<td>.000</td>
<td>.796</td>
<td>.533</td>
<td>.501</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Primary Cue Validity</td>
<td>.127</td>
<td>.213</td>
<td>.238</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validity</td>
<td>.680</td>
<td>.485</td>
<td>.434</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>13</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage Distinguishing Features</td>
<td>.971**</td>
<td>.975**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distinguishing Features</td>
<td>.000</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Distinctiveness</td>
<td>.997**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).
3.3 Discussion

3.3.1 Validity of homogeneity and functionality consensus ratings, mass-count status and Usenet ratio

Because mass and count status was assigned manually according to the intuitions of the first author, it cannot be disregarded that mass-count presentation syntax assignment may have been influenced by the homogeneity or functional consensus of the object. This may also explain a portion of the relationship between homogeneity, functionality consensus and presentation syntax in the opposite direction. Syntax may have influenced responses, but it may also have been influenced by the first author’s own would-be responses to the two questions.

Regardless of any possible bias in selection, mass-count status did seem to accurately reflect the potential syntax of the objects. One object did appear to be a slight outlier (newspaper), but it did not strongly affect the statistics. Homogeneity and functionality consensus may not have been completely removed from the pictures or syntax used to display them, but homogeneity was externally verified using a similar measure.

One problem with Usenet ratio did present itself: the high frequency phrases of some objects used a slightly different interpretation of the words, which could not be separated from the rest. The common use of “some tuba”—as an event wherein tuba is played—did not refer to the object itself but was indistinguishable from other uses of the phrase. Future efforts could be made to differentiate uses of the objects from uses of immaterial nouns with the same label.

3.3.2 Properties defining mass and count nouns

It is still relatively unclear what properties of objects had the strongest effect on choice of syntax. Although some of it appeared to be related to homogeneity and some slightly less to functionality, it was not enough to base it solely on the composition and use of the object itself. As was seen with traditionally count objects like a tuba and a radio, there were additional contexts that allowed them to be used as mass nouns, by referring to them as arbitrarily divisible units of musical experience. In order to effectively predict possible syntax, they must be treated more like concepts and less like objects.
3.3.3 Distinctiveness Measures

The pattern of distinctiveness of an object’s primary features was quite different from the pattern of the means. From the analyses into the sources of Usenet ratio and mass-count status, the primary features of the objects appeared to be more important to syntax than an average over all features.

4 Experiment 2

4.1 Method

4.1.1 Participants

282 participants began the study, and of these 135 completed at least 18 of the 20 objects in the task. 110 of the initial 282 participants were psychology students at the University of Toronto at Scarborough. Most were in their first year of a psychology program, and took part in return for course credit or entry into a draw for a $25 amazon.ca gift certificate. The remaining 172 participants covered five continents, with 15% coming from Canada, 71% from the United States, and 13% from other countries. 97 of the 109 local participants completed at least 18 of the objects, whereas only 38 of the 172 non-local participants reached the same level of completion.

(In the following paragraph, the number who participated is reported, followed by the number who completed 18 of 20 objects in brackets. Reported means are for completed participants only.) Data was recorded from participants using five different methods. Because one of the goals of this study was to assess the success and efficiency of online participation, 211 (70) users took part in the online component (10 male, 60 female), typing their answers into text boxes using an HTML form. The mean age of the online participants was 28.50 (SD = 11.93). An additional 71 (65) participants were recruited from the University of Toronto at Scarborough for an in-lab version of the task. Of these participants, 35 (29) performed the task by typing their responses into the same interface as the online participants (10 male, 18 female, 1 unreported), and 13 (13) responded using pen and paper (5 male, 8 female). 11 (11) participants (3 male, 8 female) responded using a verbalization (think aloud) protocol (giving their answers as well as any other thoughts they had), which will be described in the following section. 8 (8) participants (3 male, 5 female) responded by giving their answers verbally but not elaborating on their
thought process (talk aloud). The mean age of the in-lab typing participants was 19.31 (SD = 1.42), and the mean age of the pen and paper participants was 18.69 (SD = .75). Mean age of the think aloud participants was 18.67 (SD = 1.23), and of the talk aloud participants the mean age was 18.00 (SD = .00).

