Eureka

Eureka

Eureka is a student-founded and student-run initiative whose mission is to promote the world class undergraduate research done in the faculty of science. This journal offers undergraduate scientists the unique opportunity to share their discoveries with the scientific world, while learning the peer review process. Most undergraduate students do not get to experience the publication process as part of their scientific education. Eureka is an educational institution, with a diverse team of reviewers from many scientific backgrounds. Through clear, effective scientific writing students will be able to work together with faculty members to promote University of Alberta science.

EUREKA EDITORIAL TEAM

ABOUT THE FRONT COVER

The front cover is a modifed version of a piece called "Whole Brain" by Christopher Madan. Each of the brain images are artistic adaptations of an MRI if his own brain. Each panel combines both 'art' and 'science' features, as the common adage about the left hemisphere being logical and the right being artistic is largely incorrect. (Also, everyone uses more than 10% of their brain!) Christopher recently finished his Ph.D. in psychology. You can find out more about Chris' research on his website: http://www.cmadan.com.

ABOUT THE BACK COVER

The back cover consists of images taken with the 14″ telescope at the University of Alberta Observatory. The images depict the Sun, Saturn, Jupiter, as well as three of the surface of the Moon.

The current observatory was established in the summer of 2011, following the opening of the main CCIS building. The observatory contains 3 telescopes, each contained in their own dome. The smallest telescope is a 12″ telescope used for public night-time observing. This telescope has been in the Universitys possession since the late 1960's. Prior to the CCIS observatory being built, it was located in the old observatory on top of the Fine Arts Building, and prior to that it was on top of the old Physics building. The medium telescope is a 14″ telescope, also used for public observing. In addition to being used for night-time observing, this telescope can be outfitted with a solar filter to observe the Sun in visible light during the day. The largest telescope is a 20″ research-grade telescope. This telescope was custom machined by the Department of Physics machine shop. This telescope is intended to be used for research, including undergraudate research projects.

In addition to research, the observatory is also used for outreach. The observatory currently offers free public observing on Thursdays, both during the day (for solar observing) and in the evening (for night-time observing) when the weather is favourable. The observatory's Facebook (http://www.facebook.com/UofAObservatory) and Twitter (@UofAObservatory) post updates when the weather looks good for public observing. The observatory is also used for outreach with Edmonton-area schools, giving students a chance to learn about astronomy, as well as look at the stars through the telescopes.

Eureka

University of Alberta Science Undergraduate Research Journal

Volume 4, Issue 1 — Summer 2014

Research is a Journey of Self-Discovery

Christopher R. Madan

Department of Psychology, University of Alberta

In this issue of *Eureka*, I am proud to share with you the exciting and innovative research projects that are being lead by undergraduate students at the University of Alberta. While I will let their work speak for themselves, I also would like to take this opportunity to reflect on the beginnings of my research career and the broad reaching effects I have found research to have.

I first became involved in research in the second year of my undergraduate studies, simply because I wanted to have a summer job that I could learn from, rather than a more mundane one like stocking shelves in a store. I approached several professors and found a particularly good match with one and ended up starting as a volunteer research assistant right away. At the time my plan was to apply for medical school after my undergraduate, but that was a bit ill-conceived since I have found that I dislike dissection labs and do not take well to raw memorization, both of which are important skills in succeeding in medicine (among many other necessary skills). During my undergraduate program, I took several research courses and worked in the lab over the summers and found that my personality and aptitudes fit well with the lifestyle of a researcher. A few months ago I completed my Ph.D. in Psychology. I didn't know that I wanted to be a researcher when I started my undergraduate, but now I couldn't be more sure that this is the path for me. Soon I move to Boston to start a position as a postdoctoral research fellow.

Throughout my research career thus far, I have met many students who have taken to research. Some wanted to work in a research lab to gain experience and a reference letter for medical or law school, or other programs like occupational therapy or speech-language pathology. Some have been graduate students with a vision of the type of lab they hope to eventually lead when they become a professor. Others yet have been required to take a research course as part of an honors program and were unsure what to expect. Regardless of the reason, I have found that being involved in a research project is also a process of self-discovery.

In any good research project, the outcome is unclear. There should be hypotheses to constrain the possible outcomes, but we do not know which will occur. Even with a good research question, many important nuances need to be determine, such as what are the potential confounding variables that need to be controlled for, and how to best control for them. Reading all of the prior studies in the world will not give you definitive answers to these questions, since every project is a little bit different and scientific fields are always evolving. In my opinion, these skills that need to be developed to undertake a research project are very similar to those needed to succeed in life: You need to be able to learn what you can from books and the experiences of others. You need to learn to adapt unforseen circumstances. You need to learn to work well both independently and in teams – being able to both solve problems on your own as well as communicate your ideas effectively. I have found that research has given me an opportunity to discover myself better, what I am good at and what situations I find challenging, and to observe others doing the same and see how they act under similar circumstances.

Research is more than just a university course you take or a project you work on with others.
Research is a way of life.

Christopher R. Madan
Eureka Editor-in-Chief

A student experience in neurosurgery:
How summer research can equate to much more than a publication

Mitchell P. Wilson

Faculty of Medicine and Dentistry, University of Alberta

On November 17, 2011 I had the great pleasure of presenting to a group of some 200 students, parents, and faculty during the banquet portion of the first ever University of Alberta Undergraduate Research Symposium (URS). This symposium aimed to unite all undergraduate faculties across campus in a two-day event, providing students the opportunity to present academic research they've completed through a summer studentship, thesis project, or simply work throughout the year. Students presented their research in a poster format, while three projects including my own, were selected for oral presentations at the banquet night.

To begin the presentation, I introduced myself as a second year medical student at the University of Alberta with a background in the Neurosciences. I explained that I have a strong interest in pursuing Neurosurgery, which began several years ago but was certainly cultivated by a summer experience following my first year of medicine. During that summer, I worked with two Pediatric Neurosurgeons performing clinical research on a condition known as Hydrocephalus, a common presentation in Pediatric Neurosurgery. In my summer, I worked on database design and data collection on Mondays, Wednesdays and Fridays. On Tuesdays I assisted in surgeries while Thursdays were spent in clinic with patients. From the perspective of a medical student, the academic as well as personal development that occurred during these shadowing experiences were far more useful than any lecture could teach. It gave me the opportunity to extend my knowledge significantly in an area of great interest to me.

I spent a significant amount of time pondering how to best present my project at the URS. In the end, I had decided to instead comment on the multiple positive aspects of the overall research. I felt and still feel that many aspects beyond the research component of summer studentships are often under-recognized. These components nonetheless, deserve to be highlighted and should serve as additional intrinsic and extrinsic motivators for students considering participation in undergraduate research.

In presenting during the URS banquet, I emphasized many components of undergraduate research that extended far beyond the fundamental principles. Some topics I emphasized included:

- Many research projects can be recognized for elective and course credit.

- Many studentships are funded, providing students with a paid position.

- Research is an excellent opportunity to establish contacts, references and friendships.

- Students develop academically by learning new content in an area of interest.

- Students develop academically by learning about the research process.

- Students develop personally by developing insight into the world of academia while simultaneously establishing firm understandings of individual interests and dislikes.

- Students are often encouraged to partake in research dissemination including conferences. Students develop public speaking and knowledge translation skills necessary to convey academic knowledge. Further, funding is commonly provided for students to present across North America and indeed, the world.

- Opportunities often arise to become involved with work in addition to the assigned research. These opportu-

nities can further develop the students curriculum vitae in addition to allowing for personal growth and new experience.

While this is by no means an exhaustive list, it provides a number of suggestions for additional benefit outside the opportunity to become involved with research with a potential for publication. From a personal perspective, I was able to establish very strong relationships with surgeons in an area I am greatly interested in. This will undoubtedly serve me well when it comes time to apply for a Residency position towards the end of my medical education. Further, the connections I developed during oral and poster presentations associated with this project as well as opportunities to participate in medical rounds and shadowing experience are very worthwhile for my personal and professional development. This research was also recognized as a medical elective component of my academic development. I was able to learn significantly about clinical and surgical Neurosurgery, an area that we are not well exposed to during pre-clinical medicine. Also, I was able to lead and assist in three research papers beyond my assigned project. Finally, the personal experience of dealing with complicated surgical cases and families of dying children has helped me develop and mature as an individual.

Undergraduate research provides students with several opportunities to develop in both an academic and personal nature. As such, the first annual Undergraduate Research Symposium was an excellent initiative to showcase the diversity, experiences, and excellence of research done by our students. Hopefully, interest in this experience will continue to grow among fellow colleagues across campus as increasing awareness and excitement for the opportunities and benefits of undergraduate research is disseminated.

Good luck in your journey,

Mitchell P. Wilson
MD Candidate, Class of 2014
Faculty of Medicine and Dentistry
University of Alberta

Women in Research

Helena Zakrzewski and Yvonne Y. Chen
Eureka Team, University of Alberta

From the Greek Mathematician Hypatia to Chemistry Nobel Prize winner Dr. Ada Yonath, women have made valuable contributions to Science for thousands of years. However, today some girls are held back by the gender stereotype that women are not good at science. To work to break the stereotype, we interviewed three women in research: Meagan Lyszczyk from Chemistry, Maddy Wang from Computing Science, and Yvonne Wong from Psychology. While they each come from different scientific disciplines, they are all splendid examples of women in research. We hope that these womens' stories and advice can inspire and encourage more students to become involved in research, especially those young women who want to but are afraid to.

MEAGAN LYSZCZYK
UNDERGRADUATE STUDENT

Education: B.Sc. (University of Alberta)
Lab affiliation: Dr. Alex Brown

How did you become involved in research?

Meagan became involved in research through her own initiative. She talked to other students about research opportunities available on the University of Alberta campus throughout the summer and made the decision to apply for the NSERC Undergraduate Student Research Awards (USRA) to find out through her own experience what the process of scientific investigation involves.

What was your role in your lab?

Throughout the summer of 2011, Meagan was the summer research student in Dr. Alex Brown's laboratory in the Department of Chemistry.

What were you hoping to investigate in your research?

The research Meagan conducted over the summer focussed on the photophysics of substituted benzofurazans. Benzofurazans are aromatic compounds whose fluorescence properties are strongly affected by substitutions at the 4, 5, and 7 positions. Substitutions are linked to changes in electron density and dipole moment and may improve fluorescence or lead to quenching. Specifically, she regarded two substituted benzofurazans: 4-azido-7-nitrobenzooxadiazole and 5-ethynyl-benzoozadiazole. Her research aided in bringing about a theoretical understanding of how these fluorescent probes can efficiently be designed and in developing a unified model for the use of substituted benzofurazans in fluorescent bio-labelling strategies.

What would you do on a typical day in the lab?

On a typical day, Meagan would be responsible for composing scripts that were then run of specific software programs, which analyzed chemical structures. During her summer in the lab, she spent a significant amount of time regarding the excitation wavelengths of light that were given off by those compounds investigated.

What is your favourite part of doing research? What is your least favourite part of doing research?

Above all, Meagan enjoyed her experience in the lab. "It allowed me to learn what research truly entails". The experience allowed her to get a sense of the bigger picture in science: "as an [undergraduate science] student, it is easy to get caught up with assignments, exams, etc. and you forget that you are part of a larger community".

While Meagan enjoyed seeing what research entails, she acknowledged that regarding one narrow, specific topic is very tedious. Instead, in her future endeavours, she hopes to branch out more and "understand the bigger picture" involved in a particular area of study.

In this area, what was most challenging for you? Have you ever had to overcome any challenges in the past related to your research career?

"There is certainly a learning curve involved," Meagan explains. "[Learning] computer programming and script writing are all about [becoming more familiar with them] through experience". There is a specific language involved in the field of analytical chemistry, which can be quite challenging to those previously not exposed to it. She credits the help of the graduate students in Dr. Brown's lab with helping her through this learning processes and providing her any help needed in overcoming "writing glitches" and any other problems that she incurred.

What do your future ambitions include?

Meagan prioritizes finishing her current degree and cites that it is still too early to determine whether she would pursue graduate level work.

What advice do you have for anyone who may be interested in pursuing similar opportunities?

Never be fearful of new opportunities. "Go for it!" Meagan exclaims. She expresses that undergraduate research endeavours are all about the learning experiences that come along with them. "Try something even if you're not sure you never know. [Especially, use such opportunities] to learn more and find out if its what you'd like to pursue before committing to graduate studies".

Is there anything else you believe our readers should know about becoming involved in doing research?

"It is an amazing opportunity!" Meagan cites that most students do not know what research related opportunities are available to them and emphasizes to spend some time getting familiar with what's out there. "Many people in their undergraduate [studies] don't take advantage of such experiences and they most definitely should".

MADDY WANG
GRADUATE STUDENT

Education: Ph.D. (University of Alberta), M.Eng. (University of Regina)
Lab affiliation: Dr. Hong Zhang

How did you become involved in research?

Maddy has a background in Engineering; she has always been interested in machine learning. She came to Canada to pursuit further in her studies first at University of Regina and then here at University of Alberta

What was your role in your lab?

Maddy is currently a Ph.D. candidate in the Zhang lab. She is working very hard to finish her dissertation, but she is also actively training the junior members of the lab

What were you hoping to investigate in your research?

Maddy's research focus on computer vision, a form of machine learning that specialized for image processing and object recognition. She pointed out that how human perform object recognition is very different from how a computer could do it. She wants to figure out a solution for computer vision to perform object recognition that is accurate and efficient.

What would you do on a typical day in the lab?

Maddy said a typical day of computing science Ph.D. student can be different depends on where you are in your studies. The first 2 years will have many course works where you can learn in a class setting. This sets the foundations for your research, main while you will

also read many research papers to learn the research background. The later years, which is Maddy at now, she describes that she spends majority of her time on computer alone with some interaction her colleagues.

What is your favourite part of doing research? What is your least favourite part of doing research?

Maddy most enjoys the exploration aspect of research, generating new ideas and finding solutions. Often a question will have multiple solutions, it is never as simple as $1+1=2$. Maddy finds the possibilities of various solutions very exciting. However, the uncertainty of implementing an idea can be frustration as well. Sometimes what you thought could be the solution to the problem is actually not working; it might take a month or even a year to figure out what the issue was.

In this area, what was most challenging for you? Have you ever had to overcome any challenges in the past related to your research career?

Finding the suitable research direction has been the biggest challenge in Maddy's research life. Maddy said Ph.D. can be a very short time where you can be lost in finding your topic. Maddy used her interests to guide her in selecting a research topic, she followed her heart and found a topic that fits her interest and skill sets.

What do your future ambitions include?

Since Maddy's research focus is computer vision, she really wants to apply the research to solve real life problem. For example, computer vision software can help physicians detect abnormalities in patient's brain imaging scans. She hopes she will be at a position to bridge the gap between research and industry.

What advice do you have for anyone who may be interested in pursuing similar opportunities?

It's all about passion. Maddy believes it is very important to find something you really like doing, once you find it you are more motivated to pursuit further. Maddy also advises that research is a long process, so don't be discouraged when things don't work out, it will all be worth it in the end.

Is there anything else you believe our readers should know about becoming involved in doing research?

Never give up, stay focused, and be dedicated: those are the keys qualities to have when you are involved in research.

YVONNE WONG
POSTDOCTORAL RESEARCH FELLOW

Education: Ph.D. (University of Western Ontario), M.Sc. (Univeristy of Lethbridge)
Lab affiliation: Dr. Anthony Singhal

How did you become involved in research?

Yvonne Wong became involved in research "through chance". After completing her undergraduate degree,

she worked at a call centre where she was introduced to Dr. Ian Whishaw, a professor at the University of Lethbridge, who recruited her to complete a M.Sc. under his supervision. Upon its completion, Yvonne pursued a Ph.D. at the University of Western Ontario.

What is your role in the lab?

Yvonne is currently a post-doctoral associate and lab manager of the Singhal lab. As well, she manages a joint electroencephalogram (EEG) laboratory in the Department of Psychology. Her work is carried out in collaboration with both Drs. Singhal and Varnhagen.

What are you hoping to investigate in your research?

The main focus of Yvonne's research is to use EEG and fMRI (functional magnetic resonance imaging) methods to examine the neural correlates of human attention and actions. Through the use of simultaneously recording EEG and fMRI, the Singhal lab examines the neural circuitry of emotion-attention interactions during various tasks. The work of the lab also uses EEG to investigate the relationship between haptics, vision, and audition on attention to stimuli in peripersonal and extrapersonal space.

What do you do on a typical day in the lab?

On a typical day, Yvonne plays an integral role in assisting with data analysis as well as being primarily responsible for maintaining several laboratories' equipment in the Department of Psychology.

What is your favourite part of doing research? What is your least favourite part of doing research?

Being "Sherlock Holmes" is what Yvonne enjoys most about completing research; she finds the inquisitive process incredibly fulfilling. Particularly, Yvonne enjoys reviewing past data: "I enjoy looking at data from previous experiments and finding new methodologies to make sense of research questions that remain unanswered". She finds this to be most exciting when new results that are obtained are supported by the work of individuals who have worked on similar projects in the past.

But, Yvonne acknowledges that the "stress of publishing" is not always agreeable and that there is a significant amount of pressure: "in the scientific world, it sometimes seems that you need to continually publish in order to remain competitive".

In this area, what is most challenging for you? Have you ever had to overcome any challenges in the past related to your research career?

Time is always a challenge! In order to be successful in this field, Yvonne describes that you need to be efficient. Keeping up with current working being done in your area is essential but time consuming and there "is always heavy competition within the research world".

What do your future ambitions include?

In the future, Yvonne hopes to primarily focus on teaching while continuing to conduct research in her spare time. Through teaching, she hopes to inspire young people who are interested in science-related fields to pursue a career in research. "If we teach young people good scientific techniques early [in their careers], it is possible to leave a successful research legacy".

What advice do you have for anyone who may be interested in pursuing similar opportunities?

To be successful in a research career, Yvonne advises that you need to "find the question that inspires you!" She strongly emphasizes that without a passion for the subject and the task at hand, it is difficult to dedicate your time to research - something that is greatly required to pursue such endeavours. Furthermore, Yvonne suggests making sure to "find a supervisor that you are able to work well with" as this will be the principle person who will guide you along in your academic path.

Is there anything else you believe our readers should know about becoming involved in research?

Yvonne emphasizes that research is filled with its challenges and struggles and you truly need to love what you do if you're going to make it. Passion is essential for success in this field and over the years she's come to appreciate that "if you wouldn't be willing to do it for free then don't do it at all".
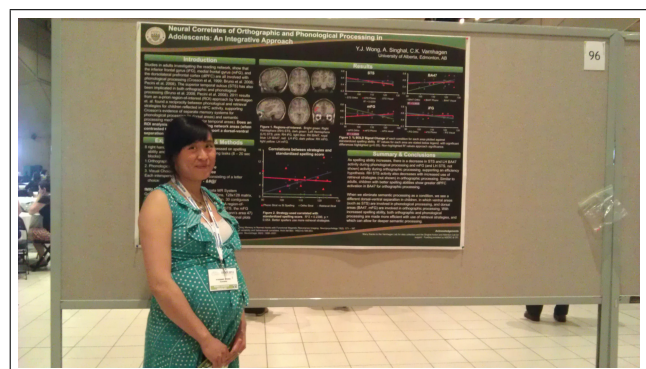
Understanding Life in Academia: A Guide for Inspired Undergraduates

Christopher R. Madan
Department of Psychology, University of Alberta

So you just got involved in a research lab and are starting to hear a lot of new vocabulary like "tenure," "postdoc," and "NSERC". You may ask your new colleagues what some of these terms mean, but you might also feel embarrassed having no idea about so many of these new terms. Instead of nodding along in bewilderment, let's go through a quick primer on life in academia and the major Canadian funding agencies. If you haven't become involved in a research, please see our earlier guide to get you started! (Madan & Chen, 2012, *Eureka*, "Getting involved in research: A guide for inspired undergraduates")

POSITIONS WITHIN THE LABORATORY ENVIRONMENT

First, let's start from where you are currently, an *undergraduate research assistant*. The position of research assistant is often abbreviated as RA. To have a position like this, you likely are volunteering, doing a 400-level research course, or have your own Honors project. Research assistants usually help with data collection, attend lab meetings, and read research papers. Depending on the field, other responsibilities may also exist. Honors students are often required to complete a *thesis* paper at the end of their degree, describing the background, methods, and results of their study. Some labs also have *lab managers*, who help with overseeing the day-to-day function of the lab, including administrative responsibilities. Lab manager is usually a full-time position and requires the completion of a bachelors degree.

The next step up is *graduate students*. Graduate students have completed a Bachelor's degree (BSc or BA), and are now either *Masters* (MSc [science] or MA [arts]) or *PhD* (*Doctor of Philosophy*) students. Students at either level can be referred to as graduate student or 'grad student', but referring to them as a Masters or PhD student would be more precise. PhD students can be further subdivided based on if they have completed an important exam called the *candidacy*, which usually involves an oral examination designed to test that they have sufficient background knowledge for their *dissertation* research. A dissertation usually consists of a number of research studies and serves as the main requirement of the PhD degree. It's quite similar to an Honors student's thesis, but larger in scope. After completing this candidacy exam, the student can then refer to themselves as a *PhD candidate*. After a PhD candidate completes their PhD, it is time to move on. Many possibilities are open to this newly minted PhD graduate. Within the realms of academia, the two most likely subsequent positions are *postdoctoral research fellow* and *sessional lecturer*. After completing either a Masters or PhD, it is also common to transition to an *industry* position. Industry positions may be tailored to the individual's specialty, such as pharmaceuticals, marketing, or software development, or may be broader such as scientific journalism.

A postdoctoral research fellow (*postdoc* or *PDF*) works with another professor, more as an equal, to exchange techniques and learn to supervise their own laboratory. This position usually is held for 2-3 years before the postdoc applies for a tenure-track professor, sessional lecturer, faculty lecturer, or industry position. An alternative path is to become a sessional lecturer, which is entirely a teaching position, with little or no required research component. Sessional lecturers are hired at universities entirely on a contractual basis. Faculty lecturers are permanent lecturing positions (i.e., long-term contracts).

At the level of *professor*, there are also a few subtle distinctions in position. In Canada, when a professor first begins, they are an *assistant professor*. After a few years, approximately six, the professor applies for *tenure*, which essentially guarantees the professor a position at the university until retirement. The review process for becoming tenured requires the professor to demonstrate that they have a strong academic record (i.e., publishing papers, supervising students), funding (described in detail later), and teaching. After obtaining tenure, the professor's rank moves up to *associate professor*. When the professor becomes more senior, they become a *full professor*, or simply *professor*. Becoming a full professor is never a gauranteed promotion by seniority, it is a very esteemed position that requires a substantial cumulative research record. At all of these levels, by supervising a research lab, the professor is also a *principle investigator* or *PI*.

Note, someone that teaches a class is not necessarily a *professor*. For instance, sessional and faculty lecturers often teach, as can PhD candidates. The correct position for someone teaching a class is *instructor*. Nearly all instructors have received their PhD, and thus should be referred to as "Dr.", though this is not in the same sense as a MD (Doctor of Medicine).

Every so often, usually every seventh year, a professor can take a year off, referred to as a *sabbatical*. During this year the professor often works to catch-up on previous projects (i.e., write papers) or visit researchers at other universities to develop collaborations and learn new techniques. This is a paid leave of absence from the university.

After a professor has retired, if they still continue to be

involved in research and affiliated with the university, they may be given the title *professor emeritus*.

FUNDING AGENCIES

In Canada, money that funds research comes from funding agencies. Most of these agencies are part of the federal or provincial government, but some may be associated with non-profit organizations. These are the funds that are used to pay for ongoing expenses such as chemical reagents or paid participants, as well as buy new equipment, support staff such as lab managers, cover conference travel and training courses, and sometimes pay for graduate stipends.

