MACHINE-LEARNING BASED CLASSIFICATION OF TEXTUAL STIMULI TO PROMOTE IDEATION IN BIOINSPIRED DESIGN

A Dissertation

by

MICHAEL W. GLIER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Daniel A. McAdams
Co-Chair of Committee,	Julie S. Linsey
Committee Members,	Richard J Malak
	Tracy A. Hammond
Head of Department,	Andreas Polycarpou

August 2013

Major Subject: Mechanical Engineering

ABSTRACT

Bioinspired design uses biological systems to inspire engineering designs. One of bioinspired design's challenges is identifying relevant information sources in biology for an engineering design task. Currently information can be retrieved by searching biology texts or journals using biology-focused keywords that map to engineering functions. However, this search technique can overwhelm designers with unusable results. This work explores the use of text classification tools to identify relevant biology passages for design. Further, this research examines the effects of using biology passages as stimuli during idea generation.

Four human-subjects studies are examined in this work. Two surveys are performed in which participants evaluate sentences from a biology corpus and indicate whether each sentence prompts an idea for solving a specific design problem. The surveys are used to develop and evaluate text classification tools. Two idea generation studies are performed in which participants generate and record solutions for designing a corn shucker using either different sets of biology passages as design stimuli, or no stimuli.

Based 286 sentences from the surveys, a k Nearest Neighbor classifier is developed that is able to identify helpful sentences relating to the function "separate" with a precision of 0.62 and recall of 0.48. This classifier could potentially double the number of helpful results found using a keyword search. The developed classifier is specific to the function "separate" and performs poorly when used for another function.

Classifiers developed using all sentences and participant responses from the surveys are not able to reliably identify helpful sentences.

From the idea generation studies, we determine that using any biology passages as design stimuli increases the quantity and variety of participant solutions. Solution quantity and variety are also significantly increased when biology passages are presented one at a time instead of all at once. Quality and variety are not significantly affected by the presence of design stimuli. Biological stimuli are also found to lead designers to types of solution that are not typically produced otherwise. This work develops a means for designers to find more useful information when searching biology and demonstrates several ways that biology passages can improve ideation.

DEDICATION

Ad majorem Dei gloriam

ACKNOWLEDGEMENTS

As with any large work, a great many people contributed to this research in ways great and small. The least I can do is to acknowledge some of those people here, even if many are unlikely to read this.

First thanks go to my advisor and co-advisor, Dr. McAdams and Dr. Linsey, for their guidance and support, but mostly for their patience. I thank my committee members, Dr. Malak and Dr. Hammond, for generously lending their time and insight to this research. I am very grateful to Dr. Matt Campbell, whose presentation at DCC'12 spurred my research in a completely new direction that led to this work, which is amusingly unrelated to that initial talk. I also appreciate the support for this work provided through the National Science Foundation through NSF EEC 1025155.

I am fortunate to have had some wonderful lab mates as I worked on this degree, and I am grateful to them all. A special thanks is due to Shraddha, for her friendship, advice, and commiseration as we both worked toward our doctorates. Ethan and Wes also deserve thanks for tackling the important but time-consuming and mind-numbing task of developing the biology corpus used in this research.

Having started as an undergrad at Texas A&M, I am grateful to the countless people who have made my time at A&M a wonderful experience, but particularly to the faculty and staff of the Mechanical Engineering Department. I must also thank Mark Horn, Ray James, Harry Hogan, Kim Moses, and Sue Geller, without whom my eighteen year-old self may not have found his way to Texas A&M. I am further indebted to Bart

Braden, Andy Long, Stephanie Parker, and Frank Shaw, for without them I may not have been *able* to study at A&M.

Finally, I will be forever grateful to my father, mother, brother, and sister for their continuous love and support while I studied engineering a thousand miles from home.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

Bioinspired design has enjoyed a growing level of interest and research from the design community. By exploring the many designs existing in the natural world, bioinspired design offers designers a vast array of highly adapted solutions to difficult design problems. Those solutions often employ approaches that are seldom encountered in traditional engineering designs. Designers seeking to leverage nature's design knowledge must, however, overcome the difficulty of identifying relevant sources of design inspiration in a domain foreign to the majority of engineering designers.

Many solutions have been proposed and developed to help designers explore nature's design knowledge. Fully discussed in section 2, these tools and methods for bioinspired design include curated databases such as AskNature [1], DANE [2], and Idea-Inspire [2, 3], a variation on the TRIZ method called BioTRIZ [4], and an application of functional modeling to bioinspired design [5]. Natural-language based tools such as automatic keyword extraction mechanisms [6-8] and the Engineering-to-Biology thesaurus [8, 9] can help direct designers when performing keyword searches through biology literature. However, these natural language tools offer no help in understanding the information retrieved and are prone to overwhelming designers with an abundance of irrelevant or incomprehensible information.

This work focuses on examining and extending natural language bioinspired design tools, specifically addressing the tendency of a bioinspired keyword searches to produce an overwhelming number of irrelevant results for bioinspired design. Prior research efforts have generally attempted to mitigate this problem by carefully selecting a search corpus likely to have easily-understood and relevant information, such as an introductory biology textbook [8, 10]. However, limiting searches to such a small corpus as a textbook necessarily reduces the scope of biological systems that could be identified for bioinspiration. Rather than limiting the corpus, this work proposes to augment simple keyword search methods with a trainable result-filtering mechanism that will determine whether or not the search results least likely to be useful to the designer.

Instead of creating a filter using some set of expert-derived rules attempting to mimic human decisions on the usefulness of some search result, the proposed filter, which will be referred to as the BID Classifier, is derived using existing machine-learning techniques for text mining. A set of surveys is performed to train the BID Classifier and to validate its effectiveness. An unstructured and a bioinspired idea generation experiment examine how the sentences found through keyword searches effect idea generation during engineering design.

The remainder of this section will discuss the specific research questions that this work seeks to answer and clearly outline the scope of this work. Sections 2 and 3 discuss research related to this work and introduce some of the key tools and concepts used in this work. Sections 4 through 7 discuss the analogy surveys and the development of the BID Classifier. Sections 8 through 10 explain the idea generation experiment and examine how using passages from biology as design stimuli effects concept generation. Finally the contributions of this work and avenues for further research are discussed in section 11.

1.1. **Research Questions**

A substantial amount of time and effort has been devoted to finding ways to allow designers to better utilize the design knowledge of natural systems. Natural-language based methods particularly seek to improve designers' ability to find useful sources of design inspiration within existing literature on biology. There is evidence to suggest that functional keyword searches are helpful for engineering design [11], but the effectiveness of this method is greatly compromised by the volume of unusable information such searches generally yield [12]. This work has two primary hypotheses

- Results from function based keyword searches can measurably improve the idea generation of engineering designers,
- The percentage of useful information in such search results can be increased by using text classifiers.

To support these hypotheses, the following research questions are explored.

- Do passages from a biology corpus found through a functional keyword search measurably improve idea generation?
- Are some passages more helpful in idea generation than others and if so, what are the characteristics of helpful and unhelpful passages?
- Can machine-learning based text classification identify passages that are likely to be helpful, or unhelpful, in generating ideas for engineering design?

1.2. Scope of Work

Figure 1 shows the components of this research and their relations. This research seeks to both develop a BID classifier to identify sentences that will act as stimuli for bioinspired design, and to measure the effects of such sentences on idea generation. This work is primarily based on four human experiments, which are listed in Table 1. The experiments are categorized as either analogy identification surveys, or idea generation experiments. These experiments will be explained in detail in sections 4 and 8, respectively. Briefly, in an analogy identification experiment, a designer is given a short design problem to consider and a set of sentences from a corpus of biology journals. The designer is asked to indicate whether each sentence prompted any idea for solving the design problem. In an idea generation experiment, participants are given a short design problem and are asked to write or sketch solution to the problem. Some design stimulus may be given as part of the ideation experiment.

Table 1. List of studies performed.

Analogy Identification Surveys
Corn Shucker Analogy Survey
Alarm Clock Analogy Survey
Idea Generation Studies (Corn Shucker Problem)
Undirected Ideation
Biology Passage Stimuli Ideation

The BID classifier is built with a supervised learning algorithm, so a set of sentences that are "known" to be good or bad design stimuli are needed. This training set is provided by corn shucker analogy survey, in which designers are given a set of randomly selected sentences, each containing a biological synonym for the function "separate", and asked to indicate which sentences generated some idea for developing a device for shucking corn.

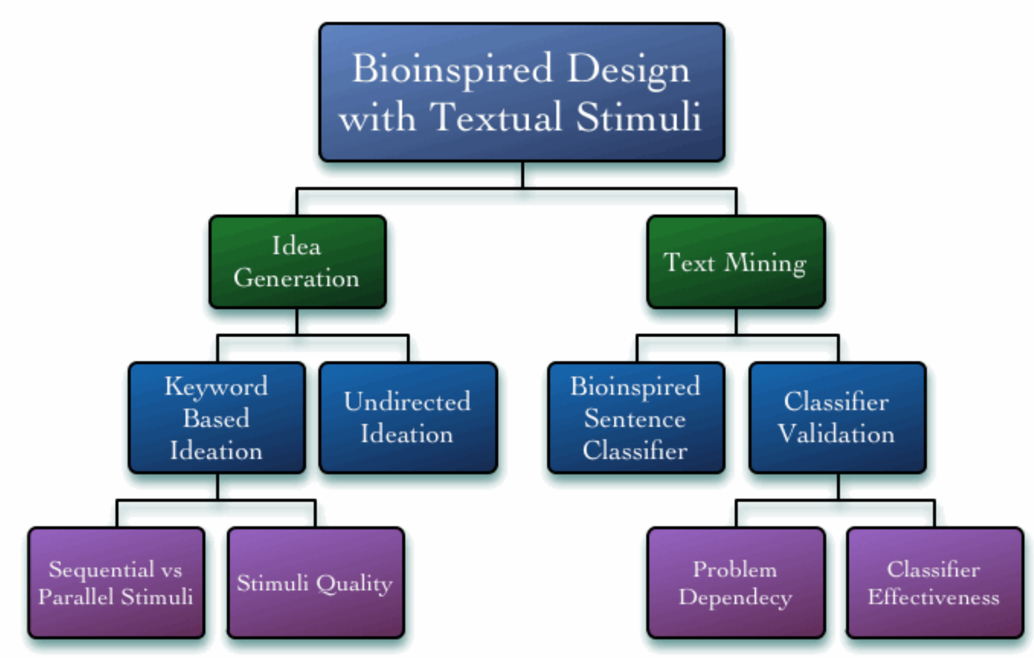


Figure 1. The scope of research is shown with the major components to be addressed.

An initial validation of the BID Classifier will be done using a standard validation procedure on the training data. Because the classifier is trained with data that relates to a single design problem, it is possible that its performance will be problem dependent. This problem dependency will be evaluated in alarm clock analogy survye. This survey gives designers a different design problem – to develop a portable alarm clock – and sentences containing a biological synonym for the function *indicate*. The sentences will be randomly selected sentences from a biology corpus. By changing the problem and function, the alarm clock survey will determine whether the BID classifier is problem dependent.

Table 2. Relations between the research questions and the proposed experiments.

Research Question	Experiment
Do passages from a biology corpus found through a functional keyword search measurably improve idea generation?	C, D
Do retrieved biology passages offer only random stimuli to aid idea generation?	A
Are some passages more helpful in idea generation than others and if so, what are the characteristics of helpful and unhelpful passages?	A
Can machine-learning based text classification identify passages that are likely to be helpful, or unhelpful, for generating ideas in engineering design?	A, B

To measure idea generation, studies are needed to quantify designers' idea generation when no method or stimuli are used and when biology passages are given as design stimuli. The unstructured and biology-stimuli idea generation experiments both ask designers to write or sketch concepts for a corn-shucking device. The unstructured ideation study establishes a baseline measure for idea generation by giving designers no method or stimuli to generate ideas. The biology-stimuli idea generation experiment gives designers a set of passages for design stimuli, but designers are directed to read all the passages at once before generating ideas, or to read a single passage and attempt to generate some ideas before moving on to the next. The biology-stimuli experiment builds on the corn shucker survey by giving designers a set of sentences that were established as either good or bad design stimuli in the survey, and asking designers to generate ideas using those stimuli. This idea generation experiment will let us compare the idea generation of designers using various configurations of biology stimuli to those using unstructured idea generation.

Each experiment in the proposed work ties directly back to the initial research questions for this investigation. Table 2 shows which experiments are relevant to each research question. Ultimately, all components of this research feed into developing and validating a text classifier for bioinspired design and developing an understanding of how passages drawn from biology can aid idea generation.

2. RELATED WORK IN BIOINSPIRED DESIGN [*]

This research seeks to improve the keyword-based information retrieval methods for bioinspired design by applying existing tools for text mining and natural language processing. As such, this section will introduce some of the existing tools and methods for bioinspired design.

One of the challenges for designers attempting to use biological systems to inspire engineering design solutions is the difficulty in drawing useful analogies between the two domains. Because biological systems and engineering systems will often accomplish similar functions through very different mechanisms, designers are prone to imitate nature rather than develop a distinct but functionally-similar engineering system [13]. Despite such difficulties, bioinspired design is a commonly acknowledged approach within the design community and is recommended by many design textbooks [14-18].

There are many approaches to bioinspired design. Methods include studying examples of bioinspired design, curated database tools, an abstraction of natural design approaches into a TRIZ format, applications of functional modeling, automated bio-keyword extraction tools, and thesauruses that help translate between engineering and biology. This section will address these methods, and discuss some studies that have been done in analogy formation in bioinspired design.

---

[*] Part of this chapter is reprinted with permission from “Concepts in Biomimetic Design: Methods and Tools to Incorporate into a Biomimetic Design Course”, M.W. Glier, D.A. McAdams, J.S. Linsey, 2011, Proceedings of the ASME 2011 International Design Engineering Technical Conferences & Computer and Information Engineering Conference, IDETC2011-48571. Copyright 2011 by ASME

2.1. **Directed Approach**

The directed approach is a common approach to bioinspired design that describes many designers' initial effort to apply bioinspiration to design. This is not a formal method, but rather a basic heuristic that directs the designer to consider nature and how it accomplishes a task similar to the designer's problem. By considering nature, the designer may be able to identify an analogous or inspirational biological system that is relevant to the design problem at hand. The directed method may be one of the simplest bioinspired design methods to implement because it requires only the designer's existing knowledge of the natural world. However, when biological knowledge is limited, teams will often recruit a biologist to expand the team's knowledge base and increase the likelihood of finding a bioinspired design solution. This multidisciplinary approach to bioinspired design has been used many times and can result in very interesting bioinspired solutions [19]. The Center for Biological Inspired Design proposes that two different directed bioinspired design processes exist: a problem-driven process and a solution-driven process [20]. The problem-driven process begins with a design problem and attempts to find a similar problem in nature so that the designer can map features of the biological solution to the engineering domain. The solution driven process instead begins by considering interesting biological phenomena and seeks engineering design problems that the principle could potentially solve.

2.2. **Case Study Approach**

The case study based method is another informal method for bioinspired design. The method proposes that if designers see enough examples of bioinspired designs, they will understand how to effectively use nature to inspire engineering design and recognize opportunities for applying bioinspired design. This method was taught through courses in Biologically Inspired Product Development at the University of Maryland. A large number of bioinspired design cases and products were collected and presented in a series of lectures. The course resulted in a successful effort by student design teams to design a bioinspired mobile robot [21]. A similar effort by Jenkins at Montana State University examined many biological systems and applied reverse engineering processes to them to inspire engineering design [22, 23]. Many resources offer collections of case studies of bioinspired designs [23-26].

2.3. **Curated Database Tools**

Two computer-based interactive design tools, called IDEA-INSPIRE [27] and Design by Analogy to Nature Engine (DANE) [2] are built around custom-built databases of biological and engineering systems. Both tools provide designers with qualitative models of biological and engineering systems, present information in multimedia format (e.g. text, photos, graphs, schematics, etc.), and allow users to search the systems by their function [2]. The tools differ primarily in how they present information to the user and how the biological and engineering systems are functionally modeled.

Chakrabarti et al. performed a preliminary examination of the IDEA-INSPIRE tool by having three designers with undergraduate engineering degrees and formal product design training use IDEA-INSPIRE to solve two design problems from a pool [27, 28]. The software, even with a limited number of entries in its database, accounted for 47% of the ideas generated. Further testing on the multimedia presentations provided by IDEA-INSPIRE demonstrated that the non-verbal cues such as images or videos were better than verbal cues such as textual system models at promoting ideation. DANE has been tested several times as optional design tool in Georgia Tech's course in biomimetic design. Unfortunately, the students had little incentive to use the software and Vattam et al. were unable to draw any firm conclusions about how and why the tool was used [2, 29].

Another more recent tool to aid information retrieval for bioinspired design is the Biologue tool [30]. This tool helps designers find relevant information by adding model-based tags to articles that are helpful to bioinspired design. The tags are applied through a community effort, with designers apply tags and adding to Biologue's database as they find new things in their own research.

DANE is currently available for general use, though IDEA-INSPIRE and Biologue are not. These tools continue to be developed, and their utility increases as their databases of natural systems expand. While these tools have the potential to provide designers with an abundance of well organized, highly accessible biomimetic design information, they are greatly limited by the time and effort that is required to add each new design to their respective databases. Unless this limitation is somehow overcome, these curated database tools will necessarily have small databases compared to the wealth of information accessible through natural language search tools.

Similar to DANE and IDEA-INSPIRE, AskNature organizes information on biological systems and bioinspired solutions in a searchable online database

AskNature.org. Additionally, information is categorized using a taxonomy that combines functional descriptions of systems with more specific mechanisms. The AskNature database currently contains articles on 1648 biological strategies and 195 bioinspired products [1]. Like DANE and IDEA-INSPIRE, AskNature presents information in a format that is accessible to designers, but it is again limited by the cost of expanding the database.

2.4. **BioTRIZ**

BioTRIZ is another tool designed to enable designers to abstract useful design information from biological systems. BioTRIZ is based on the Russian Theory of Inventive Problem Solving (TRIZ) developed by Genrich Altschuller [31]. TRIZ was developed on the premise that the vast majority of new engineering designs are simply combinations of a single set of general design principles [4] Working on this assumption, TRIZ incorporates information from nearly three million successful patents and vast stores of physical, chemical, and mathematical knowledge [32]. That information is condensed in TRIZ into 39 system parameters – things that a designer would want to maximize and minimize, and 40 inventive principles – ideas that can help solve a design problem. A designer using TRIZ would pose a design problem as a contradiction such that one of the 39 system parameters is optimized at the cost of another. The designer then consults a 39 by 39 “contradiction matrix” that lists the inventive principles that have been used to resolve that contradiction in other designs. While TRIZ can be a powerful design tool for an engineer, it was developed from a technical and engineering knowledge base; design information from the natural world is absent. BioTRIZ uses the approach of TRIZ to abstract biological design information into more general tool applicable to any form of engineering design.

The development of BioTRIZ was led by Dr. Julian Vincent at the University at Bath [33]. BioTRIZ is based on the analysis of approximately 500 biological phenomena with over 270 functions that form over 2500 contradictions [4]. Unlike Altschuller's work, BioTRIZ groups the many system parameters into only six fields of operation: *substance, structure, space, time, energy,* and *information*. Rather than a 39 by 39 contradiction matrix as in TRIZ, BioTRIZ employs a more condensed 6 by 6 matrix. BioTRIZ retains Altschuller's inventive principles, but Vincent did expand on them, giving examples of how each principle is employed in biological systems [4]. The procedure for applying BioTRIZ to a design problem is identical to that for TRIZ, except that the designer must use one of the six fields of operation instead of the 39 system parameters.

For comparison, Vincent also constructed a condensed TRIZ contradiction matrix using the six fields of operation. Comparing the inventive principles used in technology (TRIZ) and biology (BioTRIZ) for identical contradictions, Vincent found that the two matrices have a similarity of only 0.12, where identity is denoted by 1 [4]; Vincent did not explain how this similarity metric was computed. Further, he found that the at scale up to 1 meter, technological design solutions rely most heavily on the manipulation and conversion of energy while biological solutions tend to revolve around information and space [33].

BioTRIZ can be a valuable tool for biomimetic design. Like TRIZ, it provides engineers a straightforward, methodical approach to solving design problems, but BioTRIZ relies on design information not common within the technical world. The ever-increasing need to develop sustainable technology forces engineers to adopt new approaches to design; the reduced emphasis on energy inherent in BioTRIZ is a strong recommendation for this approach. From a pedagogical viewpoint, BioTRIZ is a straightforward design method, with a rigid procedure and a small set of necessary materials. With a brief introduction to the methodology and a few tables, engineering students can leverage a wealth of biological design information without needing to themselves translate biology literature into engineering term and abstract the useful information into something applicable to their design problem.

2.5. **Functional Modeling**

One method that has been successfully used to abstract engineering design information from natural systems is functional modeling. Analyzing a biological system at a purely functional level helps designers to find functionally similar technological solutions, which may or may not use the same mechanisms as their biological inspiration. Furthermore, functional modeling provides designers with a systematic, complete approach to modeling biological phenomenal [34]. Although nearly any functional language can model biological systems, many of the efforts to apply functional modeling to biomimetic design have employed the Functional Basis, as described by Hirtz et al. [35]. Tinsley et al. determined that the Functional Basis is an effective language for transferring biological design solutions into the engineering domain [36].

One of the challenges in functionally modeling biological phenomena lies in properly defining the scope of the system to be modeled. Nagel, et al. demonstrated that carefully defining the *category* and *scale* of a biological system's functional model allows designers to better isolate and model those aspects of the biological system most applicable to their design problem [37].

The authors proposed four biological categories: physiology, dealing with the vital functions of organisms; morphology, pertaining to an organism's physical form; behavior, characterized as an organism's response to stimuli; and strategy, defined as generic behavior to achieve different goals. Biological scale can range from the atomic level to populations, but the authors noted that cellular, organ, organism, and behavior scales tend to be useful and are well documented in biological literature. While biological functional models are generally restricted to a single scale, designers may employ multi-scale models to show specific biological phenomena.

Keeping in mind the importance of category and scale, Nagel et al. developed a seven-step, general methodology for functionally modeling biological systems [10]. This methodology evolved from Nagel's experience using functional modeling to conceptualize biomimetic sensor technologies [38, 39]. The suggested methodology is:

1. Identify a suitable reference (e.g. a biology textbook) for the biological system of interest.

2. Read the overview of the biological system to understand the core functionality of the system.
3. Define the design question the functional model aims to answer.
4. Define the category of the functional model.
5. Define the scale of the functional model.
6. Develop a functional model of the biological system using the functional basis modeling language within the bounds set by the design question, biological category, and biological scale.
7. Double-check and/or validate (e.g. have a biologist review model at desired biological category and scale) the functional model against the design question and black box model.

Functional modeling can be a powerful tool for biomimetic design. This method’s systematic, thorough approach can reveal aspects of biological phenomena that are not apparent with other approaches. However, creating a functional model for a natural system can challenge a designer whose background lies in the engineering domain. Creating an accurate and useful model requires the designer to have a sound understanding of the system of interest – and perhaps the insight of a biologist – to define the appropriate category and scale. Finally, even after constructing an accurate and useful biological model, the designer must still find ways to mimic the system’s functions in an engineering system.

2.6. **Automated Keyword Extraction**

Any attempt to draw analogies between the natural and engineering domains must involve some biological system as a reference. Merely identifying an appropriate biological phenomenon for inspiration may prove infeasible for a designer with no background in biology. To facilitate this task, a great deal of effort and research has been committed to developing tools that enable designers to search for relevant biological information using engineering oriented search terms. One approach to this problem is the automated keyword extraction methods developed at the University of Toronto.

Vakili and Shu published a report on their efforts to develop a natural language, biology-engineering search tool in 2001. In this initial effort, they searched for biological information relevant to a design problem by searching the index and glossary of an introductory biology textbook. They quickly discovered that using synonyms for engineering oriented keyword increases the chances of finding useful matches. They also found that a “bridge” was necessary to span the gap between engineering keywords and biological keywords. In this instance, the textbook’s glossary served as a bridge. A later paper by Hacco and Shu extended the first method by searching the entire biology text, rather than just the index. While this approach yielded more matches, the majority of those matches were irrelevant [40]. The overwhelming amount of irrelevant matches revealed the need to identify more relevant and specific search terms.

A later effort greatly improved the relevance of search results by the incorporation of WordNet, “an electronic lexical database, designed and organized according to current psycholinguistic and computational theories of how humans remember language" [41]. With WordNet's capabilities, Chiu and Shu greatly reduced the number of irrelevant search results, and designed an algorithm to identify significant verb pairs based on how often the two verbs appear together. Chiu and Shu continued to refine their BID Lab search algorithm, eventually identifying non-obvious keywords and even algorithmically producing a keyword that had been recommended by a human expert [12, 42]. Later work by Ke et al. demonstrated that adding metadata, such as part-of-speech tags and information about the meaning of ambiguous words to a natural language corpus can greatly reduce the number of irrelevant search results [43]

2.7. **Engineering-to-Biology Thesaurus**

Cheong et al. used the BID Lab search tool to identify biologically relevant keywords for terms in the Functional Basis [7]. Nagel, Stone, and McAdams [9] combined Cheong's functional basis keywords with similar sets produced at the Indian Institute of Science [3] and Oregon State University [9]. The combination of these word sets constitutes an "engineering-to-biology thesaurus" that provides designers with a more comprehensive set of biologically meaningful keywords for the terms of the Functional Basis. This thesaurus allows designers to more effectively use conventional search tools to find biological systems for design stimulus based on their function. Because the thesaurus is built around the terms of the Functional Basis, it can be used with conventional functional models of engineered systems or conceptual designs to find sources of biological inspiration.

The Engineering-to-Biology thesaurus identifies biological correspondents to terms in the Functional Basis. The thesaurus was developed by Nagel et al. [9] by combining a list of correspondent terms produced by the BID lab search tool [7] with similar lists generated at the Indian Institute of Science [3] and Oregon State University [38]. The thesaurus allows designers to more effectively use conventional search tools to find biological systems for design stimuli based on their function. Because the thesaurus is built around the terms of the Functional Basis, it can be used with functional models of engineered systems or conceptual designs to find sources of biological inspiration.

2.8. Studies on Analogies in Bioinspired Design

One of the key elements and great stumbling blocks in biomimetic design is drawing analogies between the biological and engineering domains. The disparity between those two domains can help designers find novel solutions to engineering problems, but designers (especially such novices as engineering students) can have difficulty in finding useful analogies between the domains. Here we examine several studies that examine the creation and use of analogies in biomimetic design.

A study by Vakili et al. examined an interdisciplinary engineering team using functional models of biological processes [44]. Based on observations of, and interviews with, the participants, the authors made several recommendations on how to employ functional models of biological phenomena. Vakili et al. recommended that functional models be used to extract useful design analogies, and that a third party (i.e. not the designers) perform the extraction with a systematic method [45].

A later study by Mak and Shu examined students who were asked to conceptualize a biomimetic system, given selected textual descriptions of biological systems [46]. Mak and Shu found that the students tended to fixate on employing their design stimulus to specific, and often inappropriate, aspects of their design problem. The authors also found that many students could not effectively transfer knowledge between the biological and engineering domain. Mak and Shu suggested that ambiguous or "general" biological design stimuli may be more conducive to ideation than more detailed stimuli [47]. Cheong and Shu later found that the presence of causal relations between two actions in a biological description can have a significant effect on the analogical transfer between engineering and biology [48]. A similar study by Cheong et al found that designers tend to focus on superficial and of the inappropriate characteristics of biological systems, and that fixation persists even after an analogy is identified [49]. In an attempt to break fixation Cheong and Shu found that abstracting nouns in descriptions of biological systems to more general forms can reduce fixation and cognitive bias in designers [50].

A study by Vattam et al. performed an *in vivo* study of a design team in Georgia Tech's introductory course on biomimetic design [51, 52]. The researchers found that the analogies employed in the project could be classified into five types: direct transfer, schema induction, problem transformation, deferred goal, and compositional analogy. Secondly, they determined that analogies were used in nearly every major phase of the design process: problem definition, solution search, initial design, design evaluation, and design analysis – only the redesign phase lacked analogies. A similar investigation by Helms and Goel found that analogies are used primarily in the early stages of the design process, and that while functions will often transfer from biology to engineering, the specifications and constraints of biological systems typically do not [53].

Glier et al. performed studies to examine the use of various bioinspired methods including BioTRIZ, functional modeling, keyword searches using the Engineering-to-Biology thesaurus, and the “directed method”[54, 55]. Designers were generally able to develop complete solutions to design problems using BioTRIZ. To use functional models effectively for bioinspired design, designer first needed to have a strong understanding of the modeling process. As the Engineering-to-Biology thesaurus only facilitates function-based searches, its effectiveness is greatly dependent on the functional model directing the search [11, 55]. The directed approach, discussed in section 2.1 was found to offer no advantage to novice designers over undirected idea generation [54].

3. TEXT CLASSIFICATION BACKGROUND

Text mining is a very large and diverse field that encompasses many aspects of traditional data mining methods and natural language processing tools. This discussion will only address the aspects of text mining that are relevant to this work, namely text classification and some aspects natural language processing. Very briefly, a text classifier is an algorithm that can examines some body of text, commonly referred to as a document, and assigns some classification to that document. Such classifiers are able to flag incoming emails as spam, or examine a news article and determine the section in which the article belongs (e.g. Politics, Sports, Business, etc.). In actuality, classifiers do not typically examine actual documents but rather a set of features that provide an abstract representation of the document. The following sections will briefly introduce the text classifiers used in this works, and some natural language processing tools that are used to build the feature set.

3.1. **Text Classifiers**

Many different types of classification algorithms exists, but this discussion will only deal with four types of classifiers that are used in this research. These classifiers are: Naïve Bayes, k-Nearest Neighbors, Support Vector Machine, and Rule Induction classifiers. Each of these four classifiers uses a very different method to make classifications, as will be explained in the following subsections.

3.1.1. **Naïve Bayes**

Naïve Bayes classifiers use training data to compute for each feature the probability that the feature is present for each class. To determine the class on a new input, its features are examined and the cumulative probability of the input belonging to each class is calculated. The input is assigned the class with the highest probability. Formally, for a document *d* with *i* features f_i, the probability *P* that the document will have class *c* is:

$$P(c|d) = \frac{P(c)\prod_i P(f_i|c)}{P(d)} \qquad [56]$$

Note that *P(d)* is constant for all classes and does not need to be computed. These classifiers are called naïve because they make the unjustified and generally false assumption that all features are independent. However, Naïve Bayes classifiers can be surprisingly robust and very computationally efficient [57]. Naïve Bayes classifiers tend to have mixed performance with text classification, but can perform quite well with an appropriate feature set [58]. Lewis discusses several variations on Naïve Bayes classifiers that can be used in text mining applications [59].

3.1.2. **k Nearest Neighbors**

The k-nearest neighbor (kNN) classifier one of the simplest classifiers and is referred to as a “lazy” or example-based classifier. Rather than forming some abstract relation to classify documents, the kNN algorithm simply compares each new document to the training documents and finds the k closest documents in the vector space of document features. The new document is assigned the class of it the highest proportion of its k nearest neighboring documents [60]. The kNN algorithm has a very high computational cost, as each new document has to be compared to the training documents. However, the kNN algorithm is one of the best performing classifiers, with the main drawback of its computational cost [56].

3.1.3. **Support Vector Machines**

Unlike Naïve Bayes and kNN classifiers, support vector machines (SVM) are binary classifiers [61]. Geometrically, SVMs seek to find the hyperplane in the vector space of document features that best separates the positive from the negative examples by the widest margin, i.e. the minimal distance between the plane and a training example is maximized. Introduced as a text classification method by Joachims [62], SVMs tend to be fairly robust to overfitting and scale well for very large feature sets [63]

3.1.4. **Decision Rule Induction**

Decision rule algorithms, along with the closely related decision tree algorithms have the advantage of producing a set of rules that can be easily interpreted by humans. Rather than employing some abstract mathematical expression to classify documents, decision rule classifiers generate a set of rules that are logical relations between the document features. Many examples of decision rule algorithms exist, but this research uses the CN2 induction algorithm, which was named for its creators, Clark and Niblett [64]. The CN2 algorithm performs well with noisy data sets and has significant performance advantages over other rule induction algorithms such as the ID3 and C4.5 [65].

3.2. **Natural Language Processing**

Transforming the text in a document into a reasonable set of features is a critical step in any text mining operation. Many research efforts have investigated the effects of feature selection on text mining algorithms [63, 66-69]. The most common representation is the bag-of-words model [56]. In this representation, documents are represented as vectors where each dimension represents one word occurring within the document. Thus, the dimension of the feature space is equal to the number of unique words within the document. More involved representations typically use synthetic features, which use various means to group similar words into similar features [56]. Various tools used in natural language processing are employed to generate the feature set used in this research. Word tokenization, part-of-speech (POS) tagging, and word stemming will all be briefly discussed.

3.2.1. **Tokenization**

The unprocessed documents analyzed in text mining typically take the form of a continuous stream of text. This text must be broken up into a set of distinct words, numbers, symbols, or other meaningful textual elements. This process is called tokenization. While tokenization can divide text into many types of elements depending on the application (e.g. book chapters, headings, paragraphs, etc.) this research only uses word tokenization. Word tokenization breaks a sentence into its component words, numbers, and special characters. This can be very simply performed by using whitespace characters as dividers. More complicated word tokenizes often employ regular expressions to recognize common patterns in the text [70]

3.2.2. **Part of Speech Tagging**

POS tagging is the process of assigning each word in a document with its part of speech. Prominent POS tagging algorithms include the rule based Brill tagger, Hidden Markov model taggers, and neural network classifiers [71-73]. Ulitmately, POS taggers are simply specialized classification algorithms trained to determine the parts of speech for words.

3.2.3. **Word Stemming**

Word stemming is the process of stripping prefixes and suffixes from words to condense the word to some base form. With stemming, closely related words can often be mapped to a common word stem. For instance, a stemming algorithm could map the words regulate, regulation, regulator, and regulator to the common stem regul. However, stemming algorithms can also confound words, for instance, if the prior words were mapped to the stem reg, words like region or regiment may be mapped to the same stem. Despite this shortcoming stemming algorithms can provide valuable dimensional reduction in a feature set by combining conceptually similar word feature into a single word stem feature. Word stemming algorithms include the Porter, Snowball, and Lancaster stemmers. The Porter stemmer is an early stemming algorithm that first appeared in 1979 [74, 75]. The Snowball stemmer is actually a refinement of the Porter stemmer using the Snowball string processing language developed by Porter for stemming in many languages [76]. The Lancaster stemmer is a more aggressive stemmer introduced in 1990 [77].

Lemmatizers are closely related to stemmers. Whereas stemmers algorithmically remove the suffix of a word to produce a word stem, a lemmatizer uses a dictionary to identify words and reduce them to their root form (lemma). For example, where a stemmer may reduce “woman” and “women” to “wom”, a lemmatizer will lookup the word “women” and output the lemma, “woman”. A lemmatizer will not change a word that is not in its dictionary.

4. DESCRIPTION OF ANALOGY SURVEYS

This section details the analogy surveys used in this work. Two types of experiments are used in this research, analogy surveys and idea generation experiments, which will be discussed in section 8. Two closely related analogy surveys are discussed here. This section first outlines the general structure of the surveys. Before discussing the details of the two survey, section 4.2 introduces the corpus of biology publications from which the information in the surveys is drawn.

4.1. **Analogy Survey Structure**

The goals of this research are to provide a significant, quantitative assessment of how keyword-based textual design stimuli encourage idea generation, and to develop a means for a computer to identify stimuli that are likely to encourage ideation. For this line of research, we use single-sentence passages as design stimuli. The analogy surveys are designed to gather a large sample of student assessments for a large number of passages found with Engineering-to-Biology Thesaurus (EBT) keyword searches. With a large sampling, we can determine with some confidence which passages are useful for providing design inspiration. These passages can then be used to inform the design of the idea generation experiments of section 8.2 and to develop automated tools for identifying passages that promote ideation. Our desire for a large number of passages to be reviewed drives our choice to restrict biology passages to single sentences. Longer excerpts containing multiple sentences or paragraphs take too long to read for a large sample to be obtained.

This section describes two very similar analogy surveys, one relating to the design of a corn shucker, and the other to the design of a portable alarm clock. While very similar in structure, the experiments serve very different purposes. The corn shucker survey gathers a total of 10,000 participant responses for the purpose of developing text classification tools and identifying suitable stimuli for the idea generation experiments discussed in section 8.2. In contrast, the alarm clock survey seeks to evaluate the classification tool's ability to identify relevant design stimuli for a new design problem with a different function. Because of the more limited scope of the alarm clock survey, only 4,500 participant responses are gathered.

The corn shucker and alarm clock surveys follow the same format. For brevity, the corn shucker survey will be explained in detail in section 4.3 while section 4.4 will discuss only the points where the alarm clock survey differs from the corn shucker. Before discussing the surveys , section 4.2 will introduce the corpus from which the passages used in the surveys are drawn.

4.2. **A Corpus of Biology Journals**

The passages used for bioinspiration in this experiment consist of single sentences selected from a corpus of biology journals. The corpus contains plain text copies of thousands of articles from five biology journals. Table 3 lists the journals included in the corpus, and the approximate word count of the text pulled from those journals. The journals in the corpus were chosen to place an emphasis on information regarding biological systems at the organ, organism, and multi-organism scales. *Current Biology* and *Basic and Applied Ecology* provide a broad coverage of natural systems that may be relevant to engineering design. The other three journals were collected as part of a related research effort that focused on animal interactions. Those journals are included in the corpus because they matched the focus on biological systems at the organ scale or larger. While some terms in the EBT relate specifically to cellular phenomena, our experience has suggested that designers are seldom able to find design inspiration in biological systems at the cellular or molecular scales [13]. Consequently, journals focusing on microbiology were not considered when building the corpus. Despite the corpus' focus on macrobiology, many passages used in the corn shucker experiment address cellular or molecular phenomena. The presence of these passages is primarily due to the EBT keywords that relate specifically to cellular biology (e.g. meiosis).

Table 3. Journals included in biology corpus and approximate word counts.

Journal Title	Word Count (millions)
Basic and Applied Ecology	3.17
Current Biology	25.43
Journal of Animal Behaviour	1.79
Journal of Animal Ecology	9.58
Journal of Zoology	4.05

4.3. Corn Shucker Analogy Survey

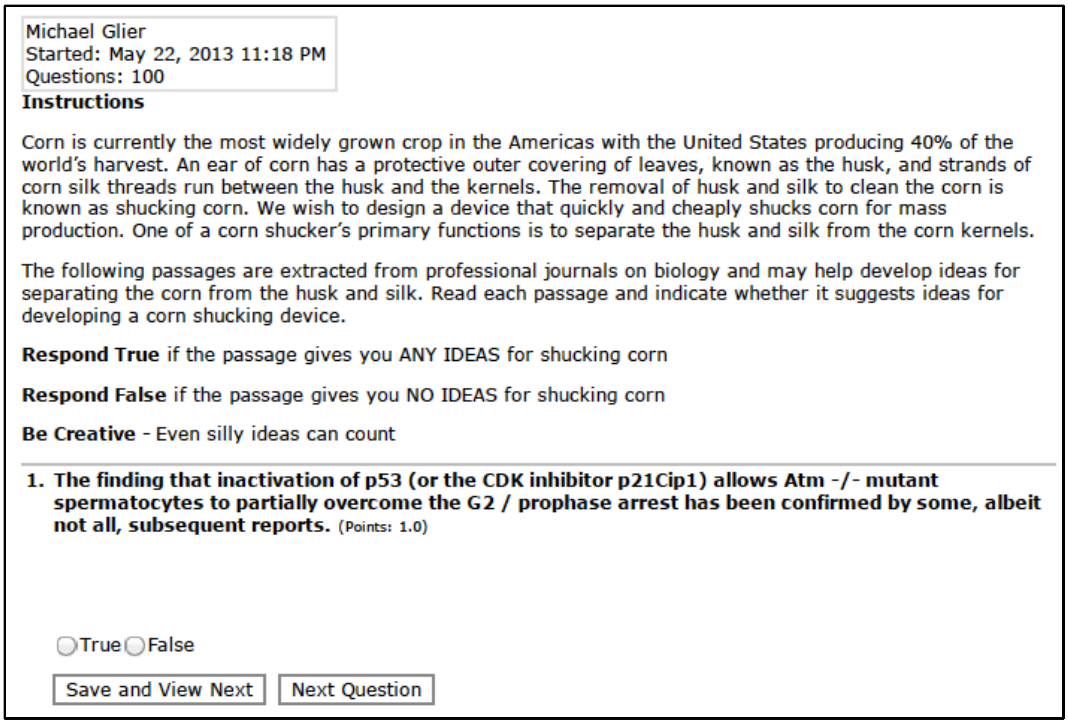
The EBT spans the entire Functional Basis, providing biological synonyms for every function and sub-function. Experimentally examining the full EBT would be impractical, so the scope of this survey is limited to biological synonyms for the function "separate." Further, while the EBT provides designers with keywords to guide their searches in biological literature, the returned results are dependent on the body of knowledge searched; a journal of biochemistry will yield significantly different search results from a high-school biology textbook.

To give some context to the passages examined in this experiment, participants are given a design prompt. The experiment asks participants to consider the problem of removing the husk and silk from an ear of corn, adapted from an idea generation experiment discussed in section 8.1 and published in Glier et al. [54]. This problem is a derivative of the peanut sheller design problem used by Linsey et al. [78]. Participants are asked to indicate whether each biology passage prompts any ideas for developing a corn shucker. Participants are instructed not to disregard any ideas, regardless of how silly the ideas seem. The full problem statement is given in Figure 2. It is important to note that while the problem statement does not direct participants to focus on the function "separate," the design problem, it does state that the passages "may help develop ideas for *separating* the corn from the husk and silk" (emphasis added).

Unlike the ideation studies from which the corn shucker problem was adapted, this experiment does not attempt to record the ideas or analogies that participants form using the biology passages. Rather, we only wish to determine if the passage prompted any ideas for solving the design problem. By not collecting a record of the ideas generated, we greatly increase the number of passages participants can examine in a relatively short time frame. The experiment was conducted as an online survey made available to participants through eLearning, Texas A&M's virtual learning environment.

Table 4. The number of participants from each recruited class.

Class	Participants
Engineering Experiments	55
Advanced Design	29
Bioinspired Design	16



Michael Glier
Started: May 22, 2013 11:18 PM
Questions: 100

Instructions

Corn is currently the most widely grown crop in the Americas with the United States producing 40% of the world's harvest. An ear of corn has a protective outer covering of leaves, known as the husk, and strands of corn silk threads run between the husk and the kernels. The removal of husk and silk to clean the corn is known as shucking corn. We wish to design a device that quickly and cheaply shucks corn for mass production. One of a corn shucker's primary functions is to separate the husk and silk from the corn kernels.

The following passages are extracted from professional journals on biology and may help develop ideas for separating the corn from the husk and silk. Read each passage and indicate whether it suggests ideas for developing a corn shucking device.

Respond True if the passage gives you ANY IDEAS for shucking corn

Respond False if the passage gives you NO IDEAS for shucking corn

Be Creative - Even silly ideas can count

1. **The finding that inactivation of p53 (or the CDK inhibitor p21Cip1) allows Atm -/- mutant spermatocytes to partially overcome the G2 / prophase arrest has been confirmed by some, albeit not all, subsequent reports.** (Points: 1.0)

○True ○False

Figure 2. A screen capture of the online survey delivered to participants.

This experiment collects responses from 100 mechanical engineering students at Texas A&M University during the Fall semester of 2012. The participants were recruited from three classes: MEEN 601, a graduate-level advanced product design course; MEEN 489, an undergraduate course in bioinspired design; and MEEN 404, a senior-level engineering experimentation course. All students had some training in engineering design. Students in the graduate-level design course and the bioinspired design course had also been introduced in the use of keyword-based search methods for bioinspired design as part of their classes. Table 4 shows the number of participants recruited from each course. Participants were given either extra credit in their recruited class or $10 as compensation for participating in the experiment.

Table 5. Biological synonyms for the function "separate" and the number of passages containing each keyword that are included in the experiment's question set. Note that some keywords were not found within the corpus and that some sentences contain multiple keywords.

Biological Keyword	No. of Included Passages
Abscission	27
Anaphase	38
Bleach	26
Cleave	28
Cytokinesis	32
Deanimate	0
Denature	8
Deoxygenate	1
Detach	0
Dialysis	27
Divide	30
Electrophoreris	0
Evacuate	3
Expel	28
Filtrate	16
Free	24
Liberate	26
Meiosis	29
Metaphase	46
Mitosis	41
Prophase	30
React	31
Release	32
Replicate	28
Segment	26

Table 6. The number of participants in each class per survey.

Survey	Engineering Experiments	Advanced Design	Bioinspired Design
Survey A	8		14
Survey B	15	6	
Survey C	13	4	
Survey D	13	5	
Survey E	7		15

As shown in the screen capture in Figure 2, participants are given a design prompt and a single biology passage. The passages were framed as a set of 100 True/False problems with participants responding True if the passage prompted any ideas for designing a corn shucker, and False if it prompted no ideas. Passages are presented one at a time, and their order is automatically randomized for each participant by the eLearning system. Because the experiment was offered online, there was no time limit for completion. However, most participants completed the experiment in less than 40 minutes.

The passages presented in this experiment are drawn from a set of 500 sentences, each sentence being pulled from the Biology Corpus and containing one or more of the biological synonyms for the function "separate" from the EBT. The biological keywords are listed in Table 5. Only the sentences containing the exact form of the listed keywords were considered as bioinspiration for the experiment. Thus, a sentence containing the word "bleach" would be included, but not "bleached." Restricting the keyword's grammatical form serves several purposes. Most importantly, the set of keywords identified in in the EBT and used in this study places an emphasis on function. Functions are commonly considered to be verbs, conveying some form of action. Accordingly, when the EBT was being constructed, the keywords were generally selected to be verbs for processes used in biology. Nouns are used in the EBT only in cases where some clearly functional process has no verb form (e.g. “mitosis”) [9]. Additionally, our restriction on the keywords limits the number of candidate passages and eliminates from the study the confounding factor of the keywords’ grammatical forms.

Because each participant reviews only 100 passages, the 500 passage set is divided into 5 surveys of 100 passages each. The passage set was divided so each survey contained approximately the same number of passages for each biological keyword. Table 6 shows how many participants in each class took each of the 5 surveys. Every participant took only a single survey. The surveys were distributed so that participants from two classes saw each survey and approximately 20 individuals took each survey.

4.4. **Alarm Clock Analogy Survey**

The overall structure of the alarm clock survey is identical to that outlined in the previous section. The differences between these two surveys lie in the problem statement, the passages used, and the participants. To begin, participants in this survey were given the problem of designing a personal alarm clock that will not disturb nearby people. This design problem is an adaptation of a similar design problem used by Genco et al. [79] as a prompt for an exercise in accessible design. The same problem is used in a slightly different format of an idea generation study by Glier et al.[54]. The survey format is the same as shown in Figure 2, but the problem description is changed. The description given in the alarm clock survey is inset in italics on page 36.

Alarm clocks are essential for college students; however they will often wake up a roommate and those around them. We wish to design an alarm clock for individual use that will not disturb others. The clock should be portable for use in a variety of situations such as on the bus, in the library, or in a classroom. One of the alarm clock's primary functions is to signal a sleeper to awaken.

Customer Needs:

- *Must wake up individual with no disturbance to others*
- *Must be portable and lightweight*
- *Electrical outlets are not available as a constant power source*
- *Low cost*

The following passages are extracted from professional journals on biology and may help develop ideas for waking a sleeping person. Read each passage and indicate whether it suggests ideas for developing a portable alarm clock.

The passages used in this experiment are drawn from the Biology Corpus and each passage contains a biological synonym for the function "signal". This survey uses a set of 100 passages. Table 7 lists the biological synonyms for "signal," along with the number of passages containing each synonym. A full list of the passages used in this experiment is included in Appendix C.

This survey had 45 participants who were recruited from MEEN 402, the second half of Texas A&M's mechanical engineering capstone design curriculum. The experiment took place between April 23rd and April 28th, 2013. It is important to note that MEEN 402 students are often co-enrolled in MEEN 404, so we can expect similar performance from these two participant groups.

Table 7. Biological synonyms for the function "signal" and the number of passages containing each keyword that are included in the experiment's question set. Note that some sentences contain multiple keywords.

Biological Keyword	No. of Included Passages
Detect	10
Locate	3
See	16
Smell	5
Observe	9
Monitor	6
Gauge	4
Watch	3
Flouresce	4
Communicate	5
React	4
Mark	7
Learn	5

5. CORN SHUCKER ANALOGY SURVEY RESULTS AND DISCUSSION

The corn shucker survey that was explained in section 4.3 is primarily intended to provide training data for developing a text classifier. However, we can learn a great deal from the survey with some statistical analyses. This section examines the results of the corn shucker survey to find trends among the experiment's participants, the passages used in the survey, the EBT keywords used to select the passages, and even the individual words within the passages.

Because results were collected using an online survey, all the results must be evaluated to ensure that participants had read and considered each sentence before indicating whether or not the sentence generated any ideas for the design problem. The virtual learning environment through which the surveys were conducted automatically records the time each participant spent to complete the survey, as well as the time spent on each question. Participants generally required between 20 and 45 minutes to complete a survey. The time spent on each sentence was examined for the results from any participant that completed a survey in less than 15 minutes. A participant's results were omitted from the evaluated data set if the timing record indicated that 10 or more sentences in a row had been rated in three seconds or less for each sentence. While a very short sentence may reasonably be read, considered, and rated in a few seconds, we assume that 10 sentences can only be rated in less than 30 seconds when a participant is not reading the sentences. Using this criterion, the results from 12 participants were excluded from the following analyses. Results from participants who took more than one hour to complete the survey were also scrutinized. A long survey duration could indicate that participants stopped and later returned to the survey.

As each sentence evaluated in this investigation was evaluated by a minimum of 17 participants, and each participant evaluated 100 sentences, there are two closely related quantities that need to be examined. To compare participants, we consider the number of sentences that each participant rated useful (a “True” response on the survey). This quantity lets us compare the three classes recruited for the experiment, as well as the five surveys used in the experiment. However, we also need to consider the percentage of participants that rated each sentence as useful. Only with this second quantity can we compare the relative usefulness of the sentences for bioinspired idea generation. For brevity, this second measure will be called “sentence quality” – ranging from 0, where no participants were able to generate ideas using the sentences, to 1, where all participants reported having some idea based on the sentence. Building on the idea of sentence quality, we can also consider the quality of an individual word. Word quality is simply the average quality of the sentences in which that word appears.

5.1. **Participant Analysis**

Among the participants themselves, we observe a very large variation in the number of useful sentences. On average, participants rated 29 sentences in a survey as being helpful for idea generation, with a standard deviation of 15 sentences. Some participants found that no sentences in their survey gave them any ideas for their design problem, while one participant was able to generate ideas using 62 sentences in the offered survey. Naturally with three classes recruited for the experiment, we must consider that number of sentences participants found helpful may vary between classes, particularly since each class contains students with a different level of design training.