4.1.2 Apparatus

Online data-collection occurred on an Apache server running PHP and MySQL. All interaction with participants occurred through a web browser. For online participants, choice of web browser was up to personal preference and operating systems also varied. For in-lab participants, all data-collection happened on a 20-inch 2.4GHz Apple iMac running the Safari web browser. Timing and sound was controlled with JavaScript. Those in the pen and paper group used the online interface, but were asked to respond on paper. Voice was recorded using a Blue Snowball microphone, set to the omni-directional polar pattern with a frequency response of 40-18kHz.

4.1.3 Stimuli

Stimuli were the same as in Experiment 1, but the 40 objects were divided into two groups and each participant was only required to respond to objects in one of the two groups.

4.1.4 Procedure

Online participants were free to participate in the task wherever they were most comfortable. There was absolutely no communication between participants and any experimenters via email or otherwise, to maintain as much anonymity as possible. Participants signed into a password-protected data-collection website where they were asked a set of demographic questions, as well as whether they considered themselves to be creative or not (a yes or no question). They were then shown 20 objects and instructed to generate as many original uses for each object as possible. An optional post hoc interview followed the task.

Those who participated in the in-lab segment only differed from online participants in the environment wherein the task was performed. The room in which participants did the task was bare except for a large desk, a coat rack and a filing cabinet. Participants responded by typing, writing or speaking. Those in the think aloud group were given an extra set of instructions on how to properly produce the protocol, and were then trained for 5 minutes while thinking aloud
about a cryptarithm puzzle (see Appendix). For simplicity and statistical power, talk aloud and think aloud participants were only given objects from group 1.

4.2 Results

4.2.1 Scoring Responses

Several scores were calculated for each participant’s responses. The number of responses was recorded as well as the number of responses per minute (with time capped at the 2-minute mark). This is one of the most common measures for the Uses of Objects task, and is usually referred to as the “raw fluency” measure (Guilford & Hoepfner, 1967). The time between the first and last response was divided by the number of responses (minus 1) for an average reaction time between responses. Responses were coded for the number of times the participant switched categories while generating uses of each object. Ten independent raters were asked to give a score from 1 to 3 of the creativity of each response, for an average subjective creativity score.

4.2.1.1 Developing Scores using Latent Semantic Analysis

Responses were roughly spellchecked to eliminate glaring errors, and then compared against a list of words in the first-year-college corpus. If words in the responses were not present in the corpus, the closest synonym present in the corpus was used.

Three scores were calculated with the help of LSA: 1) The similarity scores between successive response pairs were averaged. This was intended to be similar to the category switch score. 2) The similarity scores between every single response pair were also averaged, as a measure of the variety of responses produced by each person. 3) For each response, 25 responses produced by other people were selected at random and the similarities between the participant’s response and each of the other 25 responses were averaged for a measure of originality compared to others.

A paired t-test showed that a score comparing each response to 25 other random responses was not significantly different from a score with comparison to every single response produced (t = .99, p > .30). The correlation between the 25-comparison originality scores and all-comparison originality scores was $r = .98$ ($p < .001$). These originality scores were calculated for the scotch tape object.
4.2.2 Reliability of Creativity Scores

Because the scale is meant to test the creativity of a person, it should be independent of the objects used to test their characteristics and the correlations between the scores for the objects should be high. For the objects in Group 1, Cronbach’s alpha for semantic originality was .89 and for group 2, alpha was .80.

Correlations between the various creativity scores are included in Tables 4 and 5. It is important to note that the within-participant scores were very highly correlated, but when averaged over objects all of the LSA scores were unrelated to fluency. Averaged over users, correlations with fluency were comparatively low but still significant.

Table 4
*Correlations between LSA scores and fluency, when averaged over objects.*

<table>
<thead>
<tr>
<th>Within Participant Category Switch</th>
<th>Semantic Originality</th>
<th>Fluency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within Participant</strong></td>
<td><strong>.923</strong></td>
<td>−.097</td>
</tr>
<tr>
<td>Variety</td>
<td>.000</td>
<td>.550</td>
</tr>
<tr>
<td>Variety</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td><strong>Within Participant</strong></td>
<td><strong>.729</strong></td>
<td>−.015</td>
</tr>
<tr>
<td>Category Switch</td>
<td>.000</td>
<td>.929</td>
</tr>
<tr>
<td>Category Switch</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td><strong>Semantic</strong></td>
<td><strong>.197</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Originality</strong></td>
<td><strong>.222</strong></td>
<td></td>
</tr>
<tr>
<td><strong>40</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).
### Table 5
Correlations between LSA scores and fluency, when averaged over users.