In Canada, three funding agencies exist as part of the federal government, together are referred to as the *tri-council*. The agency most relevant to basic science and engineering research is *NSERC*, the Natural Sciences and Engineering Research Council (`http://www.nserc-crsng.gc.ca`). Health-related research is funded by CIHR, the Canadian Institutes of Health Research (`http://www.cihr-irsc.gc.ca`). The last agency is the Social Sciences and Humanities Research Council or *SSHRC* (`http://www.sshrc-crsh.gc.ca`). Professors often apply to one or more of these agencies to obtain research grants to fund their research and graduate students apply to these agencies for scholarships. While these are the main federal agencies, others exist, such as the Canadian Foundation for Innovation (*CFI*; `http://www.innovation.ca`).

At the provincial level is *Alberta Innovates* (`http://www.albertainnovates.ca`). Of the agencies that comprise Alberta Innovates, two are particularly relevant: Alberta Innovates Technology Futures (*AITF*; `http://www.albertatechfutures.ca`) has a similar scope to NSERC. This agency was formerly known as Alberta Ingenuity. Alberta Innovates Health Solutions (*AIHS*; `http://www.aihealthsolutions.ca`)

has a similar scope to CIHR and is formerly known as AHFMR (Alberta Heritage Foundation for Medical Research). Again, while these are the most broadly relevant provincial agencies, several others exist. Two of these include the Alberta Gambling Research Institute (*AGRI*; `http://www.abgamblinginstitute.ualberta.ca`) and Alberta Centre for Child, Family & Community Research (*ACCFR*; `http://www.research4children.com`).

Of these agencies, two offer undergraduate research scholarships that can pay for you to do research over the summer: *NSERC USRA* (Undergraduate Student Research Award; `http://www.nserc-crsng.gc.ca/students-etudiants/ug-pc/usra-brpc_eng.asp`) and *AIHS Summer Studentships* (`http://www.aihealthsolutions.ca/funding/training-and-early-career-funding/undergraduate-high-school/summer-studentships/`).

CONCLUSION

Hopefully this brief primer was able to orient you to your new research environment. If you hear any additional terms that were not covered here, ask your colleagues. They likely will be happy to satiate your interest in research!

FUN FACTS: PhD stands for Doctor of Philosophy, regardless of what field the student actually researches. In Germany, an RA is referred to as a HIWI (short for "Hilfswissenschaftler," which literally translates to "assistant scientist").

Development of an Ungulate Mammalian Hair Key

Patrick Jones, Sophie Sfez, and Peter T. Knamiller
Department of Ecology, University of Alberta

We have created a dichotomous hair key to differentiate between adult and juvenile ungulates seen in summer wolf diets; however we would like to stress that this key can also be used for other large carnivore diets, as it has shown merit in the identification of cougar and coyote prey items. To our knowledge there is no mammalian hair key that is able to sufficiently differentiate between juvenile and adults of the major species in the North American Wolf's (Canis lupus) diet, specifically, mule deer, white-tailed deer, elk, moose and bighorn sheep. We acquired juvenile hair samples starting at birth at bi-weekly to monthly intervals from various zoos, wildlife parks and rehabilitation centres over North America for all the major prey species except bighorn sheep. Main characters used to classify ungulate hair were basal scale margin distance, hair diameter and hair color. Scale margin distance and basal hair diameter was measured via a microscope ocular micrometer, with t-tests completed to assess differences between species. We were successful at differentiating between all juvenile ungulates, however juvenile deer species hair maybe very difficult to differentiate between and may only be accurately done with experience. Juvenile ungulate hair is usually smaller in diameter then the guard hair of adults and is more delicate in appearance. As observed in other study by De Marinis et al 2006, we too were able to differentiate juvenile ungulates by a scalloped medulla (Figure 1). We were able to distinguish juvenile moose by their ginger color appearance, large medial hair cuticle scales, and large basal hair diameter compared to other ungulate species (Figure 2). We also found the juvenile elk had smaller hair diameters then juvenile deer and appear less ridged compared to juvenile deer.

If you have any questions regarding the hair key or would like to view it, please contact Patrick Jones at `pj2@ualberta.ca`.

REFERENCES

De Marinis, A.M. and A. Asprea. 2006. Hair identification key of wild and domestic ungulates from southern Europe. - Wildl. Biol. 12: 305-320.

Harrison, R. L. 2002. Evaluation of microscopic and macroscopic methods to identify felid hair. Wildlife Society Bulletin, 30(2): 412-419

Kennedy, A.J. and L. N Carbyn. Identification of wolf prey using hair and heather remains with special reference to Western Canadian National Parks. Canadian Wildlife Service. Edmonton, AB.

Moore, T. D., Spence, L. E., Dugnolle, C. E. and W. G. Hep worth. 1974. Identification of the dorsal guard hairs of some mammals of Wyoming. Wyoming Game and Fish Department Cheyenne, Wyoming.

Wilson, D. E. and D-A. M. Reeder. 2005. A key for the identification of guard hairs of some Ontario mammals. Canadian Journal of Zoology 71:587-591.

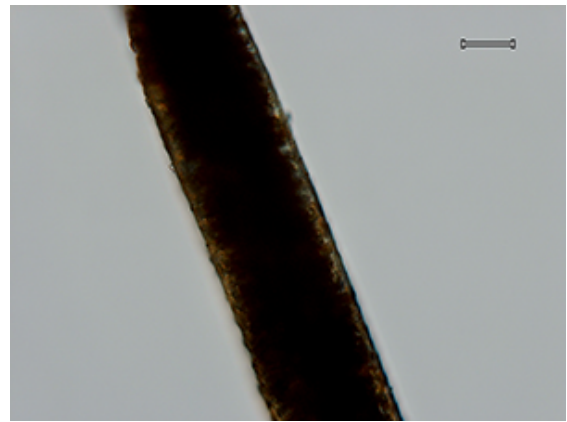FIG. 1. Juvenile elk medulla, unbroken lattice, scalloped medulla (cortex visible on either side of dark medulla), scale bar: 33.3 μm. A scalloped medulla is characteristic of most juvenile cervids and Bighorn sheep.
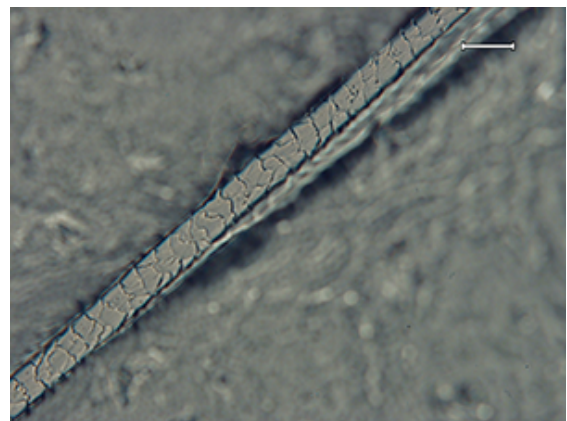

FIG. 2. Juvenile Moose tip, scales distant, margins smooth-crenate, irregular wave mosaic, scale bar: 33.3 μm.

Candle in the Wind: Goodbye Fossil

Eireen Ng Wong Hing
Faculty of Engineering, University of Alberta

After food, shelter and clothing as basic needs for the survival of man, what would be the fourth pre-requisite for survival? How about electricity? Yes, electricity. Since its discovery and the development of its numerous applications, the demand for electricity has never ceased to grow. But with the increasing repercussions of the burning of fossil fuels on health, on the environment, and on the economy, scientists and world leaders are devoting more time in considering a shift from fossil fuel to renewable sources of energy to generate electricity. So far, wind energy has the characteristics qualifying it as the best alternative to fossil fuels. Unlike other renewable sources, wind energy does not consume a lot of water and has the least carbon footprint. Wind energy is also emerging as a powerful resource with its high energy capacity, which is responsible for the wide technological support that is strengthening its development.

One of the pronounced environmental advantages of wind energy is its overall very low water consumption. In fact, no water is required for the phases of electricity generation from the wind turbines. Is this not something, in a world which is facing worsening water crises? As noted by Alina-Florentina, "wind farms virtually require no cooling, as opposed to natural gas, coal and nuclear power plants that require a tremendous amount of water for cooling" (62). According to a study investigating "the usage of water for electricity generation in the US" (Fthenakis 2039), the percentage of water consumption for wind energy accounts only for the fabrication and assembling "of steel, iron, and glass fiber for wind turbines [which] are the most significant sources of indirect, upstream water withdrawal" (Fthenakis 2042). The study led to results on wind energy having a water "withdrawal factor" of "4 liters per megawatt-hour of power produced as compared to the 7570 liters per MWh withdrawal factor for geothermal" (Fthenakis 2044). This data proves that "wind energy conserves water resources" (Alina-Florentina 62). Hence, by consuming a relatively insignificant amount of water, wind energy ensures the protection of our water resources.

Wind energy not only preserves our water resources but it also maintains the sustainability of the environment simply because it is clean. Wind turbines do not release any toxic gas, or any hazardous or radioactive waste like "sulfur, methane, carbon monoxide, nitrogen oxides, ozone, volatile organic compounds, and other particulate matter [that] wreak havoc on human and natural habitats alike by causing things like acid rain, urban ozone (caused primarily by nitrous oxide emissions, resulting in respiratory problems in humans), and global climate change" (Shoock 1019). According to research "data from the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety, approximately 67 million tonnes of CO_2 was avoided in 2006 by generating electricity through ...[renewable sources among which] wind energy [played] the most important role" (Solangi et al. 2425). Moreover, a comparative investigation was carried out to determine the extent of how much clean wind turbines are and Solangi et al. state that

> A typical turbine was installed ...with a rated capacity of 100kW and [it was] estimated that 168 tonnes of CO_2 [could] be reduced annually. According to ecological footprint, a forest absorbs approximately 3 tonnes of CO_2 per acre of trees per year. Hence, a 100kW wind turbine will prevent as much CO_2 from being emitted each year as could be absorbed by 24 acres of forest ...[thus] saving about 168 tonnes of CO_2 each year. (2426)

The investigation clearly evaluates the extent to which wind turbines are defined as clean, but one might argue that it assesses only the operation phase of the wind energy system. How about the setting up, the erection of the farms, and the implementation of the system? Ram tackles this argument in his article and states that

> The wind turbine life cycle (manufacturing, transportation, planning and construction, operations, monitoring, maintenance, and decommissioning) has fewer environmental risks than conventional energy generation because impacts are primarily temporal and localized. Life-cycle assessments show that the wind energy manufacturing and construction phase generates the largest risks owing to raw materials production and components assembly. This phase accounts for approximately 85% of [the total] emissions. (9-10)

But what is more interesting is that Ram cites a study done by Tester et al. which led to the conclusion that

> these processes take place only once during a turbines life, and so their impacts have been estimated to be relatively small. Because of these one-time life-cycle effects, wind energy appears to hold substantial promise for reducing carbon dioxide (CO_2) and other harmful emissions that pose ecological and health risks. (10)

Actually, what makes wind energy clean, is the mere fact that turbines do not use any fuel (Schiermeier et al 819). Sikri et al. explain how the power is simply and directly generated "from the kinetic energy of the mass of air moving in wind" (1). The kinetic energy trapped in the turning blades of the turbine then gets effectively converted to electrical energy when "the blades drive the shaft" which in turn "rotates the generator" of the turbine (Sikri et al.1).

In addition to their simplicity for the generation of electricity, wind turbines have a huge potential for energy capacity. Even though the electricity generation from wind farms are currently backed up by other sources of electricity generation, the prospect of it being a stand-alone source for transmitting electricity to a wide range of consumption type is not unrealistic and improbable. Oliver reports that "the amount of energy generated by the movement of Earths atmosphere is vast hundreds of terawatts" (4) and he further cites researchers from Stanford University who "calculated that at least 72 terawatts could be effectively generated using 2.5 million of todays larger turbines placed at the 13% of locations around the world that ... [are] practical sites" (4). This represents a very high efficiency of about 70% in harnessing wind for production of electricity. Similarly to the study carried out by the Stanford University researchers, a detailed investigation on wind farms was performed at the biggest wind resource in China which covers "an area of 83,000 km^2" (Lew 273). It was evaluated that the land is able to contain "nearly a million turbines with more than 500GW nameplate capacity" (Lew 273). But with intensive calculations and with assumptions of using the latest high-technology turbines which have the best features that ensure high efficiency operation, the study concludes that the "million turbines" can easily generate "an average annual wind electricity production of nearly 1800 terawatt-hour per year ... This represents 60% more energy than [what] is consumed currently in China" (Lew 273). It can hence be seen that wind farms have credible potential in generating electricity furthermore. The 60% overproduction of energy is more than satisfactory and it reflects again a very high efficiency in harnessing wind.

However, it cannot be denied that the existent potential of increasing the energy capacity of the turbines is challenged by the intrinsic characteristics of the wind. Wind is a variable and an unpredictable resource and turbines can only produce electricity when the wind blows, and it needs to blow within a specific range of speeds. Energy is generated when the wind speed reaches "16km per hour" but the turbines "shut down when the wind exceeds 90.00km per hour" (Costescu et al. 150). This is a real disadvantage that wind speeds cannot be predicted and that there is no certainty in the continuous supply of power output.

Wind does have somehow disheartening limitations in occurrence and speed, but it is alright. Thanks to technology, wind has already taken a great leap towards achievement. The wind energy industry is extensively growing with the support of new technologies specifically designed for wind farms. Turbines are getting more sophisticated to achieve a greater gain. "Computerized systems with integrated laser systems are proving effective ... [in sparing the turbine from destructive winds] while at the same time increasing efficiency" ("Contemporary Issues" 43). Modern turbine blades are carefully designed using specific type of materials. Costescu et al. mention that the blades are now made of "glass-reinforced epoxy ... and the turbines are equipped with asynchronous generators that do not create any electrical disturbance." (151). Costescu et al. discuss the "concrete foundations for the turbines" (151). Costescu et al. make a thorough analysis of "equilibrium conditions of the superstructure that transmits very large bending moments" (151) and consequently suggest detailed technological improvements to the foundations. Even the fluctuations in "wind velocity" can now be controlled by an "inverter technology" (Inoue 1313).

There is no doubt that experts and scientists are convinced that wind is the alternative to fossil fuel because the wind energy industry is undergoing mature technological developments in all its components thanks to the extraordinary on-going research and studies that are being carried out. Inoue and Miyazaki. believe that "control technology resolves a large number of problems in terms of safety and operation characteristics and is able to generate the largest possible percentage of wind power by optimizing the conversion from wind to wind power, which was previously reduced when prior technologies were used" (1308). Indeed, as Alina-Florentina states it, "wind energy has matured dramatically, making wind one of the fastest growing sources of electricity in the world today ... Wind Turbine Technology ... today are sleek and slender machines, a far cry from their bucolic wooden ancestors" (5).

Wind energy is unquestionably the alternative to fossil fuels. While being a frugal resource that saves a lot of water and a very shy polluter which has benign impacts on the environment, wind energy has outstanding and promising potential in its energy capacity. Having proven its benefits, it has attracted the attention of experts, engineers, scientists and researchers, so much so that these professionals are whole-heartedly investing their time and efforts to benefit and to further boost the wind energy industry through the medium of technology. Wind is believed to be the "cornerstone of [a] sustainable energy future" as stated by Alina-Florentina, but it is somehow too early for wind to completely take over fossil fuel because there is a need for wind-friendly policies and a need for more than just breezy funding from governments. But until the change happens, may the wind slowly blow all those candles that are sparking the burning of fossil fuel!

REFERENCES

Alina-Florentina, Cucoş. Social and Economic Impacts of Wind Power in Corelation with the Financial Crises." Annals Of The University Of Oradea, Economic Science Series 20.2 (2011): 62-68.

Costescu, Ciprian, et al. "Alternative Energy Resources: Foundation Solution For Wind Turbines." Analele Universitatii Maritime Constanta 51.13 (2010): 150-153.

"Contemporary Issues." Air Force Journal Of Logistics 33.3/4 (2009): 37-50.

Lew, Debra J. "Alternatives To Coal And Candles: Wind Power In China." Energy Policy 28.(2000): 271-286.

Fthenakis, V, and HC Kim. "Life-Cycle Uses Of Water In US Electricity Generation." Renewable & Sustainable Energy Reviews 14.7 (n.d.): 2039-2048.

Schiermeier, Quirin; Tollefson, Jeff; Scully, Tony; Witze, Alexandra; Morton, Oliver., et al. "Energy Alternatives: Electricity Without Carbon." Nature 454.7206 (2008): 816-823.

Ram, Bonnie. "Assessing Integrated Risks of Offshore Wind Projects: Moving Towards Gigawatt-Scale Deployments." Wind Engineering 35.3 (2011): 247-266.

Shoock, Corey Stephen. "Blowing In The Wind: How A Two-Tiered National Renewable Portfolio Standard, A System Benefits Fund, And Other Programs Will Reshape American Energy Investment And Reduce Fossil Fuel Externalities." Fordham Journal Of Corporate & Financial Law 12.6 (2007): 1011-1077.

Sikri, R. K., P. Harisinghaney, and N. Qureshi. "Critical Issues For Development Of Wind Energy In India." Proceedings Of World Academy Of Science: Engineering & Technology 51.(2009): 384-387.

Solangi K.H., et al. "Environmental Impact Of Wind Energy." Renewable And Sustainable Energy Reviews 15.(2011): 2423-2430.

Inoue Yoshinori, and Miyazaki Kumiko. "Technological Innovation And Diffusion Of Wind Power In Japan." Technological Forecasting & Social Change 75.(2008): 1303 1323.

Determining a Relation between X-ray Luminosity and Orbital Period of X-ray Binaries

Tyler Naffin

Department of Physics, University of Alberta

The goal of this project was to examine the relationship between the average x-ray luminosity and the orbital period of x-ray binaries. Using the data gathered by the All-Sky Monitor instrument aboard the Rossi X-ray Timing Explorer, 29 sources were selected for investigation based on the intensity of the x-rays emitted from each of the sources. A literature search was then performed to gather further details on each of the sources, including orbital period, distance, hydrogen column density, and classification of each source as either low- or high-mass x-ray binaries with either neutron star or black hole companions. Sufficient data for 22 of the sources was known in order to create plots of the average luminosity versus the period, which allowed for the period-luminosity relationships for x-ray binaries to be further examined.

I. INTRODUCTION

X-ray binaries (XRBs) are star systems that contain two gravitationally bound objects, one being a compact body; usually a neutron star or black hole, and the other a donor star. The donor star, being in close proximity to the compact body, begins to overflow its Roche lobe and matter accretes onto the compact body. Roche lobes are defined as the boundary between the two stars where the gravitational force is at an equipotential. The highly energetic nature of the mass transfer results in the emission of x-rays from the system. XRBs can be classified as either low-mass or high-mass x-ray binaries depending on the mass of the donor star[1].

Low-mass x-ray binaries (LMXBs) are defined by the mass of the donor star which typically has a mass of approximately two solar masses or less. They are typically characterized by the stars having a small orbital radii and period, with the period ranging anywhere from 11.4 minutes for the shortest known period, to 33.5 days for the longest known period, although longer periods arguably exist. The small separation between the two stars is required to facilitate the mass transfer due to the low mass of the system[2].

High-mass x-ray binaries (HMXBs) have a donor star with a mass that typically exceeds ten solar masses. As a result of the larger mass the separation of the two stars can be greater and consequently the orbital period can be longer, ranging from approximately 5 hours to 580 days in known cases. In the case of HMXBs, mass transfer can also be facilitated by strong solar winds, either in place of or in addition to the overflow of the Roche lobes[3].

XRBs that have donor stars with masses between two and ten solar masses, also known as intermediate-mass x-ray binaries (IMXBs), are less common as the donor stars of IMXBs tend to quickly transfer much of their mass onto the compact body, and mimic the properties of LMXBs[4].

XRBs are thought to be created by one of two methods. In some cases, one of the stars in a binary star system can undergo a core-collapse supernova, and thus be transformed into either a neutron star or black hole. However the supernova can result in the binary system becoming gravitationally unbound, depending on how much mass is ejected. It is believed that at least half of the total mass of the binary system must be ejected for the system to become gravitationally unbound. Therefore systems with a much more massive companion star are more likely to remain gravitationally bound. The second method involves a compact body and another star becoming gravitationally bound after chance close encounter. This is more likely to occur in areas where the density of stars is much greater, such as globular clusters or the galactic core, which is consistent with observations where relatively high numbers of XRBs are observed[5].

One method for increasing our understanding of XRBs is to create models based on theory and observation such as luminosity functions. Luminosity functions are defined by the number of sources within a given luminosity interval. They can be used compare and confirm theoretical and observational results. In the case of x-rays, XRBs are a major source of x-ray emissions in a standard galaxy, and therefore x-ray luminosity functions can place constraints on the theoretical models of XRBs[6]. Another method for understanding the physical workings of XRBs is to compare how the x-ray luminosity of XRBs relates to the orbital period of the binary system. The process of mass transferring from the donor star to the compact object in an XRB can change both the luminosity and the period, and it was the goal of this project to analyze several XRBs and their luminosity and period and to see how they relate. The relation would hopefully provide insight into the general nature and evolution of XRBs, be they LMXBs or HMXBs.

II. DATA ACQUISITION AND ANALYSIS

The data used for this project was obtained by the All-Sky Monitor (ASM) aboard the Rossi X-ray Timing Explorer (RXTE) satellite. The RXTE contains three different instruments dedicated to detecting extra-solar x-

rays, the Proportional Counter Array (PCA), the High Energy X-ray Timing Experiment (HEXTE), and the All-Sky Monitor[7]. The ASM was capable of scanning 80% of the sky for every 90 minute orbit of the earth, making it useful for gathering data from a large number of sources[8]. The RXTE was launched on December 30, 1995 and remained operational until it was decommissioned on January 5, 2012[9]. Late in the operational lifespan of the RXTE mission, the ASM instrument began to suffer from some technical faults. This was most clearly noticeable in the data from the Crab nebula where the typically constant x-ray flux dipped for a period of approximately 200 days by about 12 counts/s, down from its usual 75 counts/s. It should be noted that nothing was done to attempt to correct for this in the analysis.

The data provided by NASA's Godard Space Flight Center Guest Observer Facility offered the observations made of approximately 350 sources over the course of sixteen years, from January 1, 1996 until January 1, 2012, presenting the x-ray intensity, or count rate, for each source from 90 second dwells in a FITS format. This data has been further compiled and analyzed by a team at MIT, which offered the data in ASCII format, as well as providing plots of the data for the 2011 year for visual reference. In addition to providing the dwell-by-dwell data, the MIT team has provided one-day averages of the count rates for each of the sources.

The visual plots of the 2011 data were invaluable in selecting sources for analysis. By reviewing the plots of the data, any source that appeared to have ten or more "three-sigma" detections was noted down for further analysis. The "three-sigma" detections are the one-day averaged data points with smaller estimated uncertainties, or certainties to within three-sigma. The reasoning for this was the thought that the sources with numerous three-sigma detections would have the high or variable luminosities that were desired for further analysis.

A total of sixty-three data sources were selected from the initial review, plus an additional eight sources were selected after a brief literature search for other possible strong sources. It is however possible that some strong sources may have been passed over if the 2011 data was unremarkable, and if the literature search turned up no results.

For each of the seventy-one sources, the ASCII data for the one-day averages were copied from the MIT website, which included the date of the observation in Modified Julian Days (MJD), the count rate, and the count rate error. The ASCII files were run through a Perl script to obtain five-day averages of the count rate in an attempt to further simplify the data. From here the data was plotted using the graphing software OriginPro 8 to obtain a visual reference for all the data from each source.