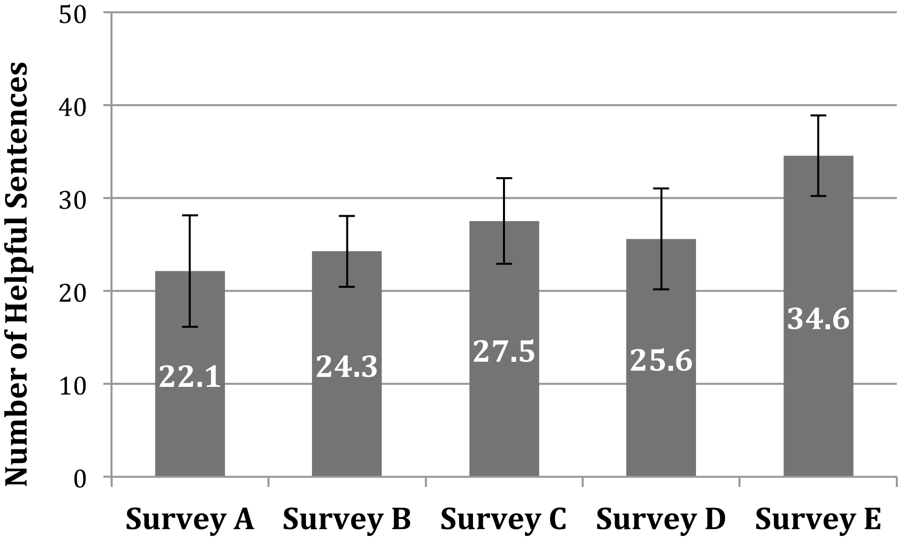


Figure 3. Number of helpful sentences identified by Engineering Experiments participants for each survey.

Before comparing classes, a one-way ANOVA was performed on number helpful sentences per participant for each survey in the Engineering Experiments class. Figure 3 shows this information as a bar plot, while Table 8 shows a standard ANOVA table for the analysis. The ANOVA demonstrates that there is no significant difference between surveys in the number of helpful sentences found (p=0.63). Further pairwise t-tests determine that there is no significant difference in the number of helpful passages between any two surveys. This is an expected result since the surveys were designed to be as similar as possible while each containing a unique set of sentences. Because of the similarity between the surveys, the survey conditions are combined to compare the three classes.

Table 8. ANOVA table for survey comparison among Engineering Experiments participants.

Source of Variation	SS	df	MS	F	p-value	F crit
Between Surveys	706.4	4	176.60	0.654	0.627	3.21
Within Surveys	13777.8	51	270.15			
Total	14484.2	55				

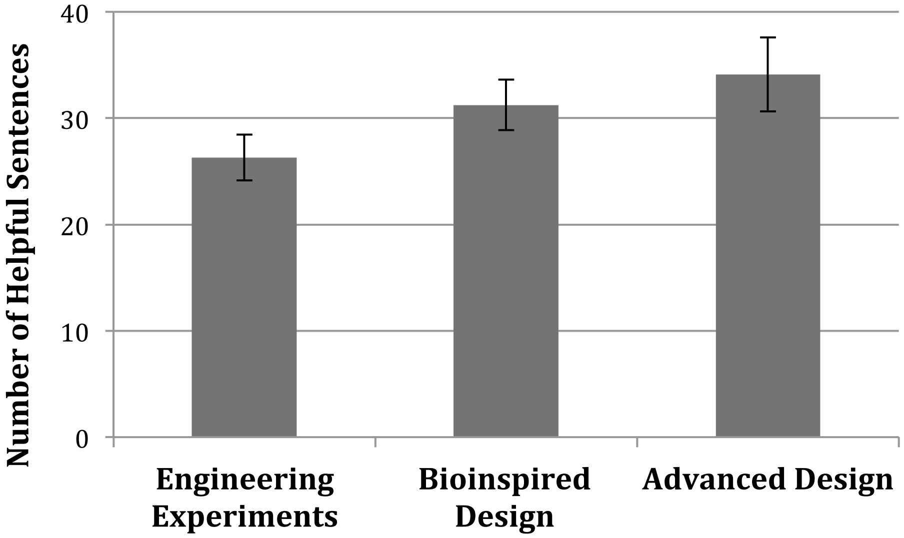


Figure 4. Average number of helpful ideas by class. Error bars denote ±1 S.E.

A one-way ANOVA between the classes, listed in Table 9, finds no significant differences between the classes (p=0.13). A pairwise comparison is also conducted to check for differences in the number of helpful sentences between classes. Figure 4 shows the average number of helpful sentences identified by each class, across all surveys. The error bars denote one standard error. A two-tailed t-test with an assumption of unequal variances is used to compare each class to the others. The differences between the Bioinspired Design course and the two others are not statistically significant using a confidence level of 0.1. However, there is a significant difference between the Engineering Experiments course and the Advanced Design course (p=0.07). It is interesting to note that students in both the Advanced Design and Bioinspired Design courses had been taught bioinspired design methods, including the use of the EBT to find sources of inspiration. As demonstrated in Figure 4, students from both these classes were able to use more biology passages to generate ideas. This finding supports the intuitive notion that the level of design training and instruction has a significant effect on the idea generation prompted by biology passages.

Table 9. ANOVA table for class comparison among corn shucker survey participants.

Source of Variation	SS	df	MS	F	p-value	F crit
Between Classes	942.1	2	471.05	2.118	0.126	2.358
Within Classes	21573.3	97	222.4			
Total	22515.4	99				

5.2. Sentence Analysis

Having compared the classes, we now examine the differences in the ratings for the evaluated sentence. As might be expected from the variations in the number of useful sentences per participant, participants offered little agreement on which sentences were helpful. The average sentence quality was only 0.29, with a standard deviation of 0.17. While several sentences had a quality of 0, the most highly rated sentence had a quality of 0.82. Out of 500 sentences, only 189 have a quality above 0.5. Table 10 lists the five highest and five lowest quality sentences as rated by participants.

Table 10. The five highest and lowest quality sentences as rated by participants. EBT keywords are italicized.

Sentence Quality	Sentence Text
0.00	If STAT92E also interacts with the Brahma complex, it might maintain SxlPe chromatin in an active state, facilitating the restoration of transcription after the 13th *mitosis*.
0.00	Cell-*free* studies show clearly that wild-type (but not mutant) CERT reconstitutes SM synthesis in the mutant cells.
0.00	Later in this unrehearsed (and obviously ill-prepared) *segment*, CNN went on to inform its viewers that West Nile differs from malaria in part because there is a vaccine for malaria.
0.00	Numbers are reported as the percentage of *metaphase* spreads that carry at least one chromosome break for each respective genotype.
0.00	We have identified novel functions for Sgs1 in *meiosis*.
0.82	It was considered that the amount of matter filtered through paper with pore size 6 was negligible because the *filtrate* was clear.
0.78	In a final step called *abscission*, cleavage of the intercellular bridge allows separation of the two daughter cells.
0.77	To *release* a food reward from the apparatus, the dogs had to either pull the wooden rod with their mouth or push it down with their paw.
0.77	Terminal shoots do not form an *abscission* zone at their base at any time, and therefore cannot be shed.
0.77	Most bacteria *divide* by assembling filaments of the tubulin-like protein FtsZ into a cytokinetic ring, which then constricts.

While high quality sentences are difficult to characterize, the low quality sentences can often be classified in one of two ways. Very low quality sentences are generally either highly technical sentences, or sentences that offered no significant information on any biological system. Sentences with very technical information, such as the first two in Table 10, cannot be readily understood without additional context or a strong background knowledge of biology. To the typical engineering designer, such sentences are often incomprehensible. Sentences offering no information on biological systems, such as the third, fourth, and fifth passages in Table 10, seldom contain enough information for designers to form any analogy and typically offer only random stimuli for ideation.

Because sentence quality varies so greatly and the mean quality was only 0.29, we must consider that participants could be responding to random stimuli in the passages or even in the participants' surroundings, and that the information in the sentences is has no detectable effect. If participant responses were governed by random factors, we would expect sentence quality to be a simple binomial random variable. Figure 5 shows a histogram of the sentence quality. A chi-square test on sentence quality, operating on the assumption that there is a probability of 0.29 that a participant will indicate that the sentence prompted some idea, demonstrates that the quality does not follow a binomial distribution ($p=0.00$). While this test offers no insight into why some sentences are more apt to prompt ideas than others, we can infer that *some* non-random properties distinguish high quality sentences from low. From here, we can try to find ways to distinguish between high and low quality sentences knowing that our participants' responses are more than random noise.

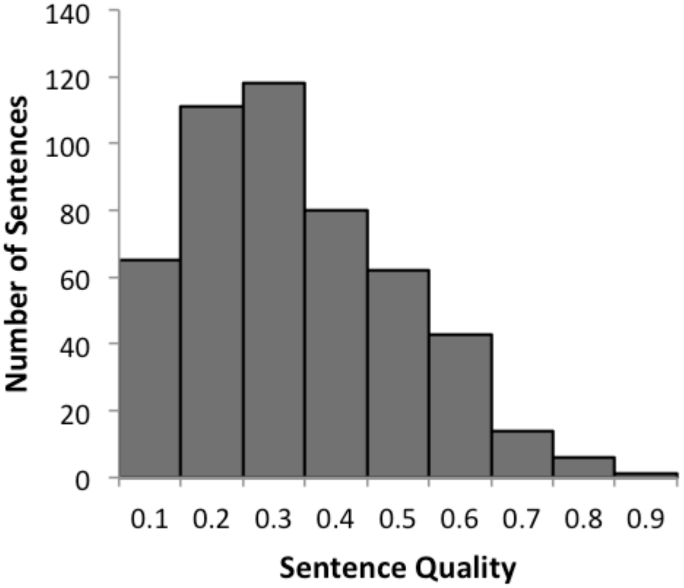


Figure 5. Histogram for sentence quality. Few high quality passages are returned using an EBT keyword search.

5.3. **Keyword Analysis**

All the passages examined in this study were drawn from a biology journal corpus using keywords from the EBT. While an examination of the sentences gives some insight into why some passages fail go promote ideation, we also need to examine the keywords used to find those sentences. Even though the EBT was designed to provide functionally meaningful keywords that would appear in biological literature, a brief look at Table 5 indicates that some of the keywords are unusual even in professional biology publications. Here we consider keyword quality: the fraction of times participants indicated that sentences containing each keyword prompted some idea. Note than sentences containing more than one keyword would be counted for each keyword.

Table 11. The quality for each keyword used in the study, along with the number of passages for each surveyed keyword. Keyword listed in order of descending quality.

Keyword	Keyword Quality	Number of Passages
deoxygenate	0.565	1
filtrate	0.429	16
abscission	0.410	27
divide	0.405	30
cleave	0.404	28
expel	0.345	28
evacuate	0.306	3
liberate	0.299	26
segment	0.291	26
dialysis	0.286	27
bleach	0.285	26
react	0.274	31
denature	0.272	8
release	0.272	32
mitosis	0.263	41
free	0.232	24
cytokinesis	0.230	32
metaphase	0.229	46
replicate	0.224	28
meiosis	0.216	29
anaphase	0.207	38
prophase	0.193	30

Table 11 lists the keyword quality for each of the keywords used in the study. While the keyword “deoxygenate” has the highest quality, only a single sentence was included in the survey, so we cannot conclude that “deoxygenate” tends to return high quality passages. Likewise, no conclusions can be drawn regarding “evacuate” and “denature”, which were used in only three and eight surveyed sentences, respectively. With the exception of “abscission,” the highest quality keywords seem to be relatively common words that would be used in many domains outside biology. In contrast, many of the lowest quality keywords refer to specific biological processes. Interestingly, five of the six lowest quality keywords pertain specifically to cell biology, suggesting that cell biology may be a poor source for design inspiration. Note however that “divide” and “cleave” also have significant meaning and usage in regards to cell biology. Examining the passages containing “divide” and “cleave” reveals that many of the higher quality passages for those keywords did not discuss cell biology or were non-technical enough for a layperson to understand.

Because high quality keywords seem to be more common words that designers would know, it seems possible that sentences with familiar keywords are more likely to encourage ideation. To explore this possibility, the relative word frequency of each keyword was measured in three domains: biology, mechanical engineering, and journalism. By examining those domains, we can examine the usage of each keyword in the domain from which our passages are drawn, biology; the domain where the designers expertise is likely to lie, engineering; and a domain familiar to all English speakers, newspapers. Word frequency was determined recording the number of results for each keyword in three different databases. For biology, abstracts of biology publications were searched using the BIOSIS Citation Index within Web of Knowledge. For mechanical engineering, abstracts of all available publications from ASME were searched using the

ProQuest Technology Research Database. Finally, to establish word frequency in newspaper publications, keyword searches were performed within the Corpus of Contemporary American English, restricting results to newspaper corpora.

The size of the searched corpus varied between domains, and the number of hits ranged between 0 and 654,000. The results from each domain were normalized with the following equation.

$$\widehat{N_\imath} = 1 - \frac{\ln(N_i + 1) - ln\big(max(N)\big)}{ln(min(N) + 1) - ln\ (max(N))}$$

where N_i is the number of hits for keyword *i,* and $\widehat{N_i}$ is the normalized number of hits that keyword. This normalization allows a logarithmic representation of frequency, forcing the most frequent keyword in a domain to have a normalized frequency very close to one, and the least frequent keyword zero. For this analysis, *min(N)*=0 in both the engineering and journalism domains. The normalized word frequencies in each domain are shown in as a stacked bar graph.

In Figure 6, the each color represents a domain, and the length of each colored segment corresponds to a keyword’s relative frequency within the domain. It is immediately apparent that “free” is the most frequent word in all domains because each segment has a length of one. “Deoxygenate” is the least frequent in all domains with each segment having a length of zero. We observe that many of the lowest quality keyword have very low frequencies in engineering and general interest publications; in fact keywords such as “anaphase” and “meiosis” did not appear outside of biology. This should match our expectations, as keywords that appear only in biology are will likely be unfamiliar to designers.

Interestingly, keywords that are very common in all domains, like “free” and “release”, also tend to have low qualities. This is likely because such words have multiple meanings and can often be used as both nouns and verbs. Such common keywords can frequently return irrelevant passages that use an alternate, and perhaps more common, meaning of the keyword. This finding supports with our observation in the sentence analysis that low quality sentences often used the EBT keyword in a non-biological context. This finding also agrees with a related study by Chiu and Shu [80] that found designers were better able to generate ideas using verbs with a specific meaning, as opposed to related verbs with a more general meaning.

We might expect that high quality keywords would be present in all three domains, but not extremely common in any domain. Keywords having such frequencies would likely be familiar to designers, but would also have a relatively narrow meaning and usage. Looking at the five highest quality keywords – disregarding “deoxygenate” because it appeared in only one sentence – we find that four of the keywords are indeed used in engineering literature, but not very commonly. Three keywords are have substantially higher relative frequency in newspaper articles while a fourth, “filtrate”, is an obvious derivative of the common word “filter”. Keyword frequency in biology literature seems to be completely independent of keyword quality.

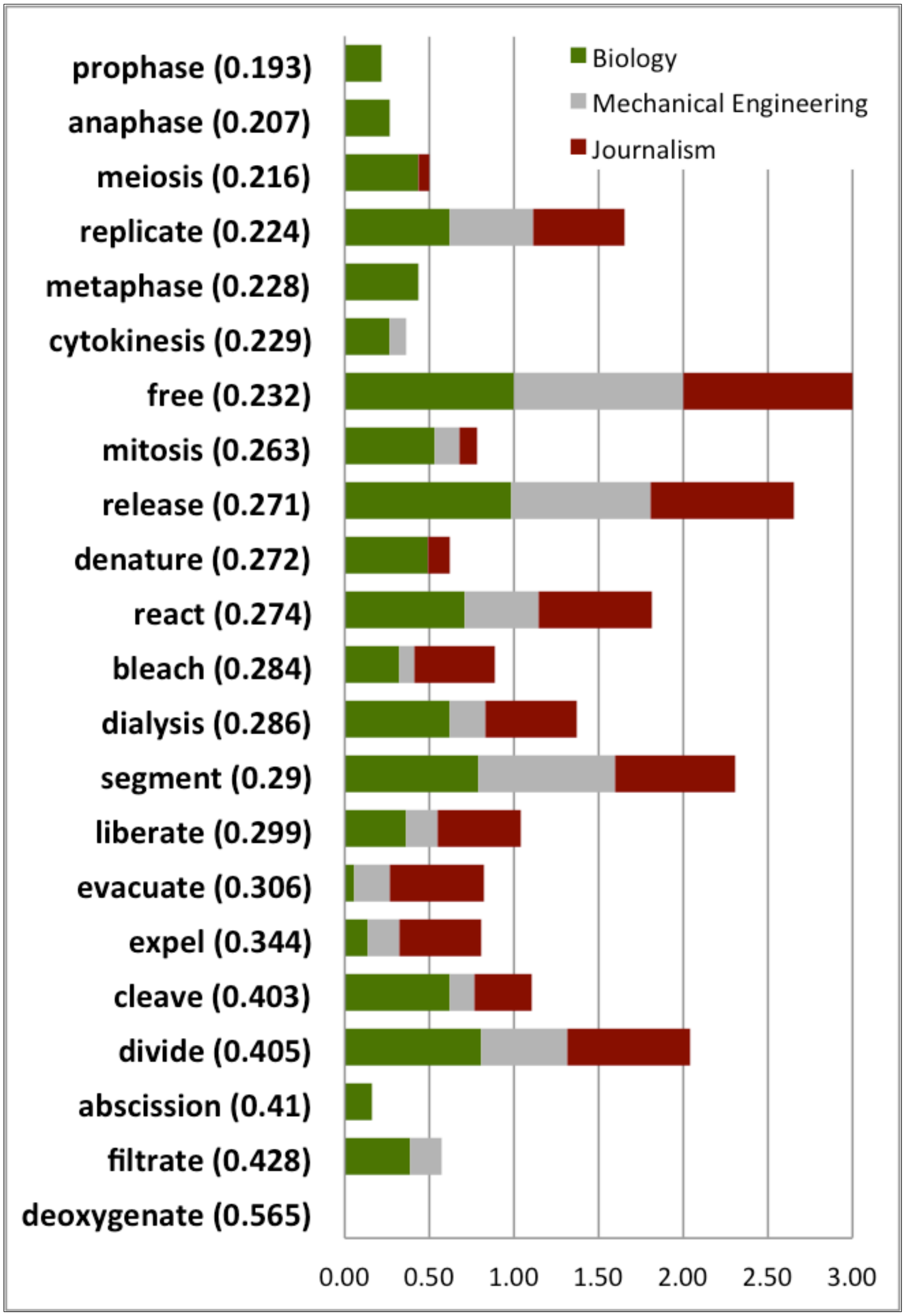


Figure 6. Relative word frequencies in biology, mechanical engineering, and journalism with quality noted beside each all keywords. More frequent keywords have normalized word frequencies closer to one. Keywords are ordered from lowest to highest quality.

While these trends in keyword frequencies may offer some guidance for directing keyword searches, important exceptions exits. “Replicate” was one of the lowest quality keywords, but its frequencies in engineering and journalism are very similar to those of “divide”, one of the best keywords. “Abscission” has the third highest quality, but it does not appear outside biology literature. Such exceptions demonstrate that ideation is more likely driven by the information within passages found using EBT keywords, rather than by the keywords themselves.

5.4. **Word Analysis**

Extending our analysis of keyword quality from the EBT, we can also examine the quality of each individual word in the passage set. Using the tokenization tools discussed out in section 6.1, we determine that the passages from the corn shucker surveys contain 3702 unique words, ignoring differences in capitalization.

It is possible that participants find one or two words within high quality sentences and use those words to generate ideas. If that is the case, words that appear in multiple, high quality sentences could be important for analogy formation. Many words in the passage set only appear in one or two passages, while only 381 words appear more in ten or more sentences. We will examine some of the highest and lowest quality words that appear in at least five sentences. lists the 20 highest and lowest quality words, along with the number of passages those words appear in, and the word qualities.

Several interesting trends can be observed in the set of high quality words. Three of the keywords from the EBT are among the highest quality words in the passage set, namely “filtrate”, “abscission”, and “cleave”. Indeed, “filtered”, the highest quality word appearing in at least five passages, is related to the keyword “filtrate”. Further, “separation” is derived from “separate”, which is the function that all the EBT synonyms are derived from. “Bacteria” and “water” could suggest specific mechanism for shucking corn, while “conditions”, “zone”, “targets”, and “bridge” may prompt designers to focus on specific aspects of the problem, such as removing only the husk at first, or specific locations on the ears of corn.

While some high quality words appear to prompt designers toward specific solution paths, others are simply common words that are meaningless outside the context of a passage. The words “along”, “another”, “able”, “few”, “then”, “they”, and “later” are all high quality keywords, but they do not suggest any action or concept that could be used to shuck corn.

Among the low quality words listed in Table 12, “mutant”, “transcription”, “wild-type”, and “antibody”, are all typically found in passages that address microbiology. For example: “It is therefore possible that Sub2p functions in conjunction with Yra1p to effect the release of mRNA from transcription site foci”. Indeed, of the 32 passages in which those words appeared, only one passage did not relate to cells, proteins, or DNA. However, beyond the relation these four keywords have to microbiology, there is no discernible pattern to low quality keywords.

Single words do not provide enough information to determine whether or not a larger passage will promote idea generation. The high quality word “form” is present in a passage with a quality of 0.12, while the low quality word “active” is in a passage with a quality of 0.41. Passages that address microbiology are often of little use to designers, but we found that the word “bacteria” has the second highest quality of any word appearing in at least five passages. Generating an idea based on a biology passage is more involved than seizing upon one or two words within the passage. To identify high or low quality passages, we need to look at more than individual words and keywords.

Table 12. List of the twenty highest and lowest quality words in the corn shucker survey passages.

High Quality Words			**Low Quality Words**		
Word	**No. Sentences**	**Quality**	**Word**	**No. Sentences**	**Quality**
filtered	6	0.517	clearly	6	0.192
bacteria	7	0.466	show	7	0.191
separation	11	0.462	reported	6	0.190
filtrate	13	0.459	mutant	11	0.188
form	14	0.456	even	5	0.188
conditions	6	0.456	mean	7	0.187
zone	7	0.456	approximately	5	0.181
along	6	0.442	male	7	0.180
targets	5	0.429	transcription	5	0.178
another	6	0.426	least	6	0.176
able	7	0.415	throughout	5	0.172
few	6	0.410	wild-type	11	0.172
then	20	0.409	fail	5	0.168
size	9	0.409	arrest	7	0.167
abscission	27	0.408	b	7	0.166
cleave	26	0.406	model	7	0.164
they	14	0.404	antibody	5	0.156
later	6	0.402	active	5	0.137
bridge	5	0.401	values	5	0.135
water	9	0.401	note	5	0.132

5.5. **Summary**

This analysis offers many important findings. Comparing the performance of the participants, we are able to find a statistically significant difference in the number sentences generating ideas between the three classes surveyed, indicating that further training in design can help a designer better use biological sources of inspiration. Importantly, participants seldom agreed on which sentences helped generate ideas; recall that the average sentence quality was only 0.29. Many sentences that prompt ideas in one individual will be meaningless to another. However, statistical analysis demonstrates that participant responses are not random, so we can infer that there are some non-random characteristics that can distinguish high and low quality sentences.

Although few high quality sentences were identified, the many low quality sentences offered insight into those factors that make a passage useless to designers. Unsurprisingly, very technical sentences filled with biology jargon are poor sources of design inspiration. Without sufficient context or background, designers cannot understand the information in such sentences to apply the information to the design problem. Analyzing the quality of the EBT keywords, we found that the lowest quality keywords included words used only in the context of biology and very common words with multiple meanings. Further, many of the highest quality keywords are uncommon, but still present in both engineering literature and newspaper articles. Examining individual words, we find that single words are insufficient for determining whether a passage containing a given word will promote idea generation. In order to predict whether a passage will promote idea generation, the next section explores the use of text classification algorithms.

6. TEXT DISCRETIZATION

Creating a text classifier based on experimental data from the Corn Shucker survey is an involved process. Developing a classifier first requires expressing sentences in terms of some feature set that can be mined. Feature selection is one the most challenging aspects of data mining [56]. This section lays out the processes used to develop a feature set from a sentences. The first four subsections will discuss each of the NLP tools used to extract feature sets. Many of the automated operations are performed using existing tools from the Natural Language Toolkit (NLTK) Project [70]. The final section will detail the feature sets used for this investigation and how they are formed using the discussed NLP tools.

6.1. **Sentence Tokenization**

Sentences are typically stored as strings of characters, but to be processed for this work those sentences strings must be broken into a set of “tokens”. Recall that a token is a distinct entity in a sentence or larger body of text. Typical examples are words, number, or punctuation mark. Tokenization is performed using NTLK’s Treebank Word Tokenizer, an automated tokenizer that forms tokens using regular expressions using rules from the Penn Treebank [81]. Specifically, the tokenizer generates tokens by performing the following steps [70]:

- Split standard contractions – e.g. don’t -> do n’t
- Treat most punctuation characters as separate tokens
- Split off commas and quotation marks when followed by white space
- Separate periods that appear at the end of the sentence

6.2. **Part-of-Speech Tagging**

A function can typically be expressed as a verb and the inputs and outputs of that function as nouns. Extending this idea, it seems possible that the parts of speech used in a sentence could play some role in the sentence’s usefulness for engineering design. A pre-built maximum entropy POS tagger from the NLTK library is used to tag each token in a sentence with a part of speech. The tagging algorithm employed applies POS tags from the Penn State Treebank Library. This application does not need the level of detail expressed by the Treebank tags, so the tags are condensed into five groups: Nouns, Verbs, Adjectives and Adverbs, Conjunctions and Prepositions, and Other. These groups allow words to be sorted into actions (Verbs), items or concepts (Nouns), modifiers (Adjectives and Adverbs), and sentence structure cues (Conjunctions and Prepositions). The Other tag primarily serves for punctuation tokens, symbols, and words that the tagger is not able to tag. Table 13 list the POS tags used by the NLTK tagger and the corresponding condensed tags used for this research.

Table 13. List of Treebank POS tags and corresponding condensed tags.

Treebank Tag	Condensed Tag	Description
CC	Conjunction/Preposition	Coordinating conjunction
CD	Other	Cardinal number
DT	Other	Determiner
EX	Other	Existential there
FW	Other	Foreign word
IN	Conjunction/Preposition	Preposition or subordinating conjunction
JJ	Adjective/Adverb	Adjective
JJR	Adjective/Adverb	Adjective, comparative
JJS	Adjective/Adverb	Adjective, superlative
LS	Other	List item marker
MD	Other	Modal
NN	Noun	Noun, singular or mass
NNS	Noun	Noun, plural
NNP	Noun	Proper noun, singular
NNPS	Noun	Proper noun, plural
PDT	Other	Predeterminer
POS	Other	Possessive ending
PRP	Noun	Personal pronoun
PRP$	Noun	Possessive pronoun
RB	Adjective/Adverb	Adverb
RBR	Adjective/Adverb	Adverb, comparative
RBS	Adjective/Adverb	Adverb, superlative
RP	Other	Particle
SYM	Other	Symbol
TO	Other	to
UH	Other	Interjection
VB	Verb	Verb, base form
VBD	Verb	Verb, past tense
VBG	Verb	Verb, gerund or present participle
VBN	Verb	Verb, past participle
VBP	Verb	Verb, non-3rd person singular present
VBZ	Verb	Verb, 3rd person singular present
WDT	Other	Wh-determiner
WP	Noun	Wh-pronoun
WP$	Noun	Possessive wh-pronoun
WRB	Adjective/Adverb	Wh-adverb

To check the accuracy of this prebuilt tagger, it was evaluated on a sample of manually tagged sentences drawn randomly from the corn shucker survey's sentence set. A set of thirty sentences, comprising 966 distinct tokens, was used for this analysis. The manual tagging was performed with the condensed tag set. Overall, the tagger agreed with 85.4% of the manual tags. Table 14 lists the performance of the tagger for each tag in the condensed set. The tagger generally performs well, but it has difficulty identifying adjectives and adverbs. Generally, the tagger misclassifies these as nouns and verbs, respectively.

Table 14. Number tokens for which the NLTK tagger agrees and disagrees with the each of the manual tags.

Condensed Tag	Same Tag	Different Tag	% Agreement
Nouns	270	27	0.909
Verbs	93	22	0.809
Adjective/Adverb	86	49	0.637
Conjunction/Preposition	166	10	0.943
Other	210	33	0.864
Total	825	141	0.854

6.3. **Stopword Removal**

Many words used in common writing typically have no significance in text classification. These words are known as stopwords. While a list of stopwords can be compiled for any text mining operation, this work employs a published list of 571 stopwords from the SMART System [82]. Word tokens are checked against the stopword list and all stopwords are removed from sentences before they are stemmed.

6.4. **Word Stemming**

A word stemmer is used to group lexically similar words into distinct word stems by stripping suffixes from the words. This process reduces the number of features in the feature set. Using a stemmer allows related words like regulate, regulatory, and regulated to be expressed as a single word stem, “reg”. The shortcoming of stemming algorithms is that very different words, like regulate and region, may be condensed into the same stem. Using word stems for text classification looses some precision to conflated word stems, but improves classifier speed, and sometimes performance, by reducing the number of features. As discussed in section 3.2.3, many different stemming algorithms can be used, but four algorithms were evaluated for use in this application.

The Porter Stemmer, Lancaster Stemmer, Snowball Stemmer, and a WordNet-based lemmatizer were evaluated using the sentence set from the classifier training data. The sentence set contains 1869 unique words, not including stopwords. Table 15 summarizes the performance of each stemmer. Ideally, a stemmer should minimize the number of word stems – condensing as many similar words into stems as possible – without condensing dissimilar words to the same stem. The accuracy of the stemmers is gauged with a sample of 150 randomly selected word stems, each of which corresponds to at least two of the 1869 words. For each of the stems, the number of distinct groups of words in the corresponding words was counted. For example, “region” and “regional” would belong to one group, while “regulate” and “regulation” would belong to another.

The average of those concept counts was averaged across the 150 stems for each stemmer.

As Table 15 shows, the accuracy of the stemmer decreases with the number of word stems. This is a natural consequence of stemming. When words are stemmed to aggressively into word stems, unrelated words can easily be reduced to the same stem. For example, “manner”, “manners” and “mannerism” are all related words that could be stemmed to “mann” with little risk of confusion. But if those words were stemmed to “man”, they could share that stem with the unrelated word “man.” In contrast, the dictionary based WordNet lemmatizer would map “manners” to the lemma “manner”, but regards “mannerisms” as a distinct lemma. For this application, the Lancaster stemmer is used to generate different feature sets. The Lancaster stemmer generates the fewest word stems while providing reasonable accuracy.

Table 15. Summary of the performance for four stemming algorithms.

	Porter	**Snowball**	**Lancaster**	**WordNet**
Word Stems	1513	1507	1425	1736
% Reduction	0.190	0.194	0.238	0.071
Average Concepts/Stem	1.007	1.013	1.120	1.000

6.5. **Features for Text Mining**

The choice of features for text classification is a critical component of the classification process. Ideally, we would like to have a feature set that contains no more that the minimum level of information needed to characterize the sentences for classification. A strong understanding of the classification problem, specifically knowledge of what features will be most important for the classification task, can help greatly reduce the number of features needed to represent a data set. Unfortunately, we have no well-defined heuristics for determining what passages will prompt idea generation. Consequently, we base our feature sets around the basic “bag-of-words” model, where each word in a passage is a feature.

Two candidate feature sets are developed and examined in this research. These sets are discussed in the following subsections will be referred to as the word feature set and the noun feature set. In the word feature set, the stop words are removed from the training sentences, remaining words stemmed, unique and very common word stems removed, and several grammatical features added. In the word feature set, all words are stemmed and grammatical features added. In the noun feature set, only stems from nouns are included. This second feature set is prompted by the tendency of designers, observed in the idea generation experiments, to focus on nouns when forming solutions. A full list of features for the both feature sets is included in the appendices.

6.5.1. **Word Feature Set**

The word feature set is formed by removing stopwords, word stemming, and POS tagging. Figure 7 shows the process for forming this feature set. The raw text of the sentence training data set is tokenized. From the tokens, all stopwords are removed. The remaining tokens are stemmed to form a list of word stems. Any stems that appear only once in the sentence set are not included in the feature set; unique stems cannot help a classifier because they do not appear in more than one sentence. Similarly word stems that appear in more than 10% of sentences in the training data are also removed because the high usage frequency indicates that the word stem relates to a very common word.

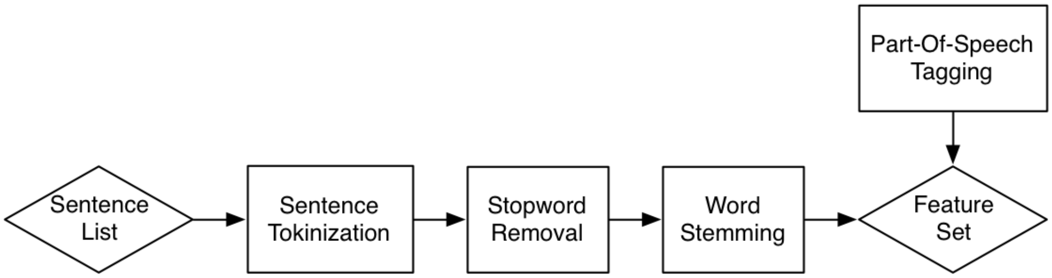


Figure 7. Flow chart of how a feature set is built from a sentence set.

Because a bag-of-words model lacks information about sentence structure, several additional features are added based on the parts of speech found in the sentences, as well as some simple length measurements. These features are:

1. Number of tokens in the sentences
2. Average token length
3. Maximum token length
4. Percentage of Noun Tokens
5. Percentage of Verb Tokens
6. Percentage of Adjective/Adverb Tokens
7. Percentage of Conjunction/Preposition Tokens
8. Percentage of Other Tokens
9. Percentage of non-alpha Tokens

The first of these features gives a simple measure of sentence length. This could be relevant because short sentences may tend to contain too little information to encourage idea generation. Alternately, very long sentences may tend to be too complex for designers to understand.

The second feature, average token length, gives some idea of the difficulty of the language in the sentence. A sentence filled with many long words is likely to be more challenging than one with a lower average token length. The third feature also relates to this concept of sentence difficulty by giving the length of the longest word in the sentence.

The next four features all relate to the grammatical composition of a sentence. A sentence a high percentage of nouns should have a larger number of “things”, physical or conceptual, from which to draw ideas. Sentences with more verbs deal more with actions or functions. Adjectives and adverbs act as modifiers, and higher values for this feature could indicate a more detailed passage. Conjunctions and prepositions are used to add structure to sentences. While we will always expect the percentage of conjunctions and prepositions in a sentence to be low, more of those parts of speech will indicate a more complex sentence structure.

The last two features provide an indicator primarily for non-word tokens. These include symbols, punctuation marks, numbers, and unusual character strings that the POS tagger cannot identify.

The word feature set is intended to offer a relatively dense feature set. The number of features reduced from a pure bag-of-words representation through stemming and stopword removal. Reducing the feature set will improve the speed of our classifiers, but sacrifices some information about the sentences with stopword removal and confounds some words through aggressive word stemming. In total, the word feature set has 565 features: the nine listed features and 556 word stems. The nine listed features are continuous numeric features, while the word stem features have integer values to indicate how many times each stem appears in the abstracted sentence.

6.5.2. **Noun Feature Set**

The noun feature set is very similar to the word feature set, but only word stems from nouns are included. This restriction is based on an observation in the passage stimuli ideation experiments. When describing their ideas, participants often seemed to focus on items or concepts within the passages, generally expressed as nouns. For instance, the noun “water” in one sentence prompted several solutions using water jets or water tanks. Another sentences with the word “spider” saw solutions involving adhesives to bond to the husk or silk. If nouns are a main driver of idea generation, then it is possible that nouns alone will provide enough information to reliably classify sentences. The process for developing this feature set is the same as that depicted in Figure 7, except an additional filtering step is performed before word stemming to remove all words except nouns, as identified by the POS tagger. The noun feature set has a total of 375 features. A full list of the features is included in the Appendix D.

7. TEXT CLASSIFICATION RESULTS AND DISCUSSION

One of the goals of this research is to automate the identification of passages in a biology corpus that will help designers generate ideas for a design problem. This section addresses the development and evaluation of text classifiers for labeling the high and low quality passages that were found in the corn shucker survey. Additionally, we will discuss the classifiers' performance identifying passages to prompt ideas for the alarm clock design problem, which is a very different problem domain from the corn shucker where the classifiers are trained. The first subsection will discuss the selection of training data from the corn shucker survey result. The development and tuning process for the classifiers will then be discussed. Section 7.2 evaluates the classifiers in detail. Section 7.3 discusses the performance of the classifiers with the alarm clock problem. The last section reexamines the classifiers' performance when applied to the entirety of the corn shucker survey responses.

7.1. **Training Data Selection**

The corn shucker survey that was discussed in section 4.3 was designed to allow a large number of passages to be viewed and rated by a large number of participants. The survey collected a total of 10000 responses across 500 different passages. While section 0 discusses many aspects of the survey, here we address the process of extracting training data for the classifier from the raw survey results.

Text classifiers are machine-learning algorithms and as such require training data to build the classifier. This training data consists of a number of sentences, expressed a set of features, with each sentence assigned a class. In this case, the class is a Boolean value to indicate whether or not each sentence promotes idea generation. For convenience we will refer to the classes as “good” and “bad” sentence classes. As we saw in section 5, sentence quality has a wide range, so some threshold value for sentence quality is needed to label sentences as good or bad. To further complicate matters there must be significant differences between the good and bad sentences if a classifier is to tell the two groups apart.

Ultimately, we use a statistical criterion to identify sets of good and bad sentences. We treat participant ratings as a series of n binomial trials (e.g. a coin toss) with a set probability, p, of a sentence being rated “good”. In that case, the odds of a passage having less than some number, x, “good” ratings out of the passage’s n ratings will be the cumulative binomial distribution of x given n trials with probability p, which is typically written,

$$\alpha = P(X \le x) = B(x; n, p)$$

Working backwards, we can find values of x given a criterion value α. Applying this to the corn shucker survey, we established that the average odds of a single sentence being rated good is 0.289. So we can find (assuming that the sentence rating is a binomial process) that if a passage has 22 ratings there is at most a probability of 0.9 that out of 22 trials, 9 or fewer ratings will be “good”. This gives us a threshold value for assigning classes; a passage with 22 ratings must have at least 9 “good” ratings to be included in our set of good sentences that will train the classifier. We can similarly find threshold values for bad sentences. Formally, a sentence with n ratings is assigned a class as follows,

$$define\ x = f(\alpha, n, p)\ \ such\ that\ \alpha \leq B(x; n, p)$$

$$x_{min} = n - f(0.9, n, 1 - 0.289)$$
$$x_{max} = f(0.9, n, 0.289)$$

$$if \begin{cases} X < x_{min} & : class = bad \\ x_{min} < X < x_{good} & : no\ class \\ X > x_{max} & : class = good \end{cases}$$

Note that some sentences fall between the thresholds. These sentences are considered to ambiguous to be labeled as a good or bad sentence in the training data.

While the statistical criterion outlined above is a convenient and repeatable criterion for assigning classes to sentences, the underlying assumption that sentence rating is a binomial process is false, as demonstrated by the chi-square test discussed in section 5. Nevertheless, this statistical criterion is a consistent and repeatable way to assign classes to sentences when building a set of training data.

Out of the 500 sentences examined in the corn shucker survey, 126 are labeled good sentences with qualities greater than 0.408, while 160 sentences are labeled bad sentences with qualities less than 0.183. The remaining 214 sentences are considered too ambiguous to be included in the training data.

7.2. **Classifier Tuning and Evaluation**

This work examines four different classification algorithms using two different feature sets to find a good BID Classifier. The algorithms are a k-Nearest Neighors classifier, a Naïve Bayes classifier, a Support Vector Machine classifier, and the CN5 rule learning classifier. These classification algorithms, which were introduced in section 3.1, each have different strengths, and their performance will be examined in the following subsections. Each subsection will briefly explain the algorithm, discuss any tuning that was performed, and evaluate the classifiers performance with using both feature sets discussed in section 6.5.

The BID classifier makes a binary class assignment to the passages, i.e the passage is / is not likely to promote idea generation. The classes assigned to passages by the classifier can be characterized as true-positive (TP), false-positive (FP), true-negative (TN), or false-negative (FN), according to how the classifier's assignment matches the passages' classes from the training data set. For example, a false-positive result would be to classify a low quality passage as high quality. Because the classifier's intended use is to identify sentences from a biology text that will aid idea generation, we are primarily concerned with the positive classifications.

Three basic metrics will be used repeatedly when evaluating the classifiers: Precision, Recall, and an F1 Score. In the context of a binary classification task, Precision is the fraction of correctly true positives to all positives. Formally,

$$Precision = \frac{TP}{TP + FP}$$

Recall is the fraction of high quality passages that were correctly classified.

$$Recall = \frac{TP}{TP + FN}$$

The F1 score is simply the harmonic mean of Precision and Recall. The F1 score is a single value that helps us gauge the balance of precision and recall for each classifier.

$$F1 = 2\frac{(Precision)(Recall)}{Precision + Recall}$$

Relating back to our intended application, suppose that we use one of our candidate BID classifiers create a set of 100 passages that should, according to the classifier, help us generate solutions to a design problem. If the classifier has a Precision of 0.8 and Recall of 0.4, we would expect that 80 of the passages will help us and that the classifier discarded 120 other passages that would have also been helpful.

7.2.1. **k Nearest Neighbors**

The kNN algorithm is a so-called "lazy" algorithm. In order to classify each new data point, the algorithm compares to the data point to the entire training data set and assigns a class based on the classes of the *k* "closest" data points in the training data. One of the most important variables when using a kNN classifiers is the choice of *k.* If k is too small, a simple sampling bias could cause many poor classifications. However, if *k* is too large the new data point could be classified based on very distant and unrelated passages. Other adjustable factors for a kNN classifier include defining how the distance between data points is calculated and assigning different weights to the data points in the training data set.

The kNN classifier measures the distance between passages as the Euclidean distance between the passages' representation in the employed feature set. Also, all passages are weighted equally, so no one passage is considered more important than the others.

The parameter *k* was manually tuned by testing the performance of the kNN classifier with the word feature set setting different values of *k*, ranging from 10 to 100 in increments of 10. The *k* corresponding to the highest F1 score, *k* =60, was selected.

Table 16 lists the precision, recall, and F1 score for the kNN classifier under a 10-fold cross-validation using each feature set. Using the word stems in word feature set, the classifier was able to correctly classify about 62% of the passages as high-quality passages. For comparison, recall that participants in the corn-shucker survey only found and average of 29% of passages to be helpful. In the training data itself, 44% of the sentences are labeled as high-quality. Using noun stems in the noun feature set, we see that the classifier's precision drops to 47%, but its recall jumps to 87%.

Precision is generally more important that recall for a BID classifier. The problem with keyword search methods is too many irrelevant passages are returned by the searches, forcing a designer to read many irrelevant passages before finding one that helps find a design solution. Even if the classifier discards many potentially helpful passages, we would prefer to give the designer as few irrelevant passages as possible.

Table 16. 10-fold cross-validation performance summary for the kNN classifier on both feature sets, with *k*=60.

	Precision	Recall	F1 Score
Word Feature Set	0.622	0.484	0.545
Noun Feature Set	0.468	0.865	0.607

Overall, the kNN classifier performs well for our application when using the word feature set, though higher recall would be preferable. The biggest drawback of using a kNN classifier for a BID classifier it that kNN classifiers are computationally inefficient. For every new passage that is to be classified, its distance to all the training passages must first be computed. These calculation can be quick when the training data set only contains a few hundred passages, but will slow as the labeled training set grows.

7.2.2. **Naïve Bayes**

The Naïve Bayes classifier is strictly probabilistic stemming algorithm that computes the probability of a passage being high or low quality based on the cumulative probability of the passage's features being in a high or low quality passage. This classification method makes the assumption that all features are independent variables. Because the classifier is making a binary class assignment, it can use a probability threshold as a classification criterion: if the probability of a passage being high-quality exceeds the probability threshold then that passage is classified as a high-quality passage. The Naïve Bayes classifier here automatically adjusts its probability threshold to maximize the number of correctly labeled passages (TP+FP).

Table 17 lists the precision, recall, and F1 score for the Naïve Bayes classifier under a 10-fold cross-validation using each feature set. Using the word stems in either feature set, the classifier is able to correctly classify about 44% of the passages as high-quality passages, but the recall of 100% shows that all the high quality passages are correctly classified. While impressive at first, further investigation shows that the Naïve Bayes classifier is actually labeling *all* passages as high quality passages. The Naïve Bayes classifier is unable to correctly identify a single low quality passage. Rather than attempting to distinguish between the two classes, the Naïve Bayes classifier's best performance is achieved by labeling every passage as high quality. This classification method provides absolutely no improvement on the current keyword search method.

Table 17. 10-fold cross-validation performance summary for the Naïve Bayes classifier on both feature sets.

	Precision	Recall	F1 Score
Word Feature Set	0.441	1.00	0.612
Noun Feature Set	0.441	1.00	0.612

The unexpected behavior of the Naïve Bayes classifier prompted a carful examination of the feature set and training data. There are a total of 286 passages in the training data set. Using 10-fold cross validation, the classifiers are trained on 257 or 258 passages and tested on 29 or 28. However, the two feature sets employed have 565 and 375 features. There are more features in our data set than passages. This imbalance effectively forces the SVM and CN2 classifiers to overfit their models to the data. In the case of the Naïve Bayes classifier, all features are weighted equally, and the "optimal" probability threshold is sufficiently low that all passages are classified as high quality. Because the kNN classifier is an instance-based classifier, its performance is not compromised.

7.2.3. Support Vector Machine

When sentences are expressed as a set of *n* features, a set of passages can be thought of as a set of *n*-dimensional vectors in the feature space. Support vector machines find the best hyperplane in the feature space that divides high quality passages from low quality passages. The plane is defined by the set of passage that lie closest to the plane. Those passages are called support vectors. Usually, rather than looking at the actual feature space where we define our vectors, we perform some transform, a kernel trick, to the space that helps deal with nonlinearities. Support vector classification is an efficient for of classification because a new passage can be classified simply by determining on which side of the hyperplane the passage lies.

There are several different types of support vector machines. This work uses the ν-SVC support vector algorithm [83]. In ν-SVC, a single parameter ν is used to adjust the performance of the classifier. ν acts as an lower bound on the fraction of vectors used as support vectors, and an upper bound on the fraction of vectors that lie on the wrong side of the hyperplane, called margin errors. Values for ν lie in the interval [0,1]; selecting too high a value for ν tends to create a classifier that overfits the data.

There are infinitely many kernels that could be used to transform the vector space. This work uses a radial basis function (RBF) that transforms the distance between two vectors with the function,

$$K(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\gamma \|\boldsymbol{x} - \boldsymbol{x}'\|^2\right)$$

where $\|\boldsymbol{x} - \boldsymbol{x}'\|$ is the Euclidean distance between **x** and **x'**. γ is a tunable parameter that lies in the interval $(0,\infty)$. Increasing γ effectively smooths irregularities in the data.

The SVM classifier is tuned by adjusting ν and γ. A coarse grid search was performed varying values of ν between 0 and 1 in steps of 0.1, and γ from 10^{-5} to 10^{4} by powers of 10. SVM classifiers using each combination of parameters were evaluated with the word feature set in terms of there percent agreement with the training data labels using a 5-fold cross validation. The results of this search are shown as a colored scatter plot in Figure 8. Based on this search, our SVM classifier was set with $\nu = 0.6$ and γ=0.1.

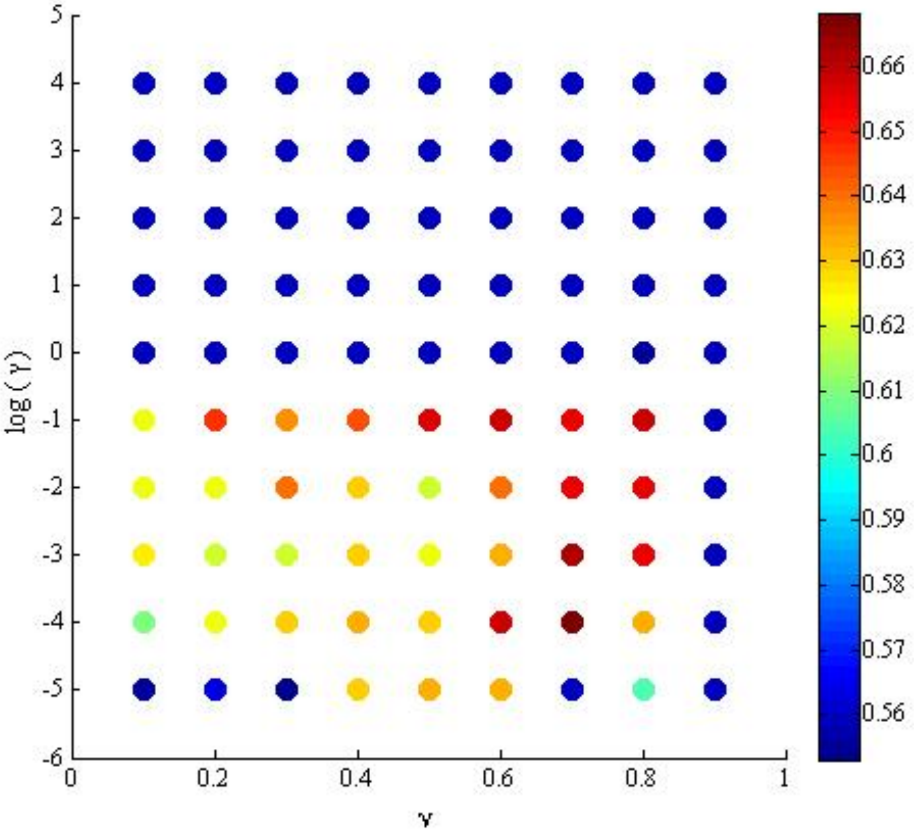


Figure 8. Tests of SVM classifier performance when varying the SVM's parameters. % agreement between passage labels and classifications is shows as color.

Table 18 lists the precision, recall, and F1 score for the SVM classifier under a 10-fold cross-validation using each feature set. SVM classification gives us the best performance with the word feature set of any classifier we have examined. If we were to take two sets of passages, one selected using the existing keyword search method, and another that were classified as high-quality passages by our SVM classifier, would could expect to find more than twice as many helpful passages in the classified passage set.

Moreover, the classifier has 58% recall, so it is correctly classifying about 10% more of the high-quality passages than the kNN classifier earlier.

When using only the noun stems of the noun feature set, the SVM classifier still performs reasonably well in terms of precision, but it has poor recall. 63% of all high-quality passages are misclassified using the noun feature set. If we restrict our features to noun stems, we get better performance with the kNN classifier seen earlier. However, as we pointed out in the discussion of the Naïve Bayes classifier, the number of features in our feature sets does force the SVM classifier to be overfitted to the data.

Table 18. 10-fold cross-validation performance summary for the SVM classifier on both feature sets.

	Precision	Recall	F1 Score
Word Feature Set	0.635	0.579	0.606
Noun Feature Set	0.573	0.373	0.433

7.2.4. CN2 Rule Induction

Rule induction algorithms are designed to generate a set of if-then rules to classify data. This format has the advantage of generating classification rules that are easily understood by the user. This work uses the CN2 (unordered) rule induction algorithm. The CN2 (unordered) algorithm uses a beam search to search through all candidate rules and a significance test to determine if a candidate rule is too specialized. For this study, we set the significance level at 95%, and the beam width at 4. Table 19 summarizes the classifiers performance for both feature sets.

Table 19. 10-fold cross-validation performance summary for the CN2 classifier on both feature sets.

	Precision	Recall	F1 Score
Word Features	0.652	0.46	0.54
Noun Features	0.542	0.357	0.431

We see that the CN2 classifier does very well with both feature sets, actually achieving better precision on that the SVM classifier with word features and nearly as well with noun features. It is somewhat surprising that such simple expression of if-then rules would allow such accurate classification. Since the CN2 classifier is a rule induction algorithm, it is worthwhile to examine some of the rules produced. Table 20 lists several rules used by the classifier.