<table>
<thead>
<tr>
<th>Within Participant</th>
<th>Category Switch</th>
<th>Semantic Originality</th>
<th>Fluency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within Participant</td>
<td>.925**</td>
<td>.702**</td>
<td>.383**</td>
</tr>
<tr>
<td>Variety</td>
<td>.000</td>
<td>.000</td>
<td>.008</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>47</td>
<td>47</td>
</tr>
<tr>
<td>Within Participant</td>
<td></td>
<td>.566**</td>
<td>.286</td>
</tr>
<tr>
<td>Category Switch</td>
<td>.000</td>
<td>.051</td>
<td></td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Semantic Originality</td>
<td></td>
<td>.368*</td>
<td>.011</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

**4.2.2.1 Validity of Creativity Scores**

The scores were validated externally using subjective ratings from three independent raters. A weak correlation was found between the semantic originality scores and averaged manual rating scores \( (r = .25, p < .12) \). Two explanations had been predicted for this low correlation: the inability of the LSA scores to reject the responses that were so far from other responses as to be unreasonable (for example, responding with “chicken” as a use for a ball), and effect of elaboration on manual responses (a more elaborate use may suggest that the responder put more thought into the idea).

A Gaussian function (centered at a score of 0.80 and with a width at half maximum of 0.50) was applied to each semantic originality score to account for the first possibility, and the new
correlation with manual scores was $r = 0.36$ ($p < .05$). The Gaussian score appeared to be more in line with manual ratings.

The second possible explanation proved to be stronger, however. A regression model of manual rating was tested with elaboration (the number of characters in the response) and semantic originality as the main factors (using standardized scores). The model was highly significant ($F(2,43) = 21.13$, $p < .0001$), and accounted for 50% of the variance in the manual scores (bringing the correlation of the manual scores and the model from the .25 reported above to .70). Both semantic originality ($b = -.12$, $p < .0005$) and elaboration ($b = .18$, $p < .0001$) were significant to the model. Most importantly, the model’s correlation with average manual rating scores was similar to the correlations between the individual manual rating scores (which were between .48 and .80).

4.2.2.2 Reducing the Scale

From the spoken recordings, it was noticed that participants began sighing and giving other indications of boredom approximately halfway through the task. To ease the load on participants, reduced scales were isolated by gradually removing objects. Those with the highest predicted alpha if deleted were deleted, after which predicted values were recalculated and the process was repeated until the list was narrowed to 5 items in each group. Alpha for the reduced set of group 1 objects was .88; the final set was composed of a book, a bottle cap, a pencil, a wheel, and scotch tape. Cronbach’s alpha for the group 2 objects was .85, and included a key, chain, cutlery, lace and paint. Quite coincidentally, this resulted in a final set of five count nouns and five mass nouns, which could be used as an abbreviated measure of creativity. Average scores for all 20 objects were then compared to the average scores for the 5 objects from each group. For fluency, the correlation was $r = .90$ ($p < .0001$). The correlation for within participant variation was $r = .74$ ($p < .0001$) and for category change it was $r = .82$ ($p < .0001$). The correlation between semantic originality scores was $r = .71$ ($p < .0001$).

Two sets of analyses were performed; the first set averaged creativity scores for each user, and the second collapsed originality scores over participants to give an average score for each object. Although averaging the scores for each user results in an expected creativity score, average scores over objects must be interpreted differently. The originality score, when averaged over participants, represents the relative variety of responses for each object; the capacity of the object
to produce a large set of original responses. Similarly, averaging fluency results in a score of how easy it is for participants to generate ideas for that object, regardless of originality and variation. Although the two within-participant scores (category switch and variety) have still relatively different meanings when averaged for each user, these differences become less informative when averaged for each object. The category switch measure was used and variety measure ignored for this reason in the analysis of stimuli.