A somewhat arbitrary selection process was used from this point to determine which sources were viable candidates for continued analysis. Numerous
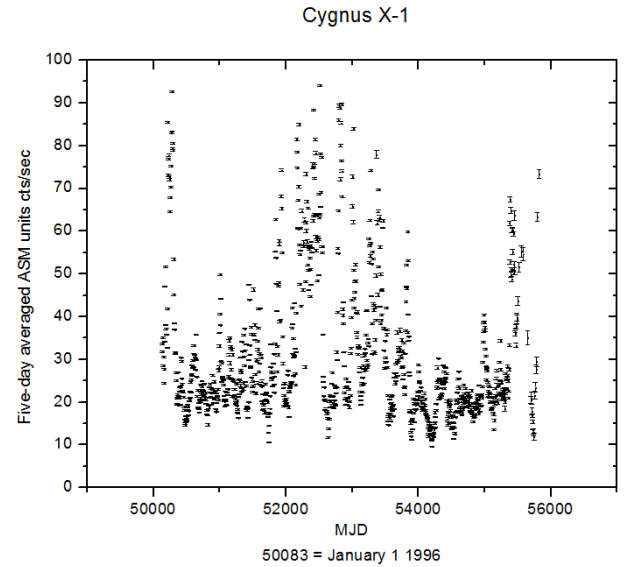
FIG. 1. Plot of the five-day averaged ASCII data showing the x-ray intensity in counts/s from the source Cyg X-1 from January 1, 1996 to January 1, 2012.

sources were rejected for their average count rates were rather low, not exceeding 10 to 15 counts per second. There were twenty-nine sources that had proportionately higher average count rates, or in some cases if the source experienced a strong, short burst of x-rays, and were selected for further analysis. In figure 1, the plot of the count rate for Cyg X-1 is shown as an example.

Once the twenty-nine sources had been selected, a literature search was then begun in order to classify the XRBs as either HMXBs or LMXBs, and whether the companion star was a neutron star or black hole. Additional facts about each source that were required were the orbital period of the XRB, the distance of the system, and the hydrogen column density (N_H), all of which are listed on table 1.

Most of the values were gathered from various catalogs, although there were notable exceptions. While most distances were known, for some it was required to estimate a distance, for which an estimate of 8kpc was used. Each of the sources with an unknown distance had the prefix GX, which indicates the source likely lies near the galactic core, at a distance of approximately 8kpc. For some sources the known distance was given as a range of values. In these cases the average between the two extremes was used as the distance. For the cases in which the hydrogen column density was unknown, the relation $N_H = 5.3 \times 10^{21} \text{cm}^{-2} \, E_{B-V}$ found in Predehl et. al. 1994 was used, where the values for interstellar reddening (E_{B-V}) were readily found in the literature, and thus enabling NH to be found.

Once the values for the hydrogen column density were known, they were used in conjunction with the average count rate, which had been determined by

taking the weighted average of the count rate for all data points, to determine the average x-ray flux of each source. Using the Portable, Interactive Multi-Mission Simulator (PIMMS v4.4) made available on-line by the Chandra X-ray Center, the count rate was converted to the unabsorbed flux of the source in units of ergs/cm^2/s. The PIMMS program is designed to enable the conversion of flux and count rates from various missions, including the RXTE ASM data. In its flux calculations for each source PIMMS requires the hydrogen column density, and the photon index, a necessary parameter for power law calculations where intensity is proportional to the energy of the photons to the negative power of the photon index. For all calculations, a photon index of 2 was used, which corresponds to the inverse square law.

Armed with the x-ray flux for each source, the luminosity could easily be determined by multiplying the flux by 4π times the distance in centimetres, giving the x-ray luminosity in ergs/s. Finally plots of the luminosity versus period could be created from the data. For seven of the twenty-nine source the orbital periods were unknown, and were therefore not included in the plots. Three plots were created, one plot displaying the neutron star LMXBs, another with the black hole LMXBs, and the final with all the HMXBs together.

III. RESULTS

The plots of the data yielded some interesting results. In the plot of the LMXBs with neutron star companions (figure 2.1), which contained twelve of the twenty-nine sources, the data did not behave entirely as expected. In Podsiadlowski et. al. 2001 an analysis done on LMXBs and IMXBs figures that the luminosity increases as the orbital period increases, although it should be noted that in the paper the mass transfer rate is used rather than the luminosity, which is proportional by the relation 10^{-8}Msun/yr$=10^{38}$ergs/s. The mass transfer rate is thought to be proportional to the luminosity by the fact that the x-rays emitted by an XRB are caused by the highly energetic in falling mass from the donor star to the compact object, with a greater mass transfer rate resulting in more x-rays being emitted, and therefore a proportional relation exists. Figure 3, taken from the Podsiadlowski paper, illustrates the relation. In contrast the results from the ASM data indicates that for periods on the order of 20 hours or less the luminosity varies from 2.44×10^{36}ergs/s in the case of Aql X-1, to $5.82\text{x}10^{38}$ergs/s in the case of Sco X-1, both of which have very similar periods. However for longer periods on the order of 100s of hours all the sources display consistently higher luminosities.

The plot of the LMXBs with black hole companions (figure 2.2), of which there were five sources, returned more predictable results, where the shorter orbital periods had lower luminosities, and longer periods had

(1) LMXBs with neutron star companions.

(2) LMXBs with black hole companions.
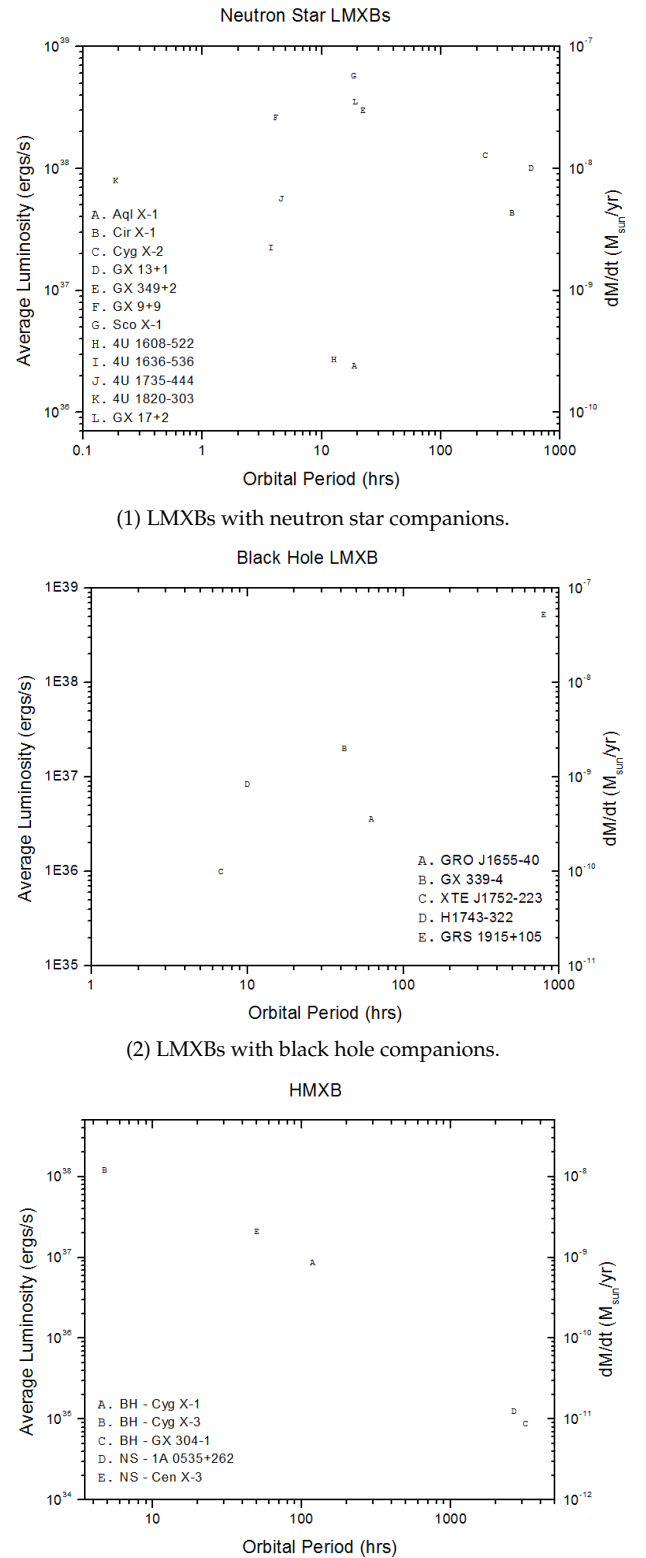
(3) HMXBs with both neutron star and black hole companions. The neutron star and black hole companions are distinguished by the NS and BH labels on the plot, respectively.

FIG. 2. The average x-ray luminosity and the mass transfer rate are plotted together on the y-axis.

Source	Period	Distance	Classification	Hydrogen Column Density N_H (cm^{-2})	Luminosity (ergs/s)
Aql X-1	18.95h (Liu 2008)	5kpc (Liu 2008)	NS, LMXB	3.60E+21 (Bodaghee 2008)	2.44E+36
Cir X-1	398.4h (Liu 2008)	5.5kpc (4-12kpc) (Liu 2008)	NS, LMXB	1.60E+22 (Bodaghee 2008)	4.36E+37
Cyg X-2	236.2h (Liu 2008)	7.2kpc (Liu 2008)	NS, LMXB	1.90E+21 (Bodaghee 2008)	1.30E+38
GX 13+1	577.6h (Liu 2008)	71kpc (Liu 2008)	NS, LMXB	3.20E+22 (Bodaghee 2008)	1.01E+38
GX 349+2	22.5h (14.9d?) (Liu 2008)	9.2kpc (Liu 2008)	NS, LMXB	6.73E+21 (Bodaghee 2008)	3.01E+38
GX 9+9	4.20h (Liu 2008)	10kpc (?)	NS, LMXB	2.60E+21 (Predehl 1995)	2.63E+38
Sco X-1	18.90h (Liu 2008)	2.8kpc (Liu 2008)	NS, LMXB	2.37E+22 (Bodaghee 2008)	5.82E+38
4U 1608-522	12.89h (Liu 2008)	3.6kpc, 4.1kpc (Liu 2008)	NS, LMXB	1.28E+22 (Bodaghee 2008)	2.73E+36
4U 1636-536	3.80h (Liu 2008)	6kpc (Liu 2008)	NS, LMXB	4.20E+21 (Bodaghee 2008)	2.27E+37
4U 1735-444	4.65h (Liu 2008)	7-9kpc (Liu 2008)	NS, LMXB	3.40E+21 (Bodaghee 2008)	5.67E+37
4U 1820-303	0.19h (Liu 2008)	7.6kpc (Liu 2008)	NS, LMXB	1.60E+21 (Bodaghee 2008)	8.00E+37
GX 17+2	19.5h (Hertz 1986)	9.8kpc (Liu 2008)	NS, LMXB	2.00E+22 (Bodaghee 2008)	3.52E+38
GX 3+1	unknown	6.5kpc (Liu 2008)	NS, LMXB	1.59E+22 (Bodaghee 2008)	7.16E+37
GX 340+0	unknown	8kpc (?)	NS, LMXB	3.90E+22 (Bodaghee 2008)	1.78E+38
GX 5-1	unknown	8kpc (?)	NS, LMXB	2.20E+22 (Bodaghee 2008)	3.73E+38
GX 9+1	unknown	5kpc (Liu 2008)	NS, LMXB	8.00E+21 (Bodaghee 2008)	3.45E+37
Ser X-1	unknown	8.4kpc (Liu 2008)	NS, LMXB	5.00E+21 (Bodaghee 2008)	8.05E+37
4U 1705-440	unknown	7.4kpc (Liu 2008)	NS, LMXB	1.42E+22 (Bodaghee 2008)	4.77E+37
GRO J1655-40	62.88h (Liu 2008)	3.2kpc (but <1.7kpc) (Liu 2008)	BH, LMXB	5.80E+21 (Bodaghee 2008)	3.61E+36
GRS 1915+105	33.5d (Liu 2008)	11.2-12.5kpc (Liu 2008)	BH, LMXB	1.98E+22 (Bodaghee 2008)	5.26E+38
GX339-4	42.14h (Liu 2008)	>6kpc (8kpc?) (Liu 2008)	BH, LMXB	3.90E+21 (Bodaghee 2008)	2.01E+37
XTE J1752-223	<6.8h (Ratti 2012)	3.5-8kpc (Ratti 2012)	BH, LMXB	5.00E+21 (Bodaghee 2008)	1.00E+36
H1743-322	>10h (Jonker 2009)	8.5kpc (Liu 2008)	BH, LMXB	1.84E+22 (Predehl 1995)	8.44E+36
Swift J1753.5-0127	unknown	6kpc (Liu 2008)	BH, LMXB	2.00E+21 (Bodaghee 2008)	1.54E+36
Cyg X-1	5d (Liu 2008)	2.14kpc (Liu 2008)	BH, HMXB	6.21E+21 (Bodaghee 2008)	8.80E+36
Cyg X-3	0.2d (Liu 2008)	9kpc (Bodaghee 2008)	BH, HMXB	8.50E+22 (Bodaghee 2008)	1.21E+38
GX 304-1	133d? (Liu 2008)	2.4kpc (Liu 2008)	BH, HMXB	1.06E+22 (Predehl 1995)	8.93E+34
1A 0535+262	111d (Liu 2008)	2-2.9kpc (Liu 2008)	NS, HMXB	6.50E+21 (Bodaghee 2008)	1.27E+35
Cen X-3	2.09d (Liu 2008)	8-10kpc (Liu 2008)	NS, HMXB	1.95E+22 (Bodaghee 2008)	2.12E+37

TABLE I. XRBs from RXTE ASM data. The twenty-nine sources are listed along with orbital period, distance, classification, and hydrogen column density values found in literature searches. The calculated average x-ray luminosity is also included.
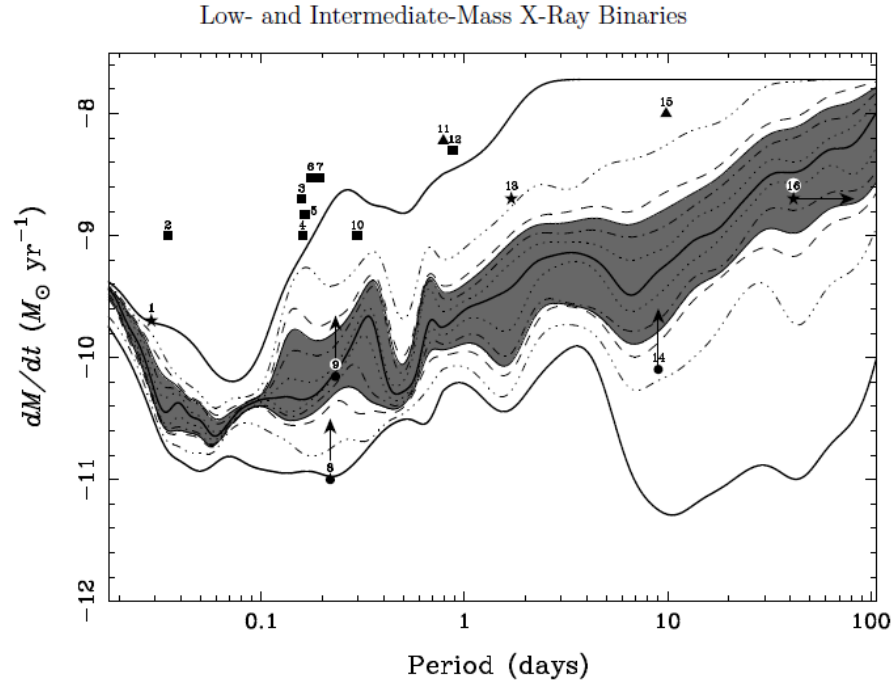
Low- and Intermediate-Mass X-Ray Binaries

FIG. 18.— Cumulative (smoothed) probability distribution for the mass-accretion rate onto the neutron-star primary as a function of orbital period for the 100 binary calculations (with equal weighting for all sequences). The thick, solid central curve gives the median mass-accretion rate; the pairs of curves moving progressively outwards from the median curve include 20, 40, 60, 80, and 98 per cent of the distribution, respectively. The shaded region contains 50 per cent of all systems around the median. The symbols indicate the mass-transfer rates of selected observed X-ray binaries (triangles: Z sources; squares: atoll sources; stars: X-ray pulsars; circles: systems with accretion-disk coronae [lower limits]). The individual systems are (in ascending order of orbital period, as given in parentheses): (1) 1626-67 (0.69 hr), (2) 1916-053 (0.83 hr), (3) 1636-536 (3.8 hr), (4) 0748-676 (3.82 hr), (5) 1254-690 (3.93 hr), (6) GX9+9 (4.2 hr), (7) 1735-555 (4.65 hr), (8) 2129+470 (5.24 hr), (9) 1822-37 (5.57 hr), (10) 1658-29 (7.11 hr), (11) Sco X-1 (18.9 hr), (12), 1624-590 (21 hr), (13) Her X-1 (40.8 hr), (14) 0921-630 (216 hr), (15) Cyg X-2 (236 hr), (16) GX1+4 (> 1000 hr) (taken from the catalogs by van Paradijs 1995 and Christian & Swank 1997).

FIG. 3. Mass transfer rate versus period plot taken from the Podsiadlowski et. al. 2001 paper. The figure description from the paper has been retained.

much higher luminosities. It is interesting to note that if the black hole LMXB data was plotted together with the neutron star LMXB data, the black hole LMXB sources fell in to clusters with the neutron star LMXB sources, indicating that the type of compact body, be it a neutron star or black hole, has no influence on the period-luminosity relation.

The plot of the HMXBs (figure 2.3), which displayed the remaining five sources, displayed results that ran somewhat contrary to the LMXB data. The HMXBs clearly tended to decrease in luminosity as the period increased. There were no prior expectations for the HMXB data; however this does clearly show that the relation between luminosity and period behaves differently between HMXBs and LMXBs. It should be noted that all the HMXBs were plotted together, regardless as to whether the compact object was a neutron star or black hole, due to a shortage of data. Much like the LMXB data though, the classification of the compact body does not seem to have any relevance to the period-luminosity relation.

IV.   CONCLUSION

In conclusion, there needs to be further investigation into the relationship between period and luminosity for XRBs. While there seems to be a definite relation for HMXBs, the case is much different for LMXBs. The vast discrepancies in luminosity despite similar orbital periods for some LMXBs suggest the period-luminosity relation may not be as straight forward as previously thought. One avenue for further investigation could include determining more details, beyond mass of donor star and classification of the compact body companion, of some of the LMXBs as it may be the case of their being more factors at play in a period-luminosity.

In short, more research is required to yield more definite results regarding the relationship between x-ray luminosity and period, however preliminary results do look promising. I believe it is likely that the work done on this project will lead to results regarding a period-luminosity relation for XRBs in the near future, or if nothing else to further our current understanding of XRBs.

REFERENCES

[1-3,5] B.W. Carroll, D.A. Ostlie. *An Introduction to Modern Astrophysics.* Pearson, 2007.

[4] P. Podsiadlowski, S Rappaport, E. Pfahl. Evolutionary Binary Sequences for Low- and Intermediate-Mass X-ray Binaries. arXiv:astro-ph/0107261v1 (December 2, 2012).

[6] M. Tremmel et.al. Modeling the Redshift Evolution of the Norma Galaxy X-ray Luminosity Function. arXiv:1210.7185v2 [astro-ph.CO]. (December 2, 2012).

[7]   http://heasarc.gsfc.nasa.gov/docs/xte/xte2.html. (December 2, 2012).

[8]   http://heasarc.gsfc.nasa.gov/docs/xte/ASM.html. (December 2, 2012).

[9]   http://heasarc.gsfc.nasa.gov/docs/xte/XTE.html. (December 2, 2012).

A Brick-Sorting LEGO Robot

Arturo Pérez and Michael R. W. Dawson
Department of Psychology, University of Alberta

Arturo Pérez is a senior undergraduate psychology student at Universidad Diego Por-tales in Santiago, Chile. In the Fall term of 2012, Arturo spent 3 months at the Universi-ty of Alberta, hosted by Dr. Michael Dawson and the Biological Computation Project (BCP). The general goal of his visit was to establish collaborative ties between this U of A laboratory and the Centro de Estudios de la Argumentación y el Razonamiento (CEAR) at UDP. A more specific purpose was to explore the BCP's approach to using simple robots to explore basic ideas in embodied cognitive science. Arturo's explora-tions involved creating, programming, and testing a new robot designed to sort ele-ments in an arena. The purpose of the current paper is to report on Arturo's robotics research at the BCP.

PURPOSE

Behavior-based robots are simple machines built from a number of basic sense-act reflexes (Brooks, 1999; Sharkey, 1997). They do not create and use complex represen-tations of their external world to plan and guide their behavior. Instead, their external world serves as its own representation; complex behaviors emerge because the robot's sensors immediately detect properties in the world, properties that trigger adaptive responses tuned to environmental stimuli.

Research on behavior-based robots is often biomimetic (Sharkey, 2006; Webb & Consi, 2001). Researchers, seeing the complex behavior of simple organisms in the natural world, explore the possibility of achieving similar behavior in simple robots by exploiting sense-act processing. For instance, robotic models of cricket phonotaxis, in which female crickets orient to particular cricket songs to choose a mate, have resulted in extraordinarily simple and plausible new theories of this phenomenon (Webb, 1996; Webb & Scutt, 2000).

One natural behavior that has inspired a great deal of robotic research is the brood sorting of ants (Franks & Sendova-Franks, 1992; Sendova-Franks, Scholes, Franks, & Melhuish, 2004). Many species of ants establish their colony in what is in essence a two-dimensional space, for instance, spread out in a plane that is covered by a large flat rock. As the eggs laid by the queen ant develop into larvae, they grow in size and require greater care. Workers in the colony facilitate this care by moving larger eggs and larvae towards the perimeter of the nest. Thus, if one quickly removed the rock covering the colony, they would see its brood organized spatially in an approximately circular arrangement; as one moved in any direction from the center of the brood one would note an increase in the size of the arranged eggs and larvae. Brood sorting has inspired several studies that attempt to achieve sorting behavior in behavior-based robots (Deneubourg et al., 1991; Holland & Melhuish, 1999; Melhuish, Sendova-Franks, Scholes, Horsfield, & Welsby, 2006; Scholes, Wilson, Sendova-Franks, & Melhuish, 2004; Wilson, Melhuish, Sendova-Franks, & Scholes, 2004).

One example of such a device is a simple Lego NXT robot named the Lemming (Dawson, Dupuis, & Wilson, 2010). The Lemming moves about a rectangular arena on wheels driven by two separate motors. It explores around its environment when its motors rotate at different speeds. An ultrasonic sensor mounted on its top detects obstacles, and causes the robot to steer away from them. An assortment of Lego bricks is scattered throughout the Lemming's environment. The robot steers towards bricks detected by a lower ultrasonic sensor. When captured by the robot's 'plow', a light sensor determines whether a brick is white or black. The robot deposits white bricks close to walls, and deposits black bricks close to other bricks. All of these behaviors are accomplished with a simple sense-act hierarchy of reflexes called a subsumption architecture (Brooks, 1991; Brooks & Flynn, 1989; Brooks, 1989). The Lemming produces some surprising and unexpected sorting behaviors, and demonstrates swarm intelligence, because sorting efficiency increases exponentially as more Lemmings are added to the arena.