The first rule tells us that if there are more than 10 symbols in a passage (not a lot of numbers, equations, or punctuation marks), along with less than 5 verbs, and there is no more than 1 "min" word stem, and the word stem "expel" is not in the passage, then passage should be classified as low quality. This very non-intuitive rule actually applies to 18 passages, about 5% of the training data. The next rule tells us that if the word stem "cel" appears less than 3 times in the passage (usually words like cell and cellular), but the stem "divid" (e.g. divide), and the average token length is greater than 3.83 (the passage probably contains some long words), the passage is good. This rule applies to 18 low quality passages times. The last two rules simply tell us that passages with the word "ring" should have high quality, and those with "state" or "stated" should have low quality.

Table 20. Selected rules generated by the CN2 algorithm.

Rule	High Quality	Low Quality
IF No. Non-Alpha Tokens>10.0 AND No. Verbs<=5.0 AND min<=1.0 AND expel<=0.0 THEN good/bad=0	0	18
IF divid>0.0 AND Average Token Length>3.83 AND cel<=3.0 THEN good/bad=1	15	0
IF ring>0.0 THEN good/bad=1	0	3
IF stat>0.0 THEN good/bad=0	0	3

The first rule in Table 20 is useful from a classification standpoint, but it does not help us understand why those 18 sentences have low quality. The other three rules suggest some interesting trends. When the word "cell" appears in a sentence without the word "divide", the sentence will likely have high quality. We might deduce then that sentences about cell division may not be as useful. From the third rule, we could suppose that the word "ring" alone could generate ideas for shucking corn. In contrast, sentences that talk about entities that are in "steady-state" or report what someone "stated" may not be as useful.

The CN2 classifier gives us interesting information about which words or combinations of words appear in high or low quality passages. However, for a set of 286 sentences in the training data, 183 rules were generated. Further, many rules only apply to one or two passages. While the CN2 classifier gives us good performance, it does overfit its induced rules to the data. Ultimately, rule induction is a poor choice for a text classification task because the structure of the if-then rules generated forces the rules to over-fit the data.

7.3. **Out-of-Domain Classifier Performance**

Section 7.2 extensively examined the performance of several classifiers with data from the corn shucker survey. All that analysis, however, focuses on the results from a single design problem, corn shucking, focusing on a single function, separating. Now we will examine the classifiers' performance with passages relating to a different design problem, with a different function. In the alarm clock survey, participant are asked to evaluate passages relating to the function "signal". We would certainly expect that the passages that suggest ideas for developing a silent alarm clock would be substantially different from passages that suggest ideas for shucking corn.

Here we examine how well our classifiers perform with passages related to a different function. To make this comparison, we train each classier with the full training data set outlined in section 7.1, and then use the trained classifiers to label passages from the alarm clock survey. All the parameters for the classifiers remain the same as detailed in section 7.2. By comparing the passage labels from the classifiers to the human responses from the alarm clock survey, we can calculate the precision and recall for each classifier in this new domain.

Recall from section 4.4 that the alarm clock survey contains 100 randomly selected passages containing biological synonyms for the function "signal". The survey collected responses from 45 people, but responses from 9 people are excluded from the analysis because the records from eLearning indicate that those 9 people were not actually reading the passages. The same criteria are used for this exclusion as were explained in the beginning of section 5, i.e. the participants moved through the survey faster than they could plausibly read the passages. On average, the remaining participants indicated 36.2% of the passages prompted some idea for creating an alarm clock. As we did with the corn shucker survey in section 7.1, we label passages as high quality, low quality, or ambiguous based on participant survey responses. Using the same binomial test, 23 passages are labeled as high quality passages, 48 as low quality, and 29 as ambiguous.

Table 21. Precision and recall of each classifier on alarm clock passages.

	SVM	**kNN**	**NB**	**CN2**
Precision	0.47	0.30	0.32	0.33
Recall	0.65	0.13	1.00	0.22
F1 Score	0.27	0.09	0.24	0.13

Table 21 lists the precision and recall of each of the BID classifiers when applied to the alarm clock passages. If we regard a classifier's precision as the percentage of high quality passages that would likely appear in a set of passages selected by the classifier, then only the SVM classifier is able to outperform the "precision" of our random sample, 37.7%. Even so, the SVM classifier is wrong about more than half of the passages it labels as high quality. Both the kNN and CN2 algorithms perform very poorly compared to their performance with the corn shucker survey. The Naïve Bayes classifier persists in labeling all passages as high quality passages.

If text classifiers are to be used to help provide designers with more relevant information for bioinspired design, the classifiers must be tied to functions. This analysis shows that a classifier trained in one problem domain with one function does not function effectively when its classification criteria are applied to a different problem and function. To demonstrate how easily a classifier can mislabel a passage, consider the following example.

One of the highest quality passages identified in the corn shucker surveys is: “It was considered that the amount of matter filtered through paper with pore size 6 was negligible because the *filtrate* was clear”. We will assume that a classifier based on the corn shucker training data would label this passage as a good passage. However, it seems unlikely that a designer would use this passage to help develop a silent alarm clock. However, if we add the phrase “see above” to the end of that passage, the passage would contain a biological synonym for the function “signal”. Neither “see” nor “above” correspond to any features in the word stem feature set, so this small change would be invisible to the classifier. The sentence would be labeled as high quality despite its inappropriateness for designing an alarm clock. If any universal structure exists that can characterize high-quality passages across multiple functions, that structure is not expressed in the feature representations used in this work.

7.4. **Full Data Classifier Performance**

In section 7.2 we discussed the problems that having a larger number of features than passages caused for the Naïve Bayes, SVM, and CN2 classification algorithms. To resolve these issues, we must either use a significantly smaller feature set, or increase the number of data points in the training data. In this section we expand the training data set by using all participant responses from the corn shucker survey rather than using the data selection procedure outlined in section 7.1. The classifiers are trained using all 500 sentences from the survey, with about 20 instances of each sentence, one for each participant response. There are a total of 10000 instances on which the classifiers are trained for this application. This analysis is based on a 5-fold cross validation with all instances of a single sentence being contained in a single fold. Sentences are all represented with the word feature set for this analysis. The classification tasks examined here are performed using the Waikato Environment for Knowledge Analysis (WEKA) software, a common data mining software package.

Because a different software package is used for this analysis, with a different selection of training data, there are some changes to the classifiers and parameters outlined in section 7.2. The Naïve Bayes classifier is unchanged. The kNN classifier is adjusted to set k=200, approximately the responses from 10 sentences. The SVM classifier is adjusted to set $\nu = 0.5$. The CN2 algorithm is not available in WEKA, so a Decision Table (DT) algorithm is used instead [84].

Table 22 summarizes the performance for each of our classifiers. For reference, a default classifier that labels every instance as high quality would have a precision of 0.29, a recall of 1.00, and an F1 score of 0.22. Not a single classifier examined has an F1 score as high as our hypothetical default classifier. However, the Naïve Bayes classifier performs better than any other examined here with an F-1 score of 0.17. Although both the SVM and Decision Table classifiers have precisions of at least 0.36, the recall for those classifiers is extremely low; most high quality sentences are mis-classified. The kNN classifier acts as a majority classifier in this application, classifying all passages as the most common class, low quality.

Participant responses from the corn shucker survey are generally inconsistent. As we saw in Figure 5, many passages have a quality between 0.3 and 0.5. This ambiguity prompted the data selection discussed in section 7.1. While the Naïve Bayes classifier was able to perform acceptably, though not well, with the full corn shucker survey responses, the other classifiers are practically unusable.

Table 22. Precision, recall, and F1-score of each classifier applied to the full corn shucker survey data.

	SVM	kNN	NB	DT
Precision	0.77	0.00	0.36	0.36
Recall	0.01	0.00	0.34	0.06
F1 Score	0.01	0.00	0.17	0.05

The kNN classifier performed quite well with the limited training data used in the previous section. Further, since the kNN is an instance based classifier, its performance is not compromised by the overly-large feature sets employed. Indeed, in many applications, the number of training documents used with kNN classifiers is deliberately limited to speed the classifications [56]. Rather than using many marginal instances, training data is limited to a smaller set of prototypical or representative examples of the classes. Based on the tests performed in this work, a kNN classifier seems to be the best algorithm for BID classification.

8. DESCRIPTION OF IDEA GENERATION EXPERIMENTS [*]

The idea generation experiments used in this research all follow a similar format that gives participants as short design problem and allows them to write or sketch solutions to the problem over a set time period. All experiments used a common design problem: designing a system or device to shuck large quantities of corn. Slight variations were added to the design problem to accommodate different experimental conditions. A copy of the design problem is given in Figure 9. All participants in these idea generation experiments first read and sign a consent form. The participants are then given a packet containing a short design problem and blank pages to record their solutions – though in one experimental condition the blank pages have a single sentence at the top to encourage idea generation. The packets also have a page immediately following the problem description asking the participants to indicate whether they had heard about the design problem prior to the experiment and whether they had generated any solutions in advance. Depending on the conditions of the experiment, the participants may receive an additional sheet of paper with sentences intended to act as design stimuli. After the packets are distributed, the experimenter reads verbal instructions from a script that varies slightly depending on the experiment's condition. Participants then are given 50 minutes to develop and record their solutions to the design prompt. Participants are notified when they have ten minutes remaining. All participants in idea generation

---

[*] Part of this chapter is reprinted with permission from "Evaluating the Directed Method for Bioinspired Design", M.W. Glier, J. Tsenn, J.S. Linsey, D.A. McAdams, 2012, Proceedings of the ASME 2012 International Design Engineering Technical Conferences & Computer and Information Engineering Conference, IDETC2012-71511. Copyright 2012 by ASME

experiments are compensated with either extra credit in the class they were recruited from, or monetarily with $20 cash.

Design Problem - Device to Aid in Shucking Corn

Problem Description:
Corn is currently the most widely grown crop in the Americas with the United States producing 40% of the world's harvest. However, only the loose corn kernels are used when bought canned or frozen in grocery stores. An ear of corn has a protective outer covering of leaves, known as the husk, and strands of corn silk threads run between the husk and the kernels. The removal of husk and silk to clean the corn is known as shucking corn.

[Additional instructions are given for specific experimental conditions]

http://www.art-photograph-gallery.com/pictures-of-corn.html

Customer Needs:
- Must remove husk and silk from corn cob with minimal damage to kernels.
- A large quantity of corn must be shucked quickly.
- Low cost.

Please sketch and note (with words) all design solutions that relate to a biology passage on the same page as that passage.

Figure 9. All ideation experiments use this corn shucker design prompt. Additional instructions are included as needed for each condition.

The following subsections detail three ideation experiments used in this research. All the ideation experiments follow the format described above, but the variations in the instructions are noted for each experiment, along with particulars such as experiment locations and the participant demographics.

8.1. **Unstructured Ideation Experiment**

This first ideation experiment establishes a baseline for idea generation with no stimuli or structured method to act as a design aid. For this experiment, participants were given the design problem shown in Figure 9, with the sole additional instruction: "Design a device that quickly and cheaply shucks corn for mass production." Participants are verbally instructed to write or sketch their ideas and to attempt to generate as many ideas as possible with the greatest quality and variety possible, even if those ideas may not be technically feasible. A full copy of the verbal instructions is included in the Appendix B.

All participants in this experiment are recruited from MEEN 401, the first of two capstone design courses in Texas A&M's mechanical engineering curriculum, during the Fall semester of 2011. The experiment was conducted during class time in a large lecture hall with no windows. By testing the students before they were taught formal ideation methods, we are able to gauge engineering students' natural abilities without their results being biased by other ideation methods.

This experiment is itself part of a larger experiment that investigates the effect of the directed approach to bioinspired design. A further discussion of that experiment can be found in Glier et al. [11]. The participants in this experiment are a group of 24 senior-level mechanical engineering students.

8.2. **Biology Passage Stimuli Ideation Experiment**

This idea generation experiment is a two-factor, three condition experiment designed to investigate how passages from biology texts are able to stimulate idea generation. The passages used in this experiment were selected from those used on the corn shucker survey described in section 4.3. The factors for this experiment are the presentation format of the passage stimuli, and the "quality" of the passages. The passages are presented to participants either all at once on a single sheet of paper, separate from the ideation packet, or one at a time with each passage being printed at the top a single sheet of the ideation packet. These formats will be referred to as a parallel and serial presentation, respectively. The quality of the passages is determined experimentally in the corn shucker survey. High quality passages are those that prompted some idea for the majority of survey participants. Conversely, low quality passages prompt an idea in few, if any, survey participants.

Each of the three conditions in this experiment presented participants with a set of passages from our biology corpus to act as design stimuli. The three conditions featured parallelly presented high-quality passages (HQP), parallelly presented low-quality passages (LQP), and serially presented high-quality passages (HQS). While a fourth condition with serially presented low-quality passages would have made this a two-factor factorial experiment, too few students participated in this experiment to allow us to have significant results with four conditions.

All conditions in this experiment follow the general ideation experiment format outlined at the beginning of this section, but slight variations occurred for each condition. For the HQP condition, participants received a packet with the design problem shown in Figure 9, with the following additional instructions:

> *You have been given a set of numbered passages extracted from professional journals on biology that may help develop ideas for separating the corn from the husk and silk.* ***Read all the provided passages, and try to use the passages to design a device that quickly and cheaply shucks corn for mass production.***

The packet contains 20 blank pages for participants to record their solutions. Additionally, the participants are given a single sheet with 20 high quality passages, listed in Table 24. The passages are ordered and numbered as they appear in Table 24. Verbally, the participants are instructed to read the design problem, read all the passages on the separate sheets, and to attempt to generate ideas or solutions based on the passages. They are also instructed to clearly note which passages were used to develop each solution and to note if any generated solutions were unrelated to the passages. The scripts used to instruct the participants in each condition are included in Appendix B

The LQP condition is identical in all respects to the HQP condition with the sole exception of the passages provided to participants. The low-quality passages used in this experiment are listed in Table 25. The HQS condition is very similar to the other conditions, but some modifications exist to encourage participants to consider each passage individually. The most obvious difference is that the passages are printed at the top of the 20 blank pages in the experiment packet. The additional instructions in the problem statement for this condition are inset below in italics. Note that the instructions are worded to prompt participants to read and generate solutions from each passage individually. Verbally, participants are instructed to read *each* passage and attempt to generate ideas or solutions based on that passage. Participants are again directed to note if any generated solutions are unrelated to the passage stimuli. Because each page in the packet had a different passage, any solutions recorded on a single page can be safely assumed to relate to the passage on that page unless otherwise noted.

> *You have been given a set of passages extracted from professional journals on biology that may help develop ideas for separating the corn from the husk and silk.* ***Read the passage on each page of this packet and try to use the passage to design a device that quickly and cheaply shucks corn for mass production.***

A total of 30 people participated in this experiment, with ten participants in each condition. The participants were recruited from MEEN 401 and MEEN 402, Texas A&M's mechanical engineering capstone design courses in the Spring semester of 2013. Twenty-five participants were recruited from MEEN 402 and five from MEEN 401. Table 23 shows the number of participants from each class that participated in each condition. This experiment was conducted late in the semester, so all participants had been taught idea generation methods as part of the capstone design courses.

The experiments are conducted outside of the recruited classes in Doherty room 309B. This room is a small conference room with two windows that were covered with Venetian blinds during the experiment. Participants are seated at individual desks spaced around the perimeter of the room, with no less than five feet between the participants. Students participated in this experiment at various times between March 25th and April 15th 2013. No more than four students took part in the experiment at any one time.

The high and low-quality passages used in this experiment are selected from the passages in first analogy survey, introduced on page 29. The high-quality passages are randomly selected from a set of 64 sentences in the corn shucker surveys. At least 50% of participants in the survey reported that the 64 candidate sentences prompted some idea for a corn shucker; that is, the candidate sentences have a quality of at least 0.5. Similarly, the low-quality passages are randomly selected from the set of 65 sentences in the survey with a quality less than 0.1.

Table 23. Number of participants in each experimental condition for each recruited class.

	MEEN 401	**MEEN 402**
Parallel – High Quality	0	10
Parallel – Low Quality	2	8
Series – High Quality	3	7

Table 24. High quality passages used in the experiment.

1	When these die off their mass decay may also briefly deoxygenate the water, which can also result in mortalities of natural local fauna and of farmed species.
2	The solid residues were filtered off, and filtrate was evaporated by a rotary evaporator at 508C under reduced pressure.
3	Although the possibility that target cell lysis could contribute to pathogenesis had been considered, lysis was assumed to have a more protective role: either CTLs could directly affect intracellular pathogens or these lymphocytes could liberate the pathogens from protective intracellular niches.
4	UBPs are thiol proteases that cleave isopeptide bonds between two ubiquitin chains or between ubiquitin and another protein.
5	But, when the metabolic biomass of species is standardized across the replicate webs, the magnitude of direct effects is much more consistent.
6	In that work, solutions containing a small amount of a test protein mixed with varying concentrations of a background macromolecule (polymer or protein) were exposed to conditions - elevated temperature or 30 % ethanol - in which the test protein would denature but the background molecule (if a protein) would not.
7	This exercise reveals that existing measures divide into a few basic groups, each of which captures a different facet of spatial turnover in the identities of species.
8	Ostrinia nubilalis larvae deliberately expel their frass from the tunnel, and must be at or near the tunnel entrance during this process.
9	There, a pattern of double segment periodicity is first generated and then subdivided to yield the final single segment repeat.
10	The minimal force required to unfold the entire straitjacket and to liberate TGF-1 is ~ 40 pN.
11	Spitting spiders are unique because they can forcefully expel a sticky gum from their cheliceral fangs onto prey or predators from as far as 10 body lengths away (Dabelow, 1958; Mcalister, 1960; Gilbert AND Rayor, 1985; Nentwig, 1985).
12	Primate cone response amplitude and sensitivity recovered substantially after the bleach.
13	We propose that GFP-Pon localization is a two-step process involving the establishment of a cortical area where the crescent will form and the progressive recruitment of protein to the predefined site until metaphase.
14	Leaf swallowing in chimpanzees to physically expel intestinal parasites appears to originate from opportunistic feeding by some individuals, which is later passed down in the form of a behavioural tradition (Huffman AND Hirata 2004).
15	AP2 was incubated with 1 mM FSBA, and the unbound FSBA was then removed by dialysis.
16	LIN-5 is a large coiled-coil protein essential for many aspects of cell division, including chromosome alignment at metaphase and sister chromatid separation.
17	They often function in cascades, with one protease cleaving the zymogen form of another, activating it to cleave the next one in the chain, and so on.
18	An increase in density may increase speed of propagation as it reduces the time to react to nearby departures (Hunter 1966; Hilton et al.)
19	Do these findings imply that separase cleavage is required to generate a form of the protease that is able to cleave cohesin.
20	This wave did not reach the tip of the leaf, and thus the last segments cleaved from the initial either did not divide at all or divided fewer times than did segments toward the base.

Table 25. Low quality passages used in the experiment.

1	We thus conclude that the CUL-3-based E3 ligase regulates cytokinesis by allowing MEL-26 to accumulate only at the right time and place.
2	The total abundance of Poecilia reticulata in the absence of Heterandria formosa over time for each replicate.
3	This region differs by only one amino acid between zMCM6a and zMCM6b, so the antibody should recognize both forms, but the region is only 55 % identical to mMCM6 and thus the antibody should cross-react poorly with mMCM6.
4	The regions near the SPBs displayed the Mtc1-GFP signals in prophase and prometaphase cells.
5	The proportion of wild-type and glu88-82 cells in metaphase, anaphase, or telophase as well as the proportion of cells with chromatin bridges are shown.
6	Rab proteins are major regulators of intracellular transport in eukaryotes, and here, we report an essential role for human Rab35 in both the stability of the bridge and its final abscission.
7	Within-subject correlation between latency to react to the gaze cue and rank difference.
8	An asterisk identifies (a) the mother cell 45 min before division, (b) the mother cell during mitosis, and (c) the daughter cells 45 min after division (supplementary movie 1, image every min).
9	Note that in the mutant lysates, the Sm proteins fail to react with either SYM10 or Y12.
10	Skuas' direct numerical response may mean that through avoidance learning by prey (Holling 1966) victims are more vigilant and, hence, react faster when skuas, and hence victims themselves are more abundant.
11	Moreover, in wild-type mice the proportion of mitotic cells in anaphase was significantly reduced after STLC treatment, indicating an arrest prior to this stage.
12	Cytokinesis in somatic cells, the germ-line and the endosperm of Arabidopsis. STZLE is required only for the cellularisation of the endosperm syncytium, and not for cytokinesis of somatic or germ-line cells.
13	Activation of components of the immune system can be energetically expensive, can lead to reductions in other immune components and reproduction ((Zuk and Stoehr, 2002) and (Martin et al., 2006)), and leads to the production of potentially hazardous free radicals (Bertrand et al.)
14	Consistent with previous studies, European and North American free-living populations are highly differentiated (Dest = 0. 162).
15	Cytokinesis in wild-type, zen-4 and air-2 mutant embryos.
16	To determine whether AIR-2 affects the timing of sister kinetochore resolution, we analyzed the distribution of HCP-3 in wild-type and air-2 (or207ts) mutant embryos during prophase, prior to NEBD.
17	Regulation of the major transitions in the cell cycle, such as G1 / S, G2 / M, and metaphase to anaphase, are increasingly well understood.
18	Some mutations that release TGS affect DNA methylation to various extents and with various specificities (ago4, cmt3, dcl3, ddm1, drd1, drm2, hda6, hog1, met1, nrpd1a, nrpd1b, nrpd2a, rdr2, suvh2, and suvh4), whereas other mutations (bru1, fas1, fas2, and mom1) have been reported to have no effect.
19	By analogy with the mitotic defects described below, we think it likely that cks-1 (RNAi) embryos fail to exit meiosis normally.
20	We assumed that the anaphase rate is the same in all experiments for a given genotype.

9. MEASURING IDEA GENERATION [*]

The ideas generated in this study are evaluated using four metrics: quantity, quality, novelty, and variety. These systematic metrics for ideation were introduced by Shah et al. [85] and further developed by Linsey et al. [78]. The metrics allow different idea generation methods to be quantitatively compared based on the ideas produced by designers using the methods. This section explains the process and metrics for evaluating the ideation experiments described in Section 8.

Briefly, the packets from the experiment are organized and preprocessed to separate discrete solutions and remove non-relevant information such as need statements and design constraints. The processed data is evaluated for the quantity of non-redundant ideas and the solution quality, and sorted into groups of similar solutions to facilitate the novelty and variety measurements. The data is coded to conceal the experimental conditions during all ratings. All ratings are performed by the author and compared against the ratings of a second, independent rater to establish inter-rater reliability for the measurements. The second rater is a PhD student also studying bioinspired design.

9.1. Preprocessing

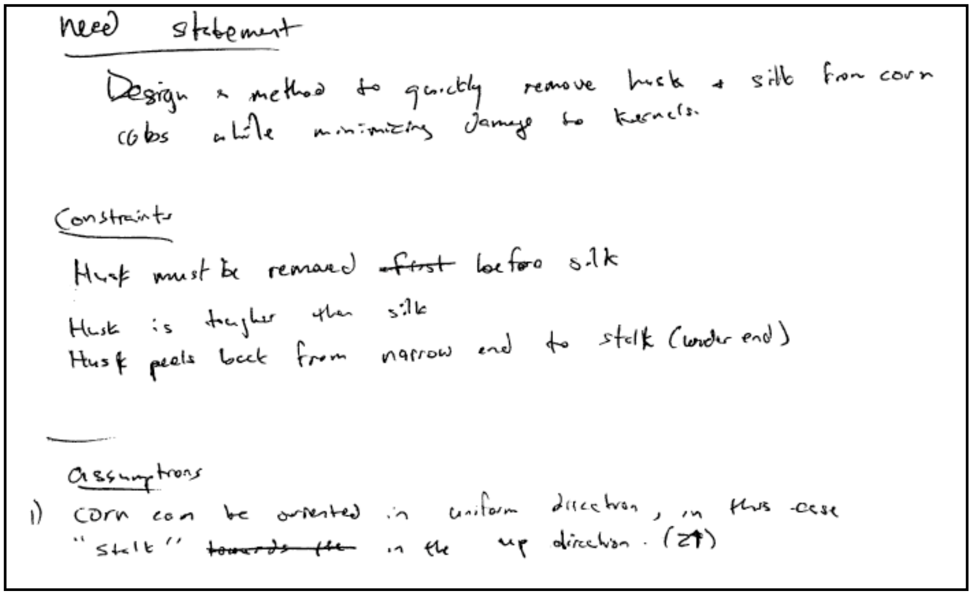
Prior to rating for quantity, quality, novelty, and variety, the raw packets are processed to remove information irrelevant to the metrics and to separate discrete solutions to facilitate the novelty and variety sorting process. This preprocessing is

---

[*] Part of this chapter is reprinted with permission from “Evaluating the Directed Method for Bioinspired Design”, M.W. Glier, J. Tsenn, J.S. Linsey, D.A. McAdams, 2012, Proceedings of the ASME 2012 International Design Engineering Technical Conferences & Computer and Information Engineering Conference, IDETC2012-71511. Copyright 2012 by ASME

necessary because the ideation metrics used in this work count, rate, or sort one of two basic, formally defined units: ideas and solutions. The definition of an “idea” in this context is fully discussed in the section on the quantity metric, but its basic definition is something that fulfills at least one function of the Functional Basis as described by Hirtz et al. [35]. A solution is commonly composed of multiple ideas, and is formally defined as the entire contents of one page of experimental results, or, in cases where the designer clearly indicated that different segments of a page contained separate solutions, the contents of the indicated segment of a page. Before participants’ concept generation can be rated for quantity, quality, novelty, and variety, the experimental data must first be separated into discrete solutions.

While metrics are concerned with ideas and solutions, as they have been defined, designers will commonly write notes to help understand the design need before attempting to solve the problem. Figure 10 is a typical example of such notes; the designer has clearly spent time to define the problem before attempting to generate solutions. Although defining needs, constraints, and assumptions are important parts of the design process, this investigation is concerned only with the ideas and solutions generated by the participants. Consequently, such notes are either ignored during the rating process if they share a page with a solution or removed from consideration when the notes fill an entire page, as in Figure 10



need statement

Design a method to quickly remove husk + silk from corn cobs while minimizing damage to kernels.

Constraints

Husk must be removed ~~first~~ before silk

Husk is tougher than silk

Husk peels back from narrow end to stalk (wider end)

Assumptions

1) corn can be oriented in uniform direction, in this case "stalk" ~~towards the~~ in the up direction. (Z↑)

Figure 10. Notes taken from an experimental packet. The notes on this page help the designer understand the problem at hand by defining the needs, constraints, and assumptions. These notes do not propose any solution to the problem.

Although participants are instructed to place only one solution on each page, occasionally multiple solutions are placed on a single page. Solutions sharing a page need to be separated before they are rated. There can be some challenge in identifying where one solution ends and another begins. In most cases, the participants clearly indicated that solutions were distinct using numbering, headings, or whitespace. Figure 11 shows a page full of solutions that need to be separated onto different pages for rating. The circled numbers indicate that there are six distinct solutions on the page. The short list in the bottom right appears to be a part of solution six. All solutions are separated by the author prior to applying the ideation metrics.

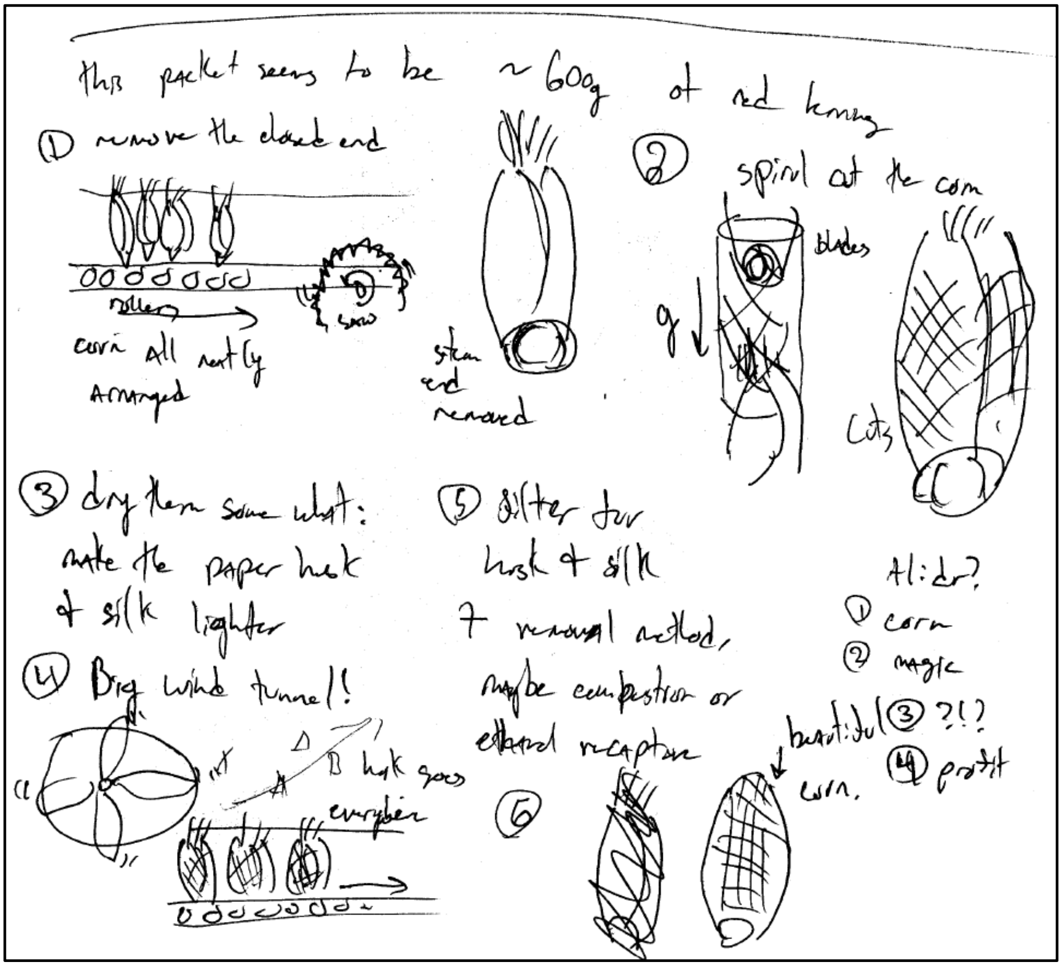


Figure 11. This page from an experiment shows six distinct solutions to the corn shucker problem. Each solution is numbered.

9.2. **Measuring Quantity**

The quantity measure is used to compare the effectiveness of each idea generation processes. Building from the work of Shah et al. [85] and implemented by Linsey et al. [86], the basic definition of a single "idea" is something that fulfills at least one function of the functional basis as described in Hirtz et al. [35]. The guidelines for this metric are listed in Table 26.

Table 26. Summary of Quantity Metric Rules [86].

1. An idea solves one or more of the functions in the Functional Basis (primary or secondary function).
2. The same idea (or component) being used in multiple places counts as one idea. Redundant ideas are only counted once.
3. Each idea counts as only a single idea even if it solves more than one function. A single component such as a motor can solve more than one function such as producing heat and increasing torque.
4. New combinations of already-counted ideas are counted in a separate measure as one new idea.
5. Categories of ideas only count as ideas when no subordinates are given, for example gear is one, but gear, helical gear, and planet gear are only two ideas not three.
6. Ideas count even if they are not needed or cause the systems to not function.
7. Ideas must be shown and not just implied.
8. For ideas that reframe the problem such as producing a slightly different product or ways to reduce waste product, count these in a category called "Problem Reframing." These are ideas that do not specifically address the problem as describe but meet the higher level customer needs.
    a. Ideas that reframe the problem usually do not fit a defined product function well.
    b. They must add something to the system.
    c. Count them if they are related to the situation such as
        i. Environmental concerns relate to the situation
        ii. Reduction in waste products resulting from solutions to the problem
        iii. Produces a different product that meets the customer needs

Counting the number of ideas produced by each participant requires the rater to identify every component that solves one of the terms in the functional basis while ignoring redundant ideas. Proceeding thoroughly through ideas, this procedure can yield different numbers of ideas for solutions that may at first seem similar. As an example, consider the corn shucking ovens shown in Figure 12 and Figure 13. Different participants proposed the two solutions, and although the solutions are practically identical, the ideas within them are quite different. Looking at Figure 12, one can identify four ideas: an oven housing, heating coils, a conveyor belt, and a basket that will catch the processed corn. This solution implies the presence of other ideas, such as an electrical source for the coils and a motor for the conveyor, but only explicitly presented ideas are counted using this quantity metric. In contrast, Figure 13 is much less detailed, and contains only a single explicit idea: an oven. Once again, other ideas may be implied, but cannot be counted. The level of detail in the solutions is strongly reflected in the number of ideas.

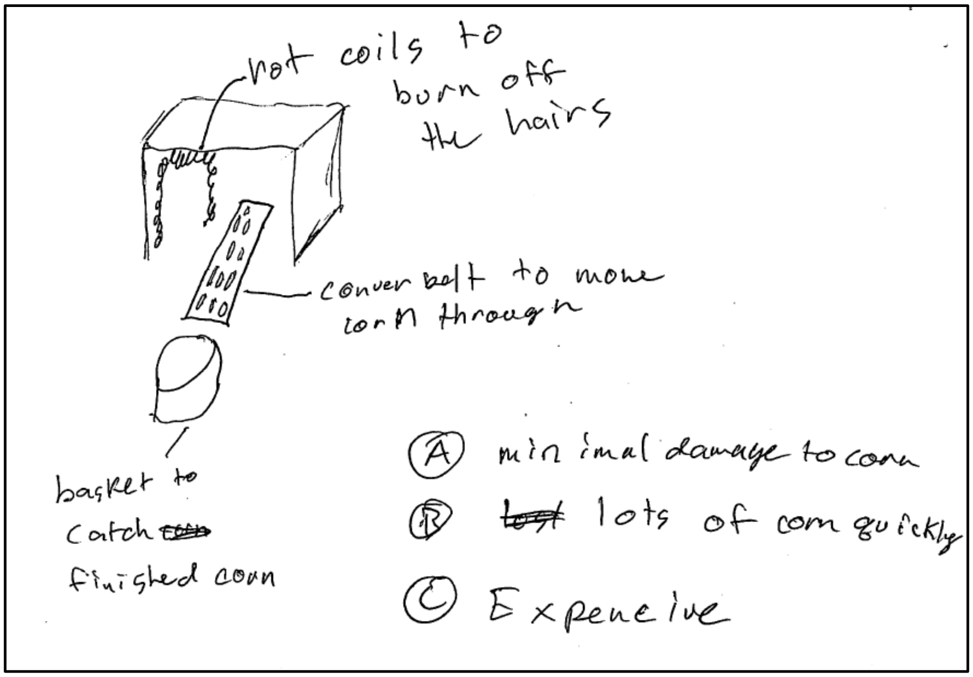


Figure 12. A proposed solution for burning the husk and silk off ears of corn. This detailed solution contains many distinct ideas.

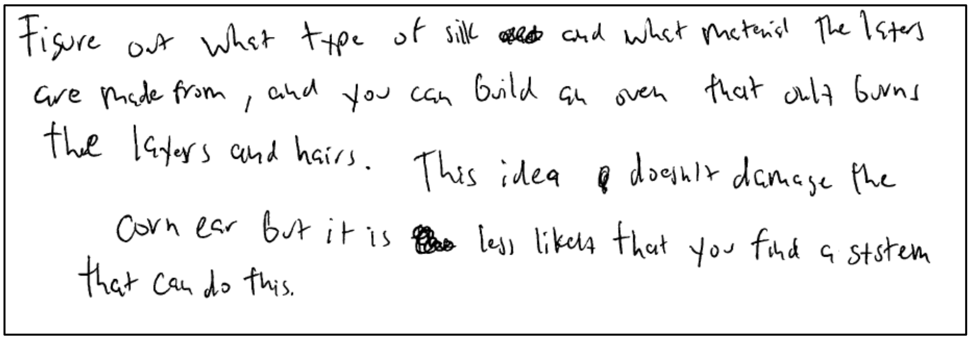


Figure out what type of silk and what material the layers are made from, and you can build an oven that only burns the layers and hairs. This idea doesn't damage the corn ear but it is less likely that you find a system that can do this.

Figure 13. Another solution for burning the husk and silk off ears of corn. This solution contains only a single explicit idea.

9.3. Measuring Quality

The quality of this study's generated solutions is measured using an anchored three point rating scale developed by Linsey et al. [86]. Unlike the quantity metric, which considers discrete ideas, quality is evaluated at the solution level. The quality rating scale, depicted in Figure 14, first asks if the solution is technically feasible. If it is not, zero points are awarded to that solution. If the solution is feasible, the rater then asks if the solution is technically difficult for the context. If it is, one point is awarded. If the solution is not technically difficult, two points are awarded. The quality rating of a solution is independent of other solutions, and thus can be evaluated while a series of experiments is in progress. Note that this quality metric for conceptual designs is a completely different entity from the sentence quality we used in earlier sections to describe the likelihood of a sentence prompting an idea for a design problem.

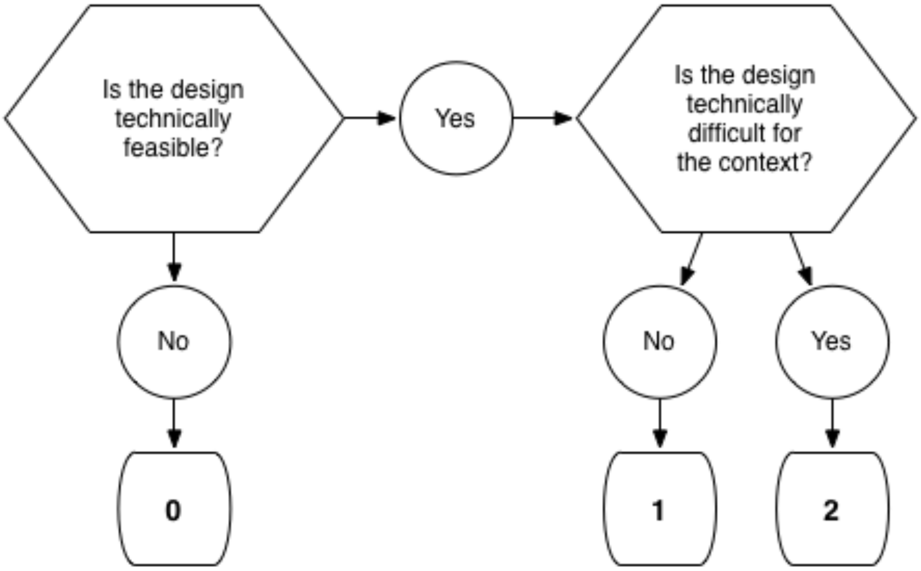


Figure 14. Three point rating scale for evaluating quality.

One of the challenges in rating solution quality is the ambiguous nature of rating a human generated solution, typically presented as a line sketch with some notation. Each rater must independently decide what is technically feasible and what constitutes technical difficulty for the proposed solution in the context of the problem posed. Infeasible solutions are generally easy to identify, as they often violate basic physical principles or exceed the limitations of current technology. Examples of solutions rated as infeasible encountered in this study are contact genetically modified strains of corn that grow without a husk and an "anti-husk gun" that would be fired at a field of standing corn.

Bioengineering husk-less corn would require a significant advancement in our understanding of genetics. The "anti-husk gun" idea is to some extent a restatement of the problem without offering any specific idea of what constitutes the gun and how it removes husks.

Determining whether an idea is technically difficult poses additional challenges in some cases, but is fairly straightforward in others. For example, some solutions, such as hiring people to shuck corn or placing corn in a tumbler, offer no technical difficulty. Other solutions, like using a water jet to strip of the husk without damaging the kernels, are less clear. To improve the consistency of quality ratings as well as inter-rater reliability, all quality ratings in this research are performed using the following guidelines. The primary purpose of these guidelines is to define what constitutes a "technical difficulty" in regards to the corn shucking problem. Note that even with this heuristic, there is substantial room of rater subjectivity, e.g. the limits of current technology, what constitutes "minimal" damage to the corn.

Quality = 0 – Technically infeasible

Does not actually remove husk or silk from corn

Is not possible with current technology

Quality = 1 – Technically feasible but difficult for the context

Likely to damage corn

Unable to process large quantities quickly

High cost

Multiple unresolved or vague technical issues

Quality = 2 – Technically feasible and not difficult for the context

Reliably removes husk and silk

Causes minimal damage

Reasonable cost to manufacture and operate

9.4. **Variety and Novelty Metrics**

Novelty reflects how unusual or unique a solution is in comparison to other generated solutions. Variety shows the diversity a participant's solutions: that is, the ratio of the solution space explored by a single participant's solutions to the space explored by all participants. The method used to evaluate this experiment is presented in Linsey, et al. [86]. Novelty and variety are evaluated on the solution level, much like the quality metric. Novelty and variety are quantitatively measured based on a sorting of solutions into self-similar groups, or "bins." This bin sort is performed using a predetermined set of bins that is established in advance.

To establish a bin set, each solution from the unstructured idea generation experiment was given to two independent raters: the author and a PhD student design researcher advised by Dr. Julie Linsey. Each rater independently sorts the solutions into groups of similar solutions. It is possible for a single solution to be in multiple groups in cases where the solution performs multiple functions that directly meet the design need. Each rater must independently decide what constitutes "similar" for the rating. The two raters then reconcile the two sets of solution groupings into a single bin set. The bin set used to rate the idea generation experiments in this work is listed in Table 27, along with short descriptions of each bin. Each solution in the ideation data can be categorized into one or more of these bins.

Table 27. Bins used to sort solutions for novelty and variety metrics with a short description of each bin.

Bin Name	Description
Abrasives	Solution uses abrasive substance or texturing to grind away husk and silk.
Adhesive	Solution uses an adhesive or "sticky substance" to bind to the husk or silk and pull it away. Simple "high friction" surfaces are excluded and should be grouped according to their removal mechanism (usually scraping or brushing).
Animal shucking	Animals are trained to remove the husk and silk with their hands/paws/claws/etc.
Automatic size adjustment	The solution is designed to automatically adjust to the size of the corn through an active or passive mechanism. (e.g. spring loaded parts, feedback control). Solutions that must be manually adjusted do not count.
Automatic solution/ black box	Solutions that remove the husk without any specified removal mechanism. Examples include ray-guns, ill-defined robots, "magic", and modified cotton gins.
Blade - cut end	Solution uses a blade to cut off the end of corn
Blade - on side	Solution uses blades to cut along the side of corn
Brushes	Solution uses a brush-like tool to remove the husk or silk. Brushes are characterized by multiple bristles or "fingers" that are flexible
Chemicals	Solution uses some chemical to dissolve, weaken, or otherwise aid in the removal of the husk or silk. Pure water is generally not counted in this bin.
Drill Flutes	Solution moves corn through a helical track to remove husk and silk. Looks like the flutes of a drill bit.
Eating	Solution uses organisms (micro or macro) to eat and digest part or all of the corn. This bin can included artificial "organisms" such as nano-scale robots when their actions are characterized as eating.
Electricity	Solution employs static electricity or electric current remove the husk or silk. Using electricity to power a motor or other device does not qualify.
Embrittle husk	Solution causes husk to become brittle (usually through freezing or dessication) to allow it to be cracked or shattered.
Entanglement	Solution causes the husk or silk to become tangled in some sort of barb, hook, etc. so that it can be pulled away. Brush type solutions are not included.

Table 27. Continued

Bin Name	**Description**
Filter	Solution uses a filter to separate the husk or silk mixed with - usually loose kernels or water
Fire	Solution uses a flame to burn away husk and silk
Impact	Solution requires a high-velocity impact to remove the husk and silk, or to cause the kernels to fall out of the ear
Humans	Solution uses unaided humans to carry out all shucking operations. Solutions that employ humans as a part of a larger process are not grouped here
Huskless/ Modified husk	Solution proposes to grow corn without a husk or with a husk that has unusual properties to ease the shucking process, often brought about by genetic modifications
Lasers	Solution uses lasers to cut or burn away the husk and silk. Lasers used for measurement are not counted in this bin
Manual clamp	Solution uses a clamp-like device to secure the corn or to grab the husk or silk. Robotic manipulators do not fall in this category
Moving fluid (air & water)	Solution uses a stream of fluid (liquid or gas) to blow, spray, or wash the husk and silk away. Common implementations include air/water jets and agitated bodies of water (e.g. a clothes washer configuration)
Popcorn	Solution requires corn kernels to be popped through some means
Remove core	Solution removes the corn cob from the center of the corn to free the kernels.
Restructure problem	Solution restructures problem to mitigate or eliminate the need for a shucking device. Examples include selling unshucked corn and shipping corn to an existing processing facility
Robotic hands	Solution uses some arrangement of robotic manipulators to shuck corn, usually imitating manual shucking processes
Rollers	Solution uses a smooth roller to remove the husk and silk. Solutions must rely on the friction of the roller rather than some surfaces feature(s) such as bristles or spikes
High-Speed Rotation/ Centrifuge	Solution spins corn very quickly to fling away husk and silk. Slow rotation used to augment other removal processes is not included
Scraper	Solution uses an implement to scrape away husk and silk. The implement is typically flexible and is not segmented like a brush. Common shapes include paddles, pads, and rings; high friction materials are often specified to improve scraping action.

Table 27. Continued

Bin Name	Description
Series of mesh wires	Solution forces whole corn through a wire mesh which cuts or scrapes away the husk and silk
Slit plates	Solution uses plates with small slits to trap husk and silk and pull or shear it away. Similar to some electric razor heads.
Soften husk	Solution causes husk to become softer to allow easier shucking. Usually this is accomplished through soaking or humidification.
Sorter	Solution employs a device designed to sort corn for processing, usually by size.
Torsion	Solution places corn in torsion to tear away the husk an silk. Action is similar to wringing out a towel.
Vibration	Solution uses vibrations to remove the husk and silk (e.g. "natural frequencies") or to separate corn kernels from a mixture.
Visual sensor	Solution employs a sensor to detect the state of the corn. Typically size or quality of shucking is measured.

A pre-established bin set would be unnecessary if a single rater were to sort the solutions to measure novelty and variety. However, if a second rater attempts to reproduce the sort, that rater's bin set will generally be different from the first. Consequently, the raters' novelty and variety scores will likely be very different and the measurements would have low inter-rater reliability. Establishing a bin set that is understood by all raters greatly improves the repeatability of novelty and variety measurements.

The following subsections will describe how novelty and variety are computed from the bin sort. A short example table for a bin sort is provided in Table 28 to clarify the discussion. The rows of the table represent the bins, while the columns show the solutions for each participant.

Table 28. An example table of a bin sort with corn shucker solutions from four hypothetical participants, sorted into four bins.

	Participant 1	Participant 2	Participant 3	Participant 4
Abrasives	2	1	0	0
Chemicals	0	1	0	0
Fire	1	1	2	0
Rollers	0	1	1	2

9.4.1. **Measuring Variety**

The variety of each participant's solutions reflects the size of the solution space spanned by their solutions. The variety of one participant's solutions is defined as the ratio of the number of bins into which the participant's solutions are sorted to the total number of bins. Accordingly, participant variety scores can range between 0 for a participant who produces no solutions at all, and 1 for the participant whose solutions fall into each of the bins. In practice, these extremes are never reached because we assume that participants who generate no solutions misunderstood the problem statement and are thus omitted from the analysis, and no single participant has ever produced enough solutions to span the entire design space.

In the context of Table 28, we have all solutions fitting in to four bins. Participant 2 has solutions that fall into each of the four bins, so that participant's variety score is 4/4 =1. In contrast, participant three generated three solutions, all of which used some roller configuration to shuck the corn. Participant 3 has a variety score of 1/4 = 0.25.

9.4.2. **Measuring Novelty**

The novelty metric measures the frequency of the occurrence of the solution based on the previously discussed groupings. For a single solution, novelty is measured by subtracting from one the number of solutions in a group divided by the total number of solutions generated by all participants with the same design problem. A single participant's novelty score is calculated by averaging the individual novelty scores of the participant's solutions. Participant novelty scores range can range between 0 for a participant who produces no solutions and $1 - 1/n$, where n is the total number of solutions in the ideation data, for the participant who generates only solutions that are the sole member in their respective bins.

Turning once again to the example in Table 28, we can first calculate the novelty of each bin. There are 12 solutions total, so solutions in the Abrasive bin would have a novelty of $1\text{-}3/12 = 0.75$, Chemicals 0.92, Fire 0.67, and Rollers 0.67. The novelty score for each participant are simply the average of that participant's solution novelties. So Participant 1 would have a novelty score of $\{2(0.75) + 1(0.67)\} / 3 = 0.72$; Participant 2, 0.75; Participant 3, 0.67; and Participant 4, 0.67.

10. IDEA GENERATION RESULTS AND DISCUSSION [*]

With the development of text classification tools discussed in section 7, it is possible to identify passages that will help idea generation, so long as the classifier has been trained with a function relevant to the design problem under consideration. In this section, we examine the effects of biology passages as stimuli for idea generation, within the context of several idea generation experiments. The following subsections will explore and discuss how and where our biological stimuli offer an advantage over unstructured idea generation. As discussed in section 8.2, we look primarily at the effects of the quality of the passages given as stimuli and the format in which the passages are presented. Overall, we examine four conditions: high quality passages presented in parallel (HQP), high quality passages presented in series (HQS), low quality passages presented in parallel (LQP), and unstructured idea generation.

Section 10.1 briefly introduces a study comparing the idea generation of participants from MEEN 401 and 402 with the corn shucker problem. This comparison produces conversion factors that will be used to adjust scores for MEEN 401 participants in the following analyses. Section 10.2 will examine the numbers of solutions produced in each condition. The next four sections will discuss the idea generation results in terms of Quantity, Quality, Novelty, and Variety introduced in section 9. Section 10.7 expands

---

[*] Part of this chapter is reprinted with permission from “Evaluating the Directed Method for Bioinspired Design”, M.W. Glier, J. Tsenn, J.S. Linsey, D.A. McAdams, 2012, Proceedings of the ASME 2012 International Design Engineering Technical Conferences & Computer and Information Engineering Conference, IDETC2012-71511. Copyright 2012 by ASME

on the bin sort used for the Novelty and Variety metrics and identifies some solution types that are more common when using biological stimuli.

Several statistical measures will be commonly used through this section. Most commonly, a t-test is employed to compare two conditions. This lets us easily identify statistically significant difference between two conditions, even with relatively small sample sizes. The t-test is used when discussing differences in quantity, quality, novelty, and variety. All t-tests are 2-tailed t-tests, with the assumption of unequal variances in the two populations; the t-test gives more conservative results when unequal variances are assumed. When assessing the results of a statistical test, we will always consider a p-value less than 0.1 to be significant. Pearson's correlation coefficient is used to establish inter-rater reliability for quantity, novelty, and variety. Cohen's kappa is used to measure inter-rater reliability for quality.

Results from 30% of participants across all conditions are evaluated by a second rater to establish inter-rater reliability. For quantity counts, the raters have very strong agreement with a Pearson's correlation of 0.92. Quality rankings have reasonably good agreement, with a Cohen's kappa of 0.40. The two raters have acceptable agreement for novelty scores, with a Pearson's correlation of 0.77. Finally, the Pearson's correlation for variety rankings is very high at 0.96.