4.2.2.3 Similarities between Groups 1 and 2

No significant differences were observed in stimulus properties or participant demographics between groups 1 and 2. There was, however, slightly different within-participant response variety and fluency between groups. Responses to objects in group 1 had less variety than objects in group 2 (t = 1.90, p < .10) but there were also a greater number of responses (higher fluency) for each object in group 1 (t = 1.92, p < .10). These differences were not statistically significant.

4.2.3 Analyzing LSA Scores by User

In the post-task interview, many of the participants reported having looked around his or her immediate surroundings for inspiration. Because of the difference in surroundings between the in-lab and online participants (in-lab participants having a much starker room in which to be inspired), it was suspected that there could have also been differences in the responses given, which may affect the between-participant semantic originality scores. For this reason, only online participants were included in the following analysis. 24 online users completed all 20 objects in group 1, and 23 online users completed all 20 objects in group 2.

4.2.3.1 What makes a person creative

Automatic backward stepwise regression revealed that first language (English vs. other), gender, self-assessment of creativity and level of education were the four strongest predictors of fluency, accounting for 58.3% of the variance and highly significant (F(4,15) = 5.25, p < .01). Self-assessment of creativity (b = 1.00, p < .005), level of education (b = -.88, p < .01) and first language (b = -.58, p < .05) showed significant effects, and gender was close to significance (b = .35, p = .08).
Backward stepwise regression was also performed for semantic originality, and once again, first language, self-assessment of creativity and level of education were the three strongest predictors. The model accounted for 39.9% of the variance in semantic originality, and was also significant ($F(3,16) = 3.54, p < .05$). Self-assessment of creativity ($b = -.87, p < .05$), level of education ($b = .65, p < .05$) and first language ($b = .61, p < .05$) all showed significant effects.

4.2.3.2 First Language

Independent samples t-tests showed an effect of first language (English or not English) on semantic originality ($t = 2.29, p < .05$) and within participant variation ($t = 2.03, p < .05$). There did not appear to be an effect on fluency, but there was a slight effect on category switch scores ($t = 1.81, p < .08$). Years speaking English did not appear to be related to any of the four scores, and nor did the number of languages spoken.

4.2.3.3 Age

An independent samples t-test showed fluency to be influenced by age ($t = 2.28, p < .05$) when age was divided into those at or below the age of 18 and those above it. The younger group ($M = 4.48, SD = 1.81$) produced significantly more responses than the older group ($M = 3.55, SD = .71$), but also had more variability. The correlation between age and fluency was almost significant after removing one outlier ($r = -.29, p < .06$), but it is still unclear whether the effect is meaningful. There were just as many 0- to 18-year-olds as 19- to 100-year-olds, which suggests that had there been a greater number of older participants with the same amount of variability as the younger group, the effect may have disappeared. However, there is still a relatively low chance ($p < .06$) that every single older participant scored lower on the task merely by chance, so this effect may be telling. There did not appear to be an effect of age on semantic originality, category change or within participant variation.

A 2x2 ANOVA (age x first language) of fluency showed a main effect of age ($F(1,43) = 4.52, p < .05$) but not first language. No interaction was observed. A 2x2 ANOVA (age x first language) of semantic originality showed a main effect of first language ($F(1,43) = 5.68, p < .05$), but not age.
When only English first language speakers were examined, a strong effect of age on fluency was observed \((t = 2.48, p < .05)\). A slight effect was also apparent for originality \((t = 1.74, p < .10)\), although this was not significant.

### 4.2.3.4 Gender

Gender had no effect on fluency, semantic originality, category change or within participant variation, although there appeared to be an effect on fluency when only English first language speakers were examined \((t = 2.72, p < .05)\). A 2x2 ANOVA (gender x first language) revealed main effects of both gender \((F(1,43) = 5.43, p < .05)\) and first language (English or not English, \(F(1,43) = 6.32, p < .05)\) on fluency, as well as an interaction \((F(1,43) = 8.47, p < .01)\).