The sorting behavior of the Lemming is interesting and successful. However, some of its abilities are in obvious need of improvement. For example, the lower ultrasonic does not provide very efficient steering towards to-be-sorted bricks. As a result, it takes a single Lemming over an hour to sort the bricks in its testing arena (Dawson et al., 2010).

The purpose of the current research was to develop a brick-sorting robot using different sensors, which require a different subsumption architecture. Initially, the motivation for the research was to improve upon the performance of the original Lemming. However, once the new robot was constructed and programmed, our interest turned to exploring new sorting behaviors that it produced, behaviors that were quite different from those exhibited by the original robot.

This paper proceeds as follows: first, we briefly describe the embodiment of the ro-bot. Second, we define the basic behaviors that make up its subsumption architecture. Third, we illustrate the interesting sorting behaviors that it produces, and which depend upon the environment in which the robot is situated. We conclude by briefly discussing the implications of our robot.

THE EMBODIMENT OF 'ARTURITO'

We named the new robot Arturito, inspired by the first name of its inventor and acknowledging its vague resemblance to the Star Wars robot R2-D2. Arturito is approximately 8 inches high, supported by a pair of wheels and by a plow mounted on its front. The body of Arturito is a Lego NXT brick. On the top of the robot is an ultrasonic sensor used to mediate obstacle avoidance. The robot's plow enables brick capture and pushing. The shape of the plow is critical: when the robot spins to avoid an obstacle, it leaves a carried brick behind. The plow also supports light sensors for detecting and steering towards bricks, as well as for measuring the color of bricks after their capture. Figure 1 illustrates Arturito's embodiment.

THE CORE BEHAVIORS OF ARTURITO

Arturito's behavior is created by employing a subsumption architecture that defines a hierarchical arrangement of simple sense-act reflexes. This subsumption architecture was programmed in the NXC language using the Bricx programming environment, which is freely available on the internet.

The most basic level of Arturito causes its two motors to run at a base speed, driving the robot forward.

The second level of its architecture uses the ultrasonic sensor to avoid obstacles. When the ultrasonic detects that an obstacle is closer than a safe threshold, motor

FIG. 1. Arturito, the brick-sorting Lego NXT robot.

directions are manipulated to turn Arturito away.

The third level of Arturito's architecture uses two lower light sensors to alter the speed of the two motors. The steering mechanism is a variation of a common robot, Braitenberg Vehicle 2 (Braitenberg, 1984; Dawson et al., 2010). Originally, we hoped that when the light sensors detected reflections, the robot would turn towards any brick. Indeed, this approach was one of the more innovative aspects of Arturito, because we had never explored the steering capability of light sensors when the desired targets were not themselves lights. However, and as is discussed below, we discovered that this level produced sharp steering towards white bricks, but steered Arturito away from black bricks. Thus one behavior that emerged from level 3 was a taste for white bricks, and an aversion to black bricks.

The final level of Arturito's subsumption architecture was invoked when a brick was captured in the robot's plow. The color of the brick altered Arturito's obstacle threshold. The threshold decreased if the brick was white, with the intent of having white bricks deposited near arena walls. In contrast, the threshold increased if the brick was black, with the intent of having black bricks deposited further from arena walls. This level also inhibited Arturito's brick-seeking behavior, with the intent of having the robot transport a brick without simultaneously seeking another.

ARTURITO'S SORTING BEHAVIOR

We studied Arturito's sorting behavior in the same arena that housed the original lemming. The arena was a small lab room, approximately 8' x 6'; a camera mounted in the ceiling recorded brick sorting. We began by placing a regular arrangement of Lego bricks on the linoleum floor of the arena in a typical study. Figure 2 illustrates an example initial configuration of bricks. We randomly placed the robot in the room, activated it, and closed the door as we left the machine to run on its own. We conducted a number of simple studies as is detailed below. We were interested in Arturito's behavior when it encountered bricks that were all the same color, as well as various mixtures and ar-rangements of different colored bricks. The text below provides brief descriptions of our results; a video that illustrates the robot and its sorting behavior is also available at the following link: http://www.youtube.com/watch?v=jTRBgAuhFaY

A. White Brick Sorting

We began by observing how the robot would behave in a room with moderate light that only contained white bricks. The performance of the robot in these conditions indicated that our embodiment and programming

FIG. 2. An example starting state for robot sorting.

FIG. 3. Arturito clears white bricks from the arena's center.

were very successful. First, the robot turned to a detected brick as if it was looking for it; this behavior was far more effective than that observed in the Lemming. After scooping up a brick in its plow, it typically drove itself to a nearby wall, ignoring the presence of other bricks along its way. Finally, it turned away from the wall, depositing the brick that it carried. Second, brick sorting behavior was much faster than that seen in the Lemming. After only about 10-15 minutes, Arturito had pushed all of the white bricks near a wall, leaving the centre of the arena empty, as shown in Figure 3.

In some instances, Arturito's behavior was surprising. Sometimes when the robot had a brick in its plow it still steered towards others. In these situations, the second brick stayed just in front of the sensor, and Arturito moved around in spirals. However, eventually another brick would attract the robot, steering it in a direction that caused it to deposit its (multiple) bricks. This returned Arturito's behavior to normal.

B. Black Brick Sorting and a Taste for White Bricks

In a second study, Arturito's arena contained only black bricks. We expected in this case that there would merely a quantitative difference between this study and the one described above. That is, we expected that Level 3 would cause the black bricks to be arranged in a similar fashion to the white bricks, with the exception that they would be further from the walls.

Instead, we surprisingly found that Arturito steered away from the black bricks. The black bricks reflected light back to the light sensors in such a way that the re-

sulting motor behaviors caused aversion instead of attraction. In the spirit of Braitenberg (1984), we might say has a "taste", it likes white bricks and dislikes black bricks. Note that we obtained this selective behavior ''for free', in the sense that we did not explicitly write different code to cause the machine to steer differently in response to different colored bricks. Instead, Arturito's "taste" emerges from an interaction between its Level 2 architecture and the way in which light reflects from different colored bricks.

The result of all of this is that in contrast to the previous study, Arturito leaves the center of the arena occupied by an expanse of black bricks. A typical final configuration is provided in Figure 4.

C. Sorting Black and White Bricks

Many of our studies examined Arturito's performance when placed in an arena that contained both white and black bricks. In this kind of situation, the robot's behavior was quite interesting because of its tendency to steer towards white bricks and to steer away from black bricks, combined with its obstacle-avoiding threshold being determined by the color of brick that was in the robot's plow. We were able to produce a wide variety of sorting behaviors by experimenting with the relative proportions of white and black bricks, as well as with their initial arrangement.

FIG. 4. Arturito leaves black bricks in the middle of the arena.

FIG. 5. Arturito has sorted the configuration of bricks from Figure 2.

When the robot was confronted with a regular arrangement of black and white bricks in a regular arrangement (Figure 2), it sorted bricks in a fashion that was similar to, but more efficient than, the Lemming. For example, the sorted pattern presented in Figure 5 required only 21 minutes, in comparison to the 70+ minutes that the lemming would require to produce a similar result.

Interestingly, Figure 5 indicates that Arturito moved

more black bricks out of the arena than was the case when only black bricks were present (Figure 4). This reflects the influence of the white bricks in the configuration, which attracted Arturito – and caused it to pick up a black brick as it approached the white one! As well, the pattern of Figure 5 is analogous to the annular arrangement that is sought in models of brood sorting (Wilson et al., 2004). This raises questions for further research. How much of annular brood sorting is due to a preference for one type of brood item over another? To what extent is the annular arrangement of sorted items dependent upon the size and shape of the environment?

In other studies, we observed how the robot would behave with different proportions of bricks of different color. In one, we selected a 3:1 proportion of black to white bricks, placing the white bricks in the center and the black ones around them, making some kind of a surrounding 'membrane'. Generally, the robot was repelled by this 'membrane', but eventually broke through it and started the sorting of white bricks. We also noted that black bricks sometimes made it harder for the robot to sort particular bricks by "hiding" them, making almost impossible for the robot to sort it because every time the robot came near it was repelled by the black bricks. Finally, particular arrangements of bricks that emerged in the middle of sorting could alter Arturito's trajectory through the arena. For example, if a straight line of black bricks emerged, then the robot would travel parallel to this line. Many of these intriguing behaviors can be seen in the video whose link was provided earlier.

IMPLICATIONS AND CONCLUSION

Our observations of Arturito suggest that we successfully achieved the initial goals of the robot project. By altering the embodiment of the Lemming, and by revising the subsumption architecture, we developed a robot that achieved brick sorting far more efficiently. The robot's unexpected preference for one color of brick over another led to some interesting sorting patterns; the Lemming would not produce the annular ar-rangement in Figure 5.

Arturito's seeking of white bricks, and avoiding of black ones, suggests future studies that explore environmental influences to a much greater degree than is reported here, or than was used to study the Lemming. Of great interest is the extent to which the ini-tial configuration of differently colored bricks affects the final pattern of sorting. In particular, some initial arrangements might either facilitate or hurt sorting; might they also determine the final arrangement of sorted items? As well, we expect that the size and the shape of the robot's arena will also have a profound influence on the final sorting result.

One aspect studied in more detail with the Lemming was the emergence of collective intelligence. Increasing the number of Lemmings from 1 to 2 to 3 produced an

exponential increase in sorting efficiency, which is a hallmark of collective intelligence (Sugawara & Sano, 1997; Sulis, 1997). When two Arturitos work together, sorting behavior is very quick. However, whether this increase is exponential or not requires studying the behavior of a collective of 3 or more robots, which we have not done to date.

One of the interesting implications of this project is the extent to which sorting behavior is affected by the arrangement of different elements that elicit different behaviors (approach or avoidance). This raises the obvious question about whether brood sorting by ants is governed by similar principles. The biomimetic nature of behavior-based robotics clearly is bidirectional, as robot behavior can raise questions to be studied using biological agents in their natural setting (Webb, 2000).

REFERENCES

Braitenberg, V. (1984). *Vehicles: Explorations In Synthetic Psychology.* Cambridge, MA: MIT Press.

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence, 47*, 139-159.

Brooks, R., & Flynn, A. M. (1989). Fast, cheap and out of control: A robot invasion of the solar system. *Journal of The British Interplanetary Society, 42*, 478-485.

Brooks, R. A. (1989). A robot that walks; emergent behaviours from a carefully evolved network. *Neural Computation, 1*, 253-262.

Brooks, R. A. (1999). *Cambrian Intelligence: The Early History Of The New AI.* Cambridge, MA: MIT Press.

Dawson, M. R. W., Dupuis, B., & Wilson, M. (2010). *From Bricks To Brains: The Embodied Cognitive Science Of LEGO Robots.* Edmonton, AB: Athabasca University Press.

Deneubourg, J. L., Goss, S., Franks, N., Sendova-Franks, A., Detrain, C., & Chretien, L. (1991). The dynamics of collective sorting robot-like ants and ant-like robots, *Proceedings of the first international conference on simulation of adaptive behavior (From animals to animats)* (pp. 356-363). Paris, France: MIT Press.

Franks, N. R., & Sendova-Franks, A. B. (1992). Brood sorting by ants: Distributing the workload over the work surface. *Behavioral Ecology and Sociobiology, 30*(2), 109-123.

Holland, O., & Melhuish, C. (1999). Stigmergy, self-organization, and sorting in collective robotics. *Artificial Life, 5*, 173-202.

Melhuish, C., Sendova-Franks, A. B., Scholes, S., Horsfield, I., & Welsby, F. (2006). Ant-inspired sorting by robots: the importance of initial clustering. *Journal of the Royal Society Interface, 3*(7), 235-242.

Scholes, S., Wilson, M., Sendova-Franks, A. B., & Melhuish, C. (2004). Comparisons in evolution and engineering: The collective intelligence of sorting. *Adaptive Behavior, 12*(3-4), 147-159.

Sendova-Franks, A. B., Scholes, S. R., Franks, N. R., & Melhuish, C. (2004). Brood sorting by ants: two phases and differential diffusion. *Animal Behaviour, 68*, 1095-1106.

Sharkey, A. J. C. (2006). Robots, insects and swarm intelligence. Artificial Intelligence Review, 26(4), 255-268.

Sharkey, N. E. (1997). The new wave in robot learning. *Robotics and Autonomous Systems, 22*(3-4), 179-185.

Sugawara, K., & Sano, M. (1997). Cooperative acceleration of task performance: Foraging behavior of interacting multi-robots system. *Physica D, 100*(3-4), 343-354.

Sulis, W. (1997). Fundamental concepts of collective intelligence. *Nonlinear Dynamics, Psychology, and Life Sciences, 1*, 35-53.

Webb, B. (1996). A cricket robot. *Scientific American, 275*, 94-99.

Webb, B. (2000). What does robotics offer animal behaviour? *Animal Behaviour, 60*, 545-558.

Webb, B., & Consi, T. R. (2001). *Biorobotics : methods and applications.* Menlo Park, CA: AAAI Press/MIT Press.

Webb, B., & Scutt, T. (2000). A simple latency-dependent spiking-neuron model of cricket phonotaxis. *Biological Cybernetics, 82*(3), 247-269.

Wilson, M., Melhuish, C., Sendova-Franks, A. B., & Scholes, S. (2004). Algorithms for building annular structures with minimalist robots inspired by brood sorting in ant colonies. *Autonomous Robots, 17*(2-3), 115-136.

Giant Steps In The Interpretation Of A Musical PDP Network

Joshua Hathaway and Michael R. W. Dawson
Department of Psychology, University of Alberta

INTRODUCTION

The key theme of research on musical cognition is discovering the mental representations used to organize and interpret music (Deutsch, 1999; Krumhansl, 1990; Lerdahl & Jackendoff, 1983; Sloboda, 1985. More recently, this approach has been extended to include the study of brain mechanisms related to such representations; that is, the cognitive neuroscience of music (Levitin, 2006; Peretz & Zatorre, 2003). Artificial neural networks are biologically inspired models of cognition (Rumelhart & McClelland, 1986), and interpretations of their internal structure can lead to insights about the nature of representation (Dawson, 2004). Thus, given the rise of both musical cognition and of the cognitive neuroscience of music, it is not surprising that artificial neural networks are frequently used to explore musical cognition (Griffith & Todd, 1999; Todd & Loy, 1991).

An artificial neural network is a computer simulation of a "brain-like" system of interconnected processing units; it generates a desired response to an input stimulus. Activities of a set of input units encode a stimulus. The response of the system is represented as activities in a set of output units. The signal sent by one input processor to an output unit is transmitted through a weighted connection. At the start of a simulation, a network has small, randomly assigned connection weights. The network is then taught by presenting it a set of training patterns, each of which is associated with a known correct response. Each pattern is presented to the networks input units, and on the basis of its existing connection weights, the network generates a response to it. An error term for each output unit is calculated by measuring the difference between the desired response of the unit and its actual response. A learning rule is used to change connection weights in such a way that network error is reduced. Ideally, with repeated presentation of the training patterns, the network learns to generate a correct response to each stimulus.

The current paper describes a musical investigation involving an artificial neural network. We taught a neural network to generate the sequence of chords in a jazz progression called the Coltrane changes. We interpreted the internal structure of this network to reveal an elegant and simple formalism that represents this progression in its entirety. We also used the structure of this network to investigate the complexity of the Coltrane changes in relation to other chord progressions.

We first introduce the notion of chord progressions by describing a particular example (the II-V-I) that is related to the Coltrane changes. Second, we describe

the Coltrane changes using a formalism derived from previous musical investigations with neural networks (Yaremchuk & Dawson, 2005, 2008). Finally, we describe how we trained a neural network to generate the Coltrane changes, how we analyzed its internal structure, and the implications of this interpretation. In particular, we discovered that a network represented transitions between chords in a fashion that could be described in terms of a new musical formalism that we had not envisioned. In short, this paper shows that the interpretation of the internal structure of a musical network can provide new formalisms for representing musical regularities, and can suggest new directions for representational research on musical cognition.

THE II-V-I CHORD PROGRESSION

Most jazz pieces are essentially song structures in which musicians play sequences of chords called chord progressions (Sudnow, 1978). Certain chord progressions are popular because the transition from chord to chord is musically pleasing, and because the progression permits moving from one musical key to another, permitting flexibility (one can start the same progression in any key) and providing musical variety.

This is illustrated in one important chord progression called the II-V-I, a progression that is likely the most commonly encountered in jazz (Levine, 1989). This progression starts in a particular musical key. For example, in the key of C major this progression starts by first playing the D minor seventh chord (Dm7), then by playing the G dominant seventh chord (G7), and ends by playing the C major seventh chord (Cmaj7). In the key of C this is a II-V-I progression because D, the root note of Dm7, is the second note of the C major scale; G, the root note of G7, is the fifth note of the C major scale; and C, the root note of Cmaj7, is the first note of the C major scale.

The II-V-I progression shifts to a new musical key by following the major seventh chord that ends the progression in one key with a minor seventh chord built on the same root that defines the II chord of the next key. For instance, after ending the progression above with Cmaj7, the next chord played is C minor seventh (Cm7), which is the II chord of a new musical key, A#. Table 1 provides a complete II-V-I progression that begins in the key of C major, and ends by returning to this key an octave lower.

When learning to play an instrument, students usually also learn to read musical notation, and are taught the elements of music theory (Martineau, 2008). In

Chord Progression For Key			
Key	**II**	**V**	**I**
C	Dm7	G7	Cmaj7
A#	Cm7	F7	A#maj7
G#	A#m7	D#7	G#maj7
F#	G#m7	C#7	F#maj7
E	F#m7	B7	Emaj7
D	Em7	A7	Dmaj7
C	Dm7	G7	Cmaj7

TABLE I. The II-V-I chord progression, descending from the key of C major. The chords in each row are played in sequence, and after playing one row, the next row is played.

learning such elements, students typically encounter the circle of fifths (Figure 1), if only to determine the number of sharps or flats associated with each musical key signature. Importantly, the circle of fifths captures a great deal of formal structure. For instance, consider a scale written in the key of C (the top of the circle, Figure 1). Scales written in the keys adjacent to C in the circle of fifths (i.e., the keys of F or G) are maximally similar to it, sharing all but one note (Krumhansl, 1990). In other words, the circle of fifths is literally a map in which the more related two keys are, the closer together they are in the circle.

Interestingly, the circle of fifths also maps the root notes of the chords in the II-V-I progression. Consider the chord that starts the Table 1 progression, Dm7. Its root note is D. If one finds that note on the circle of fifths and then moves in a counterclockwise direction, then the next note encountered is the root of the V chord, G7. The next note encountered is C, the root of the I chord, Cmaj7. If one starts again from this final C, the next three notes provide the II-V-I roots for the next key in the progression A#.

Of relevance to the current paper, one can teach the II-V-I progression to an artificial neural network. To do this, a set of input units represents possible musical notes (e.g., each unit represents a particular piano key). The output units represent another set of musical notes in a similar fashion. One presents a chord to the network by turning on the input units that define its component notes. The network is then trained with an error-correcting learning rule (Dawson, 2004, 2005) to turn on the output units that define the notes of the next chord in the progression. Interestingly, an extremely simple artificial neural network called a perceptron (Rosenblatt, 1958) can learn the II-V-I progression of Table 1 (Dawson, 2013). This network has direct connections between input units and output units, and has no hidden units that are the defining property of more powerful artificial neural networks.

The purpose of the current paper is to explore a more complicated chord progression that is an elaboration of the II-V-I progression. This second chord progression is the Coltrane changes, named after its creator, legendary saxophonist John Coltrane (Turner, 1975). We investi-

gate two different questions related to training an artificial neural network to generate the Coltrane changes. First, given that this progression is more complicated than the II-V-I, is a more complex neural network required to generate it? Second, if a neural network can learn to generate the Coltrane changes, then what musical regularities does it represent in its internal structure?

THE COLTRANE CHANGES

John Coltrane introduced the chord progression now known as the Coltrane changes on his seminal 1960 album Giant Steps, where it is central to two pieces, "Giant Steps" and "Countdown". The latter piece paid homage to Miles Davis famous composition "Tune Up" which uses the II-V-I. The Coltrane changes are an elaboration of the II-V-I; it includes the three chords of this older progression, but adds four more chords. Two of these are lead-in chords to the V, and the other two are lead-in chords to the I. Adding additional chords to existing progressions reflects Coltranes evolving musical sensitivity. In the liner notes of Giant Steps he notes, "I feel like I cant hear but so much in the ordinary chords we usually have going in the accompaniment." Table 2 provides the Coltrane changes that reflect the elaboration of the Table 1 progression.

In addition to the particular sequence of chords given in Table 2, we also assumed that chord inversions were used to minimize finger movements (assuming that the progression is being played on a keyboard). Consider the C major 7 tetrachord. In root position, its lowest note is C, followed by the notes E, G, and B. To create the first
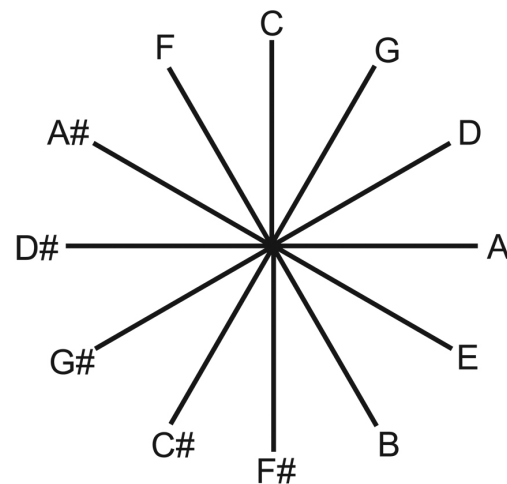

FIG. 1. The circle of fifths. Moving clockwise around the circle, adjacent notes are an interval of a perfect fifth (or seven semitones) apart. Moving counterclockwise around the circle, adjacent notes provide the root notes of chords in the II-V-I progression, as explained in the text.

Chord Progression For Key						
Key	**II**	**Lead-in 1**	**Lead-in 2**	**Lead-in 3**	**Lead-in 4**	**V**	**I**
C	Dm7	D#7	G#maj7	B7	Emaj7	G7	Cmaj7
A#	Cm7	C#7	F#maj7	A7	Dmaj7	F7	A#maj7
G#	A#m7	B7	Emaj7	G7	Cmaj7	D#7	G#maj7
F#	G#m7	A7	Dmaj7	F7	A#maj7	C#7	F#maj7
E	F#m7	G7	Cmaj7	D#7	G#maj7	B7	Emaj7
D	Em7	F7	A#maj7	C#7	F#maj7	A7	Dmaj7
C	Dm7	D#7	G#maj7	B7	Emaj7	G7	Cmaj7

TABLE II. A progression of Coltrane changes, descending from the key of C major. The chords in each row are played in sequence and after playing one row the next row is played.

inversion of the chord, the C is raised an octave, so that E is now the lowest note. To create the second inversion of the chord, the lowest note of the first inversion (E) is raised an octave, so that G is now the lowest note. To create the third inversion of the chord, the lowest note of the second inversion (G) is raised an octave, so that B is now the lowest note. To minimize finger movements in the Coltrane changes, inversions were applied to the Table 2 chords as follows: the Lead-in 1 chord is played in root position; the II chord, the Lead-in 3 chord and the I chord are played in the first inversion; the Lead-in 2 chord is played in the second inversion; the Lead-in 4 chord and the V chord are played in the third inversion. The use of inversions is evident in the musical score that is presented later as Figure 4.