10.1. **Participant Correction Factors**

Since our experimental conditions contain different ratios of MEEN 401 to MEEN 402 students (see Table 23 and section 8.1), we cannot assume that the participant groups for each condition have similar ability levels. To adjust for the differences between the classes, we compare the data drawn from the unstructured idea generation experiment using 401 students to a set of students from MEEN 402 who were given the same problem. If there is a significant difference in Quantity, Quality, Novelty, or Variety between the two classes, we will introduce a correction factor that will effectively let us treat the ideation results from a 402 student as those from a 401 student. This correction is performed by shifting the statistics of the 402 population. Specifically, if we analyze the results from 401 students in terms of quantity, their quantity will have a mean, μ_1,and standard deviation, σ_1. Similarly, results from 402 will have a mean, μ_2,and standard deviation, σ_2. To treat results from 402 students as those from 402, we can adjust the metric scores so that μ_2 and σ_2 are equal to that μ_1 and σ_1. Given a data point X (e.g. a participant's novelty score) from a 401 participant, we can find the adjusted X, using the linear equation [87]:

$$\underline{X} = \frac{\sigma_2}{\sigma_1} X + \left(\mu_2 - \frac{\sigma_2}{\sigma_1}\mu_1\right)$$

The MEEN 402 data set contains result from 11 students recruited from MEEN 402 in the Fall semester of 2011. The data set was collected under conditions nearly identical to those of the MEEN 401 participants from the unstructured ideation experiment. The participants received the same corn shucker problem, with the same instructions, and the same 50-minute idea generation period. The only significant difference between the experimental conditions is that the 401 students performed the experiment outside of class, in a windowless room, at single desks placed no less than three feet apart with partitions arranged so that participant could not see one another. No more than four participants took part in the experiment at the same time.

10.1.1. Quantity

Figure 15 shows the number of non-redundant ideas produced by MEEN 401 and MEEN 402 students with the corn shucker problem. There is a significant difference in the number of non-redundant ideas produced by the two classes (p=0.04). In our later analyses, each quantity count for a 401 participant, Q_1, will be corrected using the equation:

$$\underline{Q_1} = 1.286(Q_1) + 1.126$$

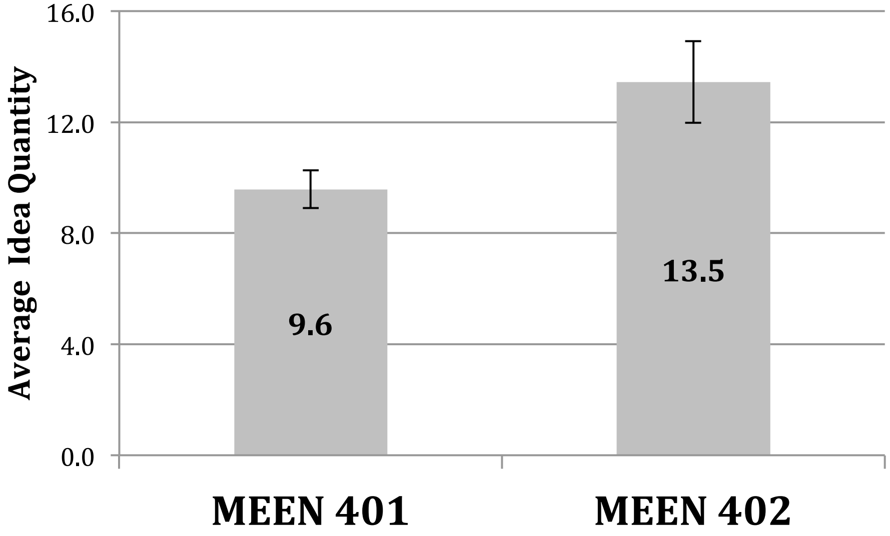


Figure 15. Quantity counts for MEEN 401 and MEEN 402 unstructured idea generation with corn shucker problem. Error bars show ±1 S.E.

10.1.2. Quality

Figure 16 shows the average quality of solutions produced by MEEN 401 and MEEN 402 students with the corn shucker problem. There is no significant difference in the solution quality produced by the two classes (p=0.25). Because there is no significant difference in quality, quality scores for MEEN 401 participants will not be adjusted.

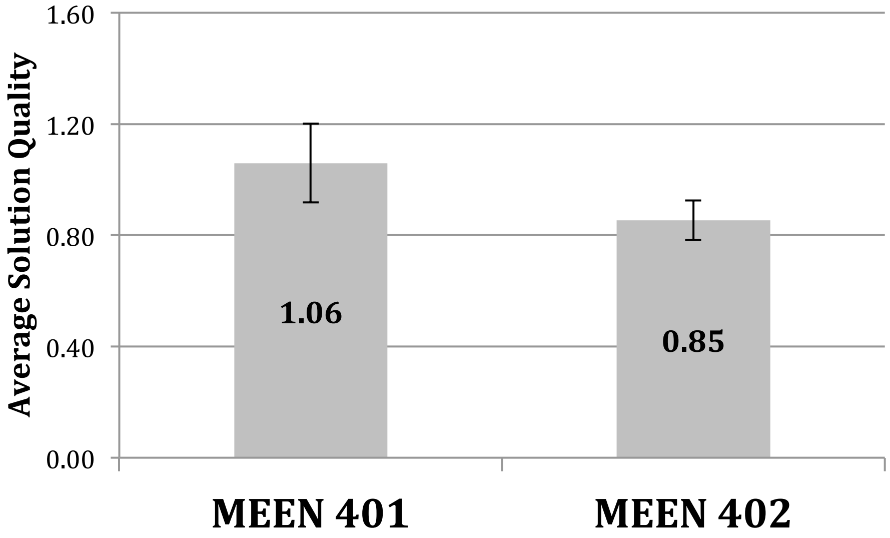


Figure 16. Average solution quality for MEEN 401 and MEEN 402 unstructured idea generation with corn shucker problem. Error bars show ±1 S.E.

10.1.3. **Novelty**

Figure 17 shows the average novelty of solutions produced by MEEN 401 and MEEN 402 students with the corn shucker problem. There is no significant difference in the solution novelty produced by the two classes (p=0.55). Because there is no significant difference in novelty, novelty scores for MEEN 401 participants will not be adjusted.

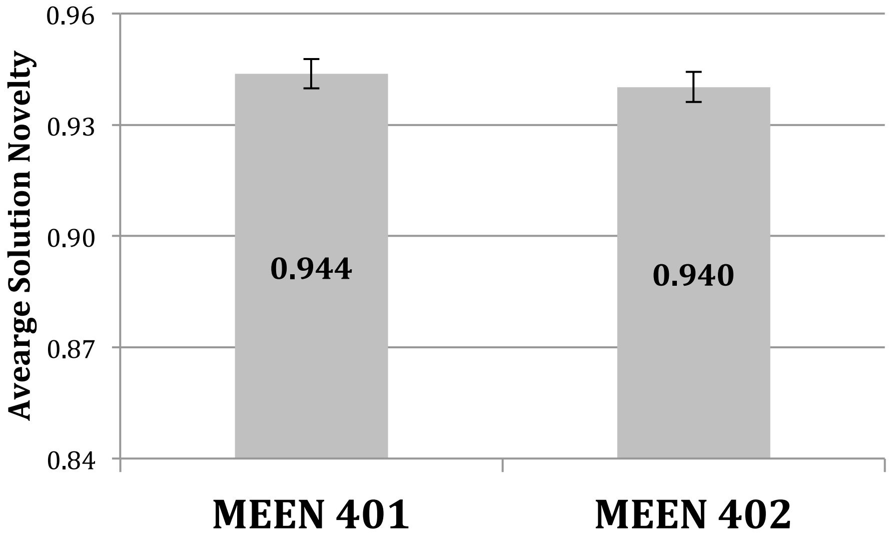


Figure 17. Novelty scores for MEEN 401 and MEEN 402 unstructured idea generation with corn shucker problem. Error bars show ±1 S.E.

10.1.4. **Variety**

Figure 18 shows the average variety of solutions produced by MEEN 401 and MEEN 402 students with the corn shucker problem. There is no significant difference in the solution variety produced by the two classes (p=0.48). Because there is no significant difference in variety, variety scores for MEEN 401 participants will not be adjusted.

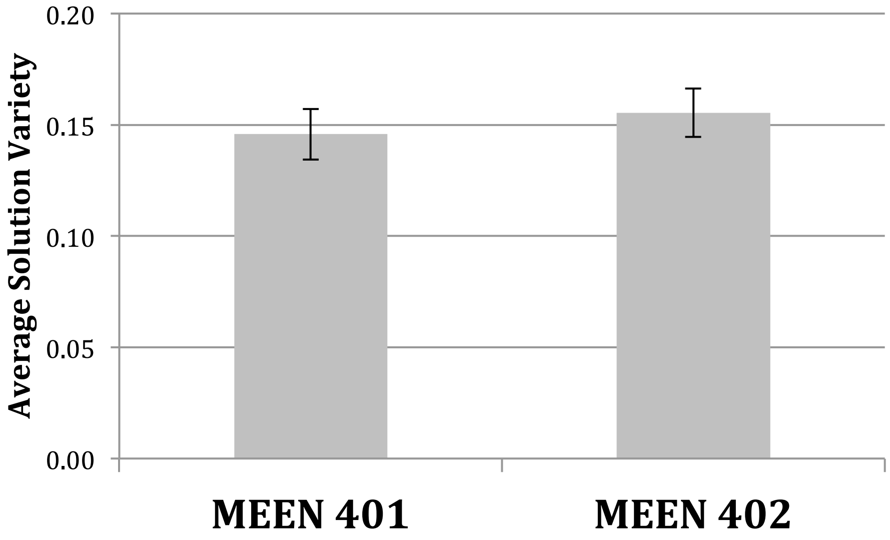


Figure 18. Variety scores for MEEN 401 and MEEN 402 unstructured idea generation with corn shucker problem. Error bars show ±1 S.E.

10.2. **Solution Counts**

As a first look at how biology passages can promote idea generation, we will examine the average number of solutions produced by participants in each condition. Looking at solution counts alone is a coarse measure, since it does consider that some solutions may be redundant, incomplete, or simply unfeasible. However, we can easily examine how often participant in the bioinspired conditions were able to use passages to generate a solution. Figure 19 shows the number of solutions participant generated based on a passage, the number based on no passage, and the total solutions for each condition. Later, Table 29 lists the p-values from t-tests on the total number of solutions between each condition, as well as on the number of bioinspired versus non-bioinspired passages within each condition. In the parallel conditions we assume that each solution was generated without the aid of a passage unless the solution explicitly indicates that a numbered passage was used. In contrast, the series condition places a passage on each page of the packet, so we assume that each solution was prompted by the passage on the page where the solution appears unless the participant explicitly states that the page's passage was not used.

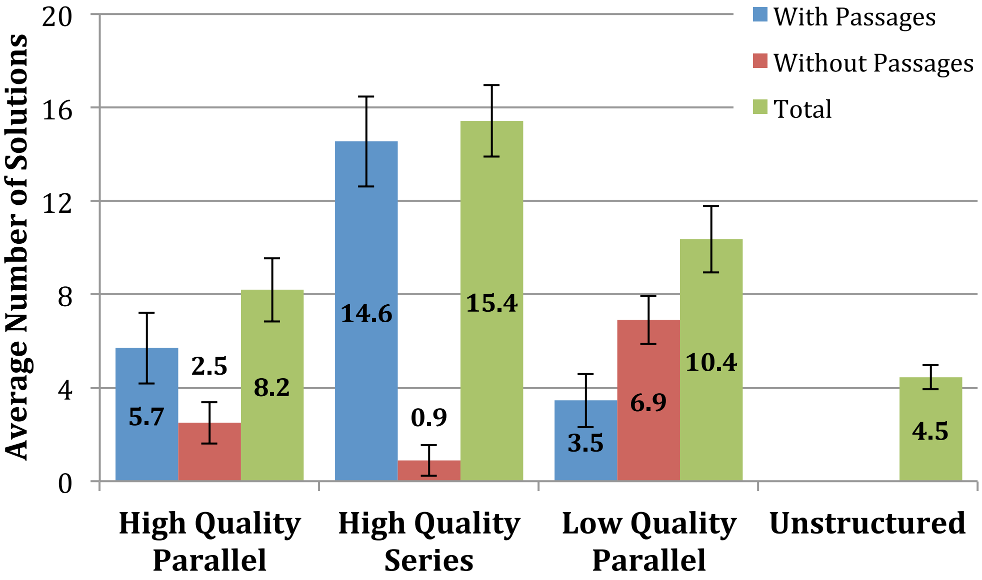


Figure 19. Average number of solutions generated in the experimental conditions. Number of solution generated with a passage, without a passage, and total solutions are shown for each condition. Error bars show ±1 S.E.

Figure 19 shows some remarkable contrast between the conditions. Firstly, we note that the number of solutions produced in every bioinspired conditions is significantly higher than those produced using unstructured idea generation. The mere presence of design stimuli seems to significantly increase designers' output. It remains to be seen, however, if the bioinspired solutions produced are measurably better, in term of quantity, quality, novelty, or variety, than those produced with unstructured ideation.

Table 29. p-values for differences in total solutions between conditions. p-values for the difference between the number of bioinspired and non-bioinspired solutions within each condition is also listed.

	High Quality Parallel	High Quality Series	Low Quality Parallel
High Quality Series	0.00	-	-
Low Quality Parallel	0.28	0.03	-
Unstructured	0.02	0.00	0.00
With vs. Without Passage	0.09	0.00	0.04

In terms of the total number of solutions, the High Quality Series (HQS) stimuli prompted significantly more solutions than any other format. Moreover participants in the HQS condition rarely produced solutions without using a passage. It may be that the rarity of non-bioinspired solutions is a result primarily of the packet format, where all the packet's pages have a passage on them and no space is readily available for non-bioinspired solutions. However, the number of solutions generated in the HQS condition suggests that participants were able to consistently use passages to generate solutions without running out of ideas.

When we compare the High Quality Parallel (HQP) and Low Quality Parallel (LQP) conditions, we first see that more solutions were produced on average in the LQP condition. However the difference is not statistically significant (p=0.28 see Table 29).

However, we do see that in the HQP condition, more solutions are generated using some passage than not using a passage. But in the LQP condition, more solutions are generated without passages. This supports our underlying assumption from the analogy surveys that designers are better able to use high quality passages to generate ideas than low quality passages.

10.3. **Quantity**

Figure 20 shows the average quantity count per condition, for bioinspired, non-bioinspired, and total ideas with error bars denoting one standard error. Table 30 lists p-values for a series of t-tests comparing the total quantity counts between each set of conditions, as well as a comparison of the quantity of bioinspired and non-bioinspired ideas within the conditions. Many of the trends observed with the solution counts are once again seen with the quantity measure. In terms of quantity we see that both the HQS and LQP conditions produce significantly more ideas than unstructured idea generation. The HQS quantity is significantly higher than the HQP count, but the difference between the HQS and LQP is not statistically significant (p=0.23). In general, the number of ideas contained in a *single* solution does not change between conditions. However, because participants the HQS and, to a lesser extent, LQP conditions have such a high average number of solutions per participant, the cumulative number of ideas grows proportionately.

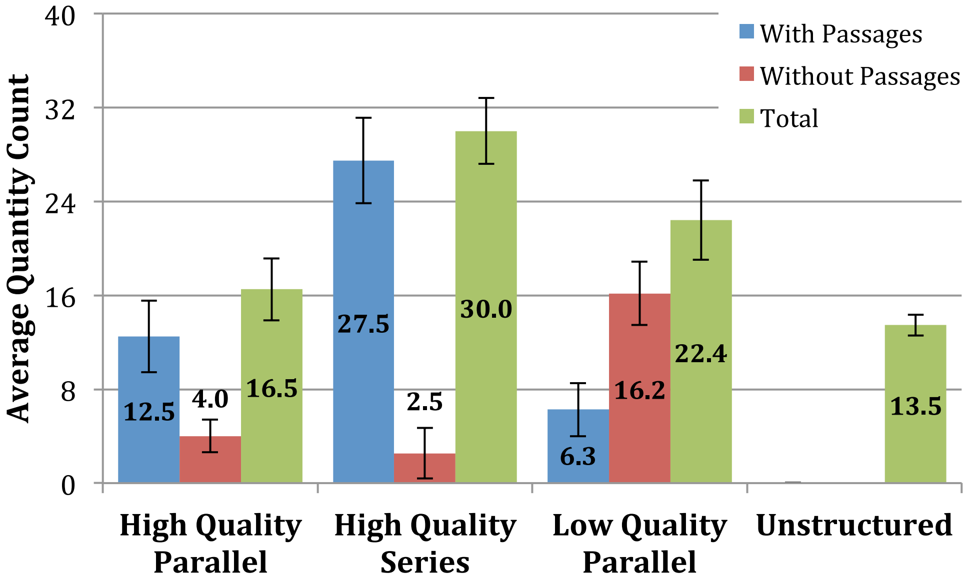


Figure 20. Average number of non-redundant ideas generated in the experimental conditions. Number of bioinspired ideas, non-bioinspired ideas, and total ideas are shown for each condition. Error bars show ±1 S.E.

Note that this quantity metric counts only non-redundant ideas, so if a participant proposes five solutions using a chemical bath, the chemical bath will only count as a single idea. Not only are more solutions being produced, but also new ideas are being introduced with the solutions. However, there is no statistically significant difference in the total number of ideas produced between the HQP and unstructured conditions, despite there being far fewer solutions per participant in the unstructured participant. Section 9.2 pointed out that quantity is often related to the level of detail in a solution. Although participants create fewer solutions without any design stimuli, the solutions produced tend to be more detailed. The greater detail of solutions from the unstructured condition is why the unstructured quantity is comparable to the HQP quantity counts.

Table 30. p-values for differences in quantity counts between conditions. p-values for the difference between the quantity of bioinspired and non-bioinspired ideas within each condition is also listed.

	High Quality Parallel	High Quality Series	Low Quality Parallel
High Quality Series	0.00	-	-
Low Quality Parallel	0.21	0.12	-
Unstructured	0.32	0.00	0.03
With vs. Without Passage	0.03	0.00	0.01

When we compare the quantity of ideas generated in bioinspired solutions to those in non-bioinspired solutions, we see the same trends observed in the solution counts: very few non-bioinspired ideas in the HQS condition, more bioinspired than non-bioinspired ideas in the HQP condition, and the reverse in the LQP condition. Once again, the number of bioinspired and non-bioinspired solution largely determines quantity counts. The average number of ideas per solution across all solutions is 1.74, 1.80, and 1.96 for the HQP, HQS, and LQP conditions, respectively. In contrast the average number of ideas per solution in the unstructured condition is 3.46.

10.4. **Quality**

Figure 21 shows the average quantity count per condition, for bioinspired, non-bioinspired, and total ideas with error bars denoting one standard error. Table 31 lists p-values for a series of t-tests comparing the total quality ratings between each set of conditions, as well as a comparison of the quantity of bioinspired and non-bioinspired ideas within the conditions.

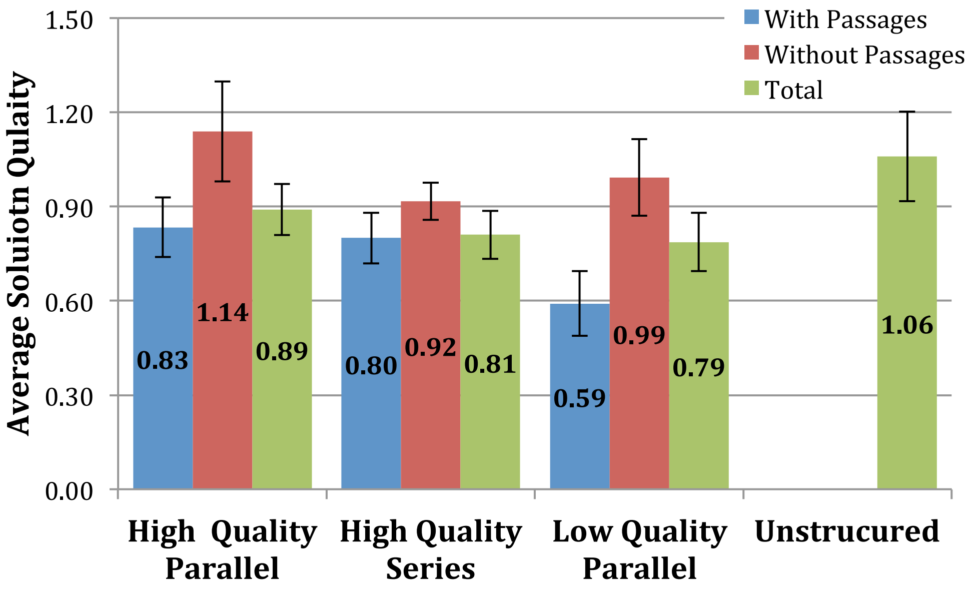


Figure 21. Average participant quality for each experimental condition. Quality for solution generated with a passage, without a passage, and total solutions are shown for each condition. Error bars show ±1 S.E.

There is no statistically significant difference in the overall quality of solutions produced in the four experimental conditions. Although the mean quality for the unstructured condition is higher than the means of the other conditions, the differences are not significant ($p \geq 0.16$). Overall, using biology passages to aid idea generation does not appear to have any noticeable effect on solution quality, with one exception.

In the LQP condition, the non-bioinspired solutions do have significantly higher quality than the bioinspired solutions. Recall that the passages used in the LQP condition were selected specifically because they seldom or never prompted ideas among participants in the corn shucker analogy survey. Solutions are likely to be infeasible or tangentially related to the design problem when they are generated using a single sentence with little relevance to the design problem, or with little comprehensible meaning to the designer. In contrast, we would expect that solutions generated without the passages will have quality rankings similar to solutions produced in the unstructured condition.

Table 31. p-values for differences in quality scores between conditions. p-values for the difference between the quality of bioinspired and non-bioinspired solutions within each condition is also listed.

	High Quality Parallel	High Quality Series	Low Quality Parallel
High Quality Series	0.50	-	-
Low Quality Parallel	0.43	0.86	-
Unstructured	0.35	0.18	0.16
With vs. Without Passage	0.17	0.39	0.04

When considering solution quality, we may also be concerned with the number of high quality ideas generated. While low quality solutions may be sufficiently creative to be worth developing, high quality solutions can conceivably be implemented immediately. Figure 22 shows the number of high quality ideas produced for each condition, averaged across participants. Table 32 lists p-values for a series of t-tests comparing the total high quality solution counts between each set of conditions, as well as a comparison of the bioinspired and non-bioinspired solutions within the conditions. In terms of overall performance, only the HQS condition produces significantly more high quality solutions than unstructured idea generation (p=0.06). Considering the number of solutions produced in each condition, we can infer that the proportion of high quality ideas is actually higher with unstructured idea generation. This may be a result of participants self-censoring their low-quality solutions when generating ideas without any stimuli. Examining the LQS condition, we find that significantly more high quality solutions are produced without passages than with (p=0.05). Typically ideas generated using low-quality passages are technically difficult, infeasible, or just silly (e.g. have a family member shuck the corn). More high quality solutions are produced using a passage in the HQS condition than without, but this difference is not statistically significant (p=0.11). However, this difference is largely a result of the scarcity of non-passage solutions generated in the HQS condition.

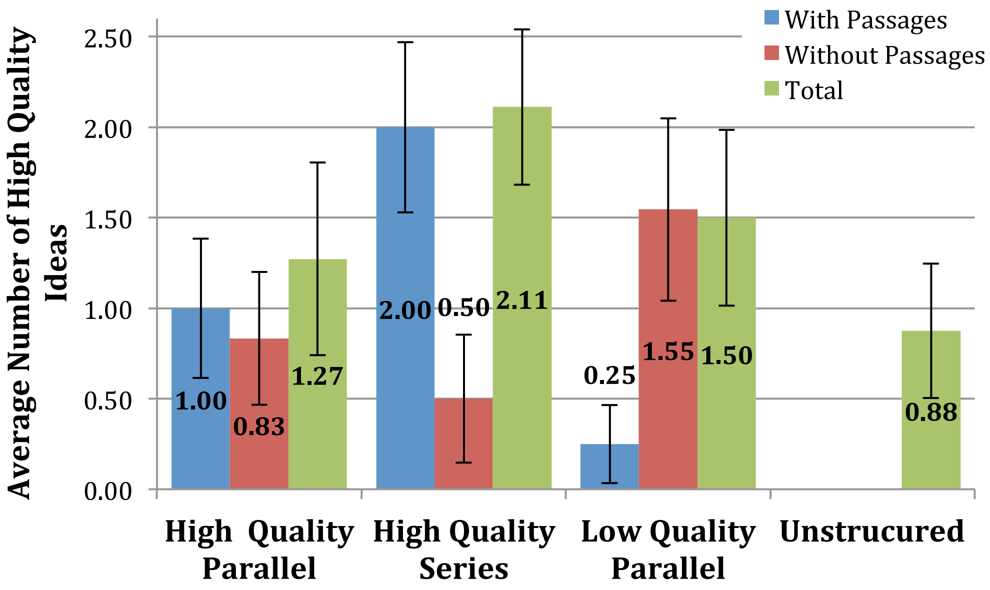


Figure 22. Average number of high quality solutions per participant for each experimental condition. Error bars show ±1 S.E.

Table 32. p-values for differences in the number of high quality solutions between conditions. p-values for the difference between the quality of bioinspired and non-bioinspired solutions within each condition is also listed.

	High Quality Parallel	High Quality Series	Low Quality Parallel
High Quality Series	0.34	-	-
Low Quality Parallel	0.76	0.50	-
Unstructured	0.46	0.06	0.26
With vs. Without Passage	0.78	0.11	0.05

10.5. **Novelty**

Figure 23 shows the average novelty per condition, for bioinspired, non-bioinspired, and total solutions with error bars denoting one standard error. Table 33 lists p-values for a series of t-tests comparing the total novelty scores between each set of conditions, as well as a comparison of the novelty of bioinspired and non-bioinspired ideas within the conditions. In terms of novelty, we actually see very little difference between the four conditions. Only the difference between the HQS and LQP conditions is close to being statistically significant (p=0.12).

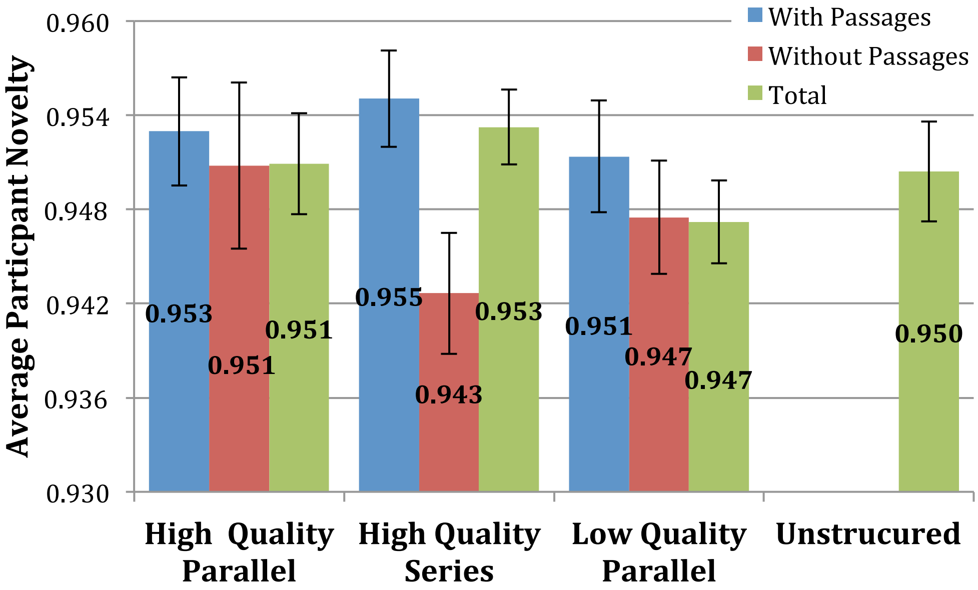


Figure 23. Average participant novelty for each experimental condition. Novelty for solution generated with a passage, without a passage, and total solutions are shown for each condition. Error bars show ±1 S.E.

One of the appeals of bioinspired design is its potential to prompt novel solutions by introducing design stimuli from biology that engineering designers would not typically consider. However, this study can find no detectable difference in the novelty of solutions regardless of the presence, quality, or presentation of biology passages. Before dismissing this approach however, we should consider two points. First, in the HQS condition, the bioinspired solutions are significantly more novel than those generated without using a passage. This could be simply the result of a sampling bias – recall from section 10.2 that HQS participant produced, on average, 1 non-bioinspired solution out of 15 solutions. However, we also observe that the mean participant novelty in the HQP and LQP conditions is higher for solutions that were generated using a passage. This seems to suggest that, while overall novelty may not improve when using biology passages as a concept generation aid, additional ideas generated without the passages will tend to be commonly proposed solutions with low novelty.

Table 33. p-values for differences in novelty scores between conditions. p-values for the difference between the novelty of bioinspired and non-bioinspired solutions within each condition is also listed.

	High Quality Parallel	High Quality Series	Low Quality Parallel
High Quality Series	0.59	-	-
Low Quality Parallel	0.41	0.12	-
Unstructured	0.92	0.50	0.46
With vs. Without Passage	0.76	0.09	0.48

While it is desirable to have idea generation methods that increase average solution novelty, we can also focus on the most novel solutions produced. In Table 34 the number of solutions from each condition is listed for bins with novelty scores greater than 0.99; each of these bins contains less than 1% of the solutions generated across all conditions. In four of the eleven listed bins (Torsion, Impact, Drill, and Vibration), at least 75% of the solutions come from the HQP or HQS conditions despite these conditions only accounting for 48% of generated solutions. Another two bins (Popcorn and Restructure Problem) appear only in conditions with biology stimuli. These highly novel solutions offer only a small sample, but it does seem that biology stimuli can help designers find novel solutions that would not be easily found through unstructured idea generation.

Table 34. Number of high novelty solutions from each condition.

Bin Name	**Bin Novelty**	**High Quality Parallel**	**High Quality Series**	**Low Quality Parallel**	**Unstructured**
Torsion	0.993	2	2	0	0
Impact	0.993	0	4	0	0
Drill	0.993	0	3	1	0
Popcorn	0.997	0	1	1	0
Restructure problem	0.993	0	1	3	0
Vibration	0.992	2	2	0	1
Lasers	0.995	1	0	1	1
Visual sensor	0.995	1	0	1	1
Remove core	0.995	0	1	1	1
Series of mesh wires	0.997	0	0	1	1
Slit plates	0.997	0	0	0	2

Novelty is a useful metric for comparing different idea generation methods, but the scores can mask potentially important information. The way novelty is calculated, it is possible that bioinspired and non-bioinspired solutions could occupy completely different bins, but if the bins are equally populated than their novelty could be the same. As an extreme example, suppose every person in the HQP condition generated only solutions that used some chemical to shuck corn while all solutions in the unstructured condition used a blade to cut away the husk and silk. If 50 chemical solutions and 50 blade-based solutions were proposed, the HQP and unstructured conditions would have equal novelty. Thus, different conditions may concentrate on different portions of the design space without noticeably affecting their novelty scores. This notion will be further explored in section 10.7.

10.6. **Variety**

Figure 24 shows the average novelty per condition, for bioinspired, non-bioinspired, and total solutions with error bars denoting one standard error. Table 35 lists p-values for a series of t-tests comparing the total novelty scores between each set of conditions, as well as a comparison of the novelty of bioinspired and non-bioinspired ideas within the conditions. It is immediately apparent that the unstructured condition has significantly lower variety than the other conditions. This is result should be expected simply based on the average solution count for the conditions. With an average of 4.5 solutions per participant in the unstructured condition, we should expect to see a variety score close to 0.125 (4.5 solutions across 36 bins) – keep in mind that a single solution may occupy multiple bins. In fact, while the HQS condition has the highest novelty, the ratio of average solutions to average variety is the lowest of any condition. This means that participants in the HQS condition are most likely to have multiple solutions that fall into the same bin (e.g. two solutions abrade away the husk and silk through different mechanisms). Despite any self-similarities in the solutions in the HQS condition, that condition's variety is significantly higher than any other condition. However, there is no significant difference between the variety of the HQP and LQP conditions.

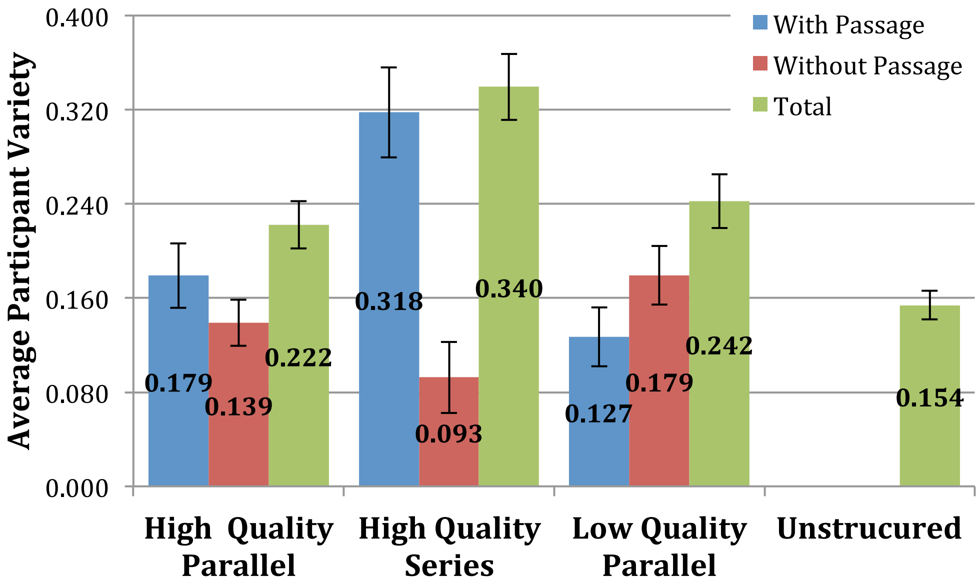


Figure 24. Average participant variety for each experimental condition. Variety for solution generated with a passage, without a passage, and total solutions are shown for each condition. Error bars show ±1 S.E.

Comparing the bioinspired and non-bioinspired conditions within conditions, we once again find a significant difference in the HQS condition between the solutions that were generated using a passage and those without. As with quantity however, this difference is very easily explained by the small number of passages that did not use a biology passage in the HQS condition. A small set of solutions cannot span a large portion of the design space. While the mean variety of bioinspired and non-bioinspired solutions is different for both the HQP and LQP conditions, the differences are not statistically significant.

Table 35. p-values for differences in variety scores between conditions. p-values for the difference between the variety of bioinspired and non-bioinspired solutions within each condition is also listed.

	High Quality Parallel	High Quality Series	Low Quality Parallel
High Quality Series	0.01	-	-
Low Quality Parallel	0.54	0.02	-
Unstructured	0.01	0.00	0.00
With vs. Without Passage	0.29	0.00	0.18

10.7. **Solution Differentiation**

The discussion of novelty in section 10.5 pointed out that different conditions can focus on different portions of the design space, but still have similar novelty. If one type of solution, i.e. solutions that would be classified in some particular bin, is only produced when participants are given biology passages, we can conclude that the passage led the participants to a different portion of the design space than they would normally explore. Figure 25 shows, for each bin used to determine novelty and variety, the relative frequency of solutions from each condition. For example, there are a total of ten solutions that use some form of automatic size adjustment. Only one of those solutions comes from the HQS condition, so the relative frequency of HQS solutions in the automatic size adjustment bin is 10%.

Examining Figure 25, we are particularly interested in bins where one or more experimental conditions are under-represented. Since both the HQP and HQS conditions use the same passages as stimuli, we will first look at the bins where the combined relative frequency of these conditions is less than 30%. These bins are: Brushes, Humans, Remove Core, and Restructure Problem.

Using a brush to remove the corn husk or silk is actually the third most generated solution in the unstructured condition. However, participants who were given passages rarely arrive at a solution using brushes. Participants from the LQP condition all indicated that their brush solutions did not use any passage. Participants from the HQP and HQS conditions all used different passages to arrive at a brush-type solution. There is no strong cue in the passages to suggest the use of a brush, so participants concentrating on the passages often overlook this type of solution. Likewise, passages associated with solutions in the bins Humans, Remove Core, and Restructure Problem, also lack any strong cue to suggest a solution in one of those bins.

A potential pitfall when using biology passages to encourage idea generation is that designers will not look beyond the passages. Designers should employ multiple idea generation methods to explore as much of the design space as possible.

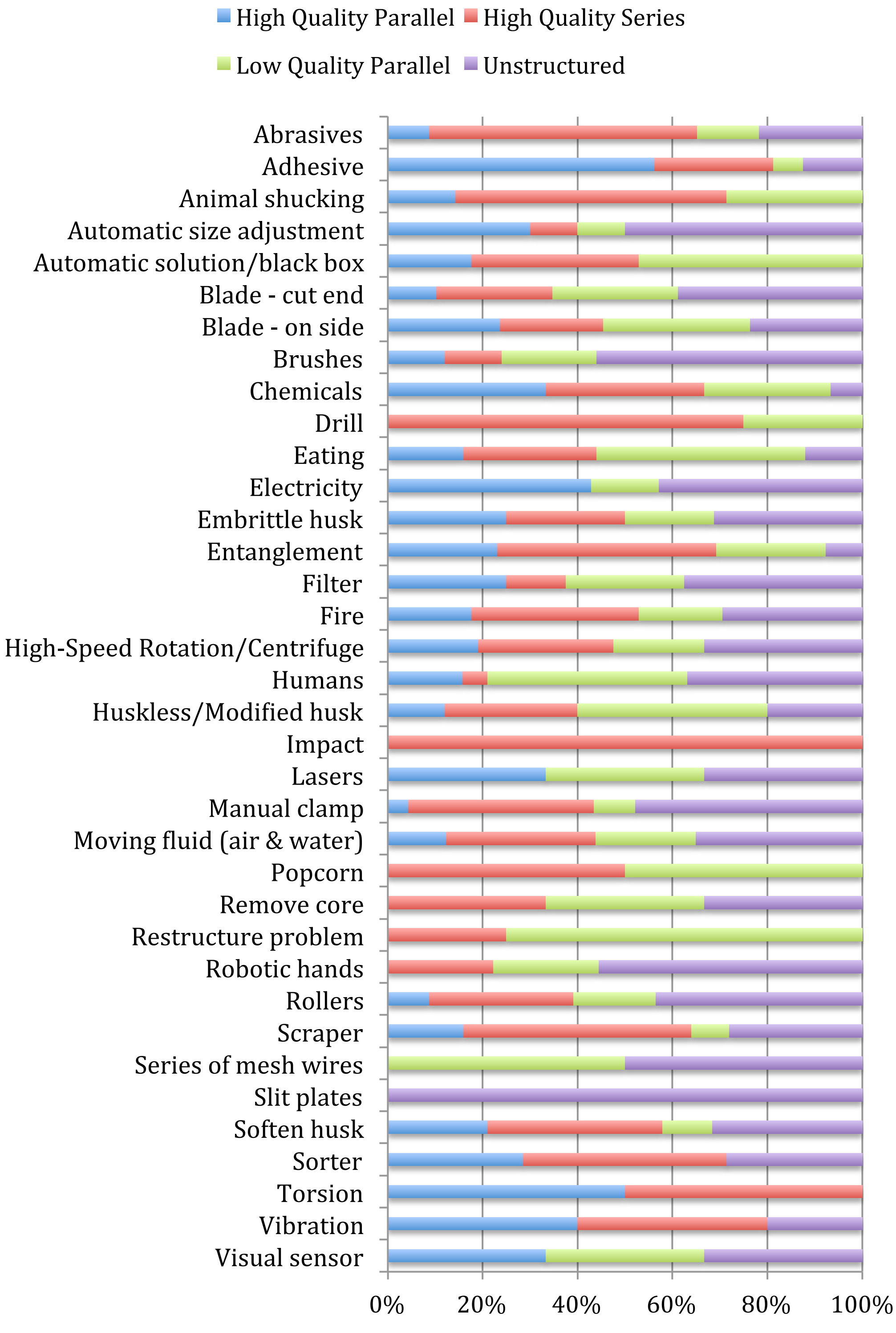


Figure 25. Relative frequency of solutions from each condition, for each bin.

11. CONCLUSIONS AND FUTURE WORK

Bioinspired design has gained much interest because of the ways natural designs differ from engineering designs. Leveraging the design knowledge within natural systems can help designers develop innovative systems. This research increases our knowledge of how sentences from biology texts can affect idea generation for engineering design and it gives us a method for identifying sentences that are likely to help designers.

From the corn shucker survey, we find that some passages are more helpful than others. While helpful passages cannot be easily characterized, many unhelpful passages are very technical sentences filled with biology jargon. Without sufficient context or background, designers cannot understand the information in such sentences to apply the information to the design problem. Additionally, we demonstrate that while participants may often disagree on which survey passages are helpful, their passage selections cannot be regarded as a random, Bernoulli process.

Using responses to a selection of 286 passages from the corn shucker survey, a k Nearest Neighbors classifier is developed to help designers find sources of inspiration in biology. This classifier can potentially double the number of helpful results, from 29% to 62%, found with a keyword search by removing many unhelpful passages from the search results. However, classifiers trained using all 10000 sentence-response data points collected in the corn shucker survey perform very poorly. This is largely because participants seldom agree on whether a given passage is helpful or not. One important caveat in applying a text classifier is that its classifications are implicitly tied to function.

With the approach used here, classifiers developed to find helpful passages pertaining to one function cannot be expected to work well for a different function, as shown by the alarm clock survey. Nevertheless, sufficient data can be gathered to build a separate classifier for each engineering function.

The idea generation experiments detailed in this work how biological passage stimuli can aid idea generation. Firstly, the presence of biology passage stimuli, regardless of stimuli quality, helps designers develop significantly more solutions that cover a larger portion of the design space than designers generating ideas with no structure. When given High Quality Series stimuli or Low Quality Parallel stimuli, the quantity of non-redundant ideas is also significantly increased. Further, by examining high quality passages one at a time instead of all at once, designers saw further increases in the quantity of ideas produced, number of solutions generated, and the variety of those solutions.

There are very few differences in the overall idea generation of designers given high quality or low quality passages. The primary difference there lies in the utilization of passages. According to their own notations, designers given low quality passages actually generated most of their solutions without using any of the passages.

The overall quality and novelty of solutions does not change regardless of whether participants are given helpful or unhelpful passage, and whether the passages are given all at once or on at a time. However, biology passages can direct designers to specific types of solutions, such as twisting the husk and silk off an ear of corn, that they are less likely to find when no stimuli are provided.

The greatest shortcoming of the text classification work presented here is the limited size of the training data. With surveys from 100 participants, 10000 passage ratings were collected, but this only provides 286 labeled sentences that can be used to train the classifiers. For data mining tasks, this is a very small data set. To build reliable and robust classifiers, we would want thousands of labels sentences for training. This desire for more sentences, as well as the desire to develop classifiers for other functions, leads to a possible extension of this work.

Text classifiers such as those presented in this research need to be integrated into a publically accessible search tool. Such a tool should allow users to retrieve passages from a biology corpus by function, and record which passages the users find helpful. By collecting usage data over many uses, the search tool could continuously build sets of training data for multiple functions. Regularly using the accumulated training data to refine text classifiers built into the tool should allow search results to become increasingly more relevant to designers over time.

The development of an integrated search and classification tool is a substantial avenue for future work. However several smaller problems also warrant further investigation. One open question that could be easily answered is how idea generation would be effected by giving a set of low quality passages to designer one at a time. Would the dramatic differences that occur when high quality passages are given one at a time instead of all at once also occur with low quality passages?

The alarm clock survey establishes that BID classifiers are functionally dependent, but there is also a question of problem dependency. The training data used for the BID classifiers was collected in the context of a corn shucker design problem, and associated with the function "separate". It would be interesting to determine whether the passages, and the classifiers by extension, are equally applicable to a different design problem focused on the same function.

REFERENCES

[1] The Biomimcry 3.8 Institue, 2013, "Ask Nature - The Bimimicry Design Portal," available from http://www.asknature.org.

[2] Vattam, S., Wiltgen, B., Helms, M., Goel, A. K., and Yen, J., 2010, "DANE: Fostering Creativity in and through Biologically Inspired Design," Proc. 1st Int. Conf. on Design Creativity, Kobe, Japan.

[3] Srinivasan, V., and Chakrabarti, A., 2009, "Sapphire - an Approach to Analysis and Synthesis," ICED (Int. Conf. on Engineering Design) 09, Stanford, CA.

[4] Vincent, J. F. V., Bogatyreva, O. A., Bogatyrev, N. R., Bowyer, A., and Pahl, A.-K., 2006, "Biomimetics: Its Practice and Theory," Journal of The Royal Society Interface, 3(9), pp. 471-482.

[5] Nagel, R. L., Midha, P. A., Tinsley, A., Stone, R. B., McAdams, D. A., and Shu, L. H., 2008, "Exploring the Use of Functional Models in Biomimetic Conceptual Design," Journal of Mechanical Design, 130(12).

[6] Cheong, H., Chiu, I., Shu, L. H., Stone, R. B., and McAdams, D. A., 2011, "Biologically Meaningful Keywords for Functional Terms of the Functional Basis," Journal of Mechanical Design, 133(2).

[7] Cheong, H., Shu, L., Stone, R., and McAdams, D., 2008, "Translating Terms of the Functional Basis into Biologically Meaningful Keywords," DETC (Design Engineering Technical Conf.) 2008, New York City, NY.

[8] Shu, L. H., 2010, "A Natural-Language Approach to Biomimetic Design," Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 24(04), pp. 507-519.

[9] Nagel, J., Stone, R., and McAdams, D. A., 2010, "An Engineering-to-Biology Thesaurus for Engineering Design," IDETC (Int. Design Engineering Technical Conf.) 2010, Toronto, Canada.

[10] Nagel, J. K. S., Nagel, R. L., Stone, R. B., and McAdams, D. A., 2010, "Function-Based, Biologically Inspired Concept Generation," Artifical Intelligence for Engineering Design, Analysis and Manufacturing, 24(04), pp. 521-535.

[11] Glier, M. W., Tsenn, J., Linsey, J. S., and McAdams, D. A., 2011, "Methods for Supporting Bioinspired Design," IMECE (Int. Mechanical Engineering Conf. and Exposition) 2011. Denver, CO

[12] Chiu, I., and Shu, L. H., 2007, "Biomimetic Design Through Natural Language Analysis to Facilitate Cross-Domain Information Retrieval," Artifical Intelligence for Engineering Design, Analysis and Manufacturing, 21(1), pp. 45-59.

[13] Vattam, S., Helms, M., and Goel, A. K., 2007, "Biologically-Inspired Innovation in Engineering Design: A Cognitive Study," Technical Report GIT-GVU-07-07, Graphics, Visualization, and Usability Center, Georgia Institute of Technology.

[14] Hyman, B., 1998, *Fundamentals of Engineering Design*, Prentice Hall, Upper Saddle River, NJ.

[15] Otto, K., and Wood, K. L., 2001, *Product Design: Techniques in Reverse Engineering, Systematic Design, and New Product Development*, Prentice Hall, Upper Saddle River, NJ.

[16] Dym, C., and Little, P., 2004, *Engineering Design: A Project-Based Introduction*, John Wiley & Sons, New York, NY.

[17] Ulrich, K. T., and Eppinger, S. D., 2004, *Product Design and Development*, McGraw-Hill/Irwin, Boston, MA.

[18] Dieter, G. E., and Schmidt, L. C., 2009, *Engineering Design*, McGraw-Hill, St. Louis, MO.

[19] Bar Cohen, Y., 2003, “Biologically Inspired Intelligent Robots”, Proc. SPIE 5051, Smart Strucutres and Material 2003, Electroactive Polymer Actuators and Devices, San Diego, CA.

[20] Center of Biologically Inspired Design, 2012, “Center for Biologically Inspired Design at Georgia Tech,” available from http://www.cbid.gatech.edu.

[21] Bruck, H. A., Gershon, A. L., Golden, I., Gupta, S. K., Gyger Jr, L. S., Magrab, E. B., and Spranklin, B. W., 2007, "Training Mechanical Engineering Students to Utilize Biological Inspiration During Product Development," Bioinspiration & Biomimetics, 2(4), pp. S198-S209.

[22] Jenkins, C. H., 2011, "Doing BiE: Lessons Learned From Teaching Bio-Inspired Engineering," IMECE (Int. Mechanical Engineering Conf. and Exposition) 2011, Denver, CO

[23] Jenkins, C. H., 2012, *Bio-Inspired Engineering*, Momentum Press, New York, NY.

[24] French, M. J., 1988, *Invention and Evolution: Design in Nature and Engineering*, Cambridge University Press, Cambridge.

[25] Vincent, J. F. V., 1982, *Structural Biomaterials*, Wiley, New York, NY.

[26] Bar-Cohen, Y., ed. , 2011, *Biomimetics: Nature-Based Innovation*, CRC Press, Boca Raton, FL.

[27] Chakrabarti, A., Sarkar, P., Leelavathamma, B., and Nataraju, B., 2005, "A Functional Representation for Aiding Biomimetic and Artificial Inspiration of New Ideas," Artifical Intelligence for Engineering Design, Analysis and Manufacturing, 19(2), pp. 113-132.

[28] Chakrabarti, A., Sarkar, P., Leelavathamma, B., and Nataraju, B., 2005, "A Behavioural Model for Representing Biological and Artificial Systems for Inspiring Novel Designs," Proc. 15th Int. Conf. on Engineering Design, Melbourne Australia, p. 1093.

[29] Goel, A. K., Vattam, S., Wiltgen, B., and Helms, M., 2011, "Cognitive, Collaborative, Conceptual and Creative—Four Characteristics of the Next Generation of Knowledge-Based CAD Systems: A Study in Biologically Inspired Design," Computer-Aided Design, 4(10), pp. 879-900.

[30] Vattam, S. S., and Goel, A. K., "Foraging for Inspiration: Understanding and Supporting the Online Information Seeking Practices of Biologically Inspired Designers," IDETC (Int. Design Engineering Technical Conf.) 2012, Chicago, IL.

[31] Altshuller, G., and Rodman, S., 1999, *The Innovation Algorithm: TRIZ, Systematic Innovation and Technical Creativity*, Technical Innovation Center Inc., Worcester, MA.

[32] Vincent, J. F. V., and Mann, D. L., 2002, "Systematic Technology Transfer from Biology to Engineering," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 360(1791), pp. 159-173.

[33] Vincent, J., Bogatyreva, O., Pahl, A., Bogatyrev, N., and Bowyer, A., 2005, "Putting Biology into TRIZ: A Database of Biological Effects," Creativity and Innovation Management, 14(1), pp. 66-72.

[34] Shu, L., Stone, R., McAdams, D., and Greer, J., 200 7, "Integrating Function-Based and Biomimetic Design for Automatic Concept Generation," Proc. of the 16th Int. Conf. on Engineering Design, Paris, France.

[35] Hirtz, J., Stone, R. B., McAdams, D. A., Szykman, S., and Wood, K. L., 2002, "A Functional Basis for Engineering Design: Reconciling and Evolving Previous Efforts," Research in Engineering Design, 13(2), p. 65-82.

[36] Tinsley, A., Midha, P. A., Nagel, R. L., McAdams, D. A., Stone, R. B., and Shu, L. H., 2008, "Exploring the Use of Functional Models as a Foundation for Biomimetic Conceptual Design," IDETC (Int. Design Engineering Technical Conf.) 2007, Las Vegas, NV.

[37] Nagel, J., Stone, R., and McAdams, D. A., 2010, "Exploring the Use of Category and Scale to Scope a Biological Functional Model," IDETC (Int. Design Engineering Technical Conf.) 2010, Toronto, Canada.