### 4.2.4 Analyzing LSA Scores by Object

A backwards regression, initially performed with every possible property collected, isolated two variables as the most significant predictors of fluency, producing a model that accounted for 33.8% of the variability. Functionality consensus was significant \((b = -0.49, p < .05)\) and BNC frequency was nearly significant \((b = 0.58, p < .10)\), and the combination produced a significant model \((F(2.16) = 4.08, p < .05)\). BNC may have been a predictor because more common words tended to be more familiar, and participants reported it being easier to generate many uses for familiar objects. A higher BNC may also have come about for objects that appeared in many different contexts, and this also may have affected the number of responses participants were capable of generating. Conversely, a high functionality consensus made it more difficult to generate responses, confirming one of the predictions of this study. Objects with a very obvious use, like pepper, tended to have a lower fluency average, likely because of functional fixedness. The obvious use appeared to be distracting participants from producing anything else.

Another regression generated a model of familiarity \((b = -0.55, p < .05)\), primary distinctiveness \((b = -1.67, p < .05)\), Usenet ratio \((b = 0.38, p < .10)\), and primary cue validity \((b = -1.22, p < .10)\), which accounted for 57.5% of the variance in semantic originality \((F(4,14) = 4.73, p < .05)\). More familiar objects made it difficult to provide original responses, and so did those with very distinctive features. Mass nouns allowed more variation than count nouns.
4.2.4.1 Mass-count Status, Usenet Ratio and Creativity

Semantic originality was not significantly correlated with mass-count presentation or with Usenet ratio, and within-participant variety was not correlated, either. There was a mild correlation between fluency and Usenet ratio, especially among basic nouns only (r = −.28, p = .23), but it was far from significant. Usenet ratio had its strongest correlation with within-participant category change (all among basic nouns), but this was still not significant (r = −.37, p = .11).

4.2.4.2 Creativity and Other Properties

Familiarity and typicality were the only two properties not directly related to feature distinctiveness that correlated with all three LSA scores. Correlations are reported in Table 6. Very familiar objects lead to very similar, unoriginal responses, whereas less common objects lead to less common ideas. This may have been an effect of functional fixedness: people would be used to interacting with familiar objects in familiar ways, inhibiting less common ideas or situations. It may also have been a question of effort, with more attempts at originality for more difficult, obscure objects.
Table 6

Correlations between LSA scores and object properties.

<table>
<thead>
<tr>
<th></th>
<th>Familiarity</th>
<th>Functionality Consensus</th>
<th>Homogeneity</th>
<th>Typicality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-participant</td>
<td>.514*</td>
<td>-0.109</td>
<td>-0.088</td>
<td>0.568</td>
</tr>
<tr>
<td>Variety</td>
<td>0.017</td>
<td>0.503</td>
<td>0.588</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>40</td>
<td>40</td>
<td>12</td>
</tr>
<tr>
<td>Category</td>
<td>.576**</td>
<td>-0.048</td>
<td>-0.063</td>
<td>0.573</td>
</tr>
<tr>
<td>Change</td>
<td>0.006</td>
<td>0.769</td>
<td>0.7</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>40</td>
<td>40</td>
<td>12</td>
</tr>
<tr>
<td>Semantic</td>
<td>.495*</td>
<td>0.227</td>
<td>-0.176</td>
<td>0.569</td>
</tr>
<tr>
<td>Originality</td>
<td>0.022</td>
<td>0.159</td>
<td>0.278</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>40</td>
<td>40</td>
<td>12</td>
</tr>
<tr>
<td>Fluency</td>
<td>0.004</td>
<td>-0.440**</td>
<td>0.267</td>
<td>-0.207</td>
</tr>
<tr>
<td></td>
<td>0.988</td>
<td>0.004</td>
<td>0.096</td>
<td>0.518</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>40</td>
<td>40</td>
<td>12</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

4.2.4.3 Effect of Ambiguity on Originality

A pattern emerges when the average semantic originality scores of ambiguous nouns are compared to that of their unambiguous brethren. Count nouns with particularly low originality scores also had the unusually low Usenet ratios particular to ambiguous nouns. Similarly, mass nouns with high scores had high Usenet ratios, behaving as count nouns. When superordinate objects were excluded, the 2 x 2 ANOVA (mass-count status x basic/ambiguous status) returned non-significant main effects but a highly significant interaction (F(1,30) = 5.51, p < .05).
Although it was previously thought that mass-count status would have a connection to creativity because of its relation to function, there was no expectation that it would be related to syntax specifically. However, it does appear as though the syntactic ambiguity of an object affected the responding of participants, suggesting that Usenet ratio (upon which ambiguity was based) was also involved in that responding.