It is clear that the progression in Table 2 is more complex than the progression in Table 1. Furthermore, the regularities that define the Coltrane changes are notoriously difficult to explain (Capuzzo, 2006; Waters, 2010). In the Giant Steps liner notes, Coltrane describes it as follows: "the bass line is kind of a loping one. It goes from minor thirds to fourths, kind of a lop-sided pattern in contrast to moving strictly in fourths or in half-steps." Even Coltranes accomplished band found his changes hard to master during the rehearsal sessions for the Giant Steps album (Turner, 1975).

To build a training set for teaching the Coltrane changes to an artificial neural network, we discovered a graphical representation of this progression, which is provided in Figure 3 below. This representation takes advantage of the fact that previous interpretations of musical neural networks (Yaremchuk & Dawson, 2008) have revealed musical circles analogous to, but different from, the circle of fifths in Figure 1. Some networks encode musical knowledge using the four circles of major thirds presented in Figure 2. In each of these circles, adjacent notes are a major third (four semitones) apart.

We created a map of the Coltrane changes by placing the circle of fifths in the center, and by then attaching a circle of major thirds to each note of this inner circle, as is shown in Figure 3. (In hindsight, our ability to do this should not have been a surprise, because intervals of major thirds are crucial to the key changes in Coltranes progressions (Waters, 2010)). A more elaborate path through this figure (in comparison to the path
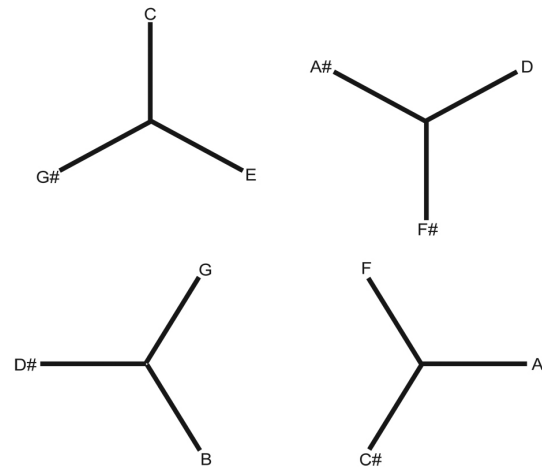

FIG. 2. The four circles of major thirds that Yaremchuk and Dawson (2008) discovered when interpreting the internal structure of an artificial neural network trained to classify musical chords.

through the circle of fifths for the II-V-I) provides the root notes for the seven chords of the Coltrane changes for a given key.

Consider, for example, the key of C. The Coltrane changes begin with Dm7; the root of this chord is in the inner circle of fifths. From this, the next root is the furthest note away on the circle of major thirds attached to the neighboring inner note (i.e. the B attached to the G). The next root is the furthest note away on the circle of major thirds attached to the next note on the inner circle (i.e. the E attached to the C). The progression then returns to the neighboring inner note (the G); the next root is the other outer note attached to it (i.e. the D# attached to the G). Next, the progression returns to the inner C; the next root is the other outer note attached to it (i.e. the G# attached to the C). Finally, the progression moves back to the inner circle of fifths and finished with the root notes G and then C. The Coltrane changes for the next key (A#) begins by keeping on the C, but playing the Cmin7 chord, and then following the identical path from this note to that followed from the D and that was just described. Although this path is complex, it is very easy to grasp when using Figure 3; Figure 3 serves
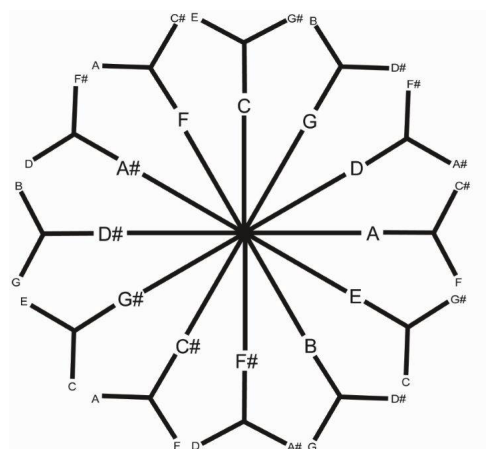
FIG. 3. The Coltrane changes emerge from a map that combines the circle of fifths on the inside with the circles of major thirds on the outside. See text for explanation.

FIG. 4. The inverted versions of the chords used to train a perceptron to generate the Coltrane changes.

as a musical guide for performing the entire Coltrane changes!

The Coltrane changes are systematic, in the sense that its sequence of root notes lies in a map like Figure 3. However, it seems much more complex than the II-V-I progression described earlier. Clearly, Table 2 involves many more chords than does Table 1, and Figure 3 (and the path through it) is much more complicated than Figure 1 (and the path through it). Is this new progression more complicated from the perspective of music cognition? One way to answer this question is to train a neural network to generate the Coltrane changes. If this progression is indeed more complex from an information processing perspective, then a more complicated network than a perceptron – a network that includes hidden units is required.

PERCEPTRONS CAN LEARN COLTRANE CHANGES

To begin to answer the complexity question, the first obvious step is to determine whether a simple network, the perceptron, is capable of learning the Coltrane changes. We created a training set to train a perceptron that had 22 input and 22 output unit. Each input and output unit encoded a particular note (i.e. each unit was analogous to a key on a piano). Each output unit used a non-linear, Gaussian activation function, as was the case in previous studies of this type (Yaremchuk & Dawson, 2008). We trained this network in a similar fashion to the method used to train the II-V-I network mentioned earlier: the network was presented a chord in the progression, and an error-correcting rule was used to teach the network to output the next chord in the progression. The entire sequence of chords from Table 2 defined the training set; importantly, the chords were presented in

their inverted form (i.e., the notes of the chord were rearranged to produce the most efficient piano fingering); chord inversions are named in the caption of Table 2. Musical notation for the sequence of chords used to train the perceptron is provided in Figure 4.

Given the complexity of the Coltrane changes, we expected that a perceptron would not learn this progression, and that we would then proceed to train a more sophisticated network. To our surprise, the perceptron successfully learned the Coltrane changes in short order. From an information processing perspective, the Coltrane changes are *not* more complex than the II-V-I progression; this is because they both can be learned by artificial neural networks that have no hidden units (i.e. both can be learned by a perceptron).

INTERPRETING THE INTERNAL STRUCTURE OF THE PERCEPTRON

How is it possible for such a simple artificial neural network to represent the complicated sequence of chords that define the Coltrane changes? To answer this question, we carefully examined the weights of the connections from each input unit to each output unit. These weights were highly systematic, and indicated that the network had learned to represent a particular chord as 1) a particular base note (i.e. the lowest note in the chord) and 2) a configuration of finger positions on keys above the base note. By changing the configuration of finger positions (i.e. the positions of the three notes) above the base note, the network could change the type of chord represented along with the chords inversion. Indeed, this makes chord changes possible even when the base note remains the same. For instance, the base note of both D#7 in root position and the base note of G#maj7 in the second version is identical (D# or Eb).

One can move from the first of these chords to the second (as required in the first line in Table 2 above) by keeping the same base note, and by changing the finger configuration above it. The connection weights of the perceptron indicated that the network had literally wired this knowledge in to the connection weights that used input unit activity to turn particular output units on.

The connection weights also revealed that the network had learned that a key aspect of the Coltrane changes was simply encoding the relationships between the different base notes used in the seven chords for the changes in any given key. These relationships only involve four different musical intervals: unison, the major second, the minor 7th, and the major 7th. Each of these intervals was defined by a unique connection weight; that is, the value of a connection weight could be used to read musical interval values from the network.

Furthermore, these relationships only involve three different base notes. Consider beginning the Coltrane changes in the key of C major. The first base note (for the first inversion of Dm7) is F. The next base note is Eb (for the root position of the Lead-in 1 chord); moving from F up to Eb is moving a musical interval of a minor 7th. In the Coltrane changes, the next three chords keep the Eb as the base note; as the note does not change, the musical interval moved is unison. Next, the base note moves on from Eb to F, which is a major second. The progression in this key ends by moving from a base note of F up to a base note of E, which is a major 7th. When the progression begins by playing the base note of the first chord of the next key, the base note shifts from E to Eb, which is another major 7th.

This pattern of shifts from one base note to another, a pattern lifted from the connection weights of the perceptron, can be represented by plotting a series of triangles. The three vertices of one triangle represent the three base notes involved in the seven chords that define the Coltrane changes in a single musical key. Importantly, the triangle that represents these changes in one key, and the triangle that represents the same changes in the next key, share a vertex (i.e., share a base note). Thus one can link the triangles together in a single manifold as is shown in Figure 5.

Figure 5, like Figure 3, provides a musical map through the various base notes of the Coltrane changes. To begin the changes in the key of C, start at F on the lower right of Figure 5. Move next to Eb, which will be repeated as a base note four times. Next, move back to F, which will be used once, and then follow the triangle inside to find that E is the final base note. Moving from E to Eb along the final side of the triangle takes the musician to the base note of the first chord played in the next musical key.

Interestingly, Figure 5 links the Coltrane changes perceptron to the structure of other musical networks studied in our lab (Yaremchuk & Dawson, 2005, 2008). While Figure 5 depicts a ring of intersecting triangles, it also
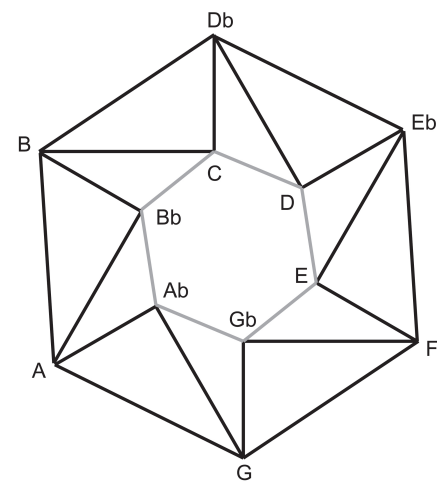
FIG. 5. A map of base note transitions for the Coltrane changes that emerges from examining the connection weights of the trained perceptron. See text for explanation.
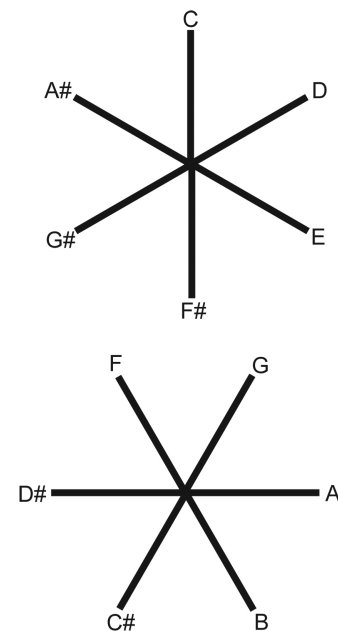
FIG. 6. The two circles of major seconds. The top circle defines the notes around the interior of Figure 5, and the bottom circle defines the notes around the exterior of Figure 5.

can be described as two circles of notes, one on the outside surrounding another on the inside. These two circles are the two circles of major seconds revealed in the connection weights of networks trained to classify chord types; for the sake of completion, the two circles of major seconds are illustrated in Figure 6. In short, the Coltrane changes reflect a specific relation (i.e., a specific orientation in relation to each other) between the two circles of major seconds.

CONCLUSION

The Coltrane changes are taken to be a more complicated elaboration of the classic II-V-I chord progression. The purpose of this paper was to explore the complexity of the Coltrane changes by training an artificial neural network to generate them. To our surprise, we found that an extremely simple network a perceptron learned the Coltrane changes, indicating that they are not more complex than the II-V-I from an information processing perspective. Furthermore, an analysis of the internal structure of this perceptron provided a simple map of base notes used in the progression. This simple map was discovered by the perceptron; the existence of this map permitted such a simple neural network to represent the complexities of the Coltrane changes.

ACKNOWLEDGEMENTS

REFERENCES

Capuzzo, G. (2006). Pat Martinos The Nature of the Guitar: An intersection of jazz theory and neo-Riemannian theory. *Music Theory Online, 12*(1), 1-17.

Dawson, M. R. W. (2004). *Minds And Machines: Connectionism And Psychological Modeling*. Malden, MA: Blackwell Pub.

Dawson, M. R. W. (2005). *Connectionism : A Hands-on Approach* (1st ed.). Oxford, UK ; Malden, MA: Blackwell Pub.

Dawson, M. R. W. (2013). *Mind, Body, World: Foundations Of Cognitive Science*. Edmonton, AB: Athabasca University Press.

Griffith, N., & Todd, P. M. (1999). *Musical Networks: Parallel Distributed Perception And Performace.* Cambridge, Mass.: MIT Press.

Krumhansl, C. L. (1990). *Cognitive Foundations Of Musical Pitch*. New York: Oxford University Press.

Levine, M. (1989). *The Jazz Piano Book*. Petaluma, CA: Sher Music Co.

Levitin, D. J. (2006). *This Is Your Brain On Music*. New York, N.Y.: Dutton.

Martineau, J. (2008). *The Elements Of Music*. New York: Walker & Company.

Peretz, I., & Zatorre, R. J. (2003). *The Cognitive Neuroscience Of Music*. Oxford ; New York: Oxford University Press.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65*(6), 386-408.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing*, V.1. Cambridge, MA: MIT Press.

Sudnow, D. (1978). *Ways Of The Hand: The Organization Of Improvised Conduct*. Cambridge, Mass.: Harvard University Press.

Todd, P. M., & Loy, D. G. (1991). *Music And Connectionism*. Cambridge, Mass.: MIT Press.

Turner, R. (1975). John Coltrane: A biographical sketch. *Black Perspective in Music, 3*(1), 3-16.

Waters, K. (2010). 'Giant Steps' and the ic4 legacy. *Intégral, 24*, 135-162.

Yaremchuk, V., & Dawson, M. R. W. (2005). Chord classifications by artificial neural networks revisited: Internal representations of circles of major thirds and minor thirds. *Artificial Neural Networks: Biological Inspirations - Icann 2005, Pt 1, Proceedings, 3696*, 605-610.

Yaremchuk, V., & Dawson, M. R. W. (2008). Artificial neural networks that classify musical chords. *International Journal of Cognitive Informatics and Natural Intelligence, 2*(3), 22-30.

BroilerBreak!: Partitioning ME in Broiler Breeders

Airell G. C. DesLauriers, Martin J. Zuidhof, Valerie Carney, and Brenda L. Schneider
Department of Agricultural, Food & Nutritional Science, University of Alberta

Feed restriction is a common tool utilized in broiler breeder production. The period of most severe feed restriction is during the rearing phase. In order to feed at the proper requirements the maintenance requirements of the birds must be estimated. Biometric models have been previously developed, however these estimates may loose accuracy due to changes in genetic selection criteria. In order to verify the estimates of metabolizable energy of maintenance (MEm), a study was conducted from 0-20 weeks with a total of 8,500 Ross 308 females placed at two separate commercial farms. The study also examined the effect of temperature on maintenance requirements of commercial broiler breeders. Although the effect of temperature was inconclusive, the estimates of maintenance requirements were consistent with other experimental findings in the literature.

INTRODUCTION

In the poultry production chickens are utilized for 2 main products, meat and eggs. Birds that are used for eggs for consumption are called leghorns. These birds have been selected for a high egg production of the large size with completely calcified shells. Broilers are birds that are raised for meat production and have a large proportion of their resources dedicated to breast meat production. Broilers are the offspring from broiler breeders, which have the same genetic growth potential as broilers. A major concern with the management of broiler breeders is the control of body weight during the rearing and reproduction phases. Broiler breeders are feed restricted due to the negative relationship between adipose tissue and egg production [1,2]. Excess energy intake is deposited largely as fat therefore to prevent excessive fat deposition broiler breeders are fed close to maintenance requirements. Broiler breeders are most intensively feed restricted during the rearing phase of development. This stage is critical in programming the egg production, health status, and longevity of the bird, making it vital to correctly determine the degree of feed restriction during this period.

As broiler breeder feeding levels are only slightly above maintenance requirements, changes in the weather can result in over- or under-feeding of broiler breeders. Since broiler breeders are feed restricted to their minimum required level to prevent excess growth, it is important to understand how common production decisions impact later growth and yield of broiler breeders. Mathematical models can be used to demonstrate the effect a production decision has on subsequent growth.

Bioeconomic models have been developed with production parameters however there is growing interest for the development of biometrical models. Biometrical models are mathematical formulas that predict biological features of a flock or individual bird [3]. There is interest in biometrical models because economic models are dependent on biometrical models. Changes in genetics have caused older biometric models to become out-dated. Biometrical models have been developed for broiler breeders [4], and broilers [5]. A broiler breeder model has been developed through raising the animals in an intensively monitored research facility [4]. In order for biometric models to be accurate enough to use in the industry, the mathematical model must be validated. One of the purposes of our study was to validate coefficients of metabolizable energy of maintenance (ME_m) obtained in previous studies. Validating the model with industry data will allow for the model to be tested against other flocks.

Historically, ME_m is expressed as a function of metabolic bodyweight ($BW^{0.75}$). However, recent studies have determined that ME_m is more accurately represented by exponents other than 0.75 in broilers [6]. Therefore, in our study, ME_m will be reported as a function of $BW^{0.67}$.

MATERIALS AND METHODS

The data were obtained through the use of datasets obtained from industry flocks. Two farms data were included in this analysis. For confidentiality, they are referred to as Farm A and Farm B. In Farm A 4,500 Ross 308 female broiler breeders were reared in 6800 square feet of floor space (stocking density of 1.51 ft^2/bird). In Farm B 4,000 Ross 308 female broiler breeders were reared separately in 7930 square feet of floor space for the first 10 weeks (stocking density of 1.98 ft^2/bird). After 10 weeks, 460 Farm B males were then cohabited with the Farm B females. Project protocol was approved by the University of Albertas Animal Care and Use Committee Livestock division prior to data collection.

Barn dimensions and configurations were recorded prior to chick placement. Each flock was housed in conventional poultry houses (Figure 1, 2) and was raised according to industry standards. Body weight (BW) of the flock was estimated through random selection of target birds weighed at placement, 3-4 weeks of age, and 9-11 weeks of age. Additional weekly weighing performed by the producer was also included in the analysis. Temperatures were recorded on a 30min interval,
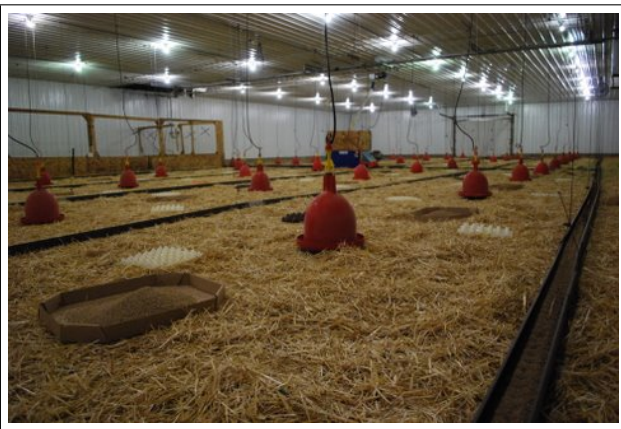
FIG. 1. In-Barn Configuration of Farm A [7]

FIG. 2. In-Barn Configuration of Farm B [7]

using automated data loggers positioned at bird head height. Historical external temperature data were obtained from Environment Canada. Using the physical location of the farms, the closest possible weather station was utilized (Three Hills and Linden). The weather stations for Farm A and Farm B were 7.83 km and 11.67 km away respectively. A regression was performed to examine the correlation between barn temperature and external temperature.

Data were analysed in a least squares linear model using the mixed procedure of SAS to estimate the metabolizable energy (ME) requirement for maintenance for age in 4 week blocks: 0-4 weeks, 4-8 weeks, 8-12 weeks, 12-16 weeks, and 16-20 weeks to the following linear model.

$$mei = A(W^{0.67}) + b(W^{0.67})(T_e - 21) + c(G) \quad [4] \quad (1)$$

Where **mei** is ME intake (kcal/day); **W** is BW (kg); $T_e - 21$ is internal barn temperature corrected to 21C; and G is gain (g/day); **A** is the coefficient representing the relationship between age specific base ME mainte-

nance requirements; **b** is the coefficient of the effect of temperature on the maintenance ME energy; **c** is the coefficient representing the ME cost for one gram of BW gain [4].

RESULTS

Our results from Farm A show a decrease in the base ME_m requirements from weeks 4-12 and then an increase in ME_m in the period 16-20 weeks (Table 1).

Farm B shows a decrease between the 4-8 week period and the 8-12 week period and then an increase in ME¡sub¿m ¡/sub¿throughout the 12-20 week period (Table 2). For both farms, the highest base ME_m requirements were during the 4-8 week period and the 16-20 week period (Table 1, 2).

The coefficient for **b**, the effect of temperature on maintenance, was 4.8348 and 4.0291 for Farm A and Farm B respectively. However the effect of temperature on maintenance requirements did not differ significantly. The coefficient for **c**, the ME requirement per gram of growth, was 0.7007 and 1.3272 respectively. However, this did not differ significantly.

The correlation between external environmental temperature and internal barn temperature was strongest when outdoor temperatures were higher than the set point of the barn. In the first few weeks, when the barn temperature was the most elevated, high external temperature was not significantly correlated to internal barn temperature (Figure 3, 4).

Excluding the first 2 weeks of rearing, the relationship between outdoor daily maximum temperature and indoor daily average temperature was found to be highly correlated for each farm (Farm A R^2=0.7650, Farm B R^2= 0.7582, Figure 5, 6).

Weeks	A	$\pm$	P-Value
0-4	97.6639	74.148	0.2203
4-8	128.91	44.6868	0.018
8-12	107.13	44.6747	0.04
12-16	76.9054	37.2027	0.0687
16-20	132.83	16.1646	<0.0001

TABLE I. Estimated coefficients values of **A** for Farm A; the coefficient representing the relationship between age specific base ME maintenance requirements.

Weeks	A	$\pm$	P-Value
0-4	-	-	-
4-8	88.3218	9.8596	<.0001
8-12	55.7727	11.6257	0.002
12-16	63.5079	8.3268	0.0001
16-20	72.9209	9.1673	<.0001

TABLE II. Estimated coefficients values of **A** for Farm B; the coefficient representing the relationship between age specific base ME maintenance requirements.
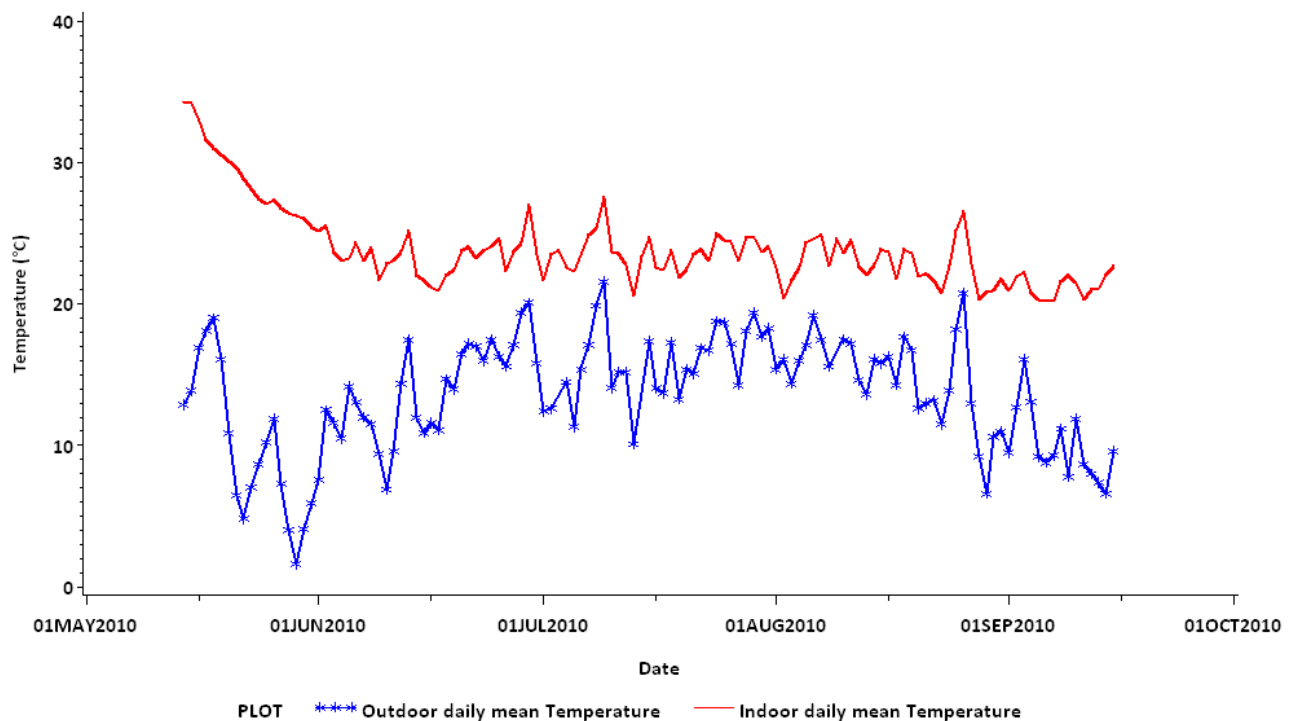
FIG. 3. Relationship between indoor and outdoor barn temperature at Farm A.