[38] Stroble, J. K., Stone, R. B., and McAdams, D. A., 2009, "Conceptualization of Biomimetic Sensors through Functional Representation of Natural Sensing Solutions," Proc. 17th Int. Conf. on Engineering Design, Stanford, CA.

[39] Stroble, J. K., Watkins, S. E., Stone, R. B., McAdams, D. A., and Shu, L. H., 2008, "Modeling the Cellular Level of Natural Sensing with the Functional Basis for the Design of Biomimetic Sensor Technology," Region 5 Conf., 2008 IEEE, Kansas City, MO.

[40] Hacco, E., and Shu, L. H., 2002, "Biomimetic Concept Generation Applied to Design for Remanufacture," DETC (Design Engineering Technical Conf.) 2002, Montreal, Canada.

[41] Chiu, I., and Shu, L. H., 2004, "Natural Language Analysis for Biomimetic Design," DETC (Design Engineering Technical Conf.) 2004, Salt Lake City, UT.

[42] Chiu, I., and Shu, L. H., 2005, "Bridging Cross-Domain Terminology for Biomimetic Design," DETC (Design Engineering Technical Conf.) 2005, Long Beach, CA.

[43] Ke, J., Chiu, I., Wallace, J., and Shu, L. H., 2010, "Supporting Biomimetic Design by Embedding Metadata in Natural-Language Corpora," IDETC (Int. Design Engineering Technical Conf.) 2010. Toronto, CA

[44] Vakili, V., Chiu, I., Shu, L., McAdams, D. A., and Stone, R., 2007, "Including Functional Models of Biological Phenomena as Design Stimuli," IDETC (Int. Design Engineering Technical Conf.) 2007, Las Vegas, NV.

[45] Vakili, V., and Shu, L. H., 2001, "Towards Biomimetic Concept Generation," DETC (Design Engineering Technical Conf.) 2001, Pittsburgh, PA.

[46] Mak, T., and Shu, L. H., 2004, "Use of Biological Phenomena in Design by Analogy," DETC (Design Engineering Technical Conf.) 2004, Salt lak City, UT.

[47] Mak, T., and Shu, L., 2008, "Using Descriptions of Biological Phenomena for Idea Generation," Research in Engineering Design, 19(1), pp. 21-28.

[48] Cheong, H., and Shu, L. H., 2012, "Automatic Extraction of Causally Related Functions from Natural-Language Text for Biomimetic Design," IDETC (Int. Design Engineering Technical Conf.) 2012, Chicago, IL.

[49] Cheong, H., Hallihan, G., and Shu, L., 2012, "Understanding Analogical Reasoning in Biomimetic Design: An Inductive Approach," Design Computing and Cognition'12. College Station, TX

[50] Cheong, H., and Shu, L., 2013, "Reducing Cognitive Bias in Biomimetic Design by Abstracting Nouns," CIRP (College International pour la Recherche en Productique) Annals-Manufacturing Technology.

[51] Vattam, S. S., Helms, M. E., and Goel, A. K., 2010, "A Content Account of Creative Analogies in Biologically Inspired Design," Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 24(04), pp. 467-481.

[52] Goel, A. K., Vattam, S., Helms, M., and Wiltgen, B., 2011, "An Information-Processing Account of Creative Analogies in Biologically Inspired Design," Proc. of the 8th ACM (Association for Computing Machniery) conference on Creativity and Cognition, Atlanta, GA, pp. 71-80.

[53] Helms, M., and Goel, A., 2012, "Analogical Problem Evolution In Biologically Inspired Design," Design Computing and Cognition '12, College Station, TX.

[54] Glier, M. W., Tsenn, J., Linsey, J. S., and McAdams, D. A., 2012, "Evaluating the Directed Method for Bioinspired Design," IDETC (Int. Design Engineering Technical Conf.) 2012, Chicago, IL.

[55] Glier, M. W., Tsenn, J., McAdams, D. A., and Linsey, J. S., 2012, "Evaluating Methods for Bioinspired Concept Generation," Design Computing and Cognition, '12, College Station, TX.

[56] Feldman, R., and Sanger, J., 2007, *The Text Mining Handbook: Advances Approaches in Analyzing Unstructured Data*, Cambridge University Press, New York.

[57] Domingos, P., and Pazzani, M., 1997, "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss," Machine learning, 29(2), pp. 103-130.

[58] Kim, S.-B., Han, K.-S., Rim, H.-C., and Myaeng, S. H., 2006, "Some Effective Techniques for Naive Bayes Text Classification," Knowledge and Data Engineering, IEEE Transactions on, 18(11), pp. 1457-1466.

[59] Lewis, D. D., 1998, "Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval," *Machine learning: ECML (European Conf. on Machine Learning) -98*, Springer Berlin Hiedelberg, pp. 4-15.

[60] Witten, I. H., and Frank, E., 2005, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, San Francisco, CA.

[61] Vapnik, V., 1999, *The Nature of Statistical Learning Theory*, Springer, New York, NY.

[62] Joachims, T., 1998,int "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," *Machine learning: ECML (European Conf. on Machine Learning) -98*, Springer Berlin Heidelberg, pp. 137-142.

[63] Brank, J., Grobelnik, M., Milic-Frayling, N., and Mladenic, D., 2002, "Interaction of Feature Selection Methods and Linear Classification Models," In Workshop on Text Learning held at ICML (Int. Conf. on Machine Learning), Sydney, Australia.

[64] Clark, P., and Niblett, T., 1989, "The CN2 Induction Algorithm," Machine learning, 3(4), pp. 261-283.

[65] Clark, P., and Boswell, R., 1991, "Rule Induction with CN2: Some Recent Improvements," *Machine learning—EWSL (European Working Session on Learning) -91*, Springer Berlin Hidelberg, pp. 151-163.

[66] Blake, C., and Pratt, W., 2001, "Better Rules, Fewer Features: A Semantic Approach to Selecting Features from Text," Data Mining, 2001, Proc. IEEE Int. Conf. on, San Jose, CA. pp. 59-66.

[67] Feldman, R., and Hirsh, H., 1996, "Mining Associations in Text in the Presence of Background Knowledge," Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining, Portland, OR, pp. 343-346.

[68] Wilcox, A., Hripcsak, G., and Friedman, C., 2000, "Using Knowledge Sources to Improve Classification of Medical Text Reports," Proc. KDD (Knowledge Discovery and Data Mining)-2000, Boston, MA.

[69] Yang, Y., and Pedersen, J. O., 1997, "A Comparative Study on Feature Selection in Text Categorization," Proc. 14th Int. Conf. on Machine Learning, Nashville, TN, pp. 412-420.

[70] Bird, S., Klein, E., and Loper, E., 2009, *Natural Language Processing with Python*, O'Reilly Media, Sabastopol, CA.

[71] Brill, E., 1992 "A Simple Rule-Based Part of Speech Tagger," Proc. of the Workshop on Speech and Natural Language, Association for Computational Linguistics, pp. 112-116.

[72] Cutting, D., Kupiec, J., Pedersen, J., and Sibun, P., 1992, "A Practical Part-of-Speech Tagger," Proc. of the Third Conf. on Applied Natural Language Processing, Association for Computational Linguistics, Columbus, OH, pp. 133-140.

[73] Schütze, H., 1993, "Part-of-Speech Induction from Scratch," Proc. of the 31st Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, pp. 251-258.

[74] Van Rijsbergen, C. J., Robertson, S. E., and Porter, M. F., 1980, *New Models in Probabilistic Information Retrieval*, Computer Laboratory, University of Cambridge, Cambridge

[75] Porter, M. F., 1980, "An Algorithm for Suffix Stripping," Program: Electronic Library and Information Systems, 14(3), pp. 130-137.

[76] Porter, M. F., 2001, "Snowball: A Language for Stemming Algorithms," available from http://snowball.tartarus.org/texts/introduction.html.

[77] Paice, C. D., 1990, "Another Stemmer," SIGIR (Special Interest Group on Information Retrieval) Forum, 24(3), pp. 56-61.

[78] Linsey, J. S., Clauss, E. F., Kurtoglu, T., Murphy, J. T., Wood, K. L., and Markman, A. B., 2011, "An Experimental Study of Group Idea Generation Techniques: Understanding the Roles of Idea Representation and Viewing Methods," Journal of Mechanical Design, 133(3).

[79] Genco, N., Hölttä-Otto, K., and Seepersad, C., 2010, "An Experimental Investigation of the Innovation Capabilities of Engineering Students," Proc. of the ASEE (American Society of Engineering Educators) Annual Conf. and Exposition, Louisville, KY.

[80] Chiu, I., and Shu, L. H., 2007, "Using Language as Related Stimuli for Concept Generation," Artifical Intelligence for Engineering Design, Analysis and Manufacturing, 21(2), pp. 103-121.

[81] Marcus, M., 2013, "Treebank Tokenization," avaliable from http://www.cis.upenn.edu/~treebank/tokenization.html.

[82] Salton, G., 1971, *The SMART Retrieval System – Experiments in Automatic Document Processing*, Prentice-Hall, Englewood Cliffs, NJ.

[83] Schölkopf, B., Smola, A. J., Williamson, R. C., and Bartlett, P. L., 2000, "New Support Vector Algorithms," Neural computation, 12(5), pp. 1207-1245.

[84] Kohavi, R., 1995, "The Power of Decision Tables," *Machine Learning: ECML (European Conf. on Machine Learning) -95*, Springer Berlin Heidelberg, pp. 174-189.

[85] Shah, J., 2003, "Metrics for Measuring Ideation Effectiveness," Design Studies, 24(2), pp. 111-134.

[86] Linsey, J., Green, M. G., Murphy, J., Wood, K. L., and Markman, A. B., 2005, "Collaborating to Success: An Experimental Study of Group Idea Generation Techniques," DETC (Design Engineering Technical Conf.) 2005, Long Beach, CA.

[87] Livingston, S. A., 2004, "Equating Test Scores (without IRT)," Princeton, NJ: ETS.

APPENDIX A. IDEA GENERATION PROBLEMS

This appendix contains copies of the design problems given to participants in the various idea generation experiments used for this research. In order, the design problems are: Unstructured Idea Generation, Biology Passage Stimuli Ideation with High Quality Parallel passages, Biology Passage Stimuli Ideation with Low Quality Parallel passages, and Biology Passage Stimuli Ideation with High Quality Series passages.

Unstructured Idea Generation – Packet page 1

Design Problem - Device to Aid in Shucking Corn

Problem Description:
Corn is currently the most widely grown crop in the Americas with the United States producing 40% of the world's harvest. However, only the loose corn kernels are used when bought canned or frozen in grocery stores. An ear of corn has a protective outer covering of leaves, known as the husk, and strands of corn silk threads run between the husk and the kernels. The removal of husk and silk to clean the corn is known as shucking corn. Design a device that quickly and cheaply shucks corn for mass production.

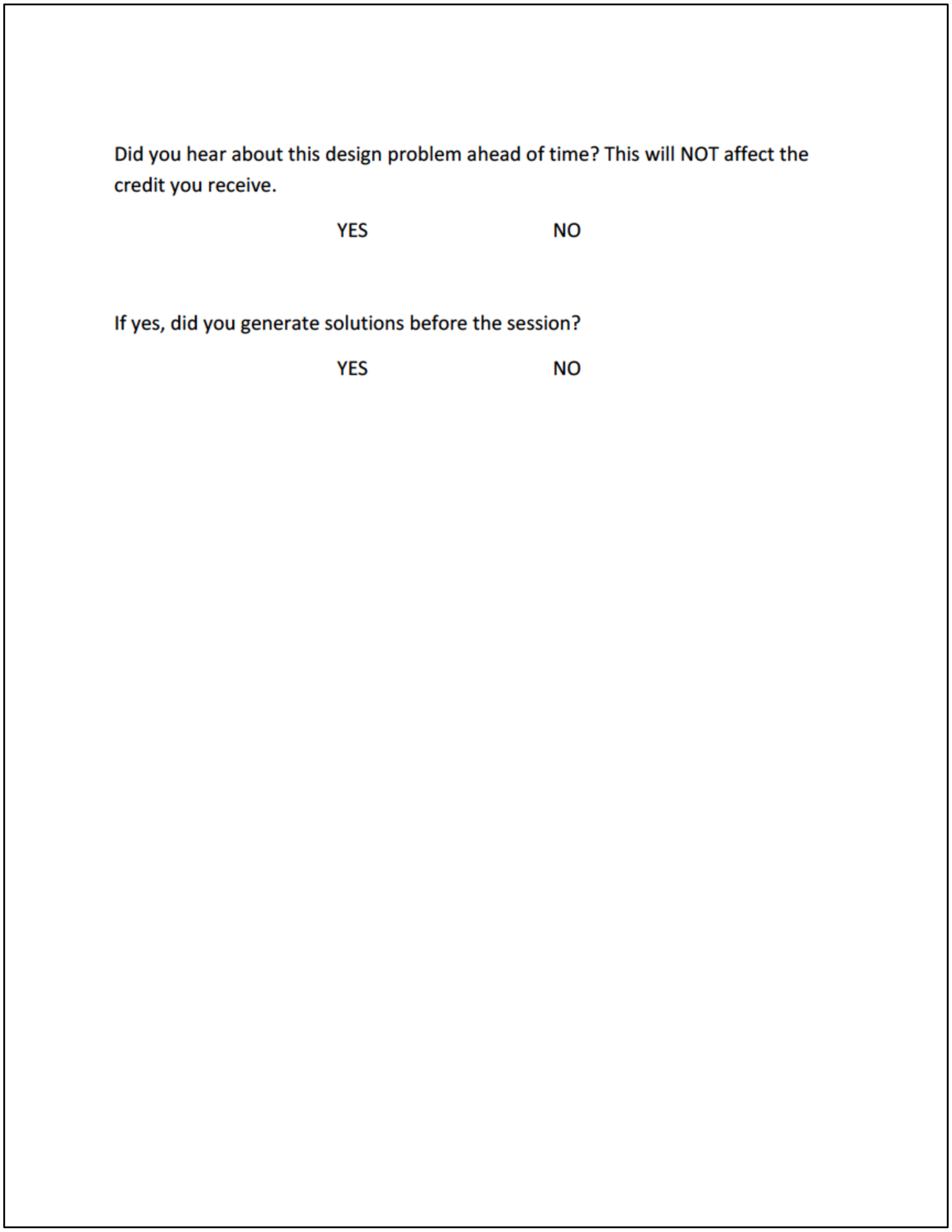
http://www.art-photograph-gallery.com/pictures-of-corn.html

Customer Needs:

- Must remove husk and silk from corn cob with minimal damage to kernels.
- A large quantity of corn must be shucked quickly.
- Low cost.

Please sketch and note (with words) one design solution per page.

Unstructured Idea Generation – Packet page 2



Did you hear about this design problem ahead of time? This will NOT affect the credit you receive.

YES NO

If yes, did you generate solutions before the session?

YES NO

Twenty blank pages make up the reminder of the packet.

Biology Passage Stimuli Ideation with High Quality Parallel passages -

Packet Page 1

Design Problem - Device to Aid in Shucking Corn

Problem Description:
Corn is currently the most widely grown crop in the Americas with the United States producing 40% of the world's harvest. However, only the loose corn kernels are used when bought canned or frozen in grocery stores. An ear of corn has a protective outer covering of leaves, known as the husk, and strands of corn silk threads run between the husk and the kernels. The removal of husk and silk to clean the corn is known as shucking corn.

You have been given a set of numbered passages extracted from professional journals on biology that may help develop ideas for separating the corn from the husk and silk. **Read all the provided passages, and try to use the passages to design a device that quickly and cheaply shucks corn for mass production.**

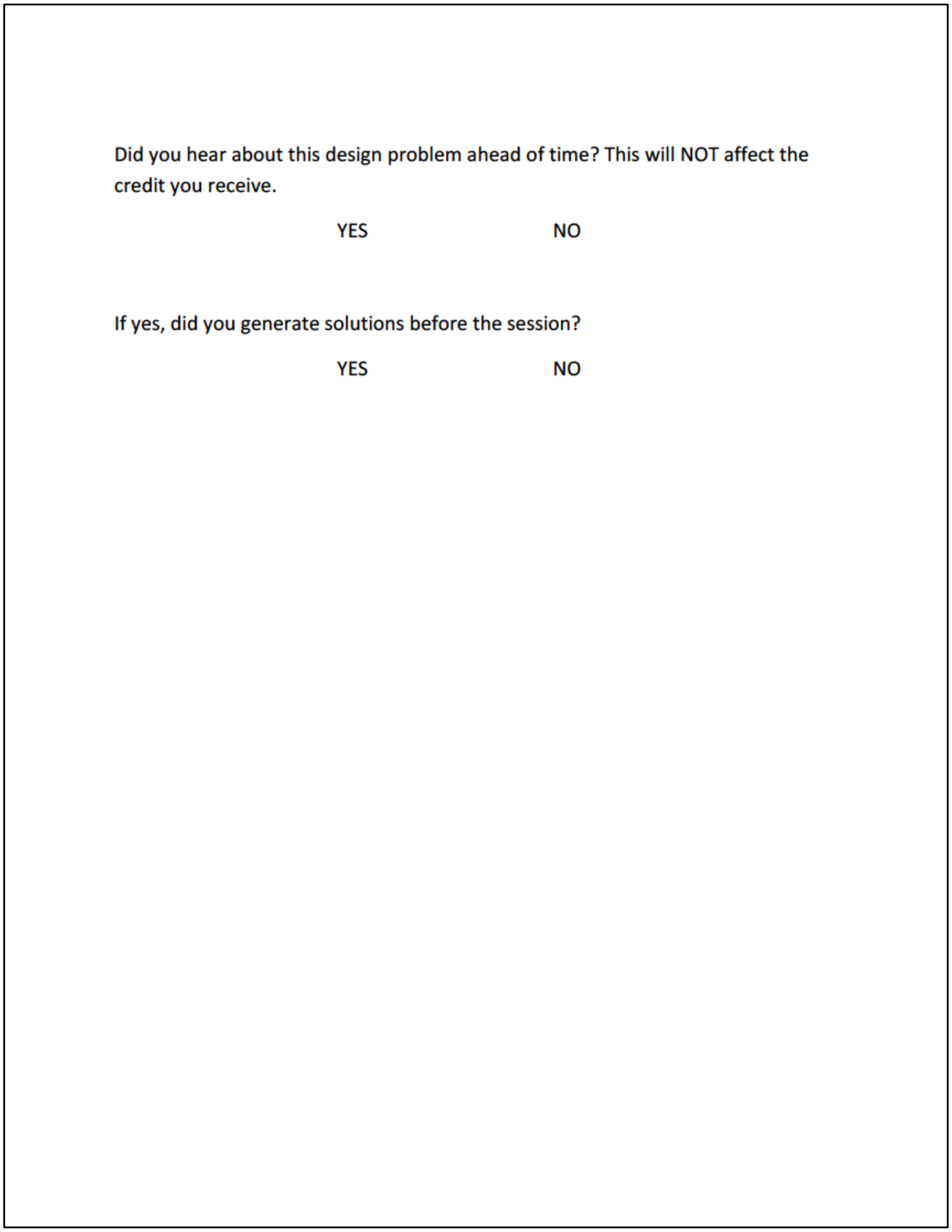
http://www.art-photograph-gallery.com/pictures-of-corn.html

Customer Needs:
- Must remove husk and silk from corn cob with minimal damage to kernels.
- A large quantity of corn must be shucked quickly.
- Low cost.

Please sketch and note (with words) one design solution per page.
Indicate which (if any) numbered passages were used to generate each solution.

Biology Passage Stimuli Ideation with High Quality Parallel passages -

Packet Page 2



Did you hear about this design problem ahead of time? This will NOT affect the credit you receive.

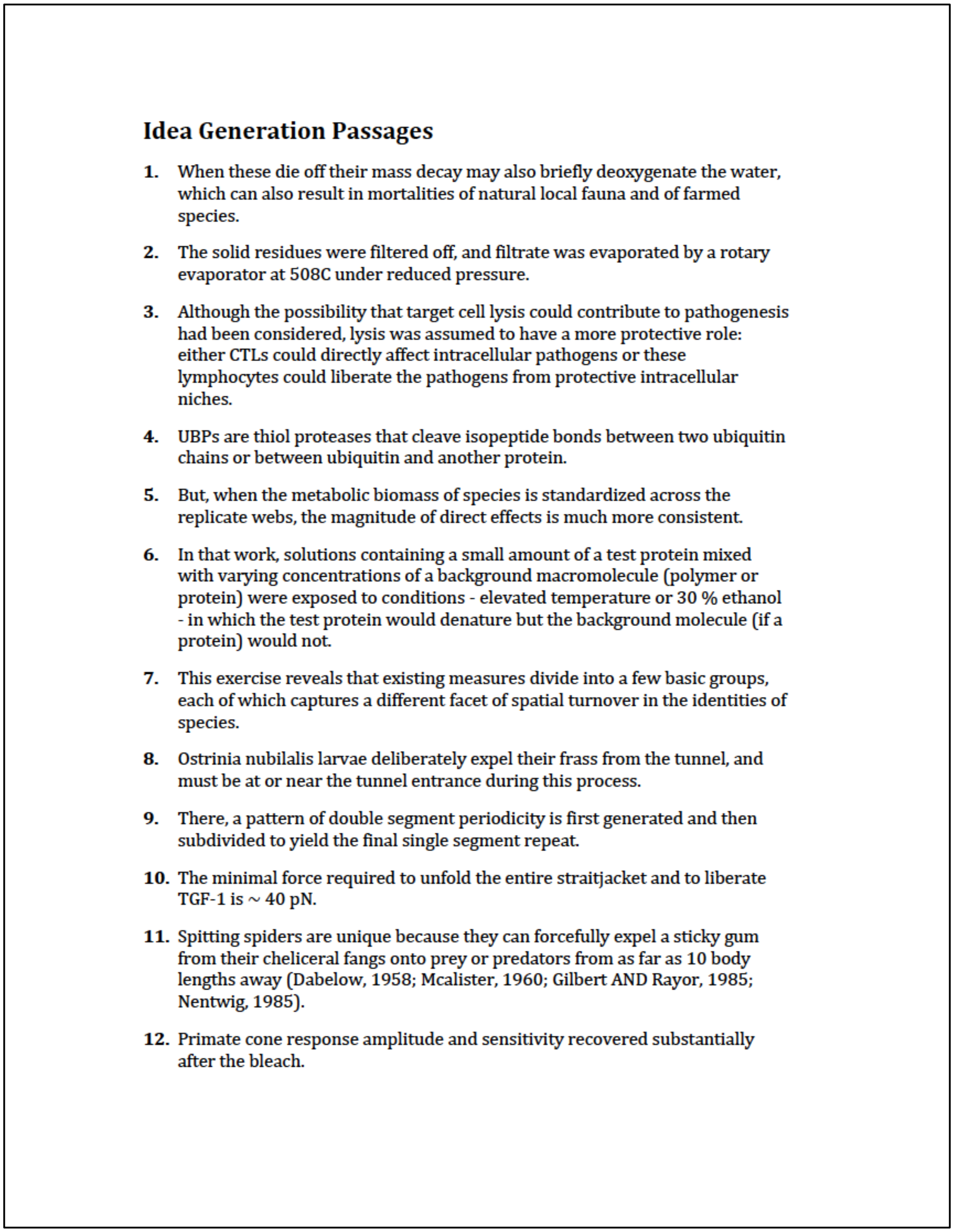
YES NO

If yes, did you generate solutions before the session?

YES NO

Twenty blank pages make up the reminder of the packet.

Biology Passage Stimuli Ideation with High Quality Parallel passages - Supplement Page 1

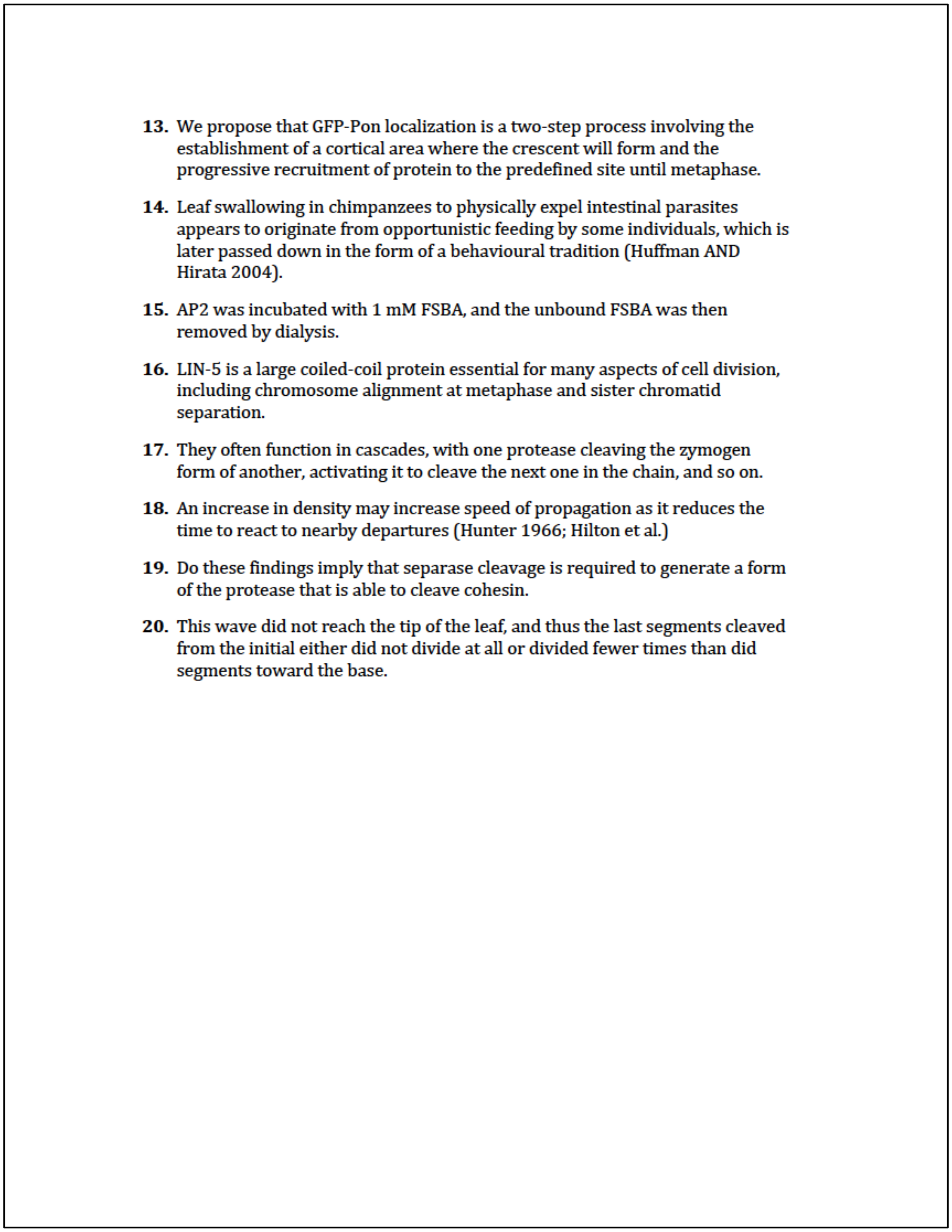


Idea Generation Passages

1. When these die off their mass decay may also briefly deoxygenate the water, which can also result in mortalities of natural local fauna and of farmed species.
2. The solid residues were filtered off, and filtrate was evaporated by a rotary evaporator at 508C under reduced pressure.
3. Although the possibility that target cell lysis could contribute to pathogenesis had been considered, lysis was assumed to have a more protective role: either CTLs could directly affect intracellular pathogens or these lymphocytes could liberate the pathogens from protective intracellular niches.
4. UBPs are thiol proteases that cleave isopeptide bonds between two ubiquitin chains or between ubiquitin and another protein.
5. But, when the metabolic biomass of species is standardized across the replicate webs, the magnitude of direct effects is much more consistent.
6. In that work, solutions containing a small amount of a test protein mixed with varying concentrations of a background macromolecule (polymer or protein) were exposed to conditions - elevated temperature or 30 % ethanol - in which the test protein would denature but the background molecule (if a protein) would not.
7. This exercise reveals that existing measures divide into a few basic groups, each of which captures a different facet of spatial turnover in the identities of species.
8. Ostrinia nubilalis larvae deliberately expel their frass from the tunnel, and must be at or near the tunnel entrance during this process.
9. There, a pattern of double segment periodicity is first generated and then subdivided to yield the final single segment repeat.
10. The minimal force required to unfold the entire straitjacket and to liberate TGF-1 is ~ 40 pN.
11. Spitting spiders are unique because they can forcefully expel a sticky gum from their cheliceral fangs onto prey or predators from as far as 10 body lengths away (Dabelow, 1958; Mcalister, 1960; Gilbert AND Rayor, 1985; Nentwig, 1985).
12. Primate cone response amplitude and sensitivity recovered substantially after the bleach.

Biology Passage Stimuli Ideation with High Quality Parallel passages -

Supplement Page 2



13. We propose that GFP-Pon localization is a two-step process involving the establishment of a cortical area where the crescent will form and the progressive recruitment of protein to the predefined site until metaphase.

14. Leaf swallowing in chimpanzees to physically expel intestinal parasites appears to originate from opportunistic feeding by some individuals, which is later passed down in the form of a behavioural tradition (Huffman AND Hirata 2004).

15. AP2 was incubated with 1 mM FSBA, and the unbound FSBA was then removed by dialysis.

16. LIN-5 is a large coiled-coil protein essential for many aspects of cell division, including chromosome alignment at metaphase and sister chromatid separation.

17. They often function in cascades, with one protease cleaving the zymogen form of another, activating it to cleave the next one in the chain, and so on.

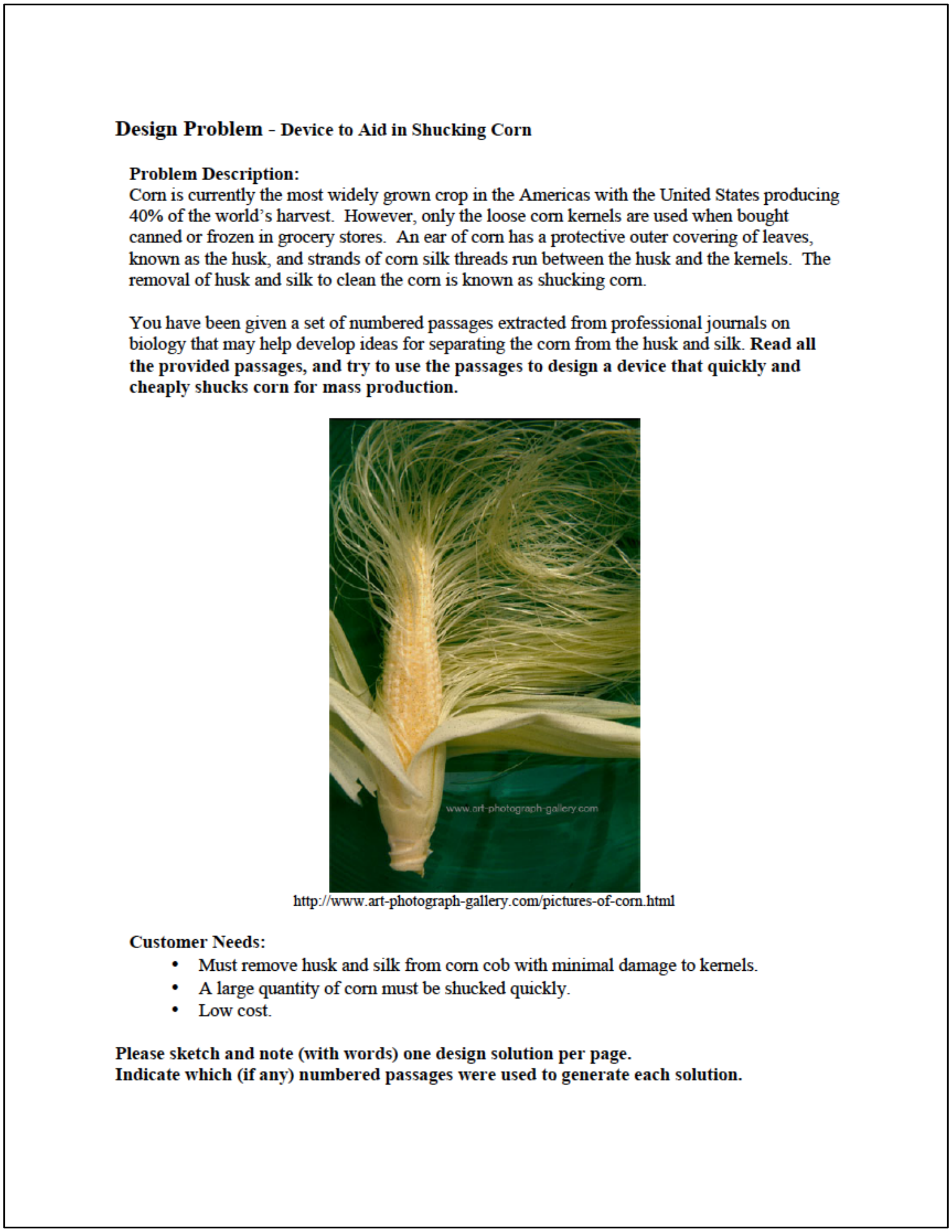
18. An increase in density may increase speed of propagation as it reduces the time to react to nearby departures (Hunter 1966; Hilton et al.)

19. Do these findings imply that separase cleavage is required to generate a form of the protease that is able to cleave cohesin.

20. This wave did not reach the tip of the leaf, and thus the last segments cleaved from the initial either did not divide at all or divided fewer times than did segments toward the base.

Biology Passage Stimuli Ideation with Low Quality Parallel passages -

Packet Page 1



Design Problem - **Device to Aid in Shucking Corn**

Problem Description:
Corn is currently the most widely grown crop in the Americas with the United States producing 40% of the world's harvest. However, only the loose corn kernels are used when bought canned or frozen in grocery stores. An ear of corn has a protective outer covering of leaves, known as the husk, and strands of corn silk threads run between the husk and the kernels. The removal of husk and silk to clean the corn is known as shucking corn.

You have been given a set of numbered passages extracted from professional journals on biology that may help develop ideas for separating the corn from the husk and silk. **Read all the provided passages, and try to use the passages to design a device that quickly and cheaply shucks corn for mass production.**

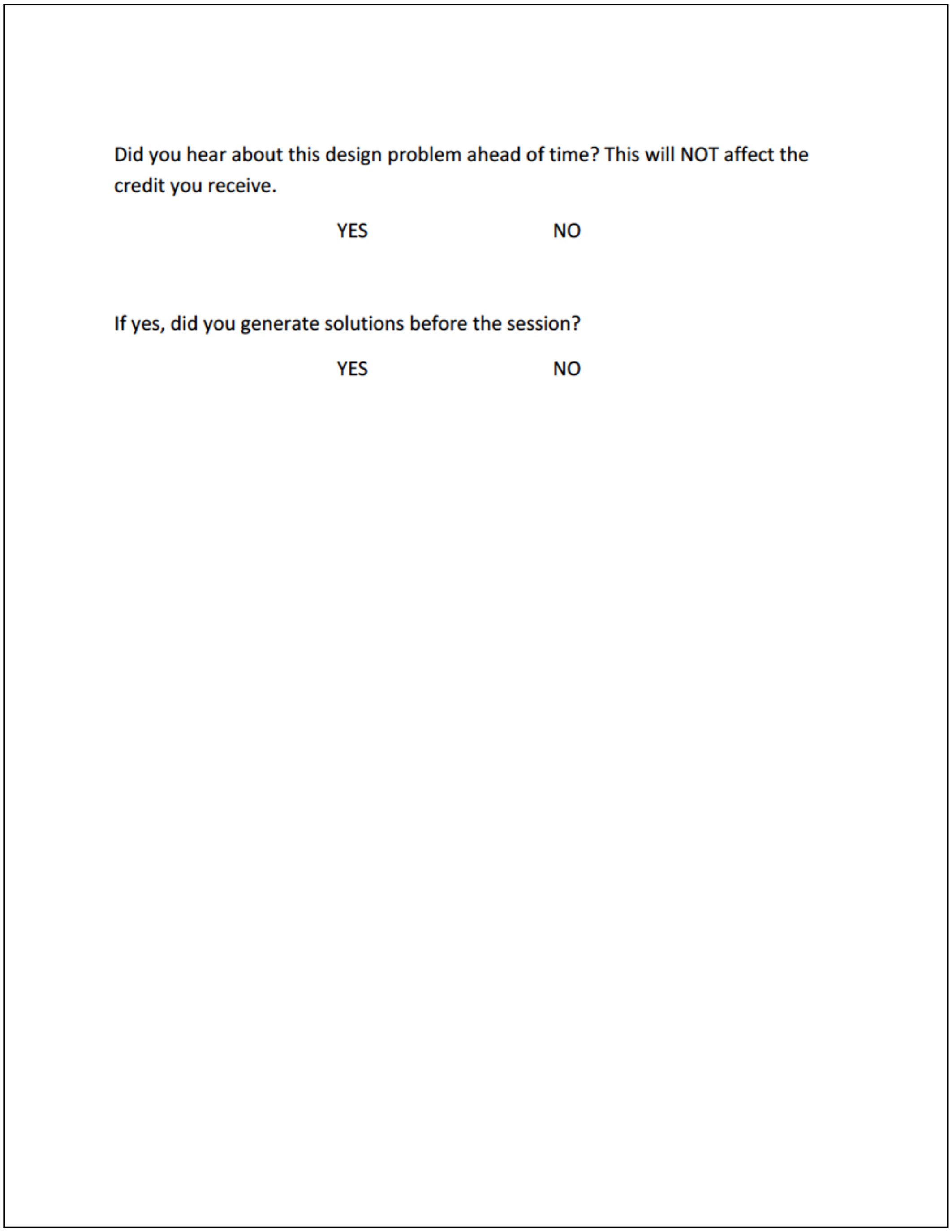
http://www.art-photograph-gallery.com/pictures-of-corn.html

Customer Needs:
- Must remove husk and silk from corn cob with minimal damage to kernels.
- A large quantity of corn must be shucked quickly.
- Low cost.

Please sketch and note (with words) one design solution per page.
Indicate which (if any) numbered passages were used to generate each solution.

Biology Passage Stimuli Ideation with Low Quality Parallel passages -

Packet Page 2



Did you hear about this design problem ahead of time? This will NOT affect the credit you receive.

YES NO

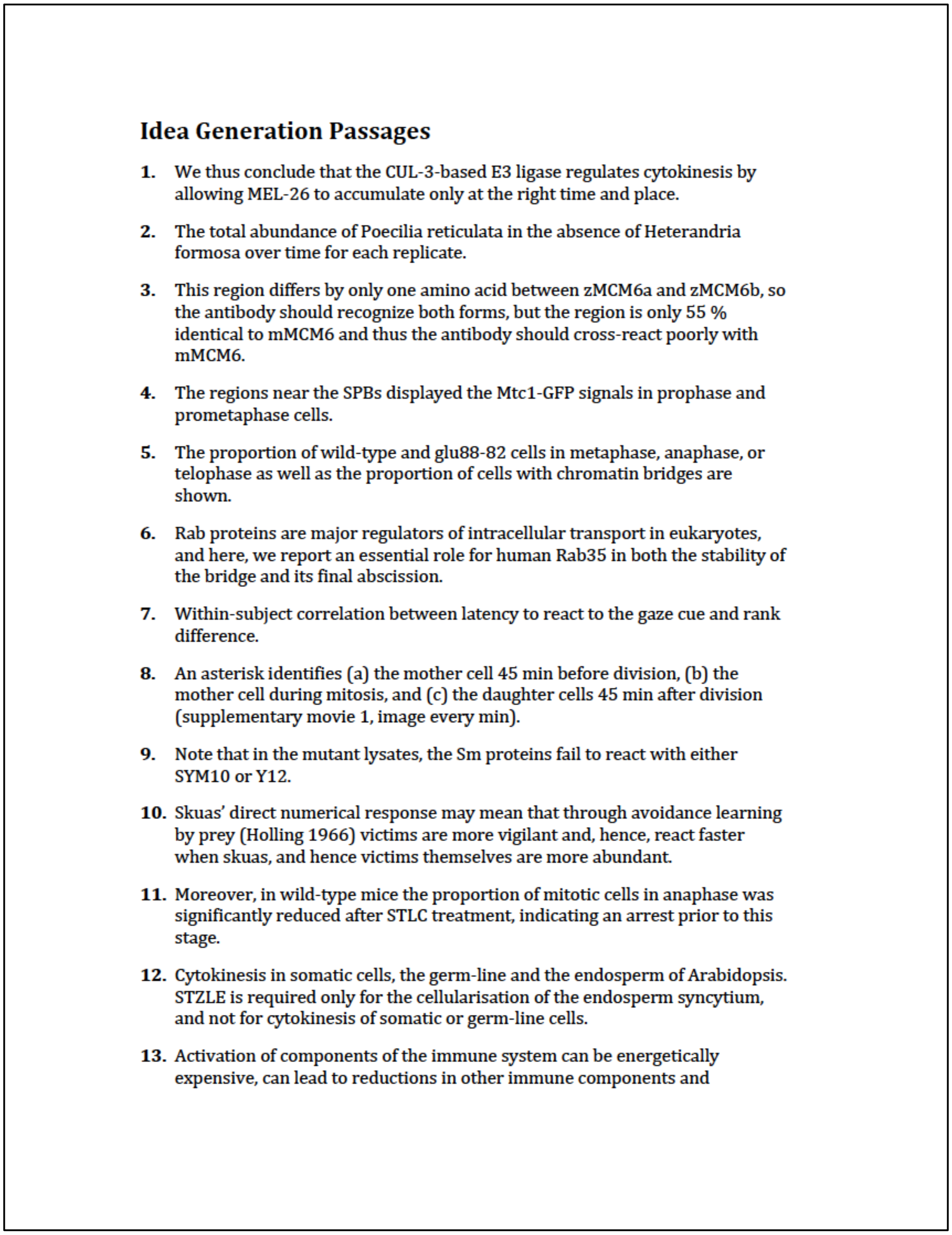
If yes, did you generate solutions before the session?

YES NO

Twenty blank pages make up the reminder of the packet.

Biology Passage Stimuli Ideation with Low Quality Parallel passages -

Supplement Page 1

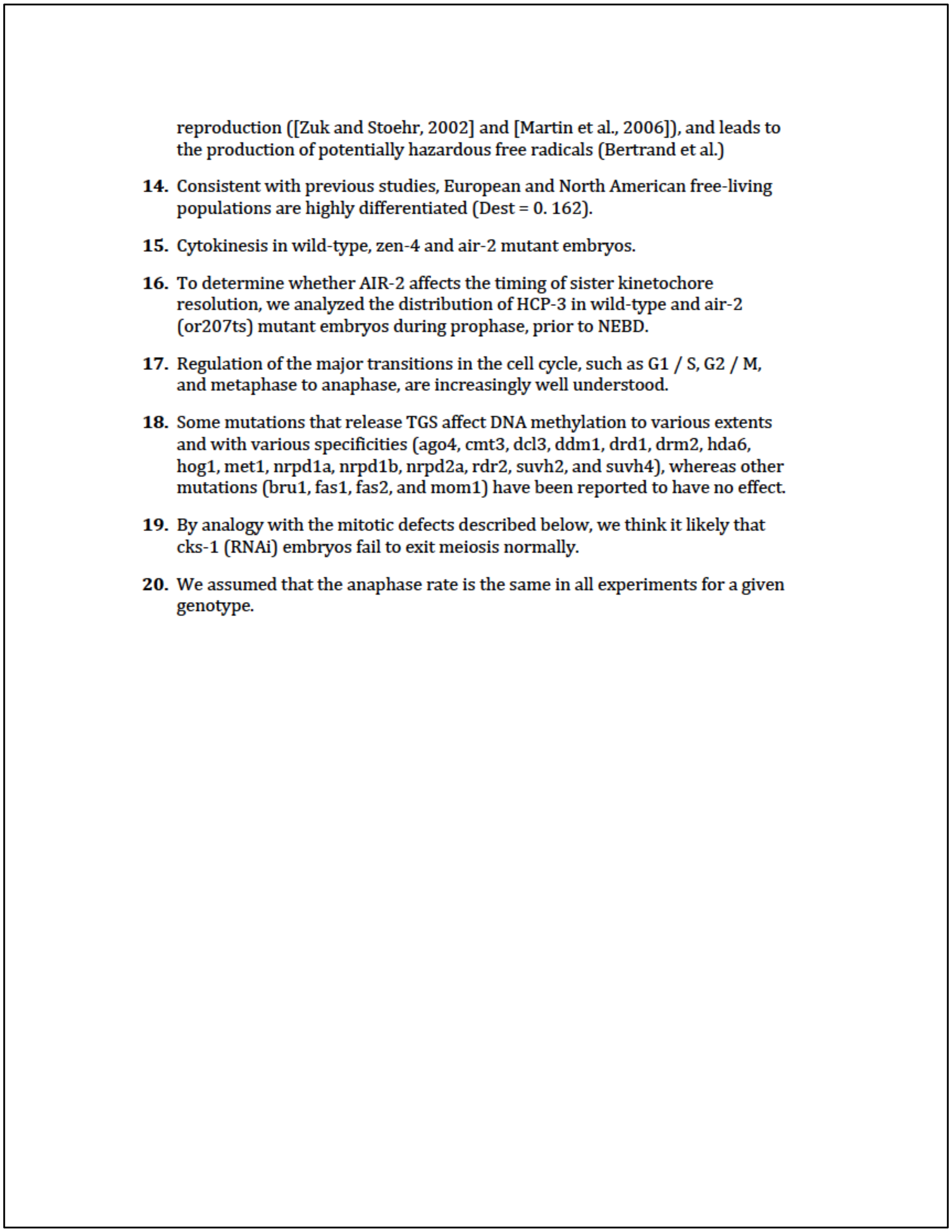


Idea Generation Passages

1. We thus conclude that the CUL-3-based E3 ligase regulates cytokinesis by allowing MEL-26 to accumulate only at the right time and place.
2. The total abundance of Poecilia reticulata in the absence of Heterandria formosa over time for each replicate.
3. This region differs by only one amino acid between zMCM6a and zMCM6b, so the antibody should recognize both forms, but the region is only 55 % identical to mMCM6 and thus the antibody should cross-react poorly with mMCM6.
4. The regions near the SPBs displayed the Mtc1-GFP signals in prophase and prometaphase cells.
5. The proportion of wild-type and glu88-82 cells in metaphase, anaphase, or telophase as well as the proportion of cells with chromatin bridges are shown.
6. Rab proteins are major regulators of intracellular transport in eukaryotes, and here, we report an essential role for human Rab35 in both the stability of the bridge and its final abscission.
7. Within-subject correlation between latency to react to the gaze cue and rank difference.
8. An asterisk identifies (a) the mother cell 45 min before division, (b) the mother cell during mitosis, and (c) the daughter cells 45 min after division (supplementary movie 1, image every min).
9. Note that in the mutant lysates, the Sm proteins fail to react with either SYM10 or Y12.
10. Skuas' direct numerical response may mean that through avoidance learning by prey (Holling 1966) victims are more vigilant and, hence, react faster when skuas, and hence victims themselves are more abundant.
11. Moreover, in wild-type mice the proportion of mitotic cells in anaphase was significantly reduced after STLC treatment, indicating an arrest prior to this stage.
12. Cytokinesis in somatic cells, the germ-line and the endosperm of Arabidopsis. STZLE is required only for the cellularisation of the endosperm syncytium, and not for cytokinesis of somatic or germ-line cells.
13. Activation of components of the immune system can be energetically expensive, can lead to reductions in other immune components and

Biology Passage Stimuli Ideation with Low Quality Parallel passages -

Supplement Page 2



reproduction ([Zuk and Stoehr, 2002] and [Martin et al., 2006]), and leads to the production of potentially hazardous free radicals (Bertrand et al.)

14. Consistent with previous studies, European and North American free-living populations are highly differentiated (Dest = 0. 162).

15. Cytokinesis in wild-type, zen-4 and air-2 mutant embryos.

16. To determine whether AIR-2 affects the timing of sister kinetochore resolution, we analyzed the distribution of HCP-3 in wild-type and air-2 (or207ts) mutant embryos during prophase, prior to NEBD.

17. Regulation of the major transitions in the cell cycle, such as G1 / S, G2 / M, and metaphase to anaphase, are increasingly well understood.

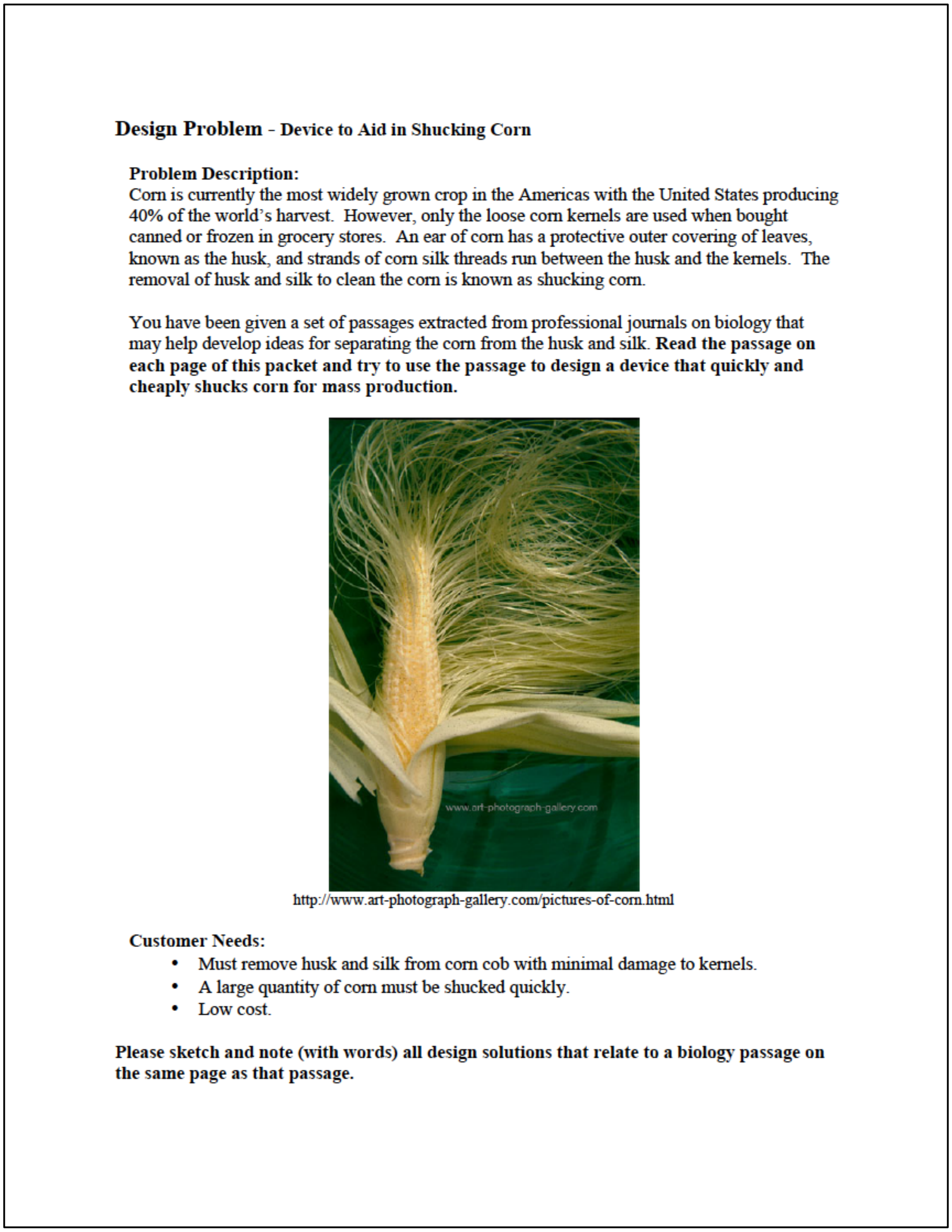
18. Some mutations that release TGS affect DNA methylation to various extents and with various specificities (ago4, cmt3, dcl3, ddm1, drd1, drm2, hda6, hog1, met1, nrpd1a, nrpd1b, nrpd2a, rdr2, suvh2, and suvh4), whereas other mutations (bru1, fas1, fas2, and mom1) have been reported to have no effect.