4.2.4.4 Feature Distinctiveness of Objects and its relation to creativity

The mean cue validity and primary cue validity of the objects had a significant correlation ($r = .43$, $p < .05$), but percentage of distinguishing features correlated just above significance with whether the primary feature was distinguishing ($r = -.43$, $p < .06$) and mean and primary distinctiveness had a much lower correlation ($r = .35$, $p < .15$).

This pattern continued for the LSA response scores. While none of them had very high correlations with mean cue validity, mean distinctiveness or percentage of distinguishing features, they had an interesting pattern of correlation with their primary counterparts. Within-participant variety and category change correlated most strongly with primary distinctiveness ($r = .31$ and $r = .30$, respectively, $p < .20$). Weaker correlations appeared with primary cue validity, and correlations with primary distinguishing status were weaker still. Originality, however, did appear to have stronger correlations with the three primary variables, and in the same order of strength. The correlation with primary distinctiveness was highly significant ($r = .48$, $p < .05$), and correlations with primary cue validity ($r = .39$, $p < .10$) and primary distinguishing status ($r = -.31$, $p < .20$) were also relatively high.

Mass-count status had the opposite pattern. Primary distinguishing status had the highest correlation with mass-count status, at $r = .48$ ($p < .05$). The next highest correlation was with primary cue validity ($r = -.43$, $p < .05$), and then primary distinctiveness ($r = -.43$, $p < .06$). This pattern was not apparent for Usenet Ratio and no correlations were significant.

An ANCOVA of mass-count status x primary distinctiveness had an interaction just above significance ($F(2,18) = 3.52$, $p < .06$). From the limited data available (only 21 distinctiveness values were available for the 40 objects, and only a few were mass nouns), it appeared as though objects presented as mass nouns with low distinctiveness values and those presented as count nouns with high distinctiveness were those that produced a wide variety of responses.
4.3 Discussion

4.3.1 Reliability and Validity of creativity scores

Scores derived from the 40 objects were highly reliable, and a reduction to a 10-object scale only changed Cronbach’s alpha by –.01 in one group and by +.05 in another. Human ratings of creativity were also closely correlated with LSA scores, with a model of semantic originality and elaboration correlating best with manual scores. This model appeared to be a much better predictor of the human rating than the standard method of scoring fluency by number of responses. Because elaboration had such a strong effect on manual ratings, it may be that manual raters were using the length of responses as a heuristic for creativity and artificially elevating the ratings of longer responses. If this is the case, it may be that the LSA-derived scores were actually more appropriate measures of creativity, because it did not take length into account. In fact, originality was more reliable than the human ratings, likely because the score was more reliable in its ability to make comparisons from one person to the next.

4.3.2 Similarity of Group 1 and 2 stimulus properties and user responses

Few differences were observed between concept groups, for both stimulus properties and quality of responses. What differences existed (in within-participant variety and fluency) were very slight and non-significant.

4.3.3 Differences between environment and task conditions

In-lab participants had much lower creativity scores than those participating online. Several reasons could account for this. The greater anonymity provided by the Internet may have afforded creativity, as has been reported in other studies of creativity testing over the Internet. Another source may have been the higher level of stimulation provided by any familiar, meaningful objects present in the environments of online participants. With only a desk, chair, filing cabinet and coat rack in the lab room, very little stimulation was available to in-lab participants.

No differences in semantic originality were observed between typing and pen and paper participants in the lab. Speaking appeared to be a minor handicap (perhaps because of shyness, despite participants being alone in the room). The think aloud protocol proved to be a huge hindrance in comparison to speaking. Although fluency was only slightly affected, explaining
their thought processes was apparently distracting enough to suppress creative thinking. Typing lead to a higher fluency than pen and pencil responding, likely because in-lab participants were younger and were more likely to have grown up with computers.