DISCUSSION

The reduction in required ME_m in the period of 8-16 weeks is likely due to the increase in the severity of feed restriction during this period. Broiler breeder flocks are fed ad libitum for the first 2-5 weeks which coincides with the decreased feed efficiency observed during this time. Feed efficiency increases during periods of feed restriction compared to ad libitum feeding [8]. The highest degree of feed restriction is during the 10-14 weeks of age [4]. Studies estimated that at 9-20 weeks of age, 95-96% of total ME intake was used for maintenance whereas from 5-8 weeks 72.5% was utilized for maintenance [9]. Feed intake has been found to be positively correlated to ME_m [10]. Through reduced feed intake ME_m is decreased [10]. This theory supports that the reduction in ME_m during the 8-16 period is due to feed restriction relative to ad libitum.

Decreases in the efficiency of the flock are seen during the 16-20 week period in both Farm A and Farm B (Table 1, 2). During the 16-20 week period, the degree of feed restriction is reduced causing an increase in the feed intake relative to metabolic BW and therefore a higher degree of diet induced thermogenesis. The differences between efficiencies can be explained in the changes in basal metabolic rate (BMR) that happen during the rearing period. Studies in rats have shown that the basal metabolic rate of restricted rats was significantly lower than that of ad libitum fed rats [11]. Work by Forsum et al. [11] can be compared to broiler breeders which begin

with ad libitum feeding early in life. Broiler Breeders are feed restricted later in life to meet the required growth curve and therefore their basal metabolic rate may decrease, as seen in the Forsum et al. [11] study.

It is likely that, because of the small variation in temperature over the course of the trials, the relationship between temperature and metabolic BW was not significant. Although our study showed inconclusive results for the effect of temperature on maintenance requirements, temperature has been shown to affect ME_m [4,12,13]. Due to the severity of feed restriction, which is very close to requirements, any change in ME_m must be accounted for as it will affect energy available for growth. In feed restricted broiler breeders at 16 wks of age, 93% of metabolizable energy intake is utilized for maintenance [4]. Since broiler breeders are homeothermic, temperature fluctuations from 15 to 17 C may require a 5% change in feed allocation [4]. Higher outdoor temperatures have more of an influence on indoor barn temperatures except during the first few weeks of rearing where indoor barn temperatures are high (Figure 3, 4). In order to adjust for these fluctuations, the weather forecast should be taken into consideration prior to making feeding decisions.

Interestingly, the data obtained from industry sources was similar to other studies of the Ross 708 (Zuidhof nutrition conference) and the Hubbard-Hi-Yield broiler breeder [9]. However, previous research has determined slightly higher coefficients for ME_m than our study. In our study, temperatures were controlled as per broiler
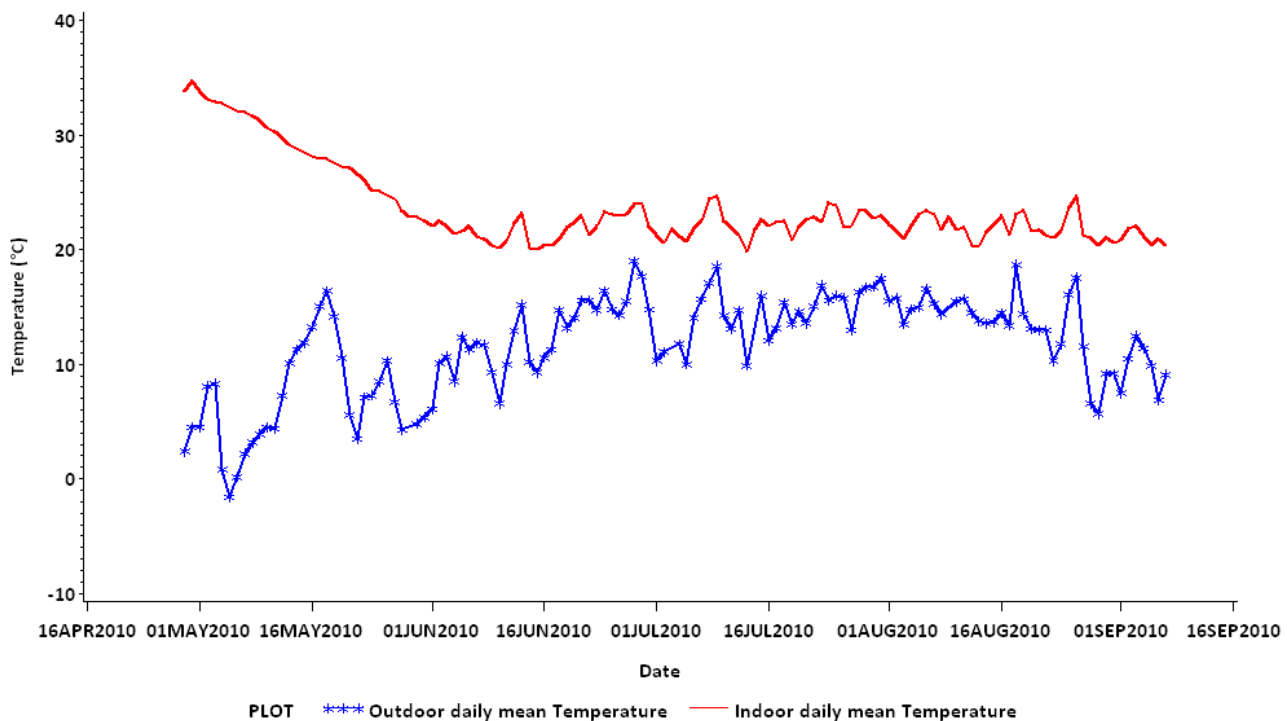
FIG. 4. Relationship between indoor and outdoor barn temperature at Farm B.
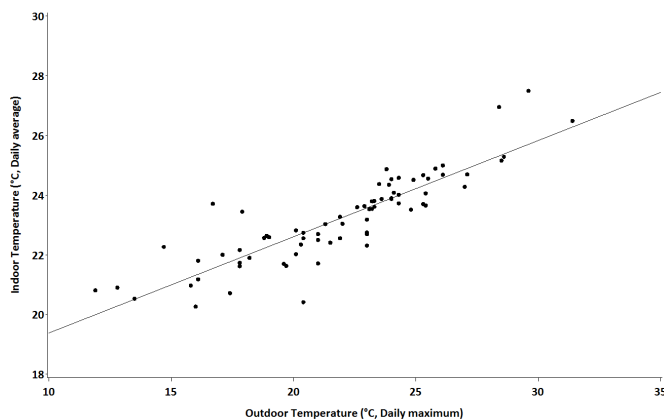
FIG. 5. Correlation between Indoor and Outdoor temperature at Farm A ($R^2 = 0.7650$).
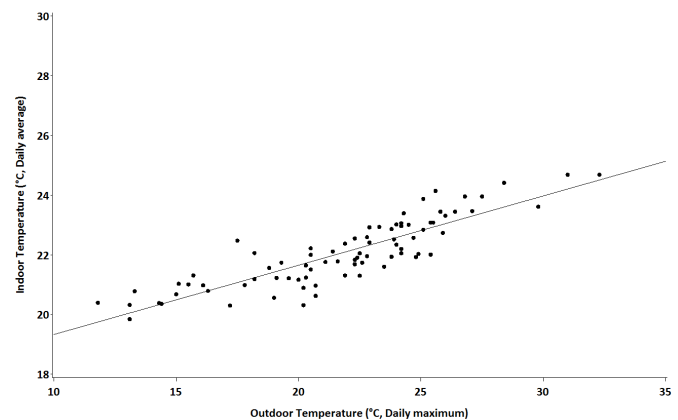
FIG. 6. Correlation between Indoor and Outdoor temperature at Farm B ($R^2 = 0.7582$).

breeder handbook recommendations and the poultry were not subjected to differing temperatures. The inconsistencies may also be due to the differences in adipose and muscle tissue ratios that are maintained by broiler breeders. Further composition of body tissues could be examined as lipid tissue is less expensive to maintain than muscle tissue.

Another possible inconsistency is the energy density and source of ME in each study. This will leave the efficiency in which different energy sources are utilized as a confounding variable [14].

Unlike Zuidhof et al. [4] the sources of error in our study may be resultant from errors in our estimates of flock BW. Variations in sample BW may have been introduced when selecting birds at random in the barn. A possible bias would arise because heavier breeders may move slower making them easier to catch.

It is interesting to note that, after a researcher weighing, the average daily gain of the flock had a significant decrease. However this decrease was not observed

after the producer weighed the flock. The decrease may have been caused through inconsistency in BW obtained between the producer and the researcher leading to the over estimation of BW by the researcher and subsequently a reduction in feeding by the producer. Due to bodyweight fluctuations the broiler breeders metabolism will be constantly interchanging from a positive energy balance to a negative energy balance impacting the BMR of the flock.

The importance of understanding how production decisions impact later growth in broiler breeders is extremely important. Through the continued development and improvement of mathematical models these decisions can be accurately tailored to increase overall efficiency.

REFERENCES

[1] Brody, T., Y. Eitan, M. Soller, I. Nir, and Z. Nitsan. 1980. Compensatory growth and sexual maturity in broiler females reared under severe food restriction from day of hatching. Br. Poult. Sci. 21:437-446.

[2] Whitehead, C. 2002. Nutrition and poultry welfare. World's Poult. Sci. J. 58:349-356.

[3] Zoons, J., J. Buyse, and E. Decuypere. 1991. Mathematical models in broiler raising. World's Poult. Sci. J. 47:243-255.

[4] Zuidhof, M.J., A. Pishnamazi, C. Ouellette, D. Paul, and R. A. Renema. 2010. Diet x temperature interactions in broiler breeders and broilers. Pages 128136 in Proc. Of Calif. Anim. Nutr. Conf., Fresno, CA.

[5] Zuidhof, M. J. 2005. Mathematical characterization of broiler carcass yield dynamics. Poult. Sci. 84:1108-1122.

[6] van Milgen, J., J. Noblet, S. Dubois, B. Carr, and H. Juin. 2001. Utilization of metabolizable energy in broilers. Poult. Sci. 80 (Suppl.1), 170 (Abstr.).

[7] Schneider, B. 2011. Agric. and Rural Dev. Alberta, Edmonton. Personal communication.

[8] Pinchasov, Y., I. Nir, and Z. Nitsan. 1985. Metabolic and anatomical adaption of heavy-bodied chicks to intermediate feeding. 1. Food Intake, growth rate, organ weight, and body composition. Poult. Sci. 64:2098-2109.

[9] Sakomura, N. K., R. Silva, H. P. Couto, C. Coon, and C. R. Pacheco. 2003. Modeling metabolizable energy utilization in broiler breeder pullets. Poult. Sci. 82:419-427.

[10] Pishnamazi, A., R. A. Renema, M. J. Zuidhof, and F. E. Robinson. 2008. Effect of initial full feeding of broiler breeder pullets on carcass development and body weight variation. Poult. Sci. 17:505-514.

[11] Forsum, E., P. Hillman, and M. Nesheim 1981. Effect of energy restriction on total heat production, basal metabolic rate, and specific dynamic action of food in rats. J. Nutr. 111:16911697.

[12] Costa, M. S. 1981. Fundamental principles of broiler breeders nutrition and the design of feeding programs. World's Poult. Sci. J. 37:177-192.

[13] Luiting, P., 1990. Genetic variation of energy partitioning in laying hens: causes of variation in residual feed consumption. World's Poult. Sci. J. 46:133152.

[14] Noblet, J., X. Shi, and S. Dubois. 1993. Metabolic utilization of dietary energy and nutrients for maintenance energy requirements in sows: basis for a net energy system. Br. J. Nutr. 70:407419.

Rainbow on a Chip: Experimental Observation of the Trapped Rainbow Effect Using Tapered Hollow Bragg Waveguides

Aaron Melnyk

Faculty of Engineering, University of Alberta

Experimental observation of the 'trapped rainbow' in the visible is demonstrated using tapered hollow Bragg waveguides. These waveguides spatially disperse an input spectrum into its various frequency components and vertical out of plane radiation was observed at wavelength dependant positions along the entire length of the waveguide. The experimental observation is corroborated by a brief theoretical analysis and simulation. These devices form the foundation for future work involving integration into a micro-spectrometer for eventual lab-on-chip use.

I. OVERVIEW AND MOTIVATION

Slow light phenomena and light trapping has been an area of interest for many years. In 2007 Tsakmakidis *et al.* coined the term 'trapped rainbow' for the slowing and spatial dispersion of white light in metamaterial structures [1]. This unique effect has led many to propose its use in a wide variety of optoelectronic devices including photonic logic circuits [2], communication networks [3], signal processors [3] and optical data processing [1]. While in principle the trapped rainbow phenomena can be realized, currently no such material exists that can stop and trap light indefinitely in this manner over a broad range of wavelengths [2]. Nonetheless there have been a few reports of experimental observation of the effect, notably Qiaoqiang Gan and Filbert J. Bartoli [2] and V.N. Smolyaninova *et al.* [4]. The former utilized chirped nanogratings while the latter emulates the necessary metamaterial properties using coated convex lenses. This paper, however, focuses on the application of the 'trapped rainbow' for on-chip spectroscopy, where the spatial separation and stopping of an input spectrum is extremely attractive.

In regards to this proposition, the global spectrometer market is expected to reach a value of \$ 4 billion this year [5]. This is in parallel with the lab-on-chip (LOC) market which is expected to reach a value of \$ 3 billion in the United States this year with an annual growth rate of 25% [6]. LOCs often employ the use of fluorescent signals to indicate the presence or absence of biomarkers of interest [7], making a micro-spectrometer a perfect sensing device for these diagnostics tools. The burgeoning spectrometer and analytical biochip markets present a unique research opportunity, where two industries are experiencing simultaneous growth while at the same time still require considerable amounts of development and innovation.

While prior results have been successful to varying degrees, they do not provide optimal geometries or properties for use in a on-chip spectroscopy. However, prior research by Dr. DeCroby's group in tapered waveguides using omnidirectional dielectric reflectors (ODRs) in the near IR region has offered promising results for microspectrometer use [8]. Here the ODRs confine in-coupled light toward the narrow end of the taper.

Wavelength dependant cut-off positions exists causing the in-coupled spectrum to be spatially disperesed and out of plane radiation to occur. This externally radiated light can then be coupled to a photo detector array for use as a spectrometer. It is the focus of this paper to transfer this design to the visible light region and realize the 'trapped rainbow' effect.

II. THEORY AND DESIGN

The proposed spectrometer relies on ODRs which have the advantage over metallic mirrors of being low loss while still achieving extremely high reflectance [9, 10]. These attributes allow guided light to propagate with minimal energy loss down the tapered waveguide to wavelength dependant cut-off positions. The structure of ORDs consist of alternating high and low index of refraction materials tuned to a specific wavelength, the Bragg wavelength (λ_{BRAGG}), defined as:

$$\lambda_{BRAGG} = 4n_1d_1 = 4n_2d_2 \tag{1}$$

Where n_1 and n_2 are the indices refraction and d_1 and d_1 are the thicknesses of alternating layers 1 and 2 respectively. This equation defines the quarter-wave stack (QWS) condition; light centered around λ_{BRAGG} experience high reflectance due to constructive interference of the partially reflected rays as they propagate through each layer. Reflectance of these structures can be optimized by increasing the number of layers, using an odd number of layers, and when the high-index material is outermost [9].

Unfortunately dielectric mirrors only offer high reflectance at all angles for a small range of wavelengths; the omnidirectional band. This band can be maximized by increasing the index contrast between the alternating layers in the QWS structure [11, 12] (Fink *et al.*[11] presents an in depth theoretical analysis of this condition). Moreover, by setting the index of the low index material around 1.45, the omnidirectional band can be maximized [13].

Based on prior familiarity and ease of use with available e-gun deposition systems, TiO_2 / SiO_2 were chosen for use in our devices. Setting λ_{BRAGG} to 550nm,
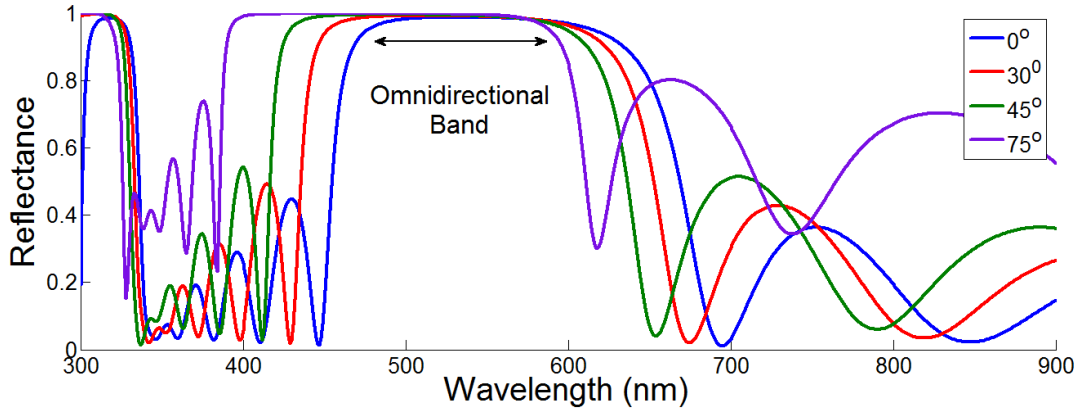
FIG. 1. Modeled reflectance of a 6.5 period TiO_2 / SiO_2 Bragg mirror at various angles. The omnidirectional band spans wave-lenghts $\sim$ 490nm - 570nm.

thicknesses of TiO_2 ($n \sim 2.35$) and SiO_2 ($n \sim 1.48$) were determined to be 61nm and 93nm respectively.

MATLAB code was used to simulate properties of the ODRs and then the tapered waveguide structures using the standard transfer matrix model [14], which is based on the electromagnetic continuity equations across boundaries. This method when applied to a QWS structure will produce a system matrix which provides a mathematical description of the light as is passes through the entire structure. An in depth analysis of this method as used in tapered waveguides with ODR claddings can be found elsewhere [15, 16].

Taking into account device mass, deposition time, and cost, 6.5 period (13 layer) Bragg reflectors were selected and modeled with TiO_2 as the terminating layer using the thicknesses derived above. Figure 1 displays the reflectance of these mirrors at a variety of angles with the omnidirectional band presented. The reflectance in the band exceeds 98%, with the band itself spanning $\sim$ 80nm (from $\sim$ 490 - 570nm).

Figure 2 considers a ray-optics representation of a tapered waveguide clad with ODRs described above. A
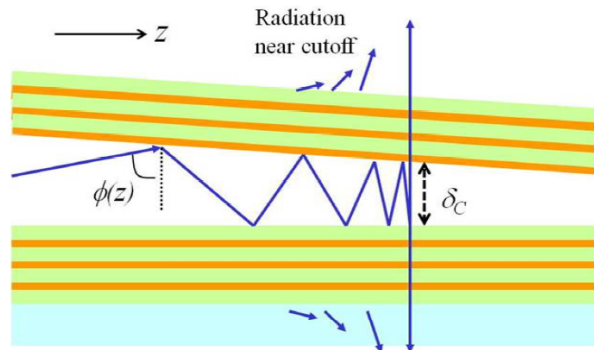
given mode of light will propagate through the waveguide core until it reaches a specific cut-off thickness, δ_{cn}. This cut-off position is defined as the core thickness of the waveguide that realizes a Fabry-Perot cavity [15] for that specific wavelength and mode. The general equation for cut-off the nth order mode is thus:

$$\delta_{cn} = (n+1)\lambda_o/2 \qquad (2)$$

With λ_o the free space wavelength and mode ordering $n = 0, 1, 2, ...$ At this cut-off position the "trapped" light will begin to radiate vertically outward. A microspectrometer can be realized by coupling a photodetector array the cut-off location of light of interest. Such a device holds key advantages over other microspectrometers, namely: scalability, the lack of complex moving parts, and a high potential throughput due to the low loss and dispersive nature of the waveguide.

III. FABRICATION

Electron beam evaporation was used to deposit the TiO_2 / SiO_2 layers on separate silicon and glass substrates. SiO_2 was deposited at a pressure of $1.5 * 10^{-6}$ Torr at a deposition rate of $\sim 3\text{Å}/s$. TiO_2 was deposited with a partial pressure of oxygen (to achieve the required stoichiometry) at a pressure of $5 * 10^{-5}$ Torr and a deposition rate of $0.5\text{Å}/s$. The optical properties of the deposited mirrors were then verified using variable angle spectroscopic ellipsometry (VASE) and profilometer measurements and were found to be in close accordance with the simulation for a wide variety of angles, confirming the omnidirectionality criteria.

The mirrors were then fabricated into tapered waveguides, the final result of which is illustrated in Fig 3(a). First the glass and silicon substrates were cleaved and diced into small dimensions in order to create waveguides with lengths of 4mm. Drops of a UV curable

FIG. 2. Ray diagram of a tapered waveguide clad with omni-directional reflectors [8].
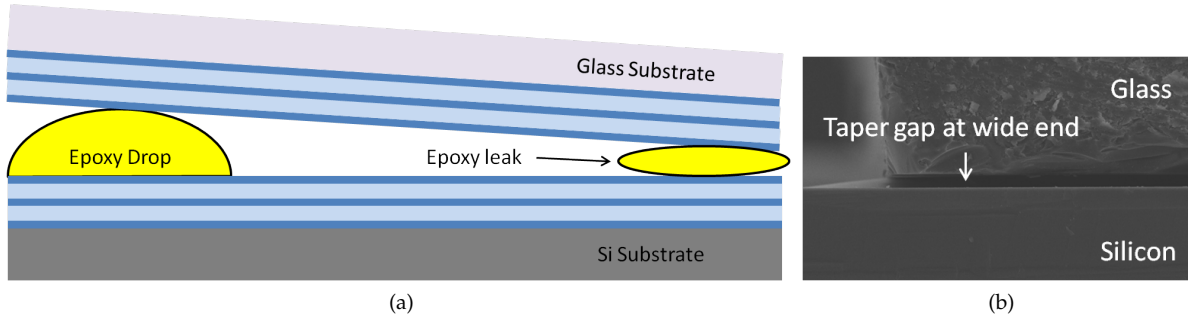
FIG. 3. (a) Illustrative side view of the fabricated waveguide with the epoxy leak visible. (b) SEM image taken of the device at the wide end of the taper.

epoxy (NOA 61, Norland) were next deposited and cured along one side of the cleaved pieces of silicon to act as permanent 'posts'. A diced piece of the glass substrate was then aligned on top of the silicon substrate (deposited layers facing one another) where the epoxy 'posts' prop up the glass at one end. In this way the silicon substrate acts as the bottom waveguide cladding while the glass substrate provides the top cladding. The peices were then bonded together at the narrow end using the NOA 61. The taper slope is thus defined at the height of the epoxy 'posts' divided by the length of the waveguide (4mm).