19. By analogy with the mitotic defects described below, we think it likely that cks-1 (RNAi) embryos fail to exit meiosis normally.

20. We assumed that the anaphase rate is the same in all experiments for a given genotype.

Biology Passage Stimuli Ideation with High Quality Series passages -

Packet Page 1



Design Problem - Device to Aid in Shucking Corn

Problem Description:
Corn is currently the most widely grown crop in the Americas with the United States producing 40% of the world's harvest. However, only the loose corn kernels are used when bought canned or frozen in grocery stores. An ear of corn has a protective outer covering of leaves, known as the husk, and strands of corn silk threads run between the husk and the kernels. The removal of husk and silk to clean the corn is known as shucking corn.

You have been given a set of passages extracted from professional journals on biology that may help develop ideas for separating the corn from the husk and silk. **Read the passage on each page of this packet and try to use the passage to design a device that quickly and cheaply shucks corn for mass production.**

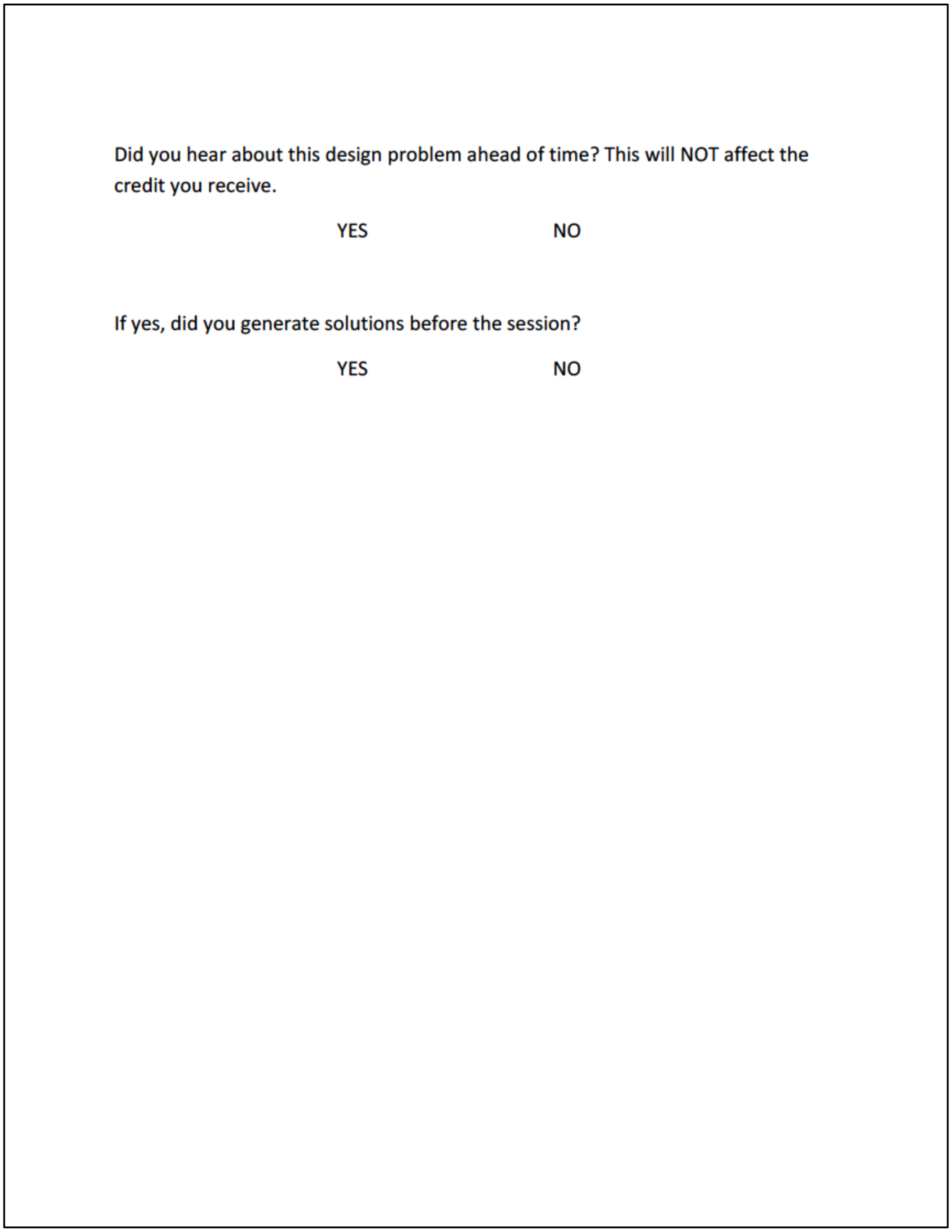
http://www.art-photograph-gallery.com/pictures-of-corn.html

Customer Needs:
- Must remove husk and silk from corn cob with minimal damage to kernels.
- A large quantity of corn must be shucked quickly.
- Low cost.

Please sketch and note (with words) all design solutions that relate to a biology passage on the same page as that passage.

Biology Passage Stimuli Ideation with High Quality Series passages -

Packet Page 2



Did you hear about this design problem ahead of time? This will NOT affect the credit you receive.

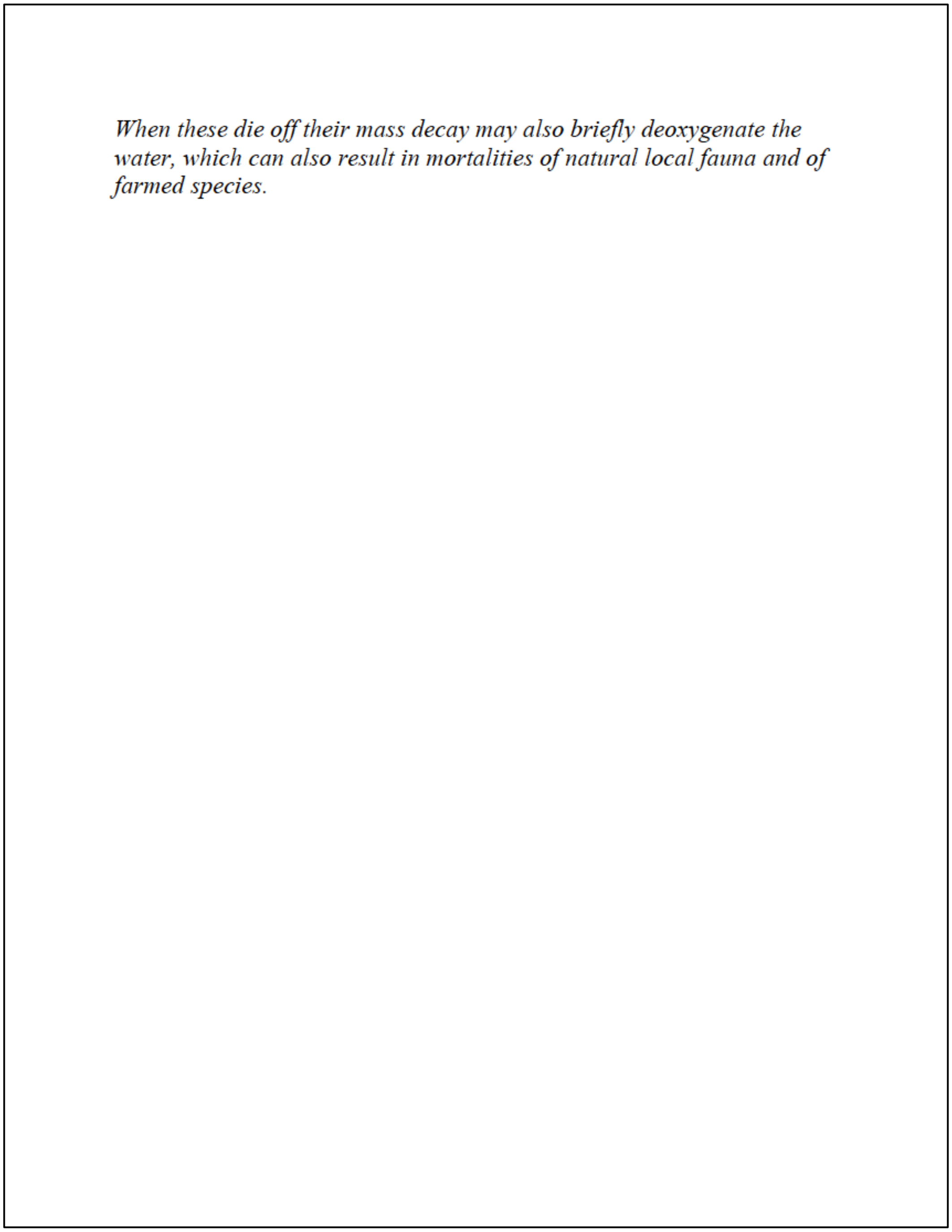
YES NO

If yes, did you generate solutions before the session?

YES NO

Biology Passage Stimuli Ideation with High Quality Series passages -

Packet Page 3



When these die off their mass decay may also briefly deoxygenate the water, which can also result in mortalities of natural local fauna and of farmed species.

The remaining 20 pages in the High Quality Series packet are identical to page 3, except that the passages listed below are used instead. The passages are listed in the order they appear within the packet.

The solid residues were filtered off, and filtrate was evaporated by a rotary evaporator at 508C under reduced pressure.

Although the possibility that target cell lysis could contribute to pathogenesis had been considered, lysis was assumed to have a more protective role: either CTLs could directly affect intracellular pathogens or these lymphocytes could liberate the pathogens from protective intracellular niches.

UBPs are thiol proteases that cleave isopeptide bonds between two ubiquitin chains or between ubiquitin and another protein.

But, when the metabolic biomass of species is standardized across the replicate webs, the magnitude of direct effects is much more consistent.

In that work, solutions containing a small amount of a test protein mixed with varying concentrations of a background macromolecule (polymer or protein) were exposed to conditions - elevated temperature or 30 % ethanol - in which the test protein would denature but the background molecule (if a protein) would not.

This exercise reveals that existing measures divide into a few basic groups, each of which captures a different facet of spatial turnover in the identities of species.

Ostrinia nubilalis larvae deliberately expel their frass from the tunnel, and must be at or near the tunnel entrance during this process.

There, a pattern of double segment periodicity is first generated and then subdivided to yield the final single segment repeat.

The minimal force required to unfold the entire straitjacket and to liberate TGF-1 is ~ 40 pN.

Spitting spiders are unique because they can forcefully expel a sticky gum from their cheliceral fangs onto prey or predators from as far as 10 body lengths away (Dabelow, 1958; Mcalister, 1960; Gilbert AND Rayor, 1985; Nentwig, 1985).

Primate cone response amplitude and sensitivity recovered substantially after the bleach.

We propose that GFP-Pon localization is a two-step process involving the establishment of a cortical area where the crescent will form and the progressive recruitment of protein to the predefined site until metaphase.

Leaf swallowing in chimpanzees to physically expel intestinal parasites appears to originate from opportunistic feeding by some individuals, which is later passed down in the form of a behavioural tradition (Huffman AND Hirata 2004).

AP2 was incubated with 1 mM FSBA, and the unbound FSBA was then removed by dialysis.

LIN-5 is a large coiled-coil protein essential for many aspects of cell division, including chromosome alignment at metaphase and sister chromatid separation.

They often function in cascades, with one protease cleaving the zymogen form of another, activating it to cleave the next one in the chain, and so on.

An increase in density may increase speed of propagation as it reduces the time to react to nearby departures (Hunter 1966; Hilton et al.)

Do these findings imply that separase cleavage is required to generate a form of the protease that is able to cleave cohesin.

This wave did not reach the tip of the leaf, and thus the last segments cleaved from the initial either did not divide at all or divided fewer times than did segments toward the base.

APPENDIX B. IDEA GENERATION VERBAL INSTRUCTIONS

Script for Unstructured Ideation Experiments

<<Hand out consent form and self efficacy survey>>

Hear are two copies of the consent form - one is for you to keep, and one is to be returned to us. Fill out the self-efficacy and GPA survey and hold on to it

<<Hand out packet>>

We are doing an individual activity. The amount of extra credit received will be dependent on the effort you put into it. Please sketch and note all designs you can think of, even if they are not technically ffeasile. Aim for as many ideas as possible at as high of quality and variety as possible. Use only one side of the paper and if you need more paper, let me know. You have 50 minutes for his activity.

Please fill out the form stating if you've heard about the design problem ahead of time. We realize that students talk about it and it's possible to accidently overhear. This will not affect the credit you receive.

Script for Biology Passage Stimuli Ideation Experiments with High Quality Parallel passages or with Low Quality Parallel passages

<<Hand out consent form>>

You have each received two copies of a consent form. One copy is for you to keep, the other needs to be signed and returned to me. Please read over the form and let me know if you have any questions. Sign and date the form where indicated.

<<Hand out packet>>

Today you will be doing a concept generation experiment. You have each been given a packet with a short design problem and a supplemental page. Please read over the design problem and the instructions. On the separate page is a series of passages from biology texts that may help develop solutions to the design problem. Read all the passages and attempt to generate ideas or solutions based on the passages. Sketch or write your any ideas you have, even if they may not be feasible. Aim for as many ideas as possible at as high of quality and variety as possible. Clearly note which passages were used to develop each solution you generate. If you generate solutions unrelated to any passage, please note that as well.

You will have 50 minutes for this experiment. Please begin.

Script for Biology Passage Stimuli Ideation Experimentswith High Quality Series passages

<<Hand out consent form>>

You have each received two copies of a consent form. One copy is for you to keep, the other needs to be signed and returned to me. Please read over the form and let me know if you have any questions. Sign and date the form where indicated.

<<Hand out packet>>

Today you will be doing a concept generation experiment. You have each been given a packet with a short design problem. Please read over the design problem and the instructions. On the following pages are a series of passages from biology texts that may help develop solutions to the design problem. Read each passage and attempt to generate ideas or solutions based on the passage. Sketch or write your any ideas you have, even if they may not be feasible. Aim for as many ideas as possible at as high of quality and variety as possible. If you generate solutions unrelated to any passage, please note that as well.

You will have 50 minutes for this experiment. Please begin.

APPENDIX C. SURVEYS FOR ANALOGY SURVEYS

A total of six sets of 100 passages were used for the two analogy surveys in this work. The full sentence set for each survey is listed in this appendix. Corpus index numbers are attached to each passage.

Corn Shucker Survey A

Passage 623366

When these die off their mass decay may also briefly deoxygenate the water, which can also result in mortalities of natural local fauna and of farmed species.

Passage 214923

Primary o and p blast cells divide with differing degrees of asymmetry.

Passage 2019150

Using MS, two clusters (Table 1) did not divide species exactly as expected in the two-mode paradigm.

Passage 267948

Most bacteria divide by assembling filaments of the tubulin-like protein FtsZ into a cytokinetic ring, which then constricts.

Passage 515430

Centrosomes Continue to Divide Despite Nuclear Cell-Cycle Arrest (A) Embryos expressing TACC-GFP were injected with mitotic cyclin dsRNA.

Passage 1982394

Such nest-count studies have suggested that western gorilla groups often have multiple silverbacks and these multimale groups occasionally divide into smaller subgroups.

Passage 1811392

Of 11 such helping males, seven remained in the same territory in the following year, while eight of 10 non-helping 1-year-old males did ($P = 0.635$, Fisher’s exact probability test), indicating that breeders did not expel non-helping auxiliaries.

Passage 1456144

Leaf swallowing in chimpanzees to physically expel intestinal parasites appears to originate from opportunistic feeding by some individuals, which is later passed down in the form of a behavioural tradition (Huffman & Hirata 2004).

Passage 604880

We revised our computational simulation to include two polar inhibitors that expel mid1p from the membrane.

Passage 1406093

Aphaenogaster cockerelli queens expel the contents of their Dufour's gland onto reproductive workers, which in turn elicits the aggression of nonreproductive workers.

Passage 524285

Detwiler threatened to expel him from the graduate program (1).

Passage 1344188

Monoclonal antibody are still unable to expel N. brasiliensis, the loss of expulsion is not due to increased IFN.

Passage 838639

Thus, during the third cleavage of Lymnaea stagnalis, helical spindle inclination (SI) and spiral blastomere deformation (SD) are observed only in the dominant dextral embryos at metaphase-anaphase, whereas in the recessive sinistral embryos, helicity emerges during the furrow ingression.

Passage 862112

Bub1, Bub3, Cdc20, and to a lesser extent, Mps1, remain detectable at kinetochores throughout anaphase whereas Mad1, Mad2, and BubR1 deplete during metaphase.

Passage 521914

First, as suggested by the analysis in fixed cells, the assembly of a metaphase plate was slow and abnormal in cells lacking Moesin.

Passage 715262

We propose that GFP-Pon localization is a two-step process involving the establishment of a cortical area where the crescent will form and the progressive recruitment of protein to the predefined site until metaphase.

Passage 334678

We thank I. Hickson for sharing unpublished results, M. Terradas for help with MO59K and J cell lines, M. Woodcock for FACS analyses, and M. Eid for initial metaphase spread preparations.

Passage 929393

The appearance of microtubules at sites proximal to chromosomes and in the cytoplasm during mitosis, and their incorporation into spindles, has also been directly observed.

Passage 905041

In cycles without the leeway provided by a G2-phase, these functions are perhaps needed to ensure that mitosis does not initiate until after DNA synthesis is completed (20.

Passage 363332

Etd1 mRNA accumulates periodically, peaking in late G2 / mitosis; this periodicity is reflected in the steady-state level of etd1p / ˙HA (a tagged form of etd1p), which is degraded at the end of cytokinesis (6).

Passage 657420

Synchronized control or Scc4-depleted cells were arrested in mitosis by nocodazole treatment.

Passage 1132527

If chromosomal passengers do not first concentrate at centromeres, it seems they are not able to move on to the central spindle and midbody later in mitosis.

Passage 1361923

Whereas the levels of CDH1 / HCT1 RNA and Cdh1 / Hct1 protein are constant throughout the cell cycle, CDC20 RNA and Cdc20 protein are present only during late S phase and mitosis and Cdc20 protein is unstable throughout the entire cell cycle.

Passage 195362

Dig-labeled Rab27 ORF probe was generated by PCR using the following primers : 5'-TTGACGTTGGCGCCGGTGCA-3', 5'- TGAGCCTCTGCAATTAGCCGGAT-3', labeled with Dig using Klenow (NEB) with labeling mix (NEB), boiled for 5 min to denature, and added to the membrane for hybridization overnight at 458C.

Passage 1480892

AFDM was estimated as the difference between the combined weight of the filter and filtrate before and after combustion at 500 C.

Passage 73814

The control treatment was supplied with the same quantity of heat-sterilised mixed inoculum plus 250 ml of inoculum-filtrate to get a similar quantity of organic matter and bacterial conditions in all treatments.

Passage 2040580

It was considered that the amount of matter filtered through paper with pore size 6 was negligible because the filtrate was clear.

Passage 475863

The fractions were analyzed by SDS-PAGE for protein content, pooled, and concentrated by dialysis (overnight, 48C) against storage buffer (50 mM Tris / HCl, pH 7. 3, 500 mM NaCl, 50 % (v / v) glycerol).

Passage 569372

After whole-cell dialysis of an RBL-1 cell with a pipette solution containing a strong buffer (10 mM ethyleneglycol-bis-aminoethyl)- N, N, N', N'- tetraacetic acid (EGTA)) and the sarcoplasmic or endoplasmic reticulum Ca2 +- ATPase pump blocker thapsigargin (2 µM) to empty stores, ICRAC developed slowly (Figure 1A).

Passage 188035

Second, the spike-generating mechanism could be destroyed by whole-cell dialysis.

Passage 1100921

When constitutively active RhoA protein (3 µg / ml) was included in the patch pipette, the ERG current was reduced to 19 % of control values (n = 5 / 5) within 10 min after beginning dialysis of the cell (Figure 2B).

Passage 977683

In addition, btuB RNA was shown to shift the distribution of radiolabelled Ado-Cbl in an equilibrium dialysis chamber.

Passage 497975

Dialysis with 20 mM 1, 2-Bis (2-aminophenoxy) ethane-N, N, N', N'- tetraacetic acid (BAPTA), which consistently prevented extracellular pH-induced increases in intracellular Ca2 + concentration, failed to prevent activation of the acid-induced currents (Figure 4D and Figure S2D).

Passage 114028

Terminal shoots do not form an abscission zone at their base at any time, and therefore cannot be shed.

Passage 433552

Several studies have demonstrated that targeted traffic of secretory and endocytic recycling vesicles at the ingressing furrow and within the bridge is required for abscission (reviewed in (2) and (3)).

Passage 114049

In addition, we observed a time lag of 3 weeks between the occurrence of drought stress and increased levels of abscission.

Passage 1105241

The mechanism of fruit dehiscence may be relevant to related processes such as abscission, anther dehiscence, as well as other processes in which cell separation is required.

Passage 319586

After abscission, Cindr could further be detected in midbody remnants (Figure S1E).

Passage 658902

The model for lifespan and resistance to starvation stress was $Y = \Delta\mu + L + S + (L*S) + R(L*S) + E$, where $\Delta\mu$ is the overall mean, L is the random effect of line, S is the fixed effect of sex, R is the random effect of replicate vial, E is the within-vial variance, and parentheses indicate nested effects.

Passage 1235193

Experiments 2 and 3 consisted of further true-and false-belief tasks that aimed to expand on and replicate the findings of the first ToM experiment.

Passage 1480039

However, because qPCR assays involved analyses of samples in triplicate (with a positive diagnosis if at least one replicate was positive), whereas nested PCR assays involved a single analysis of each sample, the overall detection rate for qPCR was greater than for nested PCR.

Passage 1692941

Each year the population rise period in a treatment replicate was defined as the period from the occurrence of reproductively active females until peak density of the population had been reached.

Passage 1763709

Mean values for the clipping and control treatments were calculated from the four replicate plots.

Passage 1482648

But, when the metabolic biomass of species is standardized across the replicate webs, the magnitude of direct effects is much more consistent.

Passage 1211718

Thus, the late defect observed in or207ts embryos may represent the null phenotype for cytokinesis.

Passage 1316303

The cyk-1 gene encodes a protein with formin homology (FH) domains that is required for cytokinesis; FH-domain-containing proteins have an essential role in cytokinesis in both yeast and Drosophila (37).

Passage 1211429

We conclude that AIR-2 performs two distinct functions during mitosis, one in DNA segregation and a second, independent function in cytokinesis.

Passage 1098139

Thus, like KNOLLE, HIK appears to play a specific role in plant cytokinesis.

Passage 989855

Cytokinesis in somatic cells, the germ-line and the endosperm of Arabidopsis.
SP…TZLE is required only for the cellularisation of the endosperm syncytium, and not for cytokinesis of somatic or germ-line cells.

Passage 1464901

This was to ensure that fish had sufficient time to evacuate their guts and were unfed for twenty hours prior to oxygen consumption measurements; twenty hours post-feeding has been shown to be adequate for the SDA response to subside (McCarthy 2000; Cutts, Metcalfe & Taylor 2002).

Passage 104670

Implicitly, all ordination techniques assume that all species or species groups react to the same composite gradient according to a common model (Jongman, Ter Braak, & Van Tongeren, 1987).

Passage 545291

In our behavioural study, we used a playback method to understand how elephants might react to a disturbed live beehive.

Passage 1478137

There is an increasing need to understand how populations react to alteration of their natural dynamics by human exploitation.

Passage 420629

On the other, they suggest that males react to jamming with a counterstrategy of jamming avoidance.

Passage 1232385

Seven small proteins, the Sm proteins - named because they react with anti-Sm autoantibodies from patients with rheumatic disease - form a heptameric ring that functions in the biogenesis of the U snRNPs.

Passage 614601

Cells expressing the D110 / 112A mutant were superfused with nominally Ca2 +- free external solution in the presence (130 mM, n = 13) or absence of Na + (TEA + substitution, n = 5).

Passage 203107

Although the formation of dauer larvae is known from many, but not all, free-living nematodes, it is interesting to note that there are many terrestrial nematodes, including Caenorhabditis and Pristionchus, that are predominantly found as dauer stages in the wild (80) and (81).

Passage 1571591

However, our results are at odds with most of the estimates from early mark & recapture studies which support the view that stream salmonids, even when unencumbered by barriers, generally confine their feeding activities within small reaches a few tens of metres in length (Gerking 1959).

Passage 573420

As a consequence, their genome is now only one-fourth to one-third the size of their free-living counterparts.

Passage 1711527

To discriminate the effects of hoard size and PUFA content, we provided groups of free-ranging chipmunks with two equicaloric supplements that varied in PUFA content and compared their torpor patterns with an unsupplemented control group.

Passage 1057105

All of the known elements required to generate CSF activity are present and active at metaphase of Meiosis I, including Mos, MAPK, p90Rsk, and Bub1, and these are essential for suppression of S phase between the meiotic divisions.

Passage 1016206

This suggests that either meiosis can occur under some unknown condition, or the genes have been conserved to function in a modified reduction process or some other process altogether.

Passage 1257808

Genetic analysis suggests that hypermethylation most likely arose either during meiosis or somatically in small sectors of the plant.

Passage 1201366

To determine whether cks-1 (RNAi) embryos can complete meiosis I, we incubated live embryos in the DNA dye DAPI, which normally stains the first polar body but not the second polar body nor DNA inside the embryo.

Passage 370815

EGG-3 co-localizes with MBK-2 both at the cortex of oocytes and in the cytoplasmic foci present after meiosis I.

Passage 433744

Fluorescence is shown 6 s before bleach, immediately after bleaching (dashed region), and at the indicated times after bleaching.

Passage 893979

The first bleach reduced the steady-state level of fluorescence by about 43 %, indicating the existence of immobile sources of fluorescence.

Passage 1190375

Fixing and staining Embryos were dechorionated with 50 % bleach and fixed either for 5 min in a bilayer of heptane : 37 % formaldehyde (for tubulin staining) or for 20 min in a bilayer of heptane : PBS + 10 % formaldehyde (for all other antibodies).

Passage 1092979

Two of these, RGR and peropsin, are thought to play a role in recycling retinaldehyde chromophore following bleach (32) and (33).

Passage 397290

For the examination of the localization of VML-FRP, eggs were collected on yeasted apple juice agar plates, dechorionated in a 50 % solution of Clorox bleach in water, and placed directly into halocarbon 27 oil (Sigma Life Science, St. Louis) for examination and photography under bright-field or fluorescence illumination.

Passage 1020826

To confirm this directly, we treated cultures with 1-5 μM adriamycin, located prophase cells 10-20 min later, and followed them through mitosis.

Passage 1067520

In view of this, the idea that the Chfr-mediated checkpoint monitors centrosome separation can be tested by simply adding nocodazole to tubulin producing cells entering prophase in the presence of well-separated centrosomes.

Passage 1116936

We conclude that other members of this conserved checkpoint pathway are likely to act during meiotic prophase, despite being dispensable for the meiotic prophase, despite being dispensable for the chromosome pairing activity of Chk-2 in C. elegans.

Passage 1166817

The regions near the SPBs displayed the Mtc1-GFP signals in prophase and prometaphase cells.

Passage 1123491

Telomeres also function in meiosis; a characteristic clustering of telomeres beneath the nuclear membrane is observed during meiotic prophase in many organisms from yeasts to plants and humans, and the role of the telomeres in meiotic pairing and the recombination of homologous chromosomes has been demonstrated.

Passage 550469

Although we cannot exclude the possibility that these outcomes obtain from an earlier prophase defect, the simplest explanation is that relative to Mad3, BubR1 has a different, or additional, activity in maintaining sister-chromatid cohesion.

Passage 268735

The timeline shows that Alexa 546-dUTP, when injected late in S phase, will be incorporated briefly and appear in anaphase chromosomes a few minutes later.

Passage 882426

Furthermore, the amount of Fob1 associated with Cfi1 / Net1 did not significantly change as cells progressed through metaphase and anaphase (Figure 6C).

Passage 843407

This number reduces to about 10 MTs in late anaphase.

Passage 1177299

In budding yeast, approximately 300 cell-cycle-regulated genes show peak levels of expression during mitosis, and of this set 113 genes are expressed during late anaphase / early telophase (32.)

Passage 451388

The engrailed-Gal4 (en-Gal4) driver was used to additionally label the posterior compartment of each segment of the epidermis.

Passage 717309

OSM-3 does appear in the middle segment but fails to undergo directed movement.

Passage 634123

The second panel shows a single segment.

Passage 1256177

Any segment of the genome may be cloned using the polymerase chain reaction, and easily manipulated with the spectrum of techniques familiar to the molecular biologist.

Passage 293858

Thus, we assessed the effect of shifting K10 one position downwards into the lipid phase or adding an extra K to the MS segment.

Passage 1039442

Noncleavable Separase Can Cleave SCC1 (A) Sequence alignment of cleavage sites in human separase and known separase substrates in humans (Hs), budding yeast (Sc), and fission yeast (Sp).

Passage 580700

A molecular hallmark of the latter process of apoptosis is activation of a cascade of proteases called caspases which cleave a variety of cellular targets, leading to cell death with distinctive morphological characteristics (2) and (3).

Passage 1039531

Do these findings imply that separase cleavage is required to generate a form of the protease that is able to cleave cohesin?

Passage 945241

The prolonged metaphase I might thus have been a result of persistent sister chromatid cohesion resulting from the inability to cleave Rec8 and resolve chiasmata.

Passage 849030

In these cells, low levels of caspase activity are thought to function cell autonomously to cleave and modify the functions of key cell cycle regulators.

Passage 966533

These results elucidate the mechanism by which mTOR controls the phosphorylation state of 4E-BP1 within mammalian cells : the TOS motif mediates 4E-BP1 / raptor association to drive mTOR-dependent multisite phosphorylation of 4E-BP1, resulting in eIF4E release and cell growth.

Passage 1120008

It is therefore possible that Sub2p functions in conjunction with Yra1p to effect the release of mRNA from transcription site foci.

Passage 1861806

In release experiments such as this, there is always a possibility of longer distance dispersal completely out of the local field.

Passage 575387

To release a food reward from the apparatus, the dogs had to either pull the wooden rod with their mouth or push it down with their paw.

Passage 501650

Does it release nutrients that are then used by the surviving cells to complete spore formation?

Passage 1344301

It seems likely that substantial conformational changes in the SRP54 GTPase domain would be required to liberate the active-site side chains.

Passage 1725923

The large numbers of helpers associated with plural breeding groups may subsequently decrease the importance of paternal care and liberate females from possible constraints against extra-pair matings.

Passage 1325623

Although the possibility that target cell lysis could contribute to pathogenesis had been considered, lysis was assumed to have a more protective role : either CTLs could directly affect intracellular pathogens or these lymphocytes could liberate the pathogens from protective intracellular niches.

Passage 1302508

One demonstrated role of Vpu is to induce the degradation of CD4 molecules complexed with Env in the ER, and thereby to liberate the viral envelope (10).

Passage 899616

Especially at high levels of Hh, additional changes, perhaps involving apposition of Smov ± Cos2v ± Fuv ± Su (fu) v ± Ci complexes and phosphorylation of Su (fu) by Fu, liberate Ci-155 in a modified complex which is better able, alone or together with Su (fu), to enter nuclei and activate transcription.

Corn Shucker Survey B

Passage 1241868

This result, along with the presence of an amino-terminal transit sequence and evidence for import of this FtsZ into yeast mitochondria, suggests that the protein functions to divide the organelle from the inside, much like bacterial FtsZ or the imported chloroplast FtsZ.

Passage 2022775

The southern part of clade 3 is bounded by the Monarch Divide east to Mather Pass in the Middle Fork of the Kings River (Fresno County).

Passage 501871

In Paulinella, there are always two chromatophores per cell that are carefully partitioned between dividing daughter amoebae and subsequently divide at a specific and consistent point in the host cell cycle (12).

Passage 357904

If the migratory divide is twice as old as the first sightings of blackcaps wintering in the United Kingdom suggest, the microevolutionary rates would be lower (Table S3) but still above the median given in (33).

Passage 418606

The photosynthesizing vegetative cells continue to divide.

Passage 1299218

Sex allocation theory (1) predicts that, when this is the case, parents are selected to vary the way they divide (allocate) their resources between the two sexes of their offspring, most fundamentally by varying the sex ratio of offspring.

Passage 1407653

In order to produce a loud, continuous call, the caller must expel the air in its air sacs at a high rate, which rapidly depletes its air supply and results in a short call (Plummer & Goller 2008).

Passage 673085

When they are stressed they expel the algae, causing them to bleach to bone white or, in the case of corals that produce their own pigments, pink, pale blue, and yellow.

Passage 1609935

Females are attracted to male territories and will attempt to expel intruding females as well (Lang & Jaeger 2000).

Passage 1475345

Areas of range overlap may represent low-quality zones that do not merit the defence effort required to expel intruders.

Passage 611503

Macrophages expel the Cn-containing phagosome and subsequently survive and propagate.

Passage 1811468

However, breeders did not expel non-helping auxiliaries, although breeding males often attacked adult auxiliary males visiting a nest for nest-building or incubating.

Passage 1096641

Regulation of the major transitions in the cell cycle, such as G1 / S, G2 / M, and metaphase to anaphase, are increasingly well understood.

Passage 772783

Cases of precocious autosomal separation are defined as those meiotic figures in which there is no main mass of chromosomes defining a metaphase plate but rather two separate masses of chromatin (sometimes connected only by wispy threads).

Passage 806764

However, topoisomerase II activity is definitely required at anaphase onset for chromosome separation because application of topoisomerase II inhibitors in metaphase prevents resolution of the remaining catenations in anaphase, and chromatids fail to separate fully (28), (29) and (30).

Passage 288845

Centromere protein CENP-E is a dimeric kinesin (Kinesin-7 family) with critical roles in mitosis, including establishment of microtubule (MT)- chromosome linkage and movement of mono-oriented chromosomes on kinetochore microtubules for proper alignment at metaphase ((1), (2), (3), (4), (5), (6), (7), (8) and (9)).

Passage 854104

Although cohesin begins to dissociate from chromosome arms in prophase (2), (3), (6), (7) and (8), cohesion normally exists between both centromeric and arm regions of sister chromatids until metaphase.

Passage 585786

If STAT92E also interacts with the Brahma complex (32), it might maintain SxlPe chromatin in an active state, facilitating the restoration of transcription after the 13th mitosis.

Passage 1313455

This phenotype is consistent with arrest prior to mitosis, as was observed when Thr167, the activating residue of Cdc2, was changed to alanine (10), and differs from the phenotypes of either mcs6 or mcs2 null mutants, which arrest with two DNA masses (3) and (4).

Passage 1020596

DNA Damage during Mitosis in Human Cells Delays the Metaphase / Anaphase Transition via the Spindle-Assembly Checkpoint

Passage 682432

Hybrids were killed by inducing a gene contributed by the R strain which prevented chromosome separation during mitosis.

Passage 1067900

On exit from mitosis, spindle disassembly and formation of the actin-based contractile ring are coordinated with chromosome segregation, chromosome decondensation, and nuclear envelope formation.

Passage 226215

Because ClpXP and other AAA + proteases denature a highly diverse assortment of cellular proteins exposing very different structures and stabilities, it is unlikely that evolution has optimized the protease activities for any single substrate.

Passage 1375282

Cells were fixed in 100 % methanol at-20 C for 5 min and then incubated in 2 N HCl for 10 min to denature the DNA, followed by 0. 1 M sodium borate pH 8. 5 for 10 min.

Passage 175357

In brief, infected Prochlorococcus cells were filtered through polycarbonate filters (0. 2 µm pore-size) to separate extracellular phage (filtrate), from host cells containing both host and intracellular phage genomic DNA.

Passage 439781

The solid residues were filtered off, and filtrate was evaporated by a rotary evaporator at 508C under reduced pressure.

Passage 46165

The analyses were conducted at ‘Laboratoire d' Analyses des Sols' of the National Institute for Agronomic Research (INRA, 62000 Arras, France) using French standards (Afnor 1994) according to the following protocols : pH was measured in a water-filtrate suspension with a mass-to-volume ratio of 1 : 2. 5 (NF ISO 10390).

Passage 1261431

Whether the apparent lack of parvalbumin in some fast-spiking cells resulted from a real difference in phenotype, from cell dialysis during whole-cell recordings or from the immaturity of the animals used in these studies - age range 14 ñ 21 days - is unclear.

Passage 1286347

The dependence of cell death on H2O2 was examined by the addition of catalase (in dialysis bags) before the rise in the number of ROS-positive cells (Figure 1b).

Passage 274343

Later, control embryos from the same parents were injected with ~ 25 nl of p150-CC1 dialysis buffer; these embryos showed a normal cleavage pattern (data not shown).

Passage 1271260

The fusion protein was expressed in Escherichia coli and purified under denaturing conditions, renatured by dialysis and concentrated using a 1 ml HiTrap Heparin column (Pharmacia) from which it was eluted with 500 mM NaCl in 10 mM Tris-Cl pH 7. 5.

Passage 885610

They found that crescentin can self-assemble into 10 nm wide filaments after solubilization in 6 M guanidinium and subsequent removal of the denaturant by dialysis, precisely the conditions used to assemble other intermediate filaments.

Passage 1286353

The H2O2 scavenging agents catalase (Cat, 10 mg / ml, added in a dialysis bag with 12, 000 MW cutoff) or 4-hydroxy-Tempo (4HT, 0. 25 mM) and (b) protease inhibitors E-64 (10 μM), chymostatin (60 μM) and AEBSF (0. 5 mM) were added to a 14-day-old Peridinium culture maintained at near saturating light intensity.

Passage 621104

The stability of the intercellular bridge and the abscission step of cytokinesis involve targeted membrane addition close to the midbody region and the establishment of particular protein and lipid domains within the bridge (15) and (16).

Passage 113921

The analysis of almost 30, 000 twig fragments over the course of 3 years demonstrated at least partial control by the tree of the process of abscission.

Passage 216911

Both sets of events are mediated by the process of abscission, wherein programmed cell death produces a weakened abscission zone that enables the separation.

Passage 620965

Rab proteins are major regulators of intracellular transport in eukaryotes (1), and here, we report an essential role for human Rab35 in both the stability of the bridge and its final abscission.

Passage 171801

Upon reaching a diameter of ~ 1 μm (3), a midbody ring (MR) forms to stabilize the intercellular bridge until abscission ((4), (5) and (6)).

Passage 1869838

This leads to the suggestion that there is an upper constraint to the rate at which E. aedis can replicate within host cells.

Passage 377246

In total there were 30 replicate boxes used, and 40 Plodia larvae were tested at each of the three food viscosities.

Passage 1443265

Each experimental replicate consisted of two broods hatching on the same day, and each brood was used for only one replicate.

Passage 1057628

The value Ni is the initial size of the winning subpopulation in the ith replicate population, and we defined fi = Ni | N, because a significant deviation in the initial frequency was when a beneficial mutation was deemed fixed.

Passage 101037

Prior to set aside, the study site of Etzdorf had been an experimental arable field comprising two fertiliser treatments (no nitrogen and 80 kg N per hectare and year) with 10 replicate plots (10v & 10 m2).

Passage 430108

On sharp contrast, we demonstrate here that parasites devoid of secretory organelles replicate normally inside the host cell.

Passage 685528

At cytokinesis, the division plane had tilted 908.

Passage 722183

We thus conclude that the CUL-3-based E3 ligase regulates cytokinesis by allowing MEL-26 to accumulate only at the right time and place.

Passage 172358

Reducing the activities of these components does not affect the rate of furrow ingression or the extent of myosin recruitment, nor does it render the cleavage furrow sensitive to perturbations that unmask stronger cytokinesis defects in other situations ((36) and (37)).

Passage 813456

A second gene Cdc14B also exists, but its role in the central spindle and cytokinesis is unknown.

Passage 1366413

Many key aspects of the process are not fully understood, however, including the positioning of the cleavage plane, the reactions necessary for the assembly of an actomyosin ring, and the coordination of the onset of cytokinesis with the completion of mitosis.

Passage 45506

Other studies demonstrate how tree species react differently to modified environmental conditions.

Passage 1409609

As stated before, we could not test all the dyads and therefore the matrices of percentage of success and latency to react contained missing values.

Passage 394260

Furthermore, the binding of an inverted repeat by its cognate transposase is sequence specific and, for mariner elements at least, the transposases and inverted repeats of even closely related elements do not cross-react (13).

Passage 646180

Note that in the mutant lysates, the Sm proteins fail to react with either SYM10 or Y12 (Figure 2C, compare lanes 1 and 2).

Passage 1380243

This region differs by only one amino acid between zMCM6a and zMCM6b, so the antibody should recognize both forms, but the region is only 55 % identical to mMCM6 and thus the antibody should cross-react poorly with mMCM6.

Passage 1409649

Within-subject correlation between latency to react to the gaze cue and rank difference.

Passage 102975

The model was parameterized to cover the major aspects of a roach' s life under the ′ … / pikeperch-free' scenario from the beginning of the growing season 1998 at day 133 (May) to day 312 (October).

Passage 1449964

However, further research found that mice and rats could be trained to distinguish the MHC-based odours of germ-free animals ((Yamazaki et al., 1990) and (Schellinck et al., 1991)).

Passage 1596320

The availability of free water to reproductive females has been suggested in previous studies, but none of these studies have measured directly the apparent importance of water sources to bats living in natural populations.

Passage 1417508

Activation of components of the immune system can be energetically expensive ((Lochmiller and Deerenberg, 2000), (Demas, 2004) and (Klasing, 2004)), can lead to reductions in other immune components and reproduction ((Zuk and Stoehr, 2002) and (Martin et al., 2006)), and leads to the production of potentially hazardous free radicals (Bertrand et al.)

Passage 1749246

For birds, the presence of mink in the beginning of the breeding season would perhaps cause them to move to other mink-free breeding sites (Martin 1993; Kilpi 1995).

Passage 290102

These specialized chromosome acrobatics are accomplished by substituting meiosis-specific players (i. e., REC8, STAG3, and SMC1?

Passage 464335

This stabilization could be correlated to the accumulation of RacGTP in the cortex overhanging the chromosomes at the end of meiosis I, especially because this accumulation has been shown to be involved in meiotic spindle anchoring to the cortex (8).

Passage 326372

Activation of CDK1 is essential for M-phase entry both in mitosis and meiosis.

Passage 618966

Eight of 22 meiosis I embryos and 13 of 17 meiosis II embryos exhibited ZYG-1 signal at centrioles; the signal was typically weaker in meiosis I.

Passage 772774

Alternatively, it is also possible that the ald1 mutation alters a protein portion that is more critical for meiosis than it is for mitosis.

Passage 377298

In contrast, identical bleach followed by 2 hr dark incubation of cones in Tra-retinas had no effect on the dark current, and the sensitivity recovered to 50 % of its dark-adapted value, from 1. 57 ± 0. 35.

Passage 1145106

For cuticle preparations, lethal embryos were collected from plates similar to those described above and dechorionated using standard bleach treatment.

Passage 1073149

Because the average pre-bleach spectrum of the UVS cones was not fitted well by the ultraviolet visual pigment template, its running average absorbance is displayed.

Passage 936595

For the shmoo experiment, the graph shows data from a single shmoo, together with its pre-bleach image; the areas bleached are indicated by circles.

Passage 630252

Half of the nucleus (G2, prophase, telophase) or of the entire cell (metaphase, bleach boundary perpendicular to metaphase plate) was photobleached, and recovery in bleached regions was followed by time-lapse recording.

Passage 331995

Tiam1-Rac signaling is the first antagonist of centrosome separation in prophase.

Passage 332111

The durations of prophase, prometaphase, and metaphase were determined from time-lapse movies.

Passage 669994

The finding that inactivation of p53 (or the CDK inhibitor p21Cip1) allows Atm -/- mutant spermatocytes to partially overcome the G2 / prophase arrest (32) has been confirmed by some, albeit not all, subsequent reports (112) and (113).

Passage 1212346

A collage of a prophase cell (arrow) and two anaphase cells (arrowheads) with differing levels of Rec8p staining.

Passage 402783

Primary arrest (red) occurs after prophase I, but before GVBD, and is released by MPF activity.

Passage 616879

Significantly, this response only occurs during anaphase through cytokinesis, does not require microtubules, and is independent of spindle orientation, but is dependent on myosin II.

Passage 332099

Moreover, in wild-type mice the proportion of mitotic cells in anaphase was significantly reduced after STLC treatment, indicating an arrest prior to this stage (Figure 3G).

Passage 843607

In budding yeast, chromosome breakage associated with a dicentric chromosome triggers a checkpoint in mid-anaphase that delays cells with a medium-length spindle (26).

Passage 806766

If the mitotic forces and anaphase movement do not give directionality to topoisomerase II, what does?

Passage 311902

FRAP of the RLC-GFP revealed that the amount and rate of myosin turnover in the equatorial cortex dramatically decreases upon the metaphase-anaphase transition and further decreases as constriction of the contractile ring proceeds (Figures 2B and 2C).

Passage 219593

This phenomenon distinguishes CEM and AWB from amphid channel cilia, where OSM-3 and kinesin-II move jointly in the cilia middle segment (Figure 3A) ((14) and (22)).

Passage 1345166

pAT26 (90) was created by insertion of a 0. 7 kb segment of DNA from yheH between the AatII and BamHI sites of pAT12.

Passage 949633

This type of experiment with Tc-otd1 led to a high rate of defective embryos: the most commonly observed phenotype was the loss of all pregnathal head segments, along with the mandibular segment.

Passage 173948

The cell body and initial axon segment (coloured purple, myelin in grey) respond with a graded, intracellular potential, illustrated to the right.

Passage 848613

There, a pattern of double segment periodicity is first generated and then subdivided to yield the final single segment repeat.

Passage 814817

Moreover, besides animal caspases, legumains are the only family from the cysteine protease group that has been shown to cleave substrates adjacent to an aspartate residue (14), (15) and (16).

Passage 1151486

We also propose that ADAM 13 functions to cleave and remodel fibronectin and possibly other molecules involved in directing CNC migration along defined pathways.

Passage 434838

Interestingly, F-TslI is also unable to cleave the intron-containing target (Figure 2C) implying that the intron lies within the cleavage / recognition sites for both I-TslI and F-TslI.

Passage 619260

A previous report (9) has shown that activated caspase-3 can cleave and inactivate human PIP5Ka (homolog of the murine PIP5K?).

Passage 1036599

It is particularly remarkable that some of these divergent Rhomboids are specifically able to recognize and cleave Drosophila targets.

Passage 1457425

Collectively, our findings support the classification of vocal performance as an index signal, with performance scores putatively reflecting the combination of social and physiological factors that contribute to androgen release ((Vehrencamp, 2000) and (DuBois et al., 2008)).

Passage 937729

Among these targets are at least three key components of EGF receptor activation : EGF receptor itself, its ligand HB-EGF and matrix metallo proteinases that release HB-EGF from heparin at the cell surface, allowing it to bind to the EGF receptor and complete the positive feedback loop.

Passage 723147

The segregation of RBC and parasite-derived membrane fragments on release sites contradicts the fusion scenario ((6), (7) and (8)), which should result in all TVN membranes enclosed between vacuolar and erythrocyte membranes after parasite release.

Passage 758964

The average spontaneous miniature excitatory postsynaptic potential (mepsp) amplitude was quantified by measuring the amplitude of approximately 100 ¥? 200 individual spontaneous release events per synapse.

Passage 322534

Dual Role of PT Thyrotrophs in Controlling Seasonal Hormone Release The present study identifies Eya3 as the first common long-day molecular signature for birds and mammals in the PT.

Passage 181609

The minimal force required to unfold the entire straitjacket and to liberate TGF -? 1 is ~ 40 pN.

Passage 354096

View high quality image (685K) The results obtained with both CBL variants suggest that unpaired residues immediately below the loop-distal DCL1 cleavage site reduce the rate of the loop-distal cleavage but are necessary for the efficient loop-proximal cleavage to liberate the miRNA : miRNA* duplex.

Passage 822429

However, irrespective of the mechanism, the binding of Hh protein to Ptc is presumed to inhibit this activity and liberate Smo.

Passage 1055006

Thus the binding of Hh to Ptc is thought to liberate Smo from repression by Ptc.

Passage 1366680

The PCR product was digested with Sa / l and Sphl to liberate a 650 bp fragment.

Corn Shucker Survey C

Passage 2022882

The mountain ridges that separate the headwaters of the South Fork Kings River from the Middle Fork Kings River, from Mather Pass to the Monarch Divide, form the northern border of the range.

Passage 1176426

This mimicked fertilisation and triggered cell-cycle oscillations in the egg, but, unless a centrosome (normally provided by the fertilizing sperm) was co-injected into the egg, the egg would not divide .

Passage 213899

Cells in a hexagonal cell sheet may divide along a number of long axes present in each cell.

Passage 929252

In its wake, cells either differentiate into photoreceptors or divide once before exiting the cell cycle as the larva forms a pupa.

Passage 1241811

Recent findings indicate that the dynamin GTPase helps to divide animal and fungal mitochondria, and that the tubulin-like FtsZ GTPase is involved in division of, not only most bacteria, but also chloroplasts and probably mitochondria of unicellular eukaryotes.

Passage 429747

A separate mechanism was required to divide the outer membranes, and our data suggest that DrpA may have been recruited for this role (possibly from an original role in the endosomal pathway).

Passage 2068602

Spitting spiders are unique because they can forcefully expel a sticky gum from their cheliceral fangs onto prey (Monterosso, 1928; Millot, 1930; Dabelow, 1958; Mcalister, 1960; Bowden, 1991) or predators (Gilbert & Rayor, 1985) from as far as 10 body lengths away (Dabelow, 1958; Mcalister, 1960; Gilbert & Rayor, 1985; Nentwig, 1985).

Passage 261987

After washing in PBS containing 1 % Triton X-100 and 10 % normal goat serum (PBS-T) for 30 min at room temperature, the brain samples were incubated in the same solution and degassed in a vacuum chamber to expel tracheal air with six cycles (depressurize to-70 mmHg then hold for 10 min).

Passage 1380044

Previous studies have identified a correlation between T cells and worm expulsion, with severe combined immunodeficient (SCID) mice and CD4 + T-cell-depleted mice failing to expel worm infections (3); a chronological association between goblet cell hyperplasia and worm expulsion has also been reported (10).

Passage 1628215

Ostrinia nubilalis larvae deliberately expel their frass from the tunnel, and must be at or near the tunnel entrance during this process.

Passage 1368179

Nonetheless, mice deficient for IL-4 have been shown to develop residual Th2 responses (3), (4) and (5) and can expel the nematode Nippostrongylus brasiliensis (6), suggesting that there is a functional equivalent of IL-4 in these processes.

Passage 1177784

The proportion of wild-type and glu88-82 cells in metaphase, anaphase, or telophase as well as the proportion of cells with chromatin bridges are shown.

Passage 907331

Synchronous release of cdc25-22 and pcp1 400-900 cdc25-22 cells from the G2 block shows that pcp1 400-900 are delayed in progression through metaphase (n = 500 for each time point).

Passage 806742

In a meiosis I spermatocyte, one bivalent was detached from a late metaphase I spindle (Figure 5; 0 and 1 min, arrows).