4.3.4 What the test says about objects

4.3.4.1 How stimulus properties influence creativity scores

Fluency was most strongly related to word frequency in the BNC and functionality consensus. That functionality consensus was the strongest predictor of fluency confirms one of the predictions of this study. It was more difficult to shut out the more obvious responses to objects with particularly high consensus on how to use an object. It was more difficult to provide a creative response for pepper than for sand. In addition, objects that occurred more frequently in language were also given a higher number of responses. This may have come about as a result of familiarity or a greater prevalence of these objects in many different contexts. In some cases, people used word association to come up with responses, and objects with more variety in their word association may have been more frequent in the BNC corpus.

Semantic originality appeared to be higher for unfamiliar objects without highly distinctive features, and those objects with a tendency towards mass syntax produced a greater variety of responses than count nouns. This confirmed the prediction that syntax (either by way of a third variable or through a direct modification of presentation syntax) would have an effect on the originality of responses.

4.3.5 What the test says about people

First language (English vs. other), self-assessment of creativity and level of education were the strongest predictors of fluency and semantic originality, and gender also had an impact on fluency. It was expected that first language would have an impact because it would be easier for participants to express themselves if they had a stronger command of the English language, and this was confirmed.

Level of education was a surprising result, however. This may have come about because younger participants (those with a naturally lower level of education) had a higher fluency, but this does not explain the effect of education on creativity. A possible reason for this would be that higher
levels of education put a higher demand on creative thinking (or a higher level of reasoning in
general), and so only those who are highly creative had the capacity to move on to higher levels
of education.

Gender’s impact on fluency was likely because responses from women were more elaborated
than those from men. Lengths of responses (and therefore time spent typing) were approximately
the same, but men produced a higher number of responses than women did. This may have been
because women thought more deliberately about how to respond, or perhaps that some responses
were inhibited because they considered them to be inappropriate or uncreative.

Self-assessment of creativity seemed to be somewhat related to creative responding, but not
enough people listed themselves as uncreative (6 of 47) to make this difference significant.

4.3.6 Strategy use

Strategy use, especially for participants asked to verbalize their thought process during the task,
was found to be highly repetitious, with specific uses appearing consistently for each and every
object. Those people who were most creative were those who were able to successfully avoid
perseveration. They did this primarily by generating typical end states (such as having a clean
house or getting to work) and deriving the uses required to achieve those end states, whereas less
creative people generated the uses themselves and applied the uses rigidly, regardless of how
inappropriate the use might be. This is an interesting finding, and one that would benefit from
further inquiry.

4.3.7 Future Study

To determine the semantic originality of each response, cosines were calculated between the
response and 25 other randomly selected responses. An alternative approach could be to compare
each response to a set of 25-50 standard responses (typical uses), rather than a random selection.
This could be more effective for quick evaluations requiring dynamic feedback. Because coding
of category change usually counts the number of category switches as opposed to the extent of a
switch, an additional score could count the number of cosine transitions that are below a given
threshold.
Some responses appeared to be related to self-report of creativity; however, only six people out of 47 reported themselves as being uncreative. In future studies, a more specific set of questions may be used to determine self-assessment, as well as the use of a Likert scale rather than a simple yes or no question.

Additional scales from the brain region taxonomy by Cree and McRae (2003) were the number of visual form & surface features, colour features, function features, visual-motor features and tactile features that appeared in each concept. Although not enough of the objects in this study overlapped with the data from Cree and McRae (2003), this could be something to pursue in future studies of this task.

5 General Discussion

There is some level of controversy surrounding how representative psychology experiments are of cognition. Some can be very artificial, and this is especially true for experiments regarding creativity. In order to be easily measurable, situations must be carefully controlled. Multiple choice responses for self-report questionnaires and measures such as fluency may be simple to quantify, but they can sacrifice realism. This study has made some strides in rectifying this issue, by allowing participants to respond with elaboration in a familiar and anonymous environment, but also providing an objective measurement scheme with the help of Latent Semantic Analysis.

The success of this measurement technique was confirmed with a scale independently rated by humans. The score was reliable, and also provided the potential of a modular format: the creativity score of a particular person did not differ significantly when averaged across 5 objects or 20 objects. The scale was also shown to be slightly related to self-assessment of creativity, which some researchers (Hocevar, 1981, for example) consider to be a superior measurement of creativity over creativity production measurements.