The periodically deposited drops of epoxy (positioned $\sim 200\ \mu m$ apart) were measured with a profilometer and produced relatively uniform feature heights, ranging from $11 - 13\ \mu m$. The taper profile of the compiled waveguides were then confirmed using SEM images (see Fig. 3(b)). SEM analysis also revealed that the substrates did not come into complete contact at the rear, instead epoxy had leaked between the substrates prohibiting contact. This had the effect of propping up the rear at a height of $2\ \mu m$. This grievance is illustrated in Fig. 3(a).

For experimental purposes the leaked epoxy has the effect of preventing roughly the first 6-7 modes from reaching cut-off since their cut-off positions appear at core heights less than $2\ \mu m$. Given this and the known post height we can estimate the number of supported modes for light of 594nm and 532nm to be ~ 37 and ~ 41 respectively with subsequent mode spacing of 0.11mm and 0.098mm given a waveguide length of 4mm.

IV. EXPERIMENTAL RESULTS

In experiments a 532nm green laser and 594nm amber laser were free space coupled to the wide (input) end of the slab waveguide structures. A broadband white light source (Koheras SuperK Red) was also coupled to the waveguides using a multi-mode fibre. Light radiated/scattered out of the top waveguide cladding (glass substrate) was captured using a visible spectrum camera coupled to objective lenses. A schematic diagram can be
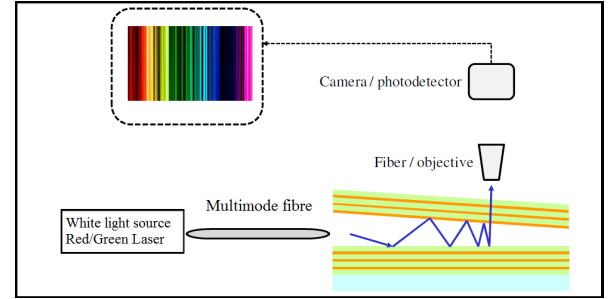
FIG. 4. Schematic view of the experimental setup.
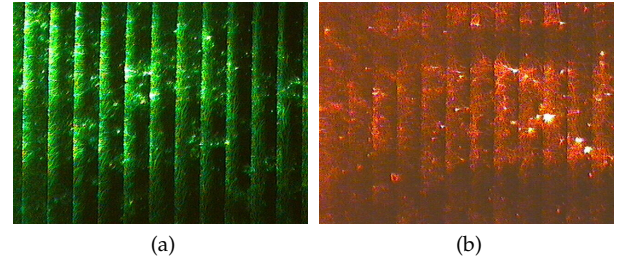
FIG. 5. Images captured of the vertically radiated light at mode cut-off for (a) the 594nm laser and (b) the 532nm laser. The bright bands correspond the standing wave patterns that form at cutoff.

seen in Fig. 4.

Images of radiated light from the 532 nm and 594nm laser are provided in Fig. 5. Periodic bands of laser light can be seen along the length of the images. Based on the previous discussions, the peak intensity lines are attributed to the various modes of the input laser source forming standing waves defined by the Fabry-Perot resonance condition, where the mode is 'trapped' and forward propagation halted. While only a portion of the waveguide is imaged here, these fringes were present along the entire 4mm length of the waveguides. Bands from the green laser appear sharper and more defined when compared to the amber laser which can be attributed to the proximity of 594nm light to the upper limit of the designed PBG which in turn negatively in-

fluences confinement and reflection in the waveguide. In total, depending on the individual sample tested, the number of modes present along the length of the waveguide numbered 39-43 and 36-39 for 532nm and 594nm light respectively, comparing quite well to the estimated number of supported modes within the waveguides.
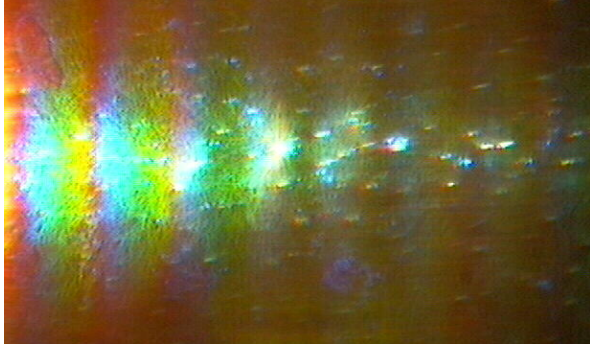
FIG. 6. 'Trapped rainbow' from broadband white light illumination. Colours visibly undergoing cut-off include blue, cyan, green and yellow indicating a wavelength range of $\lambda \sim$ 450nm-590nm.

.

Lastly the Koheras white light source was coupled to the waveguide input using a multi-mode fibre to verify the 'trapped rainbow', or mode cut-off all light within the PBG of the devices. An image of vertically radiated light from this source is presented in Fig. 6; the spatial dispersion and 'trapped rainbow' effect of the input spectrum visible. Each supported mode experiences cut-off with several colours visibly seen radiating vertically outward including blue, cyan, green and yellow. This equates to a wavelength range of $\lambda \sim$ 450nm-590nm, which is in reasonable accordance to the simulated PBG of the reflectors as presented in Fig 1.

V.   CONCLUSION AND FUTURE WORK

To conclude, 'rainbow trapping' in the range $\lambda =$ 450nm-590nm was experimentally demonstrated using a tapered waveguide structure clad with omni-directional dielectric reflectors and illuminated with 532nm/594nm lasers and a broadband white light source. Comparing results to other experimental observations of the 'trapped rainbow' [2, 4], these waveguides have the potential to offer lower loss and thus longer light "storage" characteristics due to the all dielectric structure. In respect to the LOC integration, modifying the fabrication techniques to provide an enclosed hollow waveguide core would allow for the cavity to potentially be filled with a liquid or gas analyte [7]. Potential fabrication methods to achieve this modified structure include a monolithic assembly using induced buckling [17–19] or grey-scale lithography and sacrificial etching [8]. These methods would also provide the necessary

methods to eliminate the 'epoxy leakage' issue since they do not depend on such bonding techniques.

Looking forward, we wish to amend the fabrication techniques and employ such a waveguide directly into the proposed mircospectrometer by integrating photodetector arrays. This will allow direct comparison to other microspectrometers being developed such as tunable MEMS devices [20] or filter array and wedge filter spectrometers [21, 22].

REFERENCES

[1] Kosmas L. Tsakmakidis, Allan D. Boardman & Ortwin Hess. Trapped rainbow storage of light in metamaterials. Nature 450, 397-401 (15 November 2007).

[2] Qiaoqiang Gan and Filbert J. Bartoli. Surface dispersion engineering of planar plasmonic chirped grating or complete visible rainbow trapping. Appl. Phys. Lett. 98, 251103 (2011)

[3] X. P. Zhao, W. Luo, J. X. Huang, Q. H. Fu, K. Song et al. Trapped rainbow effect in visible left-handed heterostructures. Appl. Phys. Lett. 95, 071111 (2009).

[4] Vera N. Smolyaninova, Igor I. Smolyaninov, Alexander V. Kildishev, and Vladimir M. Shalaev. Experimental observation of the trapped rainbow. Appl. Phys. Lett. 96 211121 (2010).

[5] Dr. J. Charles Forman. Spectroscopy: Market Research Report. BCC Research, 2010

[6] Steven Hener. Analytical Chip Technology Markets. Kalorama Information, 3rd Ed., 2007.

[7] O. Schmidt, M. Bassler, P. Kiesel, C. Knollenberg, and N. Johnson, Fluorescence spectrometer-on-a-fluidicchip, Lab Chip 7(5), 626629 (2007).

[8] R.G. DeCorby, N. Ponnampalam, E. Epp, T. Allen, J.N. McMullin. Chip-scale spectrometry based on tapered hollow Bragg waveguides. Optics Express, vol. 17 no.19, pp 16632-16645, Sept. 14 2009.

[9] H. A. Macleod, Thin-Film Optical Filters, Taylor & Fran- cis, 2001

[10] S.O. Kasap, Optoelectronics and Photonics: Principles and Practices, Prentice Hall, 2001

[11] Y. Fink, J. N. Winn, F. Shanhui, C. Chiping, J. Michel, J. D. Joannopoulos, and E. L. Thomas, A dielectric omnidirectional reflector, Science 282, 1679-1682 (1998).

[12] J. P. Dowling, Mirror on the Wall: Youre Omnidirectional After All, Science 282, 1841 (1998).

[13] W. H. Southwell, Omnidirectional mirror design with quarterwave dielectric stacks, Appl. Opt. 38, 54645467 (1999).

[14] E. Hecht, Optics, 4th ed. Addison Wesley, 2001.

[15] N. Ponnampalam, and R. G. DeCorby, Out-of-plane coupling at mode cutoff in tapered hollow waveguides with omnidirectional reflector claddings, Opt. Express 16(5), 28942908 (2008).

[16] N. Ponnampalam, R.G. DeCorby. Analysis and fabrica- tion of hybrid metal-dielectric omnidirectional Bragg reflectors. Applied Optics, vol. 47 no. 1pp 30-37, Dec. 20 2007

[17] N. Ponnampalam, and R. G. Decorby, Self-assembled hollow waveguides with hybrid metal-dielectric Bragg claddings, Opt. Express 15(20), 1259512604 (2007).

[18] Epp, E.; Ponnampalam, N.; Newman, W.; Drobot, B.; McFarlane, S.; McMullin, J.N.; Meldrum, A.; DeCorby, R.G.; Silicon-based integrated hollow waveguides produced by controlled thin film buck-ling, IEEE Photonics Society, 2010 23rd Annual Meeting of the , vol., no., pp.258-259, 7-11 Nov. 2010.

[19] R. G. DeCorby, N. Ponnampalam, H. T. Nguyen, M. M. Pai, and T. J. Clement, Guided self-assembly of integrated hollow Bragg waveguides, Opt. Express 15, 3902- 3915 (2007).

[20] R. A. Crocombe, D. C. Flanders, and W. Atia, Micro-optical instrumentation for process spectroscopy, Proc. SPIE 5591, 1125 (2004)

[21] S.-W. Wang, C. Xia, X. Chen, W. Lu, M. Li, H. Wang, W. Zheng, and T. Zhang, Concept of a high-resolution miniature spectrometer using an integrated filter array, Opt. Lett. 32(6), 632634 (2007).

[22] O. Schmidt, P. Kiesel, and M. Bassler, Performance of chip-size wavelength detectors, Opt. Express 15(15), 97019706 (2007).

An Introduction to Golay Complementary Sequences

Elana Kalashnikov
Department of Mathematics, University of Alberta

INTRODUCTION

Golay Complementary Sequences were first introduced by Marcel Golay in 1949, but were formally defined in his 1960 article "Complementary Series" [2]. They have many applications: Golay used them in infrared multislit spectrometry, and more recently, they have been applied to Orthogonal Frequency-Division Multiplexing [19]. Aside from the applications, they are of mathematic interest because of the "deep seated symmetries which characterize them," [2] and because of the challenge they present in both constructing them and enumerating them. There are still many open questions about both Golay sequences and their extensions. In this paper, I will discuss the development of Golay sequences over the last decades, and present the main results.

First, I will discuss Golay sequences as Golay himself defined them, presenting his results on their lengths and his direct and recursive constructions. I will then discuss the broadest generalization yet defined, Golay array pairs. It is fruitful to understand complementary sequences as a special case of Golay array pairs, rather than Golay array pairs simply as an extension of the more fundamental complementary sequences. First, I will discuss their basic properties, and then the ways in which they can be transformed into higher or lower dimensional arrays. Secondly, within the context of Golay array pairs, I will discuss the other main generalizations and developments, beginning with Jedwab and Davis' non-recursive structure, which led to the division of standard and non-standard Golay sequences, both of which I will discuss. Thirdly, I will present the basic extensions of Golay sequences: Golay sets, and multiple L-shift complementary sequences.

1. ORIGINAL GOLAY SEQUENCES

1.1. Definitions

Golay defined his "complementary series", which we call sequences, in his 1961 paper ([2]) in three equivalent ways. A proof of equivalence will follow the definitions.

For the first definition, we need the notion of a separation of a sequence. A separation of length j of sequence $a = \begin{pmatrix} a_0 & a_1 \cdots & a_{n-1} \end{pmatrix}$, are the pairs (a_i, a_{i+j}) – that is, pairs of elements of the sequence separated by j.

Definition 1.1.1: Let $a = \begin{pmatrix} a_0 & a_1 \cdots & a_{n-1} \end{pmatrix}$ and $b = \begin{pmatrix} b_0 & b_1 \cdots & b_{n-1} \end{pmatrix}$ be two length n sequences with $a_k, \ b_k \in \{-1, 1\} \ \forall \ 0 \leq k \leq n-1$. Then (a, b) are a Go-

lay complementary pairs if, under any given separation j, the number of like pairs in a is equal to the number of unlike pairs in b. a and b are then Golay sequences.

Example 1.1.2: $\begin{pmatrix} 1 & 1 & -1 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 & 1 & -1 \end{pmatrix}$ are a Golay complementary pair.

This definition, however, cannot easily accommodate important extensions (such as Golay sequences in non-binary alphabets). Thus, the second definition,

Definition 1.1.3: Let $a = \begin{pmatrix} a_0 & a_1 \cdots & a_{n-1} \end{pmatrix}$ and $b = \begin{pmatrix} b_0 & b_1 \cdots & b_{n-1} \end{pmatrix}$ be two length n sequences with $a_k, \ b_k \in \{-1, +1\} \ \forall \ 0 \leq k \leq n-1$. Let

$$C_a(j) := \sum_{k=0}^{n-1-j} a_k a_{k+j}$$

Then (a, b) are a Golay complementary pair if for every $0 < j \leq n-1$.

$$C_a(j) + C_b(j) = 0$$

For $a = \begin{pmatrix} a_0 & a_1 \cdots & a_{n-1} \end{pmatrix}$ a length n sequences, the polynomial in $z \in \mathbb{C}$ associated with a is $a(z) = \sum_{k=0}^{n-1} a_k z^k$ This leads to a third definition of the original Golay Complementary Pairs.

Definition 1.1.4: Let $a = \begin{pmatrix} a_0 & a_1 \cdots & a_{n-1} \end{pmatrix}$ and $b = \begin{pmatrix} b_0 & b_1 \cdots & b_{n-1} \end{pmatrix}$ be two length n sequences with $a_k, \ b_k \in \{-1, +1\} \ \forall \ 0 \leq k \leq n-1$. Then (a, b) are a Golay complementary pair if

$$a(z) a(z^{-1}) + b(z) b(z^{-1}) = 2n$$

For a sequence a, let o_{aj} be the number of unlike pairs, and e_{aj} the number of like pairs, of elements of sequence a under separation j for $0 < j \leq n-1$.

These three definitions are equivalent, and are useful in different contexts. The first definition is primarily useful as an intuitive way of seeing the properties of such pairs of sequences, without formulas.

Proposition 1.1.5: Definitions 1.1.1, 1.1.3, and 1.1.4 are equivalent

Proof: Let $a = \begin{pmatrix} a_0 & a_1 \cdots & a_{n-1} \end{pmatrix}$ and $b = \begin{pmatrix} b_0 & b_1 \cdots & b_{n-1} \end{pmatrix}$ be two length n sequences with $a_k, \ b_k \in \{-1, +1\} \ \forall \ 0 \leq k \leq n-1$. Assume Definition 1. By assumption, there are o_{aj} like pairs for sequence b under the same separation, and $n-1-j-o_{aj}$ unlike pairs of sequence b. There are $n-1-j-o_{aj}$ like pairs for sequence a. Then

$$\sum_{k=0}^{n-1-j} a_k a_{k+j} + \sum_{k=0}^{n-1-j} b_k b_{k+j}$$

$$= (n-1-j-o_{aj}-o_{aj}) + (o_{aj} - (n-1-j-o_{aj})) = 0$$

So *(a,b)* are a Golay pair by Definition 3. Now note that

$$a(z)\,a\left(z^{-1}\right)$$

$$= \left(\sum_{k=0}^{n-1} a_k z^k\right)\left(\sum_{k=0}^{n-1} a_k z^{-k}\right)$$

$$= \sum_{k=0}^{n-1} a_k^2 + \sum_{j=1}^{n-1}\sum_{k=0}^{n-1-j} a_k a_{k+j}(z^j + z^{-j})$$

$$= n + \sum_{j=1}^{n-1}\sum_{k=0}^{n-1-j} a_k a_{k+j}(z^j + z^{-j})$$

Thus,

$$a(z)\,a\left(z^{-1}\right) + b(z)\,b\left(z^{-1}\right)$$

$$= n + \sum_{j=1}^{n-1}\sum_{k=0}^{n-1-j} a_k a_{k+j}(z^j + z^{-j})$$

$$+ \quad n + \sum_{j=1}^{n-1}\sum_{k=0}^{n-1-j} b_k b_{k+j}(z^j + z^{-j})$$

$$= 2n + \sum_{j=1}^{n-1}\sum_{k=0}^{n-1-j} (a_k a_{k+j} + b_k b_{k+j})(z^j + z^{-j})$$

$$= 2n$$

So *a* and *b* satisfy Definition 4.

Now suppose that *a* and *b* are a Golay pair by Definition 4. I will show that this means they are a Golay Pair by Definition 1. As we see from above, if $a(z)\,a\left(z^{-1}\right) + b(z)\,b\left(z^{-1}\right) = 2n$, then

$$2n + \sum_{k=1}^{n-1}\sum_{j=0}^{n-1-j} (a_k a_{k+j} + b_k b_{k+j})(z^j + z^{-j}) = 2n$$

So

$$\sum_{j=0}^{n-1-j} (a_k a_{k+j} + b_k b_{k+j})(z^j + z^{-j}) = 0 \; for \; all \; z \; \in C$$

which implies that

$$\sum_{k=0}^{n-1-j} a_k a_{k+j} + \qquad\qquad \sum_{k=0}^{n-1-j} b_k b_{k+j}$$

$$= \quad n - 1 - o_{aj} - o_{aj} + e_{bj} - (n - 1 - e_{bj})$$

$$= \qquad\qquad 0 \; for \; every \; 0 < j \leq n-1$$

Thus, $o_{aj} = e_{bj}$, so *a* and *b* are a Golay pair by Definition 1.

1.2.   Lengths for which Golay Sequences exist

Golay sequences can only exist for certain lengths, a fact which may be somewhat surprising at first glance.

While a chief difficulty is finding Golay sequences, it is helpful to find some necessary conditions on the lengths of Golay sequences. Thus, as part of the project of discovering which length Golay sequences exist, we will present some basic restrictions on possible lengths.

Proposition 1.2.1: If *(a,b)* are a Golay pair of length *n*, then

a. *n* is the sum of two integer squares [2]

b. *n* is even [2]

c. *n* is not divisible by integer *p*, where *p* is a prime number and $p \equiv 3 \; mod \; 4$ [17].

Proof: Let *a* and *b* be a Golay complementary pair.

a. Using definition 1.1.4 of a Golay complementary pair, we find

$$a(z)\,a\left(z^{-1}\right) + b(z)\,b\left(z^{-1}\right) = 2n$$

$$a(1)^2 + b(1)^2 = 2n$$

$$\left(\frac{a(1) + b(1)}{2}\right)^2 + \left(\frac{a(1) - b(1)}{2}\right)^2 = n$$

$a(1) \quad = \quad o_{a1}(1 + (-1)) \quad + \quad n_1(1+1) \quad + \quad n_2(-1 + (-1)) = 2(n_1 - n_2), \; where \; n_1 + n_2 = n - 1, \; and \; n_1, n_2 \in \mathbb{Z} \; and \; positive.$ Similarly for $b(1)$. Thus

$$\frac{a(1) + b(1)}{2}, \; \frac{a(1) - b(1)}{2} \in \mathbb{Z}$$

So this gives us that the length of a Golay sequence must be the sum of two integer squares.

b. From [16], [5],

$$o_{aj} + e_{aj} = n - j$$

and

$$C_a(j) = e_{aj} - o_{aj}$$

and

$$\prod_{k=0}^{n-1-j} a_k a_{k+j} = (-1)^{o_a} = (-1)^{(n-j-C_a(j))/2}$$

$$\prod_{k=0}^{n-2} a_k a_{k+1} \; \prod_{k=0}^{0} a_k a_{k+n-1}$$

$$= (a_0 a_1)(a_1 a_2)(a_2 a_3)\cdots(a_{n-2}a_{n-1})(a_0 a_{n-1})$$

$$= a_0^2 a_1^2 \cdots a_{n-1}^2 = 1$$

$$= (-1)^{(n-1-C_a(1))/2} (-1)^{(n-(n-1)-C_a(n-1))/2}$$

$$= (-1)^{(n-C_a(1)-C_a(n-1))/2}$$

$$C_a(1) + C_a(n-1) = n \ (mod \ 4)$$

This also holds for b, the Golay complementary sequence which pairs with a. Thus

$$C_a(1) + C_a(n-1) + C_b(1) + C_b(n-1) = 0 = 2n \ (mod \ 4)$$

So n is even.

c. Because of its length, the proof of this is not included. Refer to [17].

1.3.   Recursive Methods of Producing Golay Sequences

Golay sequences, however, are useful only if they exist. Necessary conditions on the lengths of Golay sequences do not help with producing them. One way to actually find Golay sequences is to produce them from already known Golay sequences, that is, recursively. Below, I begin by reviewing some operations on sequences in order to make notation simpler when I discuss the recursive methods.

Let $a = \begin{pmatrix} a_0 & a_1 \cdots & a_{n-1} \end{pmatrix}$ be a length n sequence, $b = \begin{pmatrix} b_0 & b_1 \cdots & b_{m-1} \end{pmatrix}$ a length m sequence.

- Multiplication by a constant:
  $ca := \begin{pmatrix} ca_0 & ca_1 \cdots & ca_{n-1} \end{pmatrix}, \ for \ c \ a \ constant$

- Reversal:
  $\tilde{a} := \begin{pmatrix} a_{n-1} & a_{n-2} \cdots & a_0 \end{pmatrix}$

- Altering:
  $\check{a} := \begin{pmatrix} (-1)^0 a_0 & (-1)^1 a_1 \cdots & (-1)^{n-1} a_{n-1} \end{pmatrix}$

- Concatenation:
  $a|b := \begin{pmatrix} a_0 & a_1 \cdots & a_{n-1} & b_0 & b_1 \cdots & b_{m-1} \end{pmatrix}$

- Interleaving:
  $a \sim b := \begin{pmatrix} a_0 & b_0 & a_1 & b_1 \cdots & a_{n-1} & b_{m-1} \end{pmatrix}$

- Tensor product:
  $a \otimes b := \begin{pmatrix} a_0 b & a_1 b \cdots & a_{n-1} b \end{pmatrix}$ a sequence of length mn

- Addition (take $m = n$):
  $a + b := \begin{pmatrix} a_0 + b_0 & a_1 + b_1 \cdots & a_{n-1} + b_{n-1} \end{pmatrix}$

Result 1.3.1 Let $a = \begin{pmatrix} a_0 & a_1 \cdots & a_{n-1} \end{pmatrix}$ and $b = \begin{pmatrix} b_0 & b_1 \cdots & b_{n-1} \end{pmatrix}$ be two length n sequences that form a Golay Complementary pair. Then reversing, negating (multiplying by $c = -1$), or altering either a or b, or both, still gives a Golay pair. Also, pairs $(a\,|b, a\,|\,(-b))$ and $(a \sim b, a \sim (-b))$ form Golay pairs of length $2n$.