Passage 859203

Our results suggest a model for control of the dynamic interaction of RCC1 with chromatin by phosphorylation during metaphase (Figure 6).

Passage 817987

(2) found that C. elegans embryos depleted of NDC proteins exhibit premature spindle pole separation, as in the KNL phenotype, but interestingly these cells are able to make loose metaphase plates, suggesting some kinetochorevvmicrotubule attachments are present.

Passage 418177

It is known that the CBF3 complex binds to the CDE III region (25) of the centromeric DNA via Ndc10p and Cep3p in metaphase.

Passage 1065302

A timeline of events in Cyclin E mutant and Cyclin E + embryos with times indicated in minutes after a heat shock pulse that induced Rux during G214.

Passage 1190275

We describe here a novel regulatory point in metazoan mitosis that occurs in response to damaged DNA.

Passage 1252800

Sid2p, a homologue of Dbf2p, is localized to both SPBs and the cell-division site at the end of mitosis (14).

Passage 770142

During mitosis in a eukaryotic cell, the mitotic spindle microtubules are connected to chromosomes via a kinetochore, a protein complex that bridges the centromeric sequences on the DNA with the spindle.

Passage 1064261

Thus, mitosis and meiosis are two related but clearly distinct variants of a single process in which related DNA sequences are segregated from one another.

Passage 686567

For example, AAA + subunits of the proteosome in eukaryotes and the Clp protease in prokaryotes denature stably folded proteins ((2) and (3)).

Passage 1395850

The loci were amplified using polymerase chain reaction with the following protocol : denature for 2 min at 95 C, followed by 35 cycles of 30 s at 95 C, 30 s annealing at 58 C (53 C for RSD-142), 30 s extension at 72 C, then a final extension step of 2 min at 72 C.

Passage 107263

Bacteria were harvested from the filtrate by centrifugation (5500g, 30 min) at 4.8C.

Passage 148811

Those species which in-filtrate the vegetation by means of extended rhizomes, and maximize contacts with other individuals of the same or other species.

Passage 23676

The extraction water of FP samples was filtered through a filter paper to remove enchytraeids, and 0. 5 ml of the filtrate was injected into each microcosm to reinoculate microflora and-fauna (37 ± 17 nematodes and 2. 3 ± 0. 6 tardigrades per microcosm).

Passage 1102475

We converted monomeric ATP actin into the AMPPNP form by exhaustive dialysis against multiple changes of AMPPNP-actin storage buffer and polymerized it in the presence of 1 % phalloidin overnight.

Passage 1111219

One of the extracts was incubated with 100 µg / ml Ribonuclease A (Boehringer) for 1 hr at 4 ∞ C, followed by the dialysis step for 3 hr as described (2).

Passage 1127866

mAb 74. 1 to dynein intermediate chain (35) was purified from ascites, concentrated by vacuum dialysis against H-KCl, and microinjected at 4 mg / ml.

Passage 254653

Hyperdilated tubules associated with reduced renal function occur in polycystic kidney diseases, which is caused by mutation in numerous genes, and can also be induced by long-term renal dialysis (138).

Passage 1127864

GST-PAK-CRIB (residues 56 - 272 of PAK1B) was prepared as previously described (12), concentrated by vacuum dialysis against H-KCl (10 mM Hepes, 140 mM KCL.

Passage 1107227

The stimulation of seed germination, the adjustment of seedling growth to soil conditions, the rate and extent of leaf expansion and the timing of vegetative senescence and abscission are a few of the plant processes regulated by ethylene.

Passage 113976

Intra-annual variation Intra-annually, there were two peaks of abscission, the first occurring in May and June, the second much more pronounced one occurring in October and November, amounting to half of all abscised twigs.

Passage 114075

Two types of bottlenecks to water transport could be identified in shoots of old oak trees, namely nodes and abscission zones.

Passage 114128

Ten micrometre cross-sections of nodes, internodes, and abscission zones of 45 1 -, 3 -, and 6-year-old twig segments were analysed.

Passage 114012

Twigs without a terminal bud had significantly more abscission scars than those with a terminal bud.

Passage 433789

Our observation that uncoupling of ARF6 / JIP4 (but not ARF6 / FIP3) interaction with the ARF6iSW mutant interferes with endosome dynamics and abscission argues for different roles for GTP-ARF6 at the midbody.

Passage 159812

For the quantification of microorganisms on the faeces of P. scaber, three replicate populations of ten individuals each were tested.

Passage 2015498

Arcview 3. 2 Geographical Information Software was used to create a computerized replicate of the 2D grid pattern at the crevice and to depict graphically the recorded positions for each individual using a minimum convex polygon procedure.

Passage 242270

To examine this expectation, we established three replicate lines in simple and complex habitats and allowed them to evolve for ten generations.

Passage 1107458

Even in budding yeast, it has been known for over a decade that the eventual inviability associated with an inability to replicate telomeres is also accompanied by genomic instability, as evidenced by a progressive increase in the frequency of chromosome loss (11).

Passage 1557668

The total abundance of Poecilia reticulata in the absence of Heterandria formosa over time for each replicate.

Passage 522010

Although our analysis has emphasized the distinct roles of Moesin and Myosin II during early mitosis, anaphase, and cytokinesis, the orchestration of changes in mitotic cell shape and spindle morphogenesis clearly requires close functional cooperation between the activities of Moesin and Myosin II.

Passage 1157241

A mouse monoclonal antibody, 8E2 (10), shows Survivin (red) to be localized on microtubules throughout mitosis : (j) metaphase, (k) anaphase, (l) cytokinesis.

Passage 737779

In two of the cells treated at 10 µM, cleavage was reinitiated, but the furrows were again unstable and cytokinesis failed.

Passage 737799

Of cells pretreated with 5 - 10 µM ionomycin at the onset of cytokinesis and then treated with 10 µM U73122 within 1. 5- 5 min, 46. 7 % failed cytokinesis by either regression or ectopic fusion, significantly less than for cells treated with U73122 alone (p <0. 01, FET).

Passage 349077

The origin of the dextral and sinistral arrangements are shown during the metaphase (left), cytokinesis (center) and following interphase (right) of the third cleavage cycle of two strains of the snail Lymnaea.

Passage 383757

The Z ring is used for cytokinesis in most bacteria, although some families such as the Planctomycetes and Chlamydia lack FtsZ and therefore must divide using another as yet unknown mechanism (67).

Passage 520260

Results from experiment 1 show that macaque monkeys, similar to 9-to 12-month-old human infants, detect the goal of an observed motor act and, according to the physical characteristics of the context (position of the obstacle), construe expectancies about the most likely action the agent will execute in a given context and therefore react differently to the same goal (object grasping) when accomplished by different means (type of hand trajectory).

Passage 2066344

Thus, it is likely that people perceive the conflict in terms of the loss incurred upon their most “valuable” livestock and react accordingly.

Passage 2064629

Drickamer (1981) has shown that mice selected for early onset of maturation did not react to male pheromones by even further acceleration of maturation.

Passage 1826000

Skuas’ direct numerical response may mean that through avoidance learning by prey (Holling 1966) victims are more vigilant and, hence, react faster when skuas, and hence victims themselves are more abundant.

Passage 1814503

Entering male body condition did not significantly improve the above logistic regression model but there was a higher probability for older males compared to first-year males to react on the playback.

Passage 1391457

An increase in density may increase speed of propagation as it reduces the time to react to nearby departures (Hunter 1966; Hilton et al.)

Passage 893725

Cell-free studies show clearly that wild-type (but not mutant) CERT reconstitutes SM synthesis in the mutant cells.

Passage 1895884

In the ocean, the salmon are free ranging in surface waters for 1-4 years before attaining maturity and returning to fresh water for spawning (Jonsson, Hansen & Jonsson 1991b).

Passage 169104

Consistent with previous studies ((24) and (27)), European and North American free-living populations are highly differentiated (Dest = 0. 162).

Passage 1820677

A significant value for the Spearman's rank correlation coefficient between the proportion alive and free space was taken as evidence of density-dependent mortality.

Passage 173046

Gene Expression to Receptor Activation (A) Stimulation with LTC4 evokes Ca2 + oscillations in Ca2 +- free solution followed by Ca2 + entry when external Ca2 + is readmitted.

Passage 1027499

The basic paradigm for meiosis is simple : match homologous chromosomes; lock them together long enough for centromeres to orient to opposite poles at metaphase I; and then move them to opposite poles at anaphase I.

Passage 1150936

This alternative spindle assembly pathway is common in meiosis (3), consistent with the lack of both centrioles and centrosomes in meiotic cells.

Passage 1027497

In meiosis in male fruitflies, chromosome pairing events do not facilitate genetic exchange, but rather create bivalents that can be sequestered to discrete pockets of the prophase nucleus.

Passage 878246

Crucially, deletion of SGO1 restored meiosis I to mam1 monopolin mutants, which confirms that the Sgo1 protein does indeed have a role in maintaining centromeric cohesion.

Passage 1009492

The formation of the six bivalents was normal in diakinetic oocytes, and the chromosomes were aligned at the metaphase plate during metaphase of meiosis I (our unpublished data).

Passage 969616

Areas of chromatin were photobleached in nuclei prior to mitosis, then the bleach pattern followed through to the daughter nuclei.

Passage 1164132

Embryos to be fixed were washed into a basket, dechorionated with 50 % bleach, washed with water, then fixed in a 1 : 1 mixture of heptane and 37 % formaldehyde for 5 min (or 3. 7 % formaldehyde for 20 min for tubulin staining).

Passage 1912008

At the end of each trial, the vole was returned to its home cage, the number of faecal pellets and urine spots in the arena were counted and the arena was cleaned using a mild bleach solution, rinsed with clean water and dried with a paper towel.

Passage 863512

After subtracting the background (mean fluorescence intensity in the bleached region after bleach) the ratio between mean fluorescence intensity of the bleached circle and the mean fluorescence of the whole cell was expressed as a percentage of the prebleach ratio of these values.

Passage 1264768

To bleach GFP-tagged TAP1 (1 ñ 6) in living cells, a small region of the ER was scanned for 10 sec.

Passage 1221341

Thus, destroying microtubules delays late G2 cells from entering prophase and, as the MI does not increase during this time, existing prophase cells do not enter prometaphase.

Passage 441273

CAPG-2-depleted chromosomes failed to condense and individualize at mitotic prophase, as with depletion of any other condensin II subunit (Figure 6; Figure S5A) (7), (9), (10), (11), (12) and (18).

Passage 585642

DNA-repair mutants activate the meiotic checkpoint during meiotic prophase, which initiates after the formation of the 16 cell cyst, whereas cuff and aub mutants appear to act earlier in oogenesis, given that they already have effects during the mitotic cycles preceding the onset of meiosis.

Passage 883958

In immature oocytes arrested in prophase I, meiotic maturation resumes with the breakdown of the nucleus (GVBD : germinal vesicle breakdown), chromatin condensation, and reorganization of the microtubule network into a bipolar metaphase I spindle.

Passage 859660

Note that HA-CycA accumulates in the nucleus in a gradual fashion, resulting in a concentrated nuclear localization in late prophase.

Passage 1190096

Unlike in cleavage embryos, tubulin remains at the spindle poles, and anaphase always occurs after the delay.

Passage 783398

We assumed that the anaphase rate is the same in all experiments for a given genotype.

Passage 890901

In sgo2 cells, 31 % of anaphase I cells fell into abnormal categories, namely GFP dots were missing from one pole, indicating that all four chromatids had segregated to the other, or three GFP dots were found at one pole and a single dot at the opposite one (3 : 1 segregation).

Passage 1203332

We report here that fission yeast Cut8 protein is required for the normal rate of destruction of destruction-box containing Cdc13 and Cut2 in the nucleus during mitotic anaphase.

Passage 767683

ZEN-4 staining can be observed on the central spindle during anaphase and at remnants ((a) and (b), arrow; arrowhead).

Passage 962325

This segment corresponds to loop 3 in the canonical RBD, which is known to play a role in RNA binding (35).

Passage 237380

Drosophila ears are modified antennae, which consist of two functional units, feather-like hairs (arista) attached to a segment called the funiculus.

Passage 1141099

The p70 sequences flanking these two domains, i. e., an N-terminal segment of 64 - 88 amino acids and a 104 amino acid carboxyterminal tail, are entirely unique to the p70 kinases; each flank contains important regulatory motifs.

Passage 1360853

The S1 helices form the outer layer of the cone, with the S2 helices packed inside enabling them to line the intracellular segment of the pore.

Passage 933299

Sortases are transpeptidases that recognize the carboxy-terminal domain of target proteins and cleave between the threonine and glycine residue of a conserved motif (LPXTG).

Passage 1378434

Biochemical studies have shown that Tolloid can cleave Sog at three positions; Chordin is cleaved, and thereby inactivated, by Xolloid or Z-tolloid at two of these three sites (11), (15) and (17).

Passage 903980

They often function in cascades, with one protease cleaving the zymogen form of another, activating it to cleave the next one in the chain, and so on.

Passage 1151480

First, ADAM 10 and snake venom metalloproteases (SVMPs), the closest ADAM relatives, have been shown to cleave purified ECM components (22) and (23) and their receptors (24) in vitro.

Passage 1126874

Since LF is able to enter resistant cells and cleave a known target of proteolysis, resistant alleles of Kif1C cannot be responsible for a loss of the ability to internalize or activate LF.

Passage 432463

How would endonucleases be pre-adapted to cleave the correct sequences?

Passage 1163350

Thapsigargin is a cell-permeable inhibitor of the calcium ATPase of the endoplasmic reticulum (ER) and effectively prevents regulated calcium release from internal stores (32), (33) and (34).

Passage 961310

In particular, presynaptic overexpression of MUNC13-1 at the Xenopus NMJ causes an increase in transmitter release (12).

Passage 352868

This process is mediated by G protein-stimulated release of Ca2 + from intracellular stores.

Passage 191004

In fact, several other forms of cell communication remain network interactions could proceed by electrical coupling, or by transmitter release due to graded calcium-dependent potentials, or by release of non-classical transmitters like nitric oxide, to name just a few.

Passage 941707

Preventing loss of phospholipid asymmetry or blocking exposed PS with annexin V appears to abrogate both uptake of the cell corpse and release of anti-inflammatory mediators (1.)

Passage 1985915

Parental attendance The presence of females near nests, tracks at the nest site, restoration of damaged nests and obvious signs of nest opening to liberate neonates upon hatching strongly suggest that females attend nests throughout the incubation period.

Passage 1196199

However, the cleavage site is close enough to the membrane surface to liberate the amino-terminal fragment from the membrane.

Passage 713323

Thus, Ric-8A could liberate NuMA, LGN, and / or Gai1GTP to stimulate microtubule force generators.

Passage 1274516

In case 3, different processes of the neuron liberate different cotransmitters, although the targets may have receptors for both.

Passage 1099509

Bacteria use it to generate extracellular pheromones or to liberate transcription factors within the cell (18.)

Corn Shucker Survey D

Passage 422973

This wave did not reach the tip of the leaf, and thus the last segments cleaved from the initial either did not divide at all or divided fewer times than did segments toward the base.

Passage 1581750

A cluster analysis for Pacific-slope data using Bray-Curtis distances and average linkage method (cophenetic correlation = 0. 88) for the 89 points suggests a grouping of sites based on distance from continental divide.

Passage 550531

Female GSCs divide asymmetrically to produce another GSC that remains in the niche and a cystoblast that is displaced away.

Passage 1031148

In terms of the model, cell death results in few normal cells leaving the zone at early cell cycles until the remaining cells in the progress zone divide to repopulate it (Figure 3).

Passage 383664

In particular, Neisseria species divide in two alternating planes, whereas S. aureus divides in three alternating planes (12).

Passage 737115

Mutants enter mitosis precociously and divide as smaller cells than wild-type.

Passage 1909996

Perhaps females can expel the plugs of early-mating males (Koprowski, 1992).

Passage 580752

During oogenesis, the nurse cells transport maternal components to the oocyte, and ultimately expel their cytoplasmic content into the oocyte, after which the nurse cell remnants die.

Passage 1609934

Intrasexual competition is well documented in P. cinereus, where males in some populations compete for high-quality territories and expel other males from those territories (Jaeger 1984; Mathis 1991).

Passage 248571

Following infection, IJs expel their symbiotic bacteria into the host, where it plays a critical role in overcoming the host immune system ((6) and (7)).

Passage 313049

Thus, host cell dehydration diminishes parasite egress from HbAA erythrocytes in hypertonic medium (2) and from the dehydrated HbSS erythrocyte in isotonic medium, likely by preventing the vacuole from reaching the threshold of osmotic pressure required to expel parasites.

Passage 849463

All three of these fusion proteins localized to metaphase chromosomes and compellingly confirmed our observation that mutation of the middle of the protein, such as allele 5, does not abolish localization (Figure 1B, Table 1).

Passage 369426

To accomplish this, we treated metaphase cells expressing GFP-tubulin (LLC-Pk1a (7)) with monastrol to inhibit Eg5 (8) or injected them with p150-CC1 to inhibit dynein (9).

Passage 214331

The representative images show centrosomes at metaphase in one-cell embryos expressing different levels of SPD-2 :: GFP.

Passage 959586

LIN-5 is a large coiled-coil protein essential for many aspects of cell division, including chromosome alignment at metaphase and sister chromatid separation.

Passage 1085680

Numbers are reported as the percentage of metaphase spreads that carry at least one chromosome break for each respective genotype.

Passage 1164062

CYC-BS derails mitosis in mid-anaphase CYC-B antibody stains all metaphase cells but fails to stain cells that have clearly progressed to anaphase, arguing that CYC-B is normally degraded rapidly at the metaphase / anaphase transition (16).

Passage 701154

Cytokinesis during female gametogenesis is uncoupled from mitosis, resulting in cellularization of the multinucleate embryo sac.

Passage 203889

Production of new nuclei by mitosis must therefore match tip growth.

Passage 1179863

How are cells prevented from progressing into mitosis in response to checkpoint signals?

Passage 477849

Cells were defective in the association of mitochondria with the spindle poles during mitosis (Figure 2A): 23 % (5 / 22) of the mutant cells showed spindles with constant attachment to the mitochondrial network compared with 60 % (18 / 30) in wild-type cells.

Passage 1163031

An asterisk identifies (a) the mother cell 45 min before division, (b) the mother cell during mitosis, and (c) the daughter cells 45 min after division (supplementary movie 1, image every min)

Passage 1261080

In that work, solutions containing a small amount of a test protein mixed with varying concentrations of a background macromolecule (polymer or protein) were exposed to conditions - elevated temperature or 30 % ethanol - in which the test protein would denature but the background molecule (if a protein) would not.

Passage 1064659

Slides were heated to 958C for 2 min to denature the probe and chromosomal DNA and then hybridized overnight at 378C.

Passage 1140927

Parathyroid hormone (PTH), a small peptide hormone, is reabsorbed from glomerular filtrate by the proximal tubule and therefore is predicted to be more concentrated in tubular fluid when ClC-5 is inactive.

Passage 73815

Furthermore, 250 ml of filtrate from the original soil containing indigenous soil bacteria was added to all treatments.

Passage 1924281

Before screening, sample solutions were washed through a sieve (25 aperture) and the remaining filtrate diluted in a total of 100 mL dH2O and placed into a sealed plastic jar.

Passage 596174

We also found that this modified band decreased in abundance after dialysis, presumably because of endogenous phosphatases in the extract.

Passage 1357354

At room temperature and patch-clamp whole-cell recording (dialysis of intracellular space with pipette solution), glucose entry across the plasma membrane seems insufficient to have any effect on the cytoplasmic glucose level, which is clamped by the glucose concentration in the pipette solution.

Passage 1157522

In column 4, AP2 was first phosphorylated with 100 μM MgATP for 1 hr at 30 C and the excess ATP removed by dialysis.

Passage 1157529

AP2 was incubated with 1 mM FSBA, and the unbound FSBA was then removed by dialysis.

Passage 1046862

In initial tests, mounds were developed on dialysis membrane and were transferred onto agar containing 2.5 μM LatrA.

Passage 363661

Microvesicle shedding in tumor cells occurs via an actomyosin-based membrane abscission mechanism that is regulated by nucleotide cycling on ARF6.

Passage 621005

In addition, the time required for abscission was strongly increased in the cells that did not become binucleated after Rab35 S22N overexpression or after Rab35 knockdown by RNAi (Figure S3).

Passage 433636

Control cells exhibited the highest number of midbodies 2 hr after release; the number then dropped abruptly after 4 and 6 hr as cells completed abscission (data not shown).

Passage 171850

Although barely detectable at the furrow apex at CR closure (Figure 2A, 00 : 06 : 00), Anillin C-FP was still recruited to the MR during the following 10 / ˙20 min (Figure 2A) and remained there until after abscission (not shown).

Passage 433551

In a final step called abscission, cleavage of the intercellular bridge allows separation of the two daughter cells.

Passage 1689565

The variability we observed was likely because foliage mass was a function not only of growth, but also of subsequent needle abscission.

Passage 1643920

There was some evidence for relative declines in species richness in 0L and 20L treatments of at least 18 species per survey per replicate plot (Tables 1b, d and 2b, d).

Passage 1444623

Controlling for female wing length, replicate, and male wing length there was no significant effect of convergence on female longevity (GLM : W1 = 0. 94, P = 0. 33).

Passage 1553002

Data points and error bars are mean ± SE of eight values (two replicate tanks & four sampling dates).

Passage 66055

Analysis of variance (ANOVA) was used to assess the main effects of defoliation, competition, year, replicate, and their interactions on FB and CB production of grasses.

Passage 443258

To verify the effect of treatment, in-between each replicate colony, cuticular profiles were gathered (15) from pentacosane-and nonacosane-treated workers before and after treatment.

Passage 832489

To evaluate the overall evolutionary conservation of cytokinesis, it will be necessary to include representatives of these divergent branches.

Passage 433555

In addition, the exocyst complex, an ARF6 and Rab11 effector controlling docking of vesicles at the plasma membrane (14), (15) and (16), is also essential for membrane delivery during cytokinesis (10) and (17).

Passage 749916

RNAi of the C. elegans homolog (CE16015) of the DnaJ domain protein BUG7 caused a weak-spindle phenotype during early embryogenesis (31), suggesting a role in spindle organization, whereas RNAi of the Drosophila homolog (CG15081) of POC17 is reported to cause formation of binucleate cells in S2 cells (29), suggesting a cytokinesis defect.

Passage 679610

A histone-GFP-expressing Drosophila S2 cell progressing from metaphase through anaphase and cytokinesis.

Passage 1305319

Second, microtubules somehow control the positioning and assembly of the contractile actomyosin network in cytokinesis (29), single-cell wound healing (this report), and microtubule motors are required for ventral enclosure in C. elegans (35).

Passage 1485255

This ensured that fish were unfed for 20 h prior to oxygen uptake measurements and had sufficient time to evacuate their guts; 20 h post-feeding has been shown to be adequate for the specific dynamic action (SDA) response to subside (McCarthy 2000; Cutts, Metcalfe & Taylor 2002).

Passage 1900798

Amphibians are known to be very sensitive to pond drying and can rapidly react by increasing their development rate (Newman, 1992).

Passage 1316102

Oligonucleotides were gel purified, allowed to react with the N-hydroxylsuccinimidyl esters of the caged carboxyfluorescein, CMNB2AF, (a generous gift of Tim Mitchison, see also (16)) or caged rhodamine green (Molecular Probes) and labeled oligonucleotide was separated from unreacted dye using a Sephadex G-50 column (15).

Passage 104552

The biodiversity parameters that react most unambiguously to changing land-use are plant functional types.

Passage 1402269

Individuals might differ consistently in how flexibly they react to environmental changes ((Rale et al., 2007) and (Dingemanse et al., 2010)) and this consistent variation in behavioural responses could in part be mediated by individual differences in learning ability.

Passage 564218

In contrast, approximately half the miR-277 in cultured S2 cells failed to react with NaIO4 (Figure 3A), suggesting that approximately half of miR-277 is modified at its 3' terminus.

Passage 1085852

Normal subjects react more quickly to a pair of visual stimuli than to a stimulus alone.

Passage 587550

HopI1 expression resulted in a 60 % decrease in the level of the SA-inducible PR1 (PATHOGENESIS-RELATED 1) gene transcript and in approximately 50 % lower free and total SA levels (Figures 6C and 6D).

Passage 1084222

These prolonged signals might represent the kinetochores that are free from microtubules.

Passage 1185868

After unusually lengthy discussions comprising three separate debates, MPs agreed in a free vote to endorse the new regulations by a majority of almost 2 to 1.

Passage 945741

Localized minus end capping in combination with destabilization of free minus ends gives centrosomal microtubules a selective advantage.

Passage 811169

Finally, we show that Mps1 and the mitotic spindle checkpoint are responsible for the developmental cell cycle arrest of the three haploid products of female meiosis that are not used as the female pronucleus.

Passage 1057193

Previously, it was shown that the linkage of the consecutive M phases in Meiosis I and II requires the Mos / MAPK / Rsk pathway and results in only partial APC-mediated cyclin B degradation at anaphase I (12), (13), (14), (24) and (25).

Passage 918580

We have identified novel functions for Sgs1 in meiosis.

Passage 436702

Meiosis I is a unique kind of segregation event because homologs are separated, in contrast to mitosis or meiosis II during which sister chromatids are separated (Figure 1).

Passage 965100

In addition, the Mad2 checkpoint protein is required for accurate chromosome segregation as well as the spindle checkpoint in meiosis I. Homologous pairs of chromosomes have trouble orienting on the meiosis I spindle in the absence of Mad2, but not the checkpoint protein Mad3 or Cdc20p, the target of the spindle checkpoint (M. A.

Passage 960494

Collected embryos were dechorionated in 50 % bleach, washed, and then attached to a glass microsope slide coated with a thin layer of glue (made by extracting packing tape with heptane).

Passage 595547

Percentage of growth of GFP-MreB bundles on either side (arbitrarily referred to as left- or right-hand side of the bleach mark) of the bleached region is shown.

Passage 1068073

Wild-type embryo extracts were made as follows : 0.3 hr embryos were collected and washed; embryos were dechorionated by incubation in 50 % bleach for 5 min; 0. 15 g of embryos were homogenized in 450 µl of extraction buffer (EB) which consisted of 50 mM Tris (pH 8. 0), 150 mM NaCl, 0. 1 % NP-40, 0. 5 mM EGTA, 0. 5 mM EDTA, 1 mM DTT, 1 mM PMSF, 1 : 100 Protease Inhibitor Cocktail (Sigma).

Passage 893991

The bleach reduced the fluorescence by 80 % (images a and b), after which vesicles reappeared in the footprint (c and d).

Passage 377403

We find that, after an essentially complete bleach, the mammalian retina promotes cone dark adaptation at rates comparable to the rates of recovery of cone sensitivity in vivo (11) and (13).

Passage 290155

Clearly chronological age - the shear amount of time spent in prophase arrest - is an insufficient explanation, and in this respect the naturally aged mouse models provide important insight : although both studies report a linear age-related decline in chromosome associated cohesins, the expected increase in aneuploidy is only evident in reproductively senescent females.

Passage 664659

During meiotic prophase, the processes prone to go wrong are the pairing of homologous chromosomes and meiotic recombination.

Passage 1221358

Treating cultures with lumi-colcemid (Figure 3f), a non-functional analogue of colcemid (9), or cytochalasin D (Figure 3g), which inhibits actin filament dynamics (10), also failed to reduce the number of prophase cells.

Passage 1065373

Note that partially disjoined chromosomes appear similar to partially congressed chromosomes and hence are likely to be classified as prophase (see Figure 2L).

Passage 1064412

To determine whether AIR-2 affects the timing of sister kinetochore resolution, we analyzed the distribution of HCP-3 in wild-type and air-2 (or207ts) mutant embryos during prophase, prior to NEBD.

Passage 818795

XRHAMM associates with microtubules during prophase and metaphase, with higher concentrations near the spindle pole.

Passage 805895

Cells containing an HO break arrested in metaphase at the DNA damage checkpoint for 6-9 hr before undergoing adaptation and progressing into anaphase (Figure 3B).

Passage 912895

Many cells in both colchicine-treated and untreated san and deco mutant brains contained overcondensed sister chromatids exhibiting PSCS with no evidence of anaphase movement (Figures 1D and 1H; Table 1).

Passage 984327

Unlike in higher eukaryotes, anaphase spindle insertion is the only known function of dynein in budding yeast.

Passage 1365969

To ensure the accuracy of this process, all pairs of sister chromatids must achieve a state of bivalent attachment prior to their poleward movement at anaphase.

Passage 1103716

Each segment could then have been moved independently by muscles inserted into the next more proximal segment.

Passage 1230650

N-WASP contains a domain that binds PIP2, a Cdc42-binding (GBD) domain, a proline-rich region, a G-actin-binding verprolin homology (V) domain, a cofilin homology (C) domain and a carboxy-terminal acidic segment (A).

Passage 1250284

In this segment, the axoneme is surrounded by a filamentous structure known as the fibrous sheath (14).

Passage 1026939

Later in this unrehearsed (and obviously ill-prepared) segment, CNN went on to inform its viewers that West Nile differs from malaria in part because there is a vaccine for malaria.

Passage 1066248

The histogram shows the average number of gold markers / BR pre-mRNP (black bars) observed in the middle segment of the BR gene (m), in the distal segment (d), and in nucleoplasmic (np) BR mRNPs.

Passage 1230106

UBPs are thiol proteases that cleave isopeptide bonds between two ubiquitin chains or between ubiquitin and another protein (1).

Passage 1151459

Purified ADAM 13 can cleave fibronectin, and tissue culture cells that express wild-type, but not protease-defective, ADAM 13 can remodel a fibronectin substrate.

Passage 963470

By utilizing a universally conserved base pair as a cleavage determinant, I-BmoI accomplishes two tasks : it maximizes the potential to cleave related substrates no matter what arginine codon is present (CGN or AGPur), and it minimizes cleavage at intron-containing substrates, because the G at position-2 is replaced by a T (Figure 2), which is selected against by I-BmoI (Figure 4).

Passage 1104376

The proteolytic region, the light chain, of tetanus toxin (TeTxLC) cleaves synaptobrevin, while the botulinum toxins cleave either synaptobrevin, SNAP-25 or syntaxin.

Passage 1107411

Alternatively, Mus81 may cleave the regressed fork (Holliday junction), followed by recombination-dependent replication restart.

Passage 123456

This suggests two things : first, MukBEF molecules probably open and release DNA in a reproducible manner, possibly starting from one end of the DNA molecule and progressing along its length; second, that the MukBEF complex remains tightly associated with the DNA template, even when large forces (> 60 pN) are applied and the DNA molecule is completely extended.

Passage 425471

As the animal develops, Ca2 + can activate myriad responses, including neurotransmitter release, contraction, energy production, and cell growth and proliferation (1).

Passage 710337

Some mutations that release TGS affect DNA methylation to various extents and with various specificities (ago4, cmt3, dcl3, ddm1, drd1, drm2, hda6, hog1, met1, nrpd1a, nrpd1b, nrpd2a, rdr2, suvh2, and suvh4) ((2), (3), (4), (5), (6), (7), (8), (9), (10), (11), (12), (13), (14), (15), (16), (17), (19) and (20)), whereas other mutations (bru1, fas1, fas2, and mom1) have been reported to have no effect ((1) and (18)).

Passage 10939

In an entertaining way, Eisenberg explains many important ecological concepts, ideas, and hypotheses, eg trophic cascades, top-down vs. bottom-up control, the green world hypothesis, mesopreda-tor release, keystone species, the ecology of fear, the theory of island biogeography, succession, resilience, minimum viable populations, or ecological effectiveness.

Passage 723189

Instead, these two kinds of amphiphiles had the same effect, and thus intrinsic membrane curvature is unlikely to be the main driving force for release.

Passage 875178

We do not yet know whether these three groups of Abg neurons innervate different or overlapping regions of male internal reproductive organs, and it is possible that the ablation of any of these sets of Abg neurons induces the uncoordinated release of male substances.

Passage 802029

Genomic DNA isolated from HT1080 cells stably infected with the described retroviruses at regular intervals was digested with restriction enzymes to liberate telomere-containing fragments, which were detected by Southern hybridization with a telomeric probe, as previously described (26).

Passage 275802

Whereas it is well established that neurons and neuroendocrine cells liberate ATP via classical mechanisms involving Ca2 +- dependent exocytosis of specialized secretory vesicles, non-excitable cell types locally release ATP via non-lytic mechanisms that do not involve obvious exocytosis.

Passage 1274506

In such cases, do neurons that contain multiple transmitter substances release all of their cotransmitters at all of their targets, or can they liberate different subsets of their transmitter complement at different synaptic endings?

Passage 1338978

In contrast, their new study used natural chromatosomes, the positioning of which is defined by the micrococcal nuclease that was used to liberate the chromatosomes in the first place.

Passage 993394

Perhaps, paradoxically, it is the shadow of Al Qaeda and those elusive envelopes of anthrax bacilli that will now liberate the money and talent needed to combat the depredations of man' s ancient unicellular enemies.

Corn Shucker Survey E

Passage 1777569

This exercise reveals that existing measures divide into a few basic groups, each of which captures a different facet of spatial turnover in the identities of species.

Passage 1614604

This two-dimensional ordination appeared to divide sites into clusters primarily by an interaction between sampling year and prairie size along the x-axis and by sampling year along the y-axis.

Passage 824838

These embryos still failed to divide (Figure 3R), indicating that the presence of midzone microtubules was not sufficient to allow the completion of cytokinesis.

Passage 880638

How do cells move, feed, grow, divide, secrete, anticipate sunrise, find mates and avoid dangers?

Passage 276976

At high density and / or when nutrients are depleted, these cells may divide to form the more elongated caulonemal cells, which in turn produce bud initials that will develop into leafy gametophores rooted by brown-pigmented rhizoid filaments (13).

Passage 1955134

Some large orthopteran taxa do not autohaemorrhage, for example king crickets (Anostostomatidae), but do expel vile-smelling liquid faeces which may have a similar role.

Passage 611469

One reasonably clear example is highlighted : The macrophage appears to expel a yeast cell (at 0 : 40), which initially remains attached to the cell surface, but by 05 : 19, the Cryptococcus cell is clearly extracellular and floats away.

Passage 1344190

Conversely, mice deficient in the Th-1-promoting transcription factor IRF-1 expel N. brasiliensis even faster than do wild-type controls (17).

Passage 1544494

The fledglings that remain at home are the dominant brood members, and they actively expel the subordinate brood mates from the parental territory 3-6 weeks after fledging.

Passage 141590

In the lab mussels were placed in pans of lake water to filter and expel for 2 days, with water refreshed every 24 hours.

Passage 1380062

Recently, it has been reported that animals in which components of the IL-4 signaling pathway have been disrupted, notably Stat6 -/- mice and IL-4 receptor (IL-4R)-/- mice, also fail to expel gastrointestinal worms (3).

Passage 1149853

The centromere region is also a site of strong cohesion; in higher eukaryotes it is the last point of contact between sister-chromatids at metaphase prior to their separation at anaphase.

Passage 995861

As cyclin B was associated with spindles (but the chromosomes had not congressed to a metaphase plate), we concluded that these cells were in prometaphase.

Passage 902190

Spc25 Is Required to Maintain Metaphase Chromosome Alignment

Passage 806676

A metaphase I chromosome (0 min, arrow) was manipulated with one needle pulling on one homolog (1 min, arrow) and the other pulling the other homolog in the opposite direction (1 min, arrowhead).

Passage 962715

In contrast, the simultaneous loss of pins and bazooka activities results in the formation of a small symmetric spindle at metaphase, which is rather similar to the basal half of the wild-type spindle (Figures 2H and 2K).

Passage 809860

Time is after G2 release; n = nuclei which fill with NIMA-GFP as they enter mitosis; p = nuclei which display NIMA-GFP at the nuclear periphery.

Passage 783361

In human cultured cells, DNA damage also induces a delay in mitosis (7), (8) and (9).

Passage 1142887

Cdk2 activity was higher during mitosis (Figure 1a, 30-40 min) and declined thereafter, consistent with previous reports (4).

Passage 209602

In this model, newly synthesized CenH3CID becomes available for deposition when, at mitosis, CenH3CID is no longer regulated by SCFPpa.

Passage 626514

The SPB signals were invisible in the early to middle stages of mitosis of Dis1WT and Dis16E, suggesting that Cdc2 phosphorylation restrains the association of Dis1 with the SPBs in metaphase.

Passage 1283082

A synthetic peptide corresponding to residues 208-225 of human mitosis-inducing phosphatase Cdc25c was used to assay the activity of 6His-Cds1 purified from Sf9 cells in the presence or absence of caffeine.

Passage 428720

Nucleotide-free kinesin-1 is known to denature or misfold over time (10), and we interpret the KIF3A / B inactivation as an analogous process that has a greater nucleotide sensitivity.

Passage 1620459

We filtered a known fraction of the filtrate through 47 mm GF / F silica filters (pore size = 0.7 μm).

Passage 77111

HNF showed almost the same chemosensory response at 60 min to the 0. 22 m filtrate of Ps-B2 bacterial surface compounds (Ic = 3. 09) as with the whole-cell resuspension of Ps-B2 (Ic = 2. 87) (Fig.

Passage 1294087

The cells were then trypsinised, neutralised in 0. 5 % serum-containing medium, filtered through 100 µm meshes (Cell MicroSieve, BioDesing Inc.) and the filtrate passed through 35 µm meshes.

Passage 175359

The qPCR primers used to measure DNA in the filter and filtrate fractions are listed in Table S1.

Passage 1100960

Peak ERG currents elicited by steps to + 20mV after 1 min (open) and after 10 min (shaded) dialysis of Rac1 (61L) from the patch pipette into the cell.

Passage 596175

Furthermore, the presence of phosphatase inhibitors in the dialysis buffer prevented the decrease of this modified form of CEP-1 (Figure S3C).

Passage 569429

Dialysis with 5 mM succinic acid, citric acid, oxaloacetic acid, or a-ketoglutaric acid were all much less effective (Figure 3B).

Passage 1138167

In vitro telomerase reactions Extract preparation and in vitro telomerase assays were performed as described previously (48), except that fractions from the DEAE-agarose column were desalted and concentrated with a vacuum dialysis apparatus and Colloidion nitrocellulose membranes (Schleicher and Schuell).

Passage 648436

Dialysis of the mercaptoethanol and exposure to air reverses the shift, as is observed for other flavodoxins (Figure 3J).

Passage 216929

Studies of mutants in which the abscission zone and / or replum and associated tissues are reduced or eliminated have allowed for the characterization of half a dozen genes that regulate formation of these structures (6).

Passage 114072

Cladoptosis, the abscission of twigs, is the main mechanism of changes in crown structure in senescing pedunculate oak (Quercus robur L.).

Passage 216922

Tomato fruits are shed by breakage of the pedicel (p) at an abscission zone (orange) which is absent in the jointless mutant.

Passage 113935

An altered anatomical zone that differs from the surrounding tissue and is located at the base of twigs, and the abscission layer, the distal portion of the abscission zone, which forms during the abscission when the zone becomes meristematic.

Passage 695019

Consistent with this similarity, vesicles accumulate at the midbody prior to abscission in C. elegans, and abscission requires secretion (3), leading to speculation that abscission in animal cells may be similar to cytokinesis in plants.

Passage 1313677

In the in vitro simian virus 40 (SV40) replication system, minichromosomes that contain HMG-17 replicate more efficiently than those without HMG-17 (11); HMG-17 may enhance the rate of replication of a chromatin template by unfolding the higher-order chromatin structure.

Passage 107524

Three replicate sites per phase were selected.

Passage 1818788

For five species, only one replicate was obtained in the acceptable size range (Table 2).

Passage 78481

Within the experimental area at each site, we created six replicate, randomly located blocks, spaced about 50 m apart from each other.

Passage 1758781

Hence, the Racconigi site could not be used as a red - grey replicate for site RG1 and demonstrates the difficulties of site replication in experimental field studies.

Passage 1801499

This pattern of change in territory size was the same on all four replicate plots, though the difference was not significant in all years.

Passage 1295612

Note the single spot of cdc12p in two interphase cells (upper panel) and staining of the contractile ring during cytokinesis (lower panel).

Passage 1096384

This patch localizes exactly at the anterior pole of the pIIa cell (t = 4 : 06 : 02) and is redistributed apically to the pIIIb stalk after cytokinesis (t = 4 : 11 : 12).

Passage 1211445

Cytokinesis in wild-type, zen-4 and air-2 mutant embryos.

Passage 1160626

Thus, the MEN is crucial for the regulation of cytokinesis, as well as mitotic exit.

Passage 172328

Although these data reveal that CED-10 / Rac1 inhibition is not the primary function of the CYK-4 GAP domain, mutation of ced-10 / Rac1 and inhibition of ARP-2 / 3-dependent actin nucleation allow cyk-4 (or749) embryos to complete cytokinesis.

Passage 2002770

Gilbert (1986) demonstrated that surfacing dytiscids evacuate their subelytral air store through ventilatory movements of the abdominal tergites, and that these ventilatory movements only allow the exchange of approximately two-thirds of the subelytral reservoir.

Passage 708881

We do not merely react to the outside world, we predict the state of the outside world based on previous experience.

Passage 1455361

Previous playback experiments show that blue tit males react aggressively towards male song.

Passage 1618259

Species with longer incubation periods might develop higher levels of B-1 cell diversity, and with a more diverse B-1 cell repertoire, a greater number of B-1 cell lines are likely to recognize and react to auto-antigens.

Passage 1588439

We consider two alternative hypotheses : consumers are constrained to adjust sufficiently to the lower trophic level, or prey species react more strongly than their predators to reduce predation.

Passage 5753

How ecological communities react to species extinctions is a long-standing yet current question in ecology.

Passage 1209399

The arrowhead indicates a vesicle that was free from acridine orange.

Passage 2011220

In the marine environment, boats may be thought of as introducing artificial habitat structures (albeit temporary, ephemeral ones) into killer whale habitat that may impede locomotion, or increase perceived risk associated with locomotion by making whales more conspicuous to human, predation-free predators.

Passage 1791914

In studies marked with an asterisk, the infectious unit was the individual free-living infective particle, whereas the remaining studies used the infected host cadaver, for which transmission coefficients were estimated by Dwyer & Elkinton (1993).

Passage 1217756

(b) CHO-K1 cells transfected as in (a) were incubated for 18h in medium containing 10 % serum, after which the cells were transferred either to serum-free medium or maintained in serum-containing medium for an additional 12h.

Passage 627517

Cells were then perfused with Ca2 +- free intracellular solution plus 100 μM EGTA for 6 min before being subjected to the 100 μM Ca2 + test pulse.

Passage 1109315

Separase is a conserved CD clan cysteine protease known to be required for separating sister chromatids during both mitosis (10) and meiosis (22) in ascomycetes.

Passage 1201432

By analogy with the mitotic defects described below, we think it likely that cks-1 (RNAi) embryos fail to exit meiosis normally.

Passage 1361130

The chromosomes are monitored throughout this first phase of meiosis by the recombination checkpoint system, which detects the presence of unresolved recombination intermediates (5).

Passage 1249005

In fission yeast (Schizosaccharomyces pombe), the subnuclear positions of centromeres and telomeres change during vegetative growth and in the early stages of meiosis.

Passage 712990

At the onset of meiosis, actin (blue) forms long “fishnet-like˜ polymers surrounding the nucleus, while the spindle microtubules (green) form their bipolar structure, and the chromosomes (red) are scattered in the nucleus.

Passage 1073148

Each graph shows the average pre-bleach spectra (upper traces) with best-fitted visual pigment templates (solid lines) and average post-bleach spectra (lower traces) with their running averages (solid lines).

Passage 377380

Primate cone response amplitude and sensitivity recovered substantially after the bleach.

Passage 377379

Cone ERG responses from primate retina, dark adapted in the absence of pigment epithelium after enucleation in bright light (left) and after subsequent bleach and 1 hr dark incubation (right).

Passage 522336

Individual clusters were imaged and then illuminated either with a brief exposure to the Argon (488 nm) laser to bleach the GFP fluorescence (green arrowheads) or more prolonged exposure to the DPSS (561 nm) laser to excite Alexa 594 fluorescence (red arrowheads).

Passage 894014

Immediately after the bleach there was a steep fluorescence gradient (Figure 3, trace b), and this relaxed over a period of tens of seconds (Figure 3, traces cv ± e).

Passage 890488

Two prophase cells are shown as examples.

Passage 1109158

The ‘prophase pathway’ is largely, if not completely, absent in the yeast S. cerevisiae, and, as a consequence, the bulk of Scc1 remains on chromosomes until the metaphase to anaphase transition, whereupon it is cleaved by separase (10).

Passage 441278

At the first embryonic mitosis, capg-2-depleted chromosomes fail to condense into distinct rod shapes during prophase (left).

Passage 1261844

When both centrosomes are destroyed during prophase these cells form a functional bipolar spindle.

Passage 878739

Decades of cytological observation have suggested a mechanistic link between the chromosome and / or telomere clustering that occurs during early meiotic prophase and the onset of homolog alignment.

Passage 660849

Histogram shows the percentages of mitotic cells that had progressed from prophase to anaphase onset within 60 min, within 60-120 min, within 120- 240 min, and those that had required more than 240 min.

Passage 1023387

In addition, it has been reported that there is a slight delay in anaphase onset in cells overexpressing the centromere-targeting domain of MCAK (9).

Passage 311897

Additionally, expressing a constitutively phosphorylated RLC drives filament formation and premature localization of myosin to the mitotic cortex, but these filaments cannot localize to the equator until after anaphase onset, when the rest of the cytokinetic machinery is targeted to this zone.

Passage 298967

Changes in kinetochore behavior leading to Bub1, BubR1 and Mps1 re-recruitment confirm that kinetochores experience reduced tension at anaphase, and that as a minimum the removal of Aurora B from centromeres serves the purpose of preventing Bub1, BubR1 and Mps1 re-recruitment (and, possibly, re-activation).

Passage 824762

During anaphase, the spindle deviated an average 2. 58 ± 1. 748 (mean ± standard deviation) from linear (four embryos), whereas the spindle of spd-1 (oj5) embryos deviated an average 11. 88 ± 11. 058 (three embryos).

Passage 340827

They also show the highest frequency (about 20 %) of sub-lineage M2 (17) and (23), which has the highest nucleotide diversity within a fast evolving segment (HVS1) of the mtDNA compared to other sub-lineages.

Passage 696573

Misalignment, where it does occur, is always in phase each en-expressing cells in a given segment aligns with cells of comparable positional identity within the segment as it approaches from the opposite side of the embryo.

Passage 1088109

Thus, overexpression of the C-terminal segment of NORE that binds MST1 in vivo is sufficient to suppress almost completely the proapoptotic action of Ki-RasG12V and Ha-RasG12V, E37G.

Passage 534698

The several hundred thousand neurons per segment are housed within cytoarchitectonically defined anatomical layers called laminae.

Passage 1016064

Although their study on Abd-B is suggestive, the authors' general claim that other Hox genes could also be involved in the maintenance of segment boundaries, requires more extensive studies to be appropriately supported.

Passage 296673

If we assume that cells take 5 min to methylate newly synthesized DNA and that the time taken for MutH to cleave DNA near mismatches is exponentially distributed, the mean time that a MutL focus would be associated with a mismatch destined to be successfully repaired would be <70 s, assuming that MutS and MutL detect mismatches and become microscopically visible as soon as the DNA has replicated.

Passage 1314166

Apoptosis involves an intracellular proteolytic cascade, mainly mediated by members of the caspase family of cysteine proteases, which cleave one another and various key intracellular target proteins to kill the cell neatly and quickly (4).

Passage 409599

In addition, active caspases cleave off the Mst carboxy-terminal regulatory domain.

Passage 1214935

(20) hypothesized that, if Sir2p has ADP-ribosyltransferase activity, then the protein might cleave the glycosidic bond between nicotinamide and ADP-ribose in NAD.

Passage 1314202

To confirm that some caspases are activated during normal epidermal differentiation, we made cytosolic extracts of normal foreskin epidermis and tested their ability to cleave the fluorogenic peptide caspase substrate z-Asp-Glu-Val-Asp-aminotrifluoromethylcoumarin (zDEVD-AFC), as previously described (9).

Passage 844633

In both rats and mice, prolactin release is induced by the copulatory stimulus, and thereafter occurs in two daily surges that are roughly coordinated to lights-on and lights-off, suggesting that release is under circadian control and is therefore vulnerable to gene mutations that alter the circadian pacemaker (20), (21) and (22).

Passage 745168

On the basis of preceding observations showing that a correlate of visual perception is impaired in situations where dopaminergic transmission is decreased (Table 1), and knowing METH's ability to induce dopamine release (1), we hypothesized that we should be able to rescue the compromised visual response in these transgenic flies at the restrictive temperature by exposing them to METH.

Passage 1356810

Second, the actin cytoskeleton also controls the polarisation of T cells towards antigen-presenting cells, a process that is critical for efficient contact between antigen-presenting cells and T cells and for the directed release of cytokines and other effector molecules.

Passage 1106340

The cells were then shifted to 32 ∞ C in the presence of cycloheximide to allow a synchronous wave of release of the VSV G. Cells were fixed in paraformaldehyde at (a ñ

d) 0 min, (e ñ h) 30 min, and (i ñ l) 90 min after release, then stained for cell surface VSV G using an antibody specific to the extracellular domain of the viral glycoprotein.

Passage 396655

Although Pds1 was degraded within 20 min of release from the spindle checkpoint arrest in GLC7 cells, it was delayed up to 40 min in glc7-10 cells despite normal kinetics of Mps1 repression (Figure 2A and data not shown).

Passage 713641

Although the effect of PI3K on ecdysone release is likely due to its effect on PG size, it is less immediately clear how altered Raf activity regulates ecdysone release and body size.

Passage 1180251

Mutations or treatments that interfere with myosin II function would liberate more Lgl for SNARE-binding function, resulting in the rescue observed in embryos carrying maternally persistent Lgl protein.

Passage 983561

We and others identified Dynamin 2 (Dyn2) as a ubiquitously expressed isoform that can mediate multiple endocytic processes (10) and (11), liberate vesicles from the Golgi (12), and act as a modulator of the actin cytoskeleton (13), (14) and (15).

Passage 899709

In similar vein, apposition of Smo-associated complexes has not been demonstrated but might serve in Hh signaling to liberate Ci from inhibition by both Cos2 and Su (fu).

Passage 738447

These observations for MCAK have led to a proposed mechanism in which the MCAK / tubulin complex detaches from the MT, and ATP turnover is used to liberate the motor from the tubulin heterodimer ((14), (15), (16), (17) and [18)).

Passage 1283192

But most iniquitously of all, one of the missionaries of the new field has stated that it will liberate us from the ì shackles of hypothesis-dominated biological research.

Alarm Clock Survey (passage indices omitted)

Annual U and p estimates and standard error estimates for female Weddell seals in Erebus Bay, Antarctica, obtained from program mark.

All ponds were fish-free, but varied substantially in area, permanence and vegetation composition, although all were above ultra-basic serpentine geology on heathland / unimproved grassland.

Based upon the detection of song from one whale 29 kilometres from the devices, the researchers calculated that they were able to monitor sounds over an area of 2, 600 square kilometres.

The main objective of this research was to observe the occurrence of female sex cues and the effect they have on male nuptial gift construction in P. ornata.

Similar results were obtained using either a Myc epitope-tagged Drosophila epsin or a chimeric GST-epsin (Xenopus) in conjunction with HA-epitope-tagged Ub (HA-Ub).

The timing of Fuller' s study may have been premature to detect an effect.

All other flies were obtained from stock centers or previously published (see Supplemental Experimental Procedures).