There is still some question as to whether syntax is related to creative production, but this study has isolated some properties of objects that may affect responses. More importantly, fluency was shown to be less reliable than semantic originality, showing that fluency scores could vary substantially from test to test. The LSA scores developed in this study were a much better approximation of the creativity of a particular person, whereas fluency scores were dependent on creativity as well as the nature of the objects being used for the task.
References


Appendices

Appendix 1: Question posed of homogeneity raters

Objects can vary in composition. Some objects are very homogeneous, in that all the parts of the object are the same. Examples of homogeneous objects are: milk, rubber, glass, etc. Other objects are not homogeneous, and are made up of many different substances. Examples of objects that are not homogeneous are: ice cream sundaes, tennis shoes, a VCR, etc. For each of the following objects, please judge whether they are very homogeneous (a score of 7) or not at all homogeneous (a score of 1). Is the following object very homogeneous (every part the same, a score of 7) or not at all homogeneous (many parts are different, a score of 1)?

Appendix 2: Question posed of functionality raters

State the most common way you would use this object. Give only one answer and respond as quickly as you can.

Appendix 3: Instructions for the Uses of Objects Task

The following task is called the Uses of Objects Task. Please read the following instructions before continuing:

- Make sure you write every idea on a separate line.
- If you have installed any browser plug-ins that disable JavaScript or sound, the experiment may not work properly. It is strongly recommended that you disable these sorts of plug-ins. If you cannot, simply advance each time the “Time's Up!” text appears.
- List as many uses as you can possibly think of for each object you are shown.
- You'll be looking at pictures and words, but the picture is simply to clarify the meaning of the word for the object you are to use. If you are shown a red brick in a picture, this does not mean you can only use one red brick; you can use many bricks, and any kind of brick you like (red bricks, grey bricks, clay bricks... any brick used to build things).
- This is a creativity test, so we encourage you to be as creative and original as possible. Choose more interesting responses over basic ones; if you see a hammer, for example, you don't have to say “hammer nails.”
• You can take the object apart if you want to. For example, you can use just the handle of a hammer or the whole thing.
• You'll have 2 minutes for each object, so write as quickly as you can.
• After 2 minutes a bell will sound and the text box will disappear, at which point you can click “Continue” to move ahead.
• If you're stumped, you can also advance at any time by clicking the “Continue” button.

List as many creative uses for this object as you can, as quickly as you can. Use a new line for each new idea.

Appendix 4: Think Aloud Instructions

Try to think aloud. You probably do it sometimes when you’re alone and working on a problem. Say everything that passes through your head while you’re coming up with answers. The chief thing is to talk aloud constantly from the minute I present the object, because I want to know everything you happen to think of, no matter how irrelevant you may think it is. I’m not interested in the quality of your answers or how many you come up with, but in how you’re going about coming up with answers. I won’t count your wrong attempts or whether you go off topic, so say everything that goes through your head. Don’t plan what to say or talk about what you just thought, but rather let your thoughts speak, as though you were really thinking aloud.

Appendix 5: Think Aloud Cryptarithm

\[
\begin{array}{ccc}
D & O & N & A & L & D \\
+ & G & E & R & A & L & D \\
\hline
R & O & B & E & R & T \\
\end{array}
\]
Appendix 6: List of Concepts

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count Nouns</td>
<td>Mass Nouns</td>
</tr>
<tr>
<td>a book</td>
<td>carpet</td>
</tr>
<tr>
<td>a bottle cap</td>
<td>furniture</td>
</tr>
<tr>
<td>a door</td>
<td>metal</td>
</tr>
<tr>
<td>a pencil</td>
<td>pepper</td>
</tr>
<tr>
<td>a plate</td>
<td>sand</td>
</tr>
<tr>
<td>a radio</td>
<td>scotch tape</td>
</tr>
<tr>
<td>a stone</td>
<td>snow</td>
</tr>
<tr>
<td>a tuba</td>
<td>water</td>
</tr>
<tr>
<td>a vehicle</td>
<td>wood</td>
</tr>
<tr>
<td>a weapon</td>
<td>an orange</td>
</tr>
<tr>
<td>a wheel</td>
<td></td>
</tr>
</tbody>
</table>