Individual vesicles from which measurements were made (see below) are circled and labeled a, b and c. (b) Sequential images of a single vesicle observed after electrical stimulation.

The question of whether variation in species evenness or in identity of the dominant species in uneven mixtures (species abundances) influences the functioning of species combinations has received less attention.

There has been considerable interest in how these different mechanisms may operate in diverse enemy assemblages, and how positive or negative interactions among species may determine the range or direction of diversity effects.

However, any association between Tea3p and Tea1p may be fragile, because even centrifugation of cells is sufficient to delocalize proteins from the cell ends (see the Supplementary Material).

Recruitment may be indirect, because we did not detect physical association between PI4KIIIa and purified WNK11 / _491 in pull-down assays.

At each site, all burrows, ponded water, vegetation, and debris were searched for mitten crabs along a 5 m transect parallel to the stream channel; the height of these transects encompassed the distance between the low water line and the high tide mark along the bank.

F box proteins are components of E3 ubiquitin ligase complexes that add polyubiquitin tracts on selected lysine residues and thereby mark a protein for proteasome-mediated degradation.

These simulations produced a bimodal fluorescence distribution similar to that seen for PFUS1-STE4 cells, with weak positive feedback causing delayed inactivation after pheromone removal and stronger positive feedback causing an increasing fraction of persistently active cells.

We chose this statistical approach because it has higher statistical than t-tests and individual two-way anovas performed on each predator combination, the methods typically used to detect nonadditive effects of multiple predators.

In addition to these biophysical changes within individual pattern-initiating neurons, the temporal relationship between their burst activities was also modified by operant conditioning.

In a similar way, we use the light ñ dark cycle as a primary Zeitgeber, but in addition we can be driven voluntarily by our clocks and watches (see Figure 2a): daylight-saving time in the winter provides these timers with just one hour' s adjustment to the changing photoperiod, twice each year.

A robotic device applied a mechanical load to the jaw as subjects repeated words – saw, say, sass, and sane - appearing on a video monitor.

Observer bats in this study presumably had the opportunity to eavesdrop on search and approach phase echolocation calls, feeding buzzes and chewing sounds to learn about the location and nature of the tethered mealworm.

Apis florea dancers are often observed following the dances of other bees in between their own bouts of dancing.

Because Tremblaya has the smallest reported bacterial genome, it is surprising to observe an inversion in its genome.

It follows that if horizontal migrations are being made by animals that are at least partly dependent on soil moisture content, these may be associated with permanent water bodies, as typical for many amphibian species.

One can also expect to observe beetle density accumulation near the southern edge of a habitat, if the accumulation is only partially compensated by additional density-dependent movement.

Instead, we focused on bird species that were either frequent-exclusive or near-exclusive to each woodland type.

Dispatch Circadian biology : the physiology of inner retinal photoreceptors.

View high quality image (198K) Importantly, after exposure to the new contingency, all participants saw a majority of probe trials consistent with the rotation direction contingent with EVM during exposure - that is, the learning effect was highly significant (see Supplemental Data).

At this genomic location, there exists a string of approximately 30 thymidine residues.

This combination may allow them to communicate with conspecifics without changing the body colours that provide camouflage.

Within this investigation however, whether hummingbirds do not make use of colour because of overshadowing, floral variability or some other cause, this experiment highlights a case where an animal does not make use of what would appear to us to be one of a resource’s most obvious attributes, according to how we see the world.

A small proportion of the embryos (about 1 %) had one or two internal cell walls but clearly differed from kn or keu embryo.

We suggest that the smaller size in A. notoscriptus adults reared in predation cues is likely to limit their reproductive output and perhaps their ability to disperse, but see the study by Maciel-De-Freitas, Codego & Lourenco-De-Oliveira.

Relationships between the FIRE gradient and indices of bird community average niche position for 551 1 km2 squares in Catalonia (NE Iberian Peninsula).

They base their optimism on what was accomplished by such inspirational leaders and they emphasize the need for a change in attitudes because they see consumption and fundamental lifestyle choices as an underlying environmental problem that cannot be addressed by legislation.

To assess internal branch strength, ML quartet puzzling support values as well as MP bootstrap, jackknife and decay values were determined (see Supplementary material).

We thank John Dixon and Gordon Stott for helpful discussions, and Mark Carlton and Aaron Zorn for comments on the manuscript.

In fact, using the same logic that investigators apply to , accept, ontogenetic scaling, one can argue that the opposite is true.

Variation in brown trout biomass was best explained by a model including presence or absence of Arctic charr, ice breakup date and the interaction between Arctic charr and ice breakup date.

Finally, in contrast to most mammalian dispersal patterns the ultimate cause of dispersal in both male and female red squirrels is thought to be resource competition: thus, their dispersal patterns should be affected by resource patchiness or variability.

As with panneural expression , coexpressing gal and Rbf1 with mutant tau did not substantially ameliorate retinal toxicity.

As depicted in Figures 1C / _1E, PGCs expressing the various mutated CXCR4b forms were exposed to SDF-1a, which was provided by endodermal cells engineered to express high levels of the chemokine (see Supplemental Experimental Procedures).

To assess whether male courtship is dependent on the identity of the female he is courting, day-to-day correlations of courtship effort were compared between set-ups 1 and 3.

Quantitative analysis of the data (see Supplementary material) revealed that high percentages of cultured organs cycled rhythmically with a period in the circadian range of 22 - 25 hours.

Our new Bayesian tree confirmed the earlier reported convergence between bats that use constant frequency (CF bats) echolocation (horseshoe and leaf-nosed bats) and dolphins.

As we shall see, these kinds of auxiliary interactions can be used to make activation of a given gene dependent upon more than one physiological signal, and to make sensitive switches.

This vertically centered visual field, of just below 100 deg, closely matches Snell' s window (the 97 deg circular window through which an underwater observer can see the entire 180 deg of the terrestrial world compressed by refraction as the light passes through the water surface.)

See explanations in (d) for Gr5a and rp49.

However, searches of the corresponding genome sequences from humans and C. elegans, and of EST collections failed to detect candidate missing exons.

Naturally, these two tests are not independent of each others and hence the P-values in Table 2 should be compared to Dunn-corrected P-limits.

In inside-out patches excised from control transfected cells, we did not observe channel events, even at millimolar concentrations.

Remarkably, pups display the typical search and locate response to a glass rod when it is coated with 2MB2.

The results showed that English speakers were faster to locate a target from a different category to distractors (for example, a green target amidst blue distractors), than from the same category, even though the degree of physical difference of targets from distractors was equated.

Limpets may be constrained to return along an axis of travel to locate their home scar after foraging, while the variable spatial distribution of resources around the scar may be inconsistent with our assumption of a constant standing crop per unit area (see Thompson et al.

However, UVA photoreceptors are not sensitive enough to detect ambient UVB radiation in nature.

Whales began clicking at an average range of 295-539 m from the depth of the first recorded buzz, which is well within the theoretical range over which sperm whales should be able to detect squid based on the properties of their regular clicks.

Combinations of auditory, olfactory, and tactile cues were essential for males to detect rivals, and any combination of two of these three cues resulted in equivalent responses.

A new study has shown that silver Y moths migrate only on nights when winds would displace them southward, implying that they detect their direction of movement while airborne, likely by a magnetic sense.

Specifically, do male chimpanzees continue to monitor cycling females during food-poor times?

But along with these potentially undesirable aspects of fMRI is the great advantage of being able to monitor many thousands of voxels at once, throughout the entire brain.

Human leaders not only initiate group action but also motivate, plan, organise, direct, monitor, and punish to achieve group action.

We used the Observer 3. 0 (Noldus Information Technology, Wageningen, the Netherlands) as an event recorder, which allowed us to continuously monitor on which patch each crab was foraging.

This suggests a task-dependent flow of information between prefrontal and occipital cortices, where the regions communicate longer with each other in the case of more ambiguous facial expressions requiring integration such as "happy", whereas for easy expressions such as "fear" much less communication would be required.

How do the SIN and the contractile ring communicate?

Infant macaques solicit their mother' s affiliative responses and actively communicate to her.

Among social insects, it is well known that CHC function as intraspecific cues to communicate, for example, nestmate membership, task or fertility status.

Five wasps were then removed carefully blindly from the selected watch-glass using a fine paintbrush.

We used the diffuse light entering the hut through its translucent roof to observe the bees during the day, and we used red-filtered lights to watch them at night.

Condition 2 : food and caching trays versus an empty cage To determine whether observers preferred to watch the trays and food themselves, we performed a nonsocial version of condition 1, in which no birds were placed in the side cages.

When coseeded on coverslips and fused by brief (30 s) incubation in polyethylene glycol (Roche; 50 % vol : vol), the Cos and 3T3 cells fuse and fluoresce red.

Since scorpion cuticles fluoresce green under UV light, and the median eyes respond maximally to green and secondarily to UV, we hypothesize that scorpions, fluorescence may be related to their perception of light.

The non-treated cyprids when exposed to Concanavalin A had auto fluorescence but the attachment disc of the third antennular segment did not fluoresce (Plate 1e).

When a gentle touch stimulus is applied to the anterior region of the worm, the corresponding anterior mechanosensitive neurons expressing the cameleon under the mec-4 promoter fluoresce in response to an increase in intracellular Ca2 +.

Infant 3, like the others, exhibits a drop in intermembral index near the 5-month mark.

White arrowheads indicate the trailing edge of a fiduciary mark on an elongating actin cable, which is moving in a retrograde manner.

Commercial antimouse IgG-specific antibodies were tested in order to find out whether they cross-react with bank vole immunoglobulin and could be used for assaying vole plasma samples.

Species with longer incubation periods might develop higher levels of B-1 cell diversity, and with a more diverse B-1 cell repertoire, a greater number of B-1 cell lines are likely to recognize and react to auto-antigens.

As expected, subjects were quicker to react when the informant produced an obvious head movement in addition to the movement of the eyes.

Individuals might differ consistently in how flexibly they react to environmental changes and this consistent variation in behavioural responses could in part be mediated by individual differences in learning ability.

Whereas future archaeological work may reveal such earlier sites, the genetic situation may be explained by the fact that most of Island Southeast Asia has been re-populated recently by Austronesian-speaking migrants, who either displaced the initial populations or mixed with them.

Error bars represent 95 % confidence intervals across participants calculated separately for each data point of the running average (see Supplemental Experimental Procedures).

The two domains enable synaptojanins to hydrolyse PtdIns (4, 5) P2 to PtdIns without producing another signal, PtdIns4P, en route.

Ideally, an investigation of cohesin' s role in transcription should aim to observe the immediate consequences of the complex' s inactivation in cells that are neither undergoing mitosis nor replicating their DNA.

Note that this independence is within only an observation period, as we did not observe the same individual for consecutive intervals.

The browser view supports this conclusion because we observe a continuous distribution of reads throughout the locus that is also consistent with the histone methylation patterns observed by chromatin immunoprecipitation in embryonic stem cells.

Surprisingly, we observe that, in khc mutant oocytes, bcd mRNA is not tightly concentrated to the anterior cortex but is diffusely spread out in a wide cortical ring that expands toward the posterior.

Therefore, experimental studies conducted in natural ponds are necessary to quantify species interactions in more complex environments, thereby providing a means to gauge the effects of mesocosm or laboratory studies.
How do bees gauge flight direction?

We sampled each microcosm every 2 days to gauge ciliate population dynamics.

Top, accelerometer; top middle, target strain gauge; bottom middle, intermediate strain gauge; bottom, eye position.

When this happens, birds would rather be influenced by an intrinsic tendency, such as personality, because of a lack of experience, than try to learn to adjust to the unfamiliar situation.

Thus, the LMAN may not simply be a developmental remnant in the adult, but serve a broader behavioral strategy that permits an old bird to learn new tricks.

Songbirds that learn to produce and recognize complex vocalizations are commonly used in laboratory studies of natural sound processing, vocal learning and vocal motor production.

Indeed, it is very difficult to understand the mechanisms by which one individual helps another to learn if, in fact, we do not know that that is indeed what it is doing.

The largest subgroup of these receptors, those underlying olfaction, have evolved to detect and distinguish small volatile molecules from each other, and any anesthesiologist knows that halothane and isoflurane smell different!

Shifts in signal chemistry as a result of olfactory mutations may also occur in other insects in which pheromone production depends on the sense of smell.

For example, recent work with domestic horses indicates that individual recognition in these species is cross-modal, meaning that receivers have a complex, multidimensional representation of known individuals that includes information on idiosyncratic features of their vocalizations, and what they look and / or smell like.

To understand how smell is transformed into behaviour, we must understand the set of transformations from odours in physico-chemical space to odours in the successive neural spaces of different layers in the brain and eventually to perceptual space.
We therefore encourage the use of simulation procedures, which are now widely available and commonly used in other research areas, as routinely exploratory tests for any studies on animal dispersal, especially for those obtained through capture-mark-reencounter data.

Consistent with our hypothesis, 10 of the 12 subjects tested (83 %) reported smelling the same smell again (e. g., marker) when the bottles were swapped, but not when the bottles were not swapped.

Although this antibody recognized AurA at centrosomes and spindle poles in Xenopus tissue culture cells by immunostaining, it failed to detect AurA at the poles of spindles induced by RanGTP in Xenopus M phase egg extracts.

APPENDIX D. CLASSIFIER FEATURE SETS

This appendix lists the all the features in the two feature sets used for text classification.

Word Feature Set

Number of Tokens
Average Token Length
Max Token Length
No. Nouns
No. Verbs
No. Adjectives & Adverbs
No. Conjunctions
No. Unrecognized Tokens
No. Non-Alpha Tokens
cel
cleav
act
metaphas
protein
absciss
divid
seg
reg
filt
form
reply
anaphas
releas
bleach
react
cytokines
min
mut
embryo
process
prophas
effect
funct
tim
dialys
lib
result
gen
mitos
rol
sit
suggest
chromosome
expel
requir
shown
sign
study
domain
level

meios
observ
progress
show
specy
tre
yeast
chromosomes
control
cyc
homolog
mal
meiot
rat
sep
zon
addit
complet
dat
diff
direct
express
fail
high
increas
intracellul
lat
loc
membr
mm
produc
reduc
respons
spindl
wat
analys
condit
expery
fluoresc
fre
hum
init
involv
mitot
ml
pres
proteas
remain
sampl
bud
cal
checkpoint
complex
con
cont
defin
den
develop
dna
fem
host
hr
includ
inhibit
mech
midbody

mod
model
nucle
numb
part
period
pol
posit
report
simil
siz
solv
acid
antibody
arrest
arrowhead
bas
bridg
chromatin
consist
day
demonst
ess
fin
find
group
im
import
individ
interact
lead
mat
med
mic
molec
occ
onset
opposit
pair
pipet
pop
poss
prevy
pri
prim
rel
reproduc
ring
singl
smal
spec
step
target
test
vary
abs
accum
ag
align
anim
appear
approxim
bacter
body
break
cent

deplet
detect
ear
er
ev
evid
expos
forc
found
ftsz
fu
grow
hel
id
ident
immedy
inact
incub
induc
lay
maj
mass
microtub
middl
nat
nematod
neuron
norm
not
pattern
peak
phosphoryl
plant
plat
pot
prev
red
rna
separas
sequ
stabl
stain
subst
telom
tot
transcrib
twig
un
wash
year
adam
affect
air
alt
amino
amount
arrow
aspect
assembl
assum
av
background
bind
br
cascad
caspas

cellul
chain
clos
clust
comp
compon
conserv
contrast
cre
cur
cytoplasm
dark
daught
dea
dech
degrad
delay
dens
determin
dist
dot
driv
due
dynam
ecolog
ent
entir
entry
exhibit
expuls
ext
extract
fila
fix
fork
gfp
glucos
grad
gtpase
half
hypothes
infect
insid
intercellul
king
larva
leaf
lin
long
lys
maintain
meas
medy
minim
mitochondr
mot
mov
nest
nm
nuclear
org
origin
overexpress
parasit
pass
patch

perc
plan
play
proport
propos
proxim
pul
ratio
reach
receiv
record
remov
rev
riv
segreg
sensit
set
sist
sm
standard
stat
stim
su
subsequ
termin
transit
transmit
tubulin
worm
abg
abund
actin
ad
adapt
altern
analog
anatom
app
area
argu
assay
assocy
asterisk
atp
atpas
bact
bar
begin
biv
block
bond
bound
brasiliens
brood
buff
calc
cas
centrom
chang
charact
chromatids
col
collect
combin
comput
conclud

consid
contain
continu
contribut
coordin
correl
cult
cyclin
dau
declin
decreas
defect
dehisc
dehydr
dext
dispens
drosophil
ecdyson
el
en
end
endoplasm
endosom
endosperm
envelop
erg
erythrocyt
est
estim
evap
ex
exit
expand
extend
extern
extracellul
facilit
fact
fig
fiss
flank
food
formaldehyd
frag
frequ
fruit
fsba
fus
genom
grady
graph
gre
green
grey
hept
hormon
hour
ii
immun
imply
in
ind
independ
inform
inject
insert

intens
interest
intern
invert
jam
key
kinetoch
larg
las
leng
liv
loss
low
main
mam
man
mark
match
math
mcak
microcosm
miss
mix
monarch
monoclon
moth
mrna
mukbef
multipl
needl
ni
nod
nor
nucleotid
num
nutry
obtain
occur
oocyt
op
ordin
organel
ory
out
pack
pap
paramet
pathog
pathway
pcr
ph
phag
phas
phenotyp
pheromon
photobleach
phys
point
polym
por
prec
prey
prometaphas
protect
protocol
pur

qpcr

quant

random

rapid

recogn

recov

recovery

recruit

regress

rep

repres

residu

reticul

sa

secret

shed

shmoo

shoot

sid

sin

som

stor

stress

sufficy

supply

surround

task

telophas

temp

ten

term

thapsigargin

thought

tip

transfer

transport

transposas

tunnel

turnov

twenty

typ

ubiquitin

undergo

unicellul

unpubl

valu

veget

vesic

victim

view

weak

wing

work

Noun Feature Set

Number of Tokens
Average Token Length
Max Token Length
No. Nouns
sign
observ
progress
direct
fail
increas
lat
remain
fem
part
import
lead
occ
pair
poss
ring
spec
deplet
expos
id
inact
induc
peak
pot
affect
assembl
av
dark
delay
driv
dynam
hypothes
insid
lin
lys
mitochondr
mov
org
overexpress
pass
pul
record
segreg
stat
stim
termin
abg
abund
actin
adapt
area
assay
assocy
asterisk
atp
atpas
bar
begin

biv
block
bond
brasiliens
brood
buff
calc
cas
centrom
charact
chromatids
col
correl
cyclin
dau
defect
dehisc
ecdyson
el
end
endosperm
erg
erythrocyt
est
exit
extend
fact
fig
fiss
food
formaldehyd
frequ
fsba
fus
grady
grey
hept
hormon
immun
inform
las
leng
man
match
math
microcosm
monarch
moth
mrna
mukbef
ni
nor
nucleotid
nutry
occur
oocyt
ordin
organel
out
pap
pathog
ph
phag
phas
phenotyp
pheromon

point
por
prey
prometaphas
protocol
qpcr
quant
random
recovery
regress
residu
reticul
sa
secret
shmoo
shoot
sid
stress
task
temp
term
thapsigargin
tip
transport
transposas
tunnel
turnov
twenty
typ
ubiquitin
vesic
victim
view
wing
work
expel
mal
express
loc
cal
cont
den
posit
report
arrest
consist
prim
test
align
ev
evid
forc
hel
plat
stabl
tot
adam
air
amino
amount
arrow
aspect
background
bind
br
cascad

caspas
clust
compon
contrast
daught
dea
dens
dot
entry
expuls
ext
fila
fork
glucos
gtpase
king
larva
leaf
nm
parasit
patch
perc
proport
ratio
receiv
riv
set
sist
sm
su
transmit
tubulin
worm
drosophil
envelop
frag
hour
ii
intens
loss
mcak
pathway
pcr
rep
reduc
hum
pres
acid
bas
find
im
interact
vary
abs
ag
bacter
body
break
er
fu
lay
mass
microtub
middl
nematod
neuron

pattern
phosphoryl
plant
sequ
telom
twig
year
chain
extract
gfp
mot
plan
sensit
transit
stor
valu
react
addit
bud
checkpoint
complex
con
host
period
pol
antibody
arrowhead
chromatin
day
group
med
mic
molec
onset
pipet
step
target
ftsz
grow
rna
separas
subst
transcrib
homolog
sep
diff
analys
expery
proteas
sampl
hr
inhibit
mech
model
numb
bridg
mat
pop
control
produc
respons
wat
fluoresc
ml
dna
midbody

nucle
siz
solv
divid
result
dat
mm
condit
chromosome
yeast
chromosomes
cyc
rat
spindl
gen
sit
tre
zon
membr
filt
bleach
domain
level
specy
funct
rol
study
embryo
meios
act
reply
mut
tim
dialys
cleav
reg
effect
releas
min
process
mitos
form
prophas
anaphas
protein
cytokines
metaphas
seg
absciss
cel

APPENDIX E. TEXT CLASSIFICATION CODE

The following pieces of code are the main codes that have been used in this work. Some small modifications have been made to these codes for various tasks, but those modifications are not retained.

MAIN CODE

```
'''
This code covers the overall structure of the filtering engine
It imports files with the sentence data and human ratings.
The output should be a rule list, with support and confidence for rules with NLP metrics
causing good or bad fits
'''


'''
The new CollectData funciton is used to import data from csv files. A list of files can be
given to combine their results for processing. The data file should have the first
two rows containing indices and sentence texts for all the collected data. The following
rows should contain TRUE/FALSE responses for each sentence, each row representing
responses
from a single participant

With imput of the n sentences and m participants responding, input has the form
```

```
Ind_1    Ind_2    ...   Ind_n
Sent_1    Sent_2    ...    Sent_n
TF_1_1    TF_2_1    ...    TF_n_1
...
TF_1_m    TF_2_m    ...    TF_n_m

The function returns an organized list of the data from the files in the form:
( [ (Ind_i,Sent_i) ] , [(Ind_j, TF_j_k)] )
So that the sentence list will have n entries and the response list will have n*m entries.
Also, TF_j_k should have values 1/0 rather than TRUE/FALSE as in the initial data
'''


def CollectData(datafiles):
    table_in=[]
    #import csv
    for datafile in datafiles:
        readfile=csv.reader(open(datafile,'rb'))
        for line in readfile:
            table_in.append(line)
    Results=zip(*table_in)
    num_responses=len(Results[0])-2
    SentenceSet=[]
```

```
    for column in Results:
        SentenceSet.append((int(column[0]),column[1]))
    ResponseSet=[]
    for column in Results:
        for response in range(num_responses):
            if column[response+2] == "TRUE":
                trans_response=1
            else:
                trans_response=0
            ResponseSet.append((int(column[0]),trans_response))



    return (SentenceSet, ResponseSet)




'''
The QuantifySentences Module takes a list of (index,sentence) tuples and outputs a list of [index,sentence,metric1,metric2,...] lists
The metrics used are specified in the function as a list of functions
'''
```

```
def QuantifySentences(Sents):
    from SentenceCharacterizerV2 import Characterization
    (SentMet,MetList)=Characterization(Sents)
    '''SentL=[list(entry) for entry in Sents]
    import FilterMetrics
    MetricNames=[FilterMetrics.M1, FilterMetrics.M2, FilterMetrics.M3]
    for metric in MetricNames:
        metric(SentL)'''
    return (SentMet,MetList)


'''
The CombineData function is used to combine entries from the human data (good/bad rankings) and the sentence metrics. It returns an orange table data set of the form (index, good/bad, m1, m2, ...)
'''


def SaveVectorizedData(HData,MData,MetList):
    DataSet=[]
    for point in HData:
        flag=True
        for S in MData:
            if S[0]==point[0]:  #finds a matching index
```

```
            DataSet.append(S[2:]+[point[1]])
            flag=False
      if flag:
         print "There was an error matching the data"
         print point
   VData=[MetList]+DataSet
   f=open('VectorizedData.csv','wb')
   csvw=csv.writer(f)
   for row in VData:
      csvw.writerow(row)
   f.close()


def LoadVectorizedData(VecDataFile):
   f=open(VecDataFile,'rb')
   csvreader=csv.reader(f)
   datain=[]
   for line in csvreader:
      datain.append(line)
   MetList=datain.pop(0)
   Metrics=[Orange.feature.Continuous(x) for x in MetList]
   isgood=Orange.feature.Discrete('good/bad',values=['0','1'])
   domain=Orange.data.Domain(Metrics,isgood)
```

```
    data=Orange.data.Table(domain,datain)

    return data


def CombineData(HData,MData,MetList):
    DataSet=[]
    for point in HData:
        flag=True
        for S in MData:
            if S[0]==point[0]:  #finds a matching index
                DataSet.append(S[2:]+[point[1]])
                flag=False
        if flag:
            print "There was an error matching the data"
            print point

    #import Orange
    Metrics=[Orange.feature.Continuous(x) for x in MetList]
    isgood=Orange.feature.Discrete('good/bad',values=['0','1'])
    domain=Orange.data.Domain(Metrics,isgood)
    data=Orange.data.Table(domain,DataSet)
```

```
        return data



def MineData(data):
    #import Orange
    #nb = Orange.classification.bayes.NaiveLearner(name="Naive Bayes")
    #adjusted_nb = Orange.classification.bayes.NaiveLearner(adjust_threshold=True,
name="Adjusted Naive Bayes")
    #svml=Orange.classification.svm.SVMLearnerEasy()
    SVM=Orange.classification.svm.SVMLearner(nu=0.6,gamma=0.1)
    kNN=Orange.classification.knn.kNNLearner(k=60)
    NB=Orange.classification.bayes.NaiveLearner()
    Rule=Orange.classification.rules.CN2UnorderedLearner(alpha=0.95, beam_width=4)
    Learners=[SVM,kNN,NB,Rule]

SVMPlay=[Orange.classification.svm.SVMLearner(nu=0.7,C=1.0,gamma=.1),Orange.cl
assification.svm.SVMLearner(nu=0.75,C=1.0,gamma=.1),Orange.classification.svm.SV
MLearner(nu=0.8,C=1.0,gamma=.1),Orange.classification.svm.SVMLearner(nu=0.85,C
=1.0,gamma=.1)]

kNNPlay=[Orange.classification.knn.kNNLearner(k=60),Orange.classification.knn.kNN
```

```
Learner(k=50),Orange.classification.knn.kNNLearner(k=40),Orange.classification.knn.k
NNLearner(k=30),Orange.classification.knn.kNNLearner(k=20)]

NBPlay=[Orange.classification.bayes.NaiveLearner(m=1),Orange.classification.bayes.N
aiveLearner(m=2),Orange.classification.bayes.NaiveLearner(m=3),Orange.classification
.bayes.NaiveLearner(m=4),Orange.classification.bayes.NaiveLearner(m=5)]
    results = Orange.evaluation.testing.cross_validation(Learners, data,folds=10)
    return [results]

def OutputRules(classifier):
    #import Orange
    f=open('Rules.txt','w')
    for rule in classifier.rules:
        line= Orange.classification.rules.rule_to_string(rule)
        print line
        f.write(line)
        f.write('\n')
    f.close()
    return None
```

```
saved_data=1

if __name__=='__main__':
    import time
    import csv
    import nltk
    import Orange
    t=time.clock()

    if not(saved_data):
        #Specify the file names for input data
        SentFileName='Sentences.txt'
        DataFileNames='Results.txt'
        DataFiles=['StatSurveyData.csv']
        VectorSentData='VectorizedData.csv'

        #Given the file names, the CollectData funciton returns an ordered list of
(SentenceIndex,SentenceString) tuples
        #as well as a list of the human ratings, posed a list of (SentenceIndex, Rating)
tuples where rating has the form good/bad (1/0)

        (Sentences,HumanData)=CollectData(DataFiles)
```

```
print 'Data Collected'
print (t-time.clock())/-1
t=time.clock()



#The QuantifySentences function takes the list of Index/Sentences form the data collection, and returns a tuple with a list of measured sentences [[index, sent, m1,m2,...]] and a list of metrics [MetricName1, MetricName2, ...]

(OrderedMetrics,MetricList)=QuantifySentences(Sentences)
print 'Sentences Quantified'
print (t-time.clock())/-1
t=time.clock()

#SaveVectorizedData(HumanData,OrderedMetrics,MetricList)




#The CombineData funtion merges the human data with the metrics to create a data set for mining
DataSet=CombineData(HumanData,OrderedMetrics,MetricList)
print 'Data Combined'
```

```
        print (t-time.clock())/-1
        t=time.clock()



        DataSet.save('SavedProcessedData.tab')
    else:
        DataSet=Orange.data.Table('SavedProcessedData.tab')
    print 'Data Loaded'

    #The mining module takes a data set and uses the Apriori algorithm to generate a set of association rules
    [Results]=MineData(DataSet)
    print 'Data Mined'
    print (t-time.clock())/-1
    t=time.clock()



    #The OutputRules function prints the rules to the command line and saves them to a text file
    PPV=Orange.evaluation.scoring.PPV(Results)
    Sens=Orange.evaluation.scoring.Sensitivity(Results)
    FScore=Orange.evaluation.scoring.F1(Results)
```

```
   for item in range(len(PPV)):
      if not(PPV[item]):
         PPV[item]=100
   for item in range(len(Sens)):
      if not(Sens[item]):
         Sens[item]=100
   print "%10s  %5s      %5s        %5s" % ("Learner", "PPV", "Sensitivity", "F Score")
   LearnerNames=['SVM','kNN','NB','CN2']
   kNNNames=['5','10','25','50','100']
   NBNames=['1','2','3','4','5']

   Names=LearnerNames
   for L in range(len(Names)):
      print "%10s: %5.3f        %5.3f       %5.3f" % (Names[L], PPV[L],
Sens[L],FScore[L])

   cm=Orange.evaluation.scoring.confusion_matrices(Results)
   print "%10s  %5s  %5s %5s %5s" % ("Learner", "TruePos", "FalsePos", "FalseNeg",
"TrueNeg")
   for L in range(len(Names)):
      print "%10s: %5.0f    %5.0f    %5.0f    %5.0f" % (Names[L], cm[L].TP, cm[L].FP,
cm[L].FN, cm[L].TN)
```

```
        #OutputRules(FullRules)
    print'Finished'
    print (t-time.clock())/-1
    t=time.clock()


if __name__=='__main2__':
    import csv
    import nltk
    import Orange
    DataSet=Orange.data.Table('SavedProcessedData.tab')
    SVM=Orange.classification.svm.SVMLearner(nu=0.1,gamma=0.125,C=1.0)
    Classifier=SVM(DataSet)
    print 'classifier built'
    """
    #Now get the new data to evaluate
    (Sentences, bogusData)=CollectData(['AlarmPossibles.csv'])
    print 'file loaded'
    (OrderedMetrics,MetricList)=QuantifySentences(Sentences)
    print 'file processed'
    NewDataSet=CombineData(bogusData,OrderedMetrics,MetricList)
    print 'data compiled'
    NewDataSet.save('AlarmPassageData.tab')
```

```
    """

    NewDataSet=Orange.data.Table('AlarmPassageData.tab')
    Classifed=Classifier(NewDataSet[0])
    print 'finished'
```

SentenceCharacterizerV2

```
'''
This is the workhorse of the program. This code does the natural language processing
tasks to
characterize sentences. To save resources, it is a monolithic function, rather than being
broken up.

The input is a list of sentences, with indices from the corpus; [(index, sentence)]
The code iterates through every list entry and returns a new LIST with the metrics
attached to the end,
as well as a list with labels for every metric.
([[index,sentence, metric1, metric2, ...,
metricN]],[Metric1Name,Metric2Name,...,MetricNName])
'''
```

```
#NOW WITH CRUNCHY VECTORIZED SENTENCES

def Characterization(SentenceList):
    import nltk
    Processed=[]

POSList=['CC','CD','DT','EX','FW','IN','JJ','JJR','JJS','LS','MD','NN','NNS','NNP','NNPS'
,'PDT','POS',

'PRP','PRP$','RB','RBR','RBS','RP','SYM','TO','UH','VB','VBD','VBG','VBN','VBP','VB
Z','WDT','WP','WP$','WRB']

    NounTypes=['NN','NNS','NNP','NNPS','PRP','PRP$','WP','WP$']
    VerbTypes=['VB','VBD','VBG','VBN','VBP','VBZ']
    AdjAdvTypes=['JJ','JJR','JJS','RB','RBR','RBS','WRB']
    ConjTypes=['CC','IN']

    MetricList=['Number of Tokens','Average Token Length','Max Token Length','No.
Nouns','No. Verbs','No. Adjectives & Adverbs',
                'No. Conjunctions','No. Unrecognized Tokens','No. Non-Alpha Tokens']

    import Text2Vector
```

```
StopFile='/Users/Glier2/dropbox/workspace/BioCorpusFilter/RuleGenerator/SMART_Stopwords.csv'

WordFile='/Users/Glier2/dropbox/workspace/BioCorpusFilter/RuleGenerator/DataWordList.csv'
    (stems,stops)=Text2Vector.GetWordsAndStops(WordFile,StopFile)

    MetricList=MetricList+stems

    for entry in SentenceList:
        #First the sentence is tokenized and taged
        SentStr=entry[1]
        TknSent=nltk.word_tokenize(SentStr)
        TagSent=nltk.pos_tag(TknSent)
        SentPos=list(zip(*TagSent)[1])

        #Next some basic word staticsics are computed
        #Number of tokens
        NumTokens=len(TknSent)

        #Average token length
```

```
AvgTknLen=1.0*(len(SentStr)-SentStr.count(' '))/NumTokens

#Maximum token length
MaxTknLen=max([len(word) for word in TknSent])

#Grammar Counts
Nouns=0
Verbs=0
AdjAdv=0
Conj=0
Unrec=0

for pos in SentPos:
  if pos in POSList:
    if pos in NounTypes: Nouns=Nouns+1
    elif pos in VerbTypes: Verbs=Verbs+1
    elif pos in AdjAdvTypes: AdjAdv=AdjAdv+1
    elif pos in ConjTypes: Conj=Conj+1
  else:
    Unrec=Unrec+1

#Count number of non-Alpha tokens
```

```
        NonAlpha=[word.isalpha() for word in TknSent].count(False)

        WordVector=Text2Vector.VectorizeStemmedText(SentStr,stems,stops)

Processed.append([entry[0],entry[1],NumTokens,AvgTknLen,MaxTknLen,Nouns,Verbs
,AdjAdv,Conj,Unrec,NonAlpha]+WordVector)


    return (Processed,MetricList)




#Test Block
if __name__ == '__main__':
    import nltk

    Sentences= [(1,'Four score and seven years ago, our forefathers brought forth on this
continent a new nation, conceived in liberty, and dedicated to the proposition that all
men are created equal.'),
                (2,'When in the Course of human events, it becomes necessary for one people
to dissolve the political bands which have connected them with another, and to assume
```

```
among the powers of the earth, the separate and equal station to which the Laws of
Nature and of Nature\'s God entitle them, a decent respect to the opinions of mankind
requires that they should declare the causes which impel them to the separation')]
    CharSent=Characterization(Sentences)
```

Text2Vector

```
'''
This module contains 2 functions: a word list builder and a text vectorizer.
The word list builder has three modes to give varying outputs. The funciton takes an
input text file,
a stopword csv file, and an ouptut mode declaration (defaults to word stems). The text is
processed to ouptut a list of the for [(word_i),(count_i)]
for i unique "words" in the document. All words are contain ONLY alpha characters

The other function takes a wordlist as from the first function and a document and
produces a two lists: a list of words (a key to the vector),
and a list of word occurances (the document vector).

'''
```

```
# First a quick function to input the texts. This can be time consuming, so it is broken
out for convenience
def ReadFiles(input_texts):
    raw_text=''
    for text in input_texts:
        textfile=open(text,'rb')
        raw_text=raw_text+textfile.read()
        textfile.close()
        print 'File Loaded'
    raw_text=raw_text.lower()
    return raw_text


'''
The BuildWordList function takes a string, a location for a stopword file, and an ouput
mode, and an option create a lemma dictionary and outputs a list
containing a two tuples, one of all the "words" in the string, and the other of the
occurances of each "words". The word list can be a list of all alpha
tokens in the text, alpha tokens not including stopwords, or word stems not including
stopwords.
'''
```

```
def BuildWordList(input_string, stopwordfile,
output_mode="Stems",lemma_dictionary='no'):
    import nltk
    import csv
    import time
    t=time.clock()
    #First retrieve the text file and tokenize it
    tokenized_text=nltk.word_tokenize(input_string)
    del input_string

    #Get the stopwords from the stopwordfile
    stopwords=[]
    with open(stopwordfile,'rb') as csvfile:
        userinput=csv.reader(csvfile)
        for row in userinput:
            stopwords.append(row[0])



    print "Files loaded and tokenized"
    print time.clock()-t
    t=time.clock()
```

```
#Now build a word dictionary with occurances, ommiting non-alpha tokens
word_dict={}
print len(tokenized_text)
chunk=int(len(tokenized_text)/20)
count=0
while count<len(tokenized_text):
   if not tokenized_text[count].isalpha(): count=count+1
   elif tokenized_text[count] in word_dict:
      word_dict[tokenized_text[count]]=word_dict[tokenized_text[count]]+1
      count=count+1
   else:
      word_dict[tokenized_text[count]]=1
      count=count+1
   if count%chunk==0: print int(count/chunk)*5   #Makes a percent complete counter!
del tokenized_text


print 'Text dictionary built'
print len(word_dict)
print time.clock()-t
t=time.clock()
```

```
    # Now make a word list, either with stopwords, without stopwords, or stems without
stopwords
    if output_mode=="With_Stopwords":
        word_list=word_dict.items()
        word_list.sort()
        word_list=zip(*word_list)
    elif output_mode=='Without_Stopwords':
        for word in stopwords:
            if word in word_dict:
                del word_dict[word]
        word_list=word_dict.items()
        word_list.sort()
        word_list=zip(*word_list)
    elif output_mode=="Stems":
        PS=nltk.stem.PorterStemmer()
        LS=nltk.stem.LancasterStemmer()
        SS=nltk.stem.snowball.EnglishStemmer()
        WNL=nltk.stem.WordNetLemmatizer()
        stem_dict={}
        for word in word_dict:
            if word in stopwords: continue      #Moves on to the next word if the word in in
the stop-list
```

```
        stem=WNL.lemmatize(word)
        if stem in stem_dict:            #Compiles a dictionary of stems with
corresponding word values
            stem_dict[stem].append(word)
        else:
            stem_dict[stem]=[word]

    if lemma_dictionary=='yes':          #prints out a csv file with the stems and the
words they came from
        f=open('StemDict.csv','wb+')
        csvw=csv.writer(f)
        for key in stem_dict:
            csvw.writerow([key]+stem_dict[key])



    word_list=[]
    for stem in stem_dict:               #Counts the entries for each stem to put in the
wordlist
        word_list.append((stem,sum([word_dict[word] for word in stem_dict[stem]])))

    word_list.sort()
    word_list=zip(*word_list)
```

```
    else:
        word_list=[]


    print time.clock()-t
    t=time.clock()


    return word_list


def SaveOutput(WL):
    import csv
    WL=zip(*WL)


    f=open('WordList.csv','wb')
    csvw=csv.writer(f)
    for row in WL:
        csvw.writerow(row)
    f.close()




def GetWordsAndStops(word_file,stop_file):
    import csv
    #Get the stopwords from the stopwordfile
```

```
    stopwords=[]
    with open(stop_file,'rb') as csvfile:
        userinput=csv.reader(csvfile)
        for row in userinput:
            stopwords.append(row[0])

    #Get the words from the word file
    words=[]
    with open(word_file,'rb') as csvfile2:
        userinput=csv.reader(csvfile2)
        for row in userinput:
            words.append(row[0])
    return(words,stopwords)



def VectorizeStemmedText(document, word_list,stopwords):
    WL=word_list
    import nltk
    document=document.lower()
    tokens= nltk.word_tokenize(document)
```

```
    '''
    #Get the stopwords from the stopwordfile
    stopwords=[]
    with open(stopwordfile,'rb') as csvfile:
        userinput=csv.reader(csvfile)
        for row in userinput:
            stopwords.append(row[0])
    '''
    tokens=[word for word in tokens if (word.isalpha() and word not in stopwords)]
    LS=nltk.stem.LancasterStemmer()
    stems=[LS.stem(token) for token in tokens]
    vector=[0]*len(WL)
    for stem in stems:
        if stem in WL:
            vector[WL.index(stem)]=vector[WL.index(stem)]+1
#    ss=float(sum(vector)*sum(vector))
#    for entry in range(len(vector)):
#        if vector[entry]>0: vector[entry]=vector[entry]/ss

    return vector
```

```
#This chunk of code was used to stem the entire corpus


if __name__ == '__main__':
   #texts=['/Users/Glier2/Dropbox/Dissertation/BioCorpus/Basic and Applied Ecology.txt','/Users/Glier2/Dropbox/Dissertation/BioCorpus/Current Biology.txt','/Users/Glier2/Dropbox/Dissertation/BioCorpus/Journal of Animal Behaviour.txt','/Users/Glier2/Dropbox/Dissertation/BioCorpus/Journal of Animal Ecology.txt','/Users/Glier2/Dropbox/Dissertation/BioCorpus/Journal of Zoology.txt']
   texts=['RawStatText.txt']
   textstring=ReadFiles(texts)
   stops='SMART_Stopwords.csv'
   output=BuildWordList(textstring,stops,lemma_dictionary='yes')
   SaveOutput(output)



if __name__=='__main2__':


SF='/Users/Glier2/dropbox/workspace/BioCorpusFilter/RuleGenerator/SMART_Stopwords.csv'


WF='/Users/Glier2/dropbox/workspace/BioCorpusFilter/RuleGenerator/TruncatedStemList.csv'
```

```
    (word,stop)=GetWordsAndStops(WF,SF)
    sent='Both sets of events are mediated by the process of abscission, wherein
programmed cell death produces a weakened abscission zone that enables the
separation.'
    v=VectorizeStemmedText(sent,word,stop)
    out=zip(word,v)
    for stem in out:
        if stem[1]>0:
            print stem
```

EngineeringBiologyThesaurus

```
global Toronto_Thesaurus, IIS_Thesaurus, OSU_Thesaurus, Full_Thesaurus

Toronto_Thesaurus={}
IIS_Thesaurus={}
OSU_Thesaurus={}

#This section creates lookup dictionaries for the published Engineering-Biology
Thesauruses
#Blank lines indicate that a new functional class begins
```

```
#Terms from the University of Toronto Thesaurus
Toronto_Thesaurus['export']=['bind','block','breakdown','excrete','inactivate']
Toronto_Thesaurus['transport']=['circulate','conduct','diffuse','pump']
Toronto_Thesaurus['transmit']=['communicate','transduce']
Toronto_Thesaurus['transfer']=Toronto_Thesaurus['transport']+Toronto_Thesaurus['transmit']

Toronto_Thesaurus['translate']=['synthesize','transcribe']
Toronto_Thesaurus['guide']=Toronto_Thesaurus['translate']
Toronto_Thesaurus['channel']=Toronto_Thesaurus['export']+Toronto_Thesaurus['transfer']+Toronto_Thesaurus['guide']

Toronto_Thesaurus['link']=['activate','bind','project']
Toronto_Thesaurus['couple']=['extend','link','overlap','stretch']+Toronto_Thesaurus['link']
Toronto_Thesaurus['mix']=['contract','exchange','fragment']
Toronto_Thesaurus['connect']=Toronto_Thesaurus['couple']+Toronto_Thesaurus['mix']

Toronto_Thesaurus['control magnitude']=Toronto_Thesaurus['stop']=Toronto_Thesaurus['inhibit']=['cover','destroy','inhibit','surround']
```

```
Toronto_Thesaurus['convert']=['decompose','degrade','develop','grow','mutate','photosynthesize']

Toronto_Thesaurus['collect']=['breakdown','concentrate','digest','reduce']
Toronto_Thesaurus['provision']=Toronto_Thesaurus['store']=['convert','deposit','photosynthesize']

Toronto_Thesaurus['stabilize']=['bind','connect']
Toronto_Thesaurus['support']=['develop','wrap']+Toronto_Thesaurus['stabilize']

#Terms from the Indian Institute of Science Thesaurus
IIS_Thesaurus['remove']=['evacuate']
IIS_Thesaurus['separate']=['free','detatch','release']+IIS_Thesaurus['remove']
IIS_Thesaurus['distribute']=['disperse','scatter','spread','spray']
IIS_Thesaurus['branch']=IIS_Thesaurus['separate']+IIS_Thesaurus['distribute']

IIS_Thesaurus['import']=['consume','inhale','intake','absorb','attract']
IIS_Thesaurus['export']=['repel']
IIS_Thesaurus['transfer']=IIS_Thesaurus['transport']=['shift','displace','fly','swim','jump','bounce']
IIS_Thesaurus['translate']=['slide']
IIS_Thesaurus['rotate']=['oscillate','spin','turn','swivel','roll']
```

```
IIS_Thesaurus['guide']=IIS_Thesaurus['translate']+IIS_Thesaurus['rotate']
IIS_Thesaurus['channel']=IIS_Thesaurus['import']+IIS_Thesaurus['export']+IIS_Thesaur
us['transfer']+IIS_Thesaurus['guide']

IIS_Thesaurus['join']=['adhere','bond','fuse']
IIS_Thesaurus['link']=['clamp']
IIS_Thesaurus['connect']=IIS_Thesaurus['couple']=['latch','lock']+IIS_Thesaurus['join']+
IIS_Thesaurus['link']

IIS_Thesaurus['actuate']=['activate','trigger']
IIS_Thesaurus['increase']=['grow','expand','multiply']
IIS_Thesaurus['decrease']=['compress','coil','divide','fold','shorten','wrap']
IIS_Thesaurus['regulate']=['preserve','sustain','remain','stabilize','maintain']+IIS_Thesaur
us['increase']+IIS_Thesaurus['decrease']
IIS_Thesaurus['change']=['alternate','fluctuate']
IIS_Thesaurus['prevent']=['constrain','obstruct']
IIS_Thesaurus['stop']=['halt','extinguish','clog','seal','suspend']
IIS_Thesaurus['control
magnitude']=IIS_Thesaurus['actuate']+IIS_Thesaurus['regulate']+IIS_Thesaurus['change'
]+IIS_Thesaurus['stop']

IIS_Thesaurus['collect']=['absorb','catch']
```

```
IIS_Thesaurus['store']=['conserve','hold']+IIS_Thesaurus['collect']
IIS_Thesaurus['supply']=['feed']
IIS_Thesaurus['provision']=IIS_Thesaurus['store']+IIS_Thesaurus['supply']

IIS_Thesaurus['signal']=IIS_Thesaurus['measure']=IIS_Thesaurus['sense']=['observe','mo
nitor','gauge','watch']

IIS_Thesaurus['support']=['cling','hold']

#Terms from the Oregeon State University Thesaurus
OSU_Thesaurus['remove']=['deoxygenate','filtrate','deanimate','liberate','expel']
OSU_Thesaurus['divide']=['divide','prophase','metaphase','anaphase','cleave','cytokinesis'
]
OSU_Thesaurus['separate']=['bleach',
'meiosis','replicate','mitosis','segment','abscission','electrophoreris','react','dialysis','denatu
re']+OSU_Thesaurus['remove']+OSU_Thesaurus['divide']
OSU_Thesaurus['distribute']=['exchange','circulate','diffuse']
OSU_Thesaurus['branch']=OSU_Thesaurus['separate']+OSU_Thesaurus['distribute']

OSU_Thesaurus['transfer']=['migrate','transfer']
OSU_Thesaurus['allow DOF']=['articulate']
OSU_Thesaurus['guide']=OSU_Thesaurus['allow DOF']+['orient','position','tunnel']
```

```
OSU_Thesaurus['channel']=OSU_Thesaurus['transfer']+OSU_Thesaurus['guide']

OSU_Thesaurus['join']=['bind']
OSU_Thesaurus['couple']=OSU_Thesaurus['join']+['recombine','mate','build','phosphoryl
ate','bond','synthesize']
OSU_Thesaurus['mix']=['blend']
OSU_Thesaurus['connect']=OSU_Thesaurus['couple']+OSU_Thesaurus['mix']

OSU_Thesaurus['actuate']=['induce','trigger']
OSU_Thesaurus['increase']=['hyperpolarize','pinocytosis']
OSU_Thesaurus['regulate']=['gate','electrophoresis','respire','regulate','organogenisis']+O
SU_Thesaurus['increase']
OSU_Thesaurus['increment']=['attach']
OSU_Thesaurus['decrement']=['decarboxylation','constrict']
OSU_Thesaurus['shape']=['elongate','stretch','attach','spread']
OSU_Thesaurus['condition']=['osmosis','constrict']
OSU_Thesaurus['change']=['ppinocytosis','catalyze','degrade','alter','bind','contract','hydr
olysis','twist','slip','spread','mutate','adiate','charged','acclimatize'] \
+OSU_Thesaurus['increment']+OSU_Thesaurus['decrement']+OSU_Thesaurus['shape']+
OSU_Thesaurus['condition']
OSU_Thesaurus['inhibit']=['repress']
OSU_Thesaurus['stop']=['interphase']+OSU_Thesaurus['inhibit']
```

```
OSU_Thesaurus['control magnitude']=OSU_Thesaurus['actuate']+OSU_Thesaurus['regulate']+OSU_Thesaurus['change']+OSU_Thesaurus['stop']

OSU_Thesaurus['convert']=['polymerize','ionize','synthesize','hydrolysis','gluconeogenisis','metabolize','glycolysis','translation','respiration','photosynthesis','fermentation','burn']

OSU_Thesaurus['store']=OSU_Thesaurus['contain']=['absorb']
OSU_Thesaurus['supply']=['lactate']
OSU_Thesaurus['provision']=OSU_Thesaurus['store']+OSU_Thesaurus['supply']

OSU_Thesaurus['sense']=OSU_Thesaurus['detect']=['detect','locate','see','smell']
OSU_Thesaurus['indicate']=['fluoresce','mark','communicate','react']
OSU_Thesaurus['process']=['learn']
OSU_Thesaurus['signal']=OSU_Thesaurus['sense']+OSU_Thesaurus['indicate']+OSU_Thesaurus['process']

OSU_Thesaurus['stabilize']=['homeostasis']
OSU_Thesaurus['secure']=['surround','envelope']
OSU_Thesaurus['support']=OSU_Thesaurus['stabilize']+OSU_Thesaurus['secure']

#Now build the full thesaurus by adding the three smaller ones
```

```
Full_Thesaurus={}
Thesaurus_list=[Toronto_Thesaurus, IIS_Thesaurus, OSU_Thesaurus]
for thesauruses in Thesaurus_list:
    for keys in thesauruses.keys():
        if not Full_Thesaurus.has_key(keys):
            Full_Thesaurus[keys]=thesauruses[keys]
        else:
            Full_Thesaurus[keys]=Full_Thesaurus[keys]+thesauruses[keys]



def EB_Thesaurus(search_key, selected_thes='Full'):

    # This last bit actually does the lookup, matching the function's inputs (the term and
the thesaurus (optionally)) to the thesauruses defined above
    # The output is a tuple with a sentence to inform of any error in the inputs and the
biology analogs to the entered terms (an empty list is returned for bad inputs)

    if selected_thes in ['Full','OSU','IIS','Toronto']:
        output_text="Search terms drawn from %s thesaurus.\n" %(selected_thes)
Thesaurus={'Full':Full_Thesaurus,'OSU':OSU_Thesaurus,'IIS':IIS_Thesaurus,'Toronto':
Toronto_Thesaurus}[selected_thes]
```

```
if Thesaurus.has_key(search_key):
```