Thus, from Golay sequences of length n, we can find Golay sequences of length $2n$. These basic methods allow us quickly extend the number of known Golay sequences.

Theorem 1.3.2 Let (a, b) and (c, d) be Golay pairs of length n and length m respectively. Then (e, f) form a Golay pair of length mn, where

$$e(z) = \frac{a(z^m)(c(z) + d(z)) + z^{m(n-1)} b(z^{-m})(c(z) - d(z))}{2}$$

$$f(z) = \frac{b(z^m)(c(z) + d(z)) + z^{m(n-1)} a(z^{-m})(c(z) - d(z))}{2}$$

Alternatively, using tensor products,

$$e = \frac{1}{2}\left[(c + d) \otimes a + (c - d) \otimes \tilde{b}\right]$$

and

$$f = \frac{1}{2}\left[(c + d) \otimes b - (c - d) \otimes \tilde{a}\right]$$

Proof: Details are omitted, but one should show that $e(z)\,e(z^{-1}) + f(z)\,f(z^{-1}) = 2n$ by using the above construction.

There are Golay sequence pairs of length 2, 10, and 26 (see [2] for examples). From the above theorem, we then know that there are Golay sequence pairs of length $2^a 10^b 26^c, \ for \ all \ a, \ b, \ c \in \mathbb{Z}, a, \ b, \ c \geq 0$. Giving them explicitly is another problem, however. Golay, in his original paper [2], provided a direct construction using Boolean functions for length $n = 2^m$, which may be of interest to the reader.

Thus far, we have defined Golay complementary sequences, proven some necessary conditions on the length of such sequences, and given recursive methods for finding more such sequences. Much of this was done by Golay early on. Now we move on to a more general construction.

2.   GOLAY ARRAY PAIRS

In [10], Fiedler, Jedwab, and Parker argue that a "Golay complementary sequence is naturally viewed as a projection of a multi-dimensional Golay array." A Golay complementary sequence is the simplest example of a Golay array pair (for the explicit relation between arrays and sequences, see Definition 2.1.3). Because this is among the most recent and most significant generalizations, I will use the Golay array structure in order to introduce the other extensions of Golay complementary sequences. A Golay array pair generalizes a Golay sequences both in terms of its dimensions and in terms of the entries of the sequence.

2.1. Definitions

Definition 2.1.1 An $n_1 \times n_2 \times \cdots \times n_r$ array is an *r*-dimensional matrix with complex entries $\mathcal{A} = (A[i_1, \cdots, i_r])$; $i_1, \cdots, i_r \in \mathbb{Z}$ where $A[i_1, \cdots, i_r] = 0$ *if* $\exists j$ *such that* $i_j < 0$ *or* $i_j \geq n_j$. The in-range entries are $\{A[i_1, \cdots, i_r] \,|\, \forall l, \, 0 \leq i_l < n_l\}$. Often, the in-range entries are constrained to lie in a certain alphabet. If this alphabet is $S_H = \{1, \mathcal{E}, \ldots, \mathcal{E}^{H-1}\}$ for H an even integer and $\mathcal{E} = e^{\frac{2p\sqrt{-1}}{H}}$, then we say that it is an H-phase array. All in-range entries are of the form $A[i_1, \cdots, i_r] = \mathcal{E}^{a[i_1, \cdots, i_r]}, a[i_1, \cdots, i_r] \in \mathbb{Z}_H$. Then $\mathcal{A}' = a[i_1, \cdots, i_r]$ is the array over $\mathbb{Z}_H$ corresponding to $\mathcal{A}$.

For z a complex number, let $\overline{z}$ be its complex conjugate.

Definition 2.1.2 The aperiodic autocorrelation function of an $n_1 \times n_2 \times \cdots \times n_r$ H-phase array is

$$\mathcal{C}_{\mathcal{A}}(u_1, \cdots, u_r)$$
$$= \sum_{i_1} \cdots \sum_{i_r} A[i_1, \cdots, i_r] \overline{A[i_1 + u_1, \cdots, i_r + u_r]}$$

Definition 2.1.3 An $n_1 \times n_2 \times \cdots \times n_r$ Golay array pair (GAP) is two $n_1 \times n_2 \times \cdots \times n_r$ arrays $\mathcal{A}$, $\mathcal{B}$ with the property that $\mathcal{C}_{\mathcal{A}}(u_1, \cdots, u_r) + \mathcal{C}_{\mathcal{B}}(u_1, \cdots, u_r) = 0, \ \forall (u_1, \cdots, u_r) \neq (0, \ldots, 0)$. Clearly, if $H=2$ and $r=1$, then an n_1 Golay array pair is an original Golay complementary sequence pair.

Just as we defined the single variable polynomial associated with a Golay sequence, we define the polynomial associated with the $n_1 \times n_2 \times \cdots \times n_r$ array $\mathcal{A} = (A[i_1, \cdots, i_r])$ to be

$$\mathcal{A}(z_1, \ldots z_r) = \sum_{i_1} \cdots \sum_{i_r} (A[i_1, \cdots, i_r] z_1^{i_1} \ldots z_r^{i_r})$$

For an analogous reason as that for the original Golay sequences, an equivalent condition for $\mathcal{A}$, $\mathcal{B}$ (as above) to be a Golay array pair is that $\mathcal{A}(z_1, \ldots z_r) \overline{\mathcal{A}(z_1^{-1}, \ldots z_r^{-1})} + \mathcal{B}(z_1, \ldots z_r) \overline{\mathcal{B}(z_1^{-1}, \ldots z_r^{-1})}$ be a constant.

Let $\mathcal{A} = (A[i_1, \cdots, i_r])$ be an $n_1 \times n_2 \times \cdots \times n_r$ array. If $\mathcal{A}$ is an H-phase array, define $\mathcal{A}^* = (A^*[i_1, \cdots, i_r])$ where

$$A^*[i_1, \cdots, i_r] = \overline{A[n_1 - 1 - i_1, \cdots, n_r - 1 - i_r]}$$

If $\mathcal{A}$ is an array over $\mathbb{Z}_H$, define

$$A^*[i_1, \cdots, i_r] = -A[n_1 - 1 - i_1, \cdots, n_r - 1 - i_r]$$

2.2. Restrictions on Length

Golay complementary sequences were previously shown to be quite restricted in terms of possible length.

Similarly, there are restrictions on the dimensions of Golay Array Pairs. What about a simpler case: restriction on lengths in the case of *r*=1, that is, polyphase Golay complementary sequences? They are not the same as those for binary Golay sequences. Already with *H*=4 (so that $S = \{1, -1, \ i, \ -i\}$), we see that there can be odd length Golay sequences:

Example 2.2.1 [1] Sequences $\begin{pmatrix} 1 & 1 & 1 & -i & i \end{pmatrix}$ and $\begin{pmatrix} 1 & i & -1 & 1 & -i \end{pmatrix}$ form a 4-phase Golay sequence pair.

In fact, for lengths less than 13, the only (non trivial) lengths for which 4-phase Golay sequences do not exist are lengths 7 and 9 (see [1]).

There is little known about the restrictions on lengths for polyphase sequences (and arrays of any dimension), because of the difficulty of performing effective computer searches at higher lengths.

2.3. Recursive Means of Producing Golay Array Pairs

As with Golay complementary sequences, we are concerned with finding Golay array pairs, and one way to do this is to use known pairs to create more. In fact, we can use complementary sequences to create Golay array pairs, and vice versa. The results of this section are taken from [5] and [10], where proofs are also to be found.

Let $n := n_1 \times n_2 \times \cdots \times n_r$; $m := m_1 \times m_2 \times \cdots \times m_r$; $i := i_1, i_2, \cdots, i_r$; $j := j_1, j_2, \cdots, j_s$

Theorem 2.3.1 Suppose there exists binary Golay complementary sequence pairs of lengths $n_1, n_2, \ldots n_r$. Then there exists an $n_1 \times n_2 \times \cdots \times n_r$ binary Golay array pair.

Because we know that binary Golay complementary sequence pairs exists for all lengths in the form of $2^a 10^b 26^c$, $a, b, c \in \mathbb{Z}, a, b, c \geq 0$, this theorem gives us that there exists a binary $n_1 \times n_2 \times \cdots \times n_r$ Golay array pair where $n_i = 2^{a_i} 10^{b_i} 26^{c_i}$, $a_i, b_i, c_i \in \mathbb{Z}, a_i, b_i, c_i \geq 0, \forall i$.

We also have the converse:

Theorem 2.3.2 Suppose there exists an $n_1 \times n_2 \times \cdots \times n_r$ binary Golay array pair. Then there exists binary Golay complementary sequence pairs of lengths $n_1, n_2, \ldots n_r$.

This theorem leads to some obvious restrictions on the dimensions of binary Golay array pairs. The product of all the dimensions must either be even or equal to one, and no dimension can have a prime factor congruent to 3 *modulo* 4.

The next theorems are used in Jedwab's three stage process of producing families of Golay complementary sequences [10].

For a positive integer n, let $\mathbf{n}^{(r)} := \overbrace{n \times \cdots \times n}^{r \ times}$.

Theorem 2.3.3 Let $n \geq 1$ be an integer. Then let $x := x_1, x_2, \ldots x_n$, and for $0 \leq k \leq n$ an integer, let $i_k := i_{k,1}, i_{k,2}, \ldots i_{k,r_k}$ and $s_k := s_{k,1} \times s_{k,2} \times \cdots \times s_{k,r_k}$. Suppose that $\mathcal{A}_k = (a_k[i_k])$, $\mathcal{B}_k = (b_k[i_k])$ form a Golay array pair over $\mathbb{Z}_H$ of size s_k. Then the arrays $\mathcal{F}_n = (f_n[i_0, i_1, \ldots, i_n, x])$, $\mathcal{G}_n = (g_n[i_0, i_1, \ldots, i_n, x])$

form a Golay array pair of size $s_0 \times s_1 \times \cdots \times s_n \times \mathbf{2}^{(m)}$ over $\mathbb{Z}_H$ where

$f_n [i_0, i_1, \ldots, i_n, x]$
$:= \sum_{k=1}^{n-1} \left(a_k [i_k] + a_k^* [i_k] - b_k [i_k] - b_k^* [i_k] + \frac{H}{2} \right) x_k x_{k+1}$
$+ \sum_{k=1}^{n} \left(b_{k-1}^* [i_{k-1}] + b_k [i_k] - a_{k-1} [i_{k-1}] - a_k [i_k] \right) x_k$
$+ \sum_{k=0}^{n} a_k [i_k]$

$$g_n [i_0, i_1, \ldots, i_n, x] := f_n' [i_0, i_1, \ldots, i_n, x] + \frac{H}{2} x_1$$

Where $f_n' [i_0, i_1, \ldots, i_n, x]$ is $f_n [i_0, i_1, \ldots, i_n, x]$ with $a_0 [i_0]$ and $a_0^* [i_0]$ interchanged, and $b_0 [i_0]$ and $b_0^* [i_0]$ interchanged.

This theorem gives a way of taking lower dimensional arrays and using them to construct suitable higher dimensional array pairs. It is the first step by which Jedwab produces Golay complementary sequences from the original Golay array pair. From these higher dimensional arrays, the next step is to produce more such pairs using affine offsets.

Theorem 2.3.4 Let $\mathcal{A} = (A [i_1, \cdots, i_r])$ and $\mathcal{B} = (B [i_1, \cdots, i_r])$ be an $n_1 \times n_2 \times \cdots \times n_r$ Golay array pair over $\mathbb{Z}_H$. Then for all $e_k, e_0' \in \mathbb{Z}_H$, $k = 0, 1, \ldots, r$, $\mathcal{A}' = (A' [i_1, \cdots, i_r])$ and $\mathcal{B}' = (B' [i_1, \cdots, i_r])$ form a Golay array pair, where

$$A' [i_1, \cdots, i_r] = A [i_1, \cdots, i_r] + \sum_{k=1}^{r} e_k i_k + e_0$$

$$B' [i_1, \cdots, i_r] = B [i_1, \cdots, i_r] + \sum_{k=1}^{r} e_k i_k + e_0'$$

This gives us affine offsets of a higher dimensional Golay array pair.

Let $\phi_{j,k}$ be the projection mapping from r dimensions to $r - 1$ dimensions which takes an $n_1 \times n_2 \times \cdots \times n_r$ array $\mathcal{A} = (A [i_1, \cdots, i_r])$ over $\mathbb{Z}_H$ to the $n_1 \times \cdots \times n_{j-1} \times n_{j+1} \times \cdots \times n_{k-1} \times n_j n_k \times n_{k+1} \times \cdots \times n_r$ array $\phi_{j,k} (\mathcal{A}) = \mathcal{B}$, where $B [i_1, \ldots, i_{j-1}, i_{j+1}, \ldots, i_{k-1}, i_j + s_j i_k, i_{k+1}, \ldots, i_r] = A [i_1, \cdots, i_r]$.

Theorem 2.3.5 Let $\mathcal{A} = (A [i_1, \cdots, i_r])$ and $\mathcal{B} = (B [i_1, \cdots, i_r])$ be an $n_1 \times n_2 \times \cdots \times n_r$ Golay array pair over $\mathbb{Z}_H$. Then $\phi_{j,k} (\mathcal{A})$ and $\phi_{j,k} (\mathcal{B})$ are an $n_1 \times \cdots \times n_{j-1} \times n_{j+1} \times \cdots \times n_{k-1} \times n_j n_k \times n_{k+1} \times \cdots \times n_r$ Golay array pair over $\mathbb{Z}_H$.

This theorem allows Golay sequences to be produced from these higher dimensional Golay array pairs and their affine offsets. Fiedler, Jedwab, and Parker additionally prove that the resulting lower dimension array is independent of the order of the projections (of the form in Theorem 2.34). This means that a succession of $r - 1$ projection mappings applied to an r dimensional array can be uniquely described by a directed path that is formed by labeling r vertices 1, 2, $\ldots, r$, and joining vertices j and k if one of the projection mappings is $\phi_{j,k}$.

Thus, the directed path is

$$\phi (1) \longrightarrow \phi (2) \longrightarrow \cdots \longrightarrow \phi (r)$$

Where ϕ is a permutation of $\{1, 2, \ldots, r\}$.

This is the final step that Fiedler, Jedwab, and Parker present. They use it to generate infinite families of Golay complementary sequences. Of the infinite families of Golay sequences that they generate, there are two types – standard, and nonstandard. Standard Golay sequences arose from the discovery of the connection of Golay sequences to the Reed-Muller code, and were first introduced in [13].

3. STANDARD GOLAY SEQUENCES

3.1. Introduction to OFDM and Peak to Mean Envelope Power Ratio Reduction

In order to understand the motivation behind the creation of large sets of Golay complementary sequences, we must look at orthogonal frequency-division multiplexing (OFDM) and the PMEPR (peak to mean envelope power ratio) of Golay complementary sequences. OFDM is way of transmitting data using multicarriers which are all orthogonal to each other [8]. The chief practical difficulty of OFDM is that the peak envelope power over n carriers can be as much n times the mean envelope power. It is advantageous to find codes with have a low PMEPR. Codes, generally speaking, are the rules by which a piece of information is encoded. Here, they are chiefly represented as sequences. The code rate is "the ratio of the number of information bits to the number of coded bits" [13].

Suppose there are n carriers. As done in [13], let $a_i \in \mathbb{Z}_H$ denote the value of the ith carrier over a certain symbol period, and f_i the frequency of the ith carrier. Then $a_i(t)$ is constant over the symbol period. The sequence $(a_0 \ a_1 \cdots \ a_{n-1})$ is a codeword for each symbol period. Let f be a constant and $\triangle f$ an integer multiple of the code rate. In order that the f_i be orthogonal, they are constrained by the relation $f_i = f + i \triangle f$. As before, let $\mathcal{E} = e^{\frac{2p\sqrt{-1}}{H}}$. The OFDM signal that is transmitted is the real part of

$$s (t) = \sum_{i=0}^{n-1} \mathcal{E}^{a_i(t) + H f_i t}$$

The envelope power $P_a(t)$ of the sequence $a = (a_0 \ a_1 \ldots a_{n-1})$ is

$P_a (t)$
$= n + \sum_{u \neq 0} \sum_i \mathcal{E}^{a_i(t) - a_{i+u}(t) - Hu \triangle ft}$
$= n + \sum_{u \neq 0} \mathcal{E}^{-Hu \triangle ft} \mathcal{C}_a (u)$

Note that $P_a (t) = \left| s_a (t) \right|^2$ The peak envelope power (PEP) of a sequence a is the supremum of $P_a (t)$ over a symbol period. The PMEPR is PEP/n.

Claim 3.1.1 A sequence which is a member of Golay complementary pair has PMEPR less than 2.

Proof: Let $a = \begin{pmatrix} a_0 & a_1 \cdots & a_{n-1} \end{pmatrix}$ and $b = \begin{pmatrix} b_0 & b_1 \cdots & b_{n-1} \end{pmatrix}$ be a Golay complementary pair. So we have $\mathcal{C}_a(u) + \mathcal{C}_b(u) = 0$. Thus,

$$P_a(t) + P_b(t) = 2n + \sum_{u \neq 0} \mathcal{E}^{-Hu \triangle ft}[\mathcal{C}_a(u) + \mathcal{C}_b(u)] = 2n$$

As $P_a(t) = \left| s_a(t) \right|^2 \geq 0$, $P_b(t) \leq 2n$. So the PMEPR of b is less than 2.

3.2.　The Reed Muller Code and Generalizations

The Reed Muller Code is an error correcting code. They are high related to Golay complementary sequences, indeed, the Reed Muller Code gives a family of Golay complementary sequences known as Standard Golay Complementary Sequences.

Definition 3.2.1 A Generalized Boolean function is a mapping $\{0,1\}^m \rightarrow \mathbb{Z}_H$ in m $\{0,1\}$-valued variables $x_1, x_2, \ldots, x_m$. It is easily proved that all generalized Boolean functions can be represented as a $\mathbb{Z}_H$ linear combination of the monomials $x_1, x_2, \ldots, x_m, x_1 x_2, x_1 x_3, \ldots, x_1 x_2 \cdots x_m$. With every Boolean function f there is an associated sequence of length 2^m: $\begin{pmatrix} f_0 & f_1 & \cdots & f_{2^m-1} \end{pmatrix}$. For a nonnegative integer $i \leq 2^m - 1$, let $i = \sum_{k=0}^{m-1} i_k 2^k$, so that $\begin{pmatrix} i_0 & i_1 & \cdots & i_{m-1} \end{pmatrix}$ is the binary form of i. Then $f_i = f\begin{pmatrix} i_0 & i_1 & \cdots & i_{m-1} \end{pmatrix}$ in the vector associated with f. Alternatively, given a sequence of length 2^m, its algebraic normal form is the generalized Boolean function associated with it.

Definition 3.2.2 The rth order generalized Reed Muller code over is the linear code generated (over $\mathbb{Z}_H$, H even) by the Boolean monomials in $x_1, x_2, \ldots, x_m$ of degree at most r. It is denoted $RM_H(r, m)$.

For a (non-binary) code, the Lee weight of a codeword $a = \begin{pmatrix} a_0 & a_1 \cdots & a_{n-1} \end{pmatrix}$ in $\mathbb{Z}_H$ is

$$wt_L(a) = \sum_{i=0}^{n-1} min\{a_i, H - a_i \, mod \, H\}$$

The Lee distance between two codewords a and b is $d_L(a, b) = wt_L(a - b \, mod \, H)$

Theorem 3.2.3 ([4], Theorem 4) The minimum Lee distance of the $RM_H(r, m)$ code is 2^{m-r}.

3.3.　Standard Golay Sequences

Davis and Jedwab, in [13], were the first to recognize the connection between Golay complementary sequences and the Reed Muller (RM) Codes: that is, standard Golay sequences are cosets of the first order RM Code in the second order RM Code.

Theorem 3.3.1 ([10], Theorem 9) Let $m \geq 1$, $m \in \mathbb{Z}$, ϕ be a permutation of $\{1, 2, \ldots, m\}$, and $e_k, e_0' \in \mathbb{Z}_H$, $k = 0, 1, \ldots, m$. Then the sequences $\mathbf{a}$ and $\mathbf{b}$ whose algebraic normal forms are $a(x_1, x_2, \ldots, x_m)$ and $b(x_1, x_2, \ldots, x_m)$ respectively form a Golay complementary sequence pair, where

$a(x_1, x_2, \ldots, x_m)$
$= \frac{H}{2} \sum_{k=1}^{m-1} x_{\phi(k)} x_{\phi(k+1)} + \sum_{k=1}^{m} e_k x_{\phi(k)} + e_0$

$b(x_1, x_2, \ldots, x_m)$
$= \frac{H}{2} \sum_{k=1}^{m-1} x_{\phi(k)} x_{\phi(k+1)} + \sum_{k=1}^{m} e_k x_{\phi(k)} + e_0' + \frac{H}{2} x_{\phi(1)}$

For $m = 1$, there are H^2 Golay sequences of length 2^m over $\mathbb{Z}_H$ of this form, and for $m > 1$, there are $\frac{H^{m+1} m!}{2}$ such sequences.

Paterson, who in [4] generalized the results of Davis and Jedwab ([13]), used an inductive proof for this theorem, which – while less concise than that of Davis and Jedwab – suggests that standard Golay sequences can be seen as arising from an "iterative construction method applied to Golay complementary pairs of length 2". The proof of Theorem 3.13 in [10] is, in a way, an extension of this idea, because it uses the three stage process of Fiedler, Jedwab, and Parker discussed above. I won't discuss the details of the proof, but just the general outline.

Stage 1: Clearly, the sequences $\mathbf{a}$ and $\mathbf{b}$ whose algebraic normal forms are $a(x_1, x_2, \ldots, x_m)$ and $b(x_1, x_2, \ldots, x_m)$ are those that, for all $(x_1, x_2, \ldots, x_m) \in \mathbb{Z}_2^m$, satisfy
$\mathbf{a}(x_m + 2x_{m-1} + \cdots + 2^{m-1} x_1)$
$= \frac{H}{2} \sum_{k=1}^{m-1} x_{\phi(k)} x_{\phi(k+1)} + \sum_{k=1}^{m} e_k x_{\phi(k)} + e_0$

$\mathbf{b}(x_m + 2x_{m-1} + \cdots + 2^{m-1} x_1)$
$= \frac{H}{2} \sum_{k=1}^{m-1} x_{\phi(k)} x_{\phi(k+1)} + \sum_{k=1}^{m} e_k x_{\phi(k)} + e_0' + \frac{H}{2} x_{\phi(1)}$

Using Theorem 2.33, with input pairs $\mathcal{A}_k = (0)$ and $\mathcal{B}_k = (0)$ for all $k = 0, 1, \ldots, n$, we get arrays $\mathcal{F}_n$ and $\mathcal{G}_n$ of size $\mathbf{1}^{(n+1)} \times \mathbf{2}^{(n)}$ over $\mathbb{Z}_H$. Remove all dimensions of size 1. We get $\mathcal{F}' = F'[x_1, \cdots, x_n]$ and $\mathcal{G}' = G'[x_1, \cdots, x_n]$ of size $\mathbf{2}^{(n)}$ over $\mathbb{Z}_H$, which, for all $(x_1, x_2, \ldots, x_n) \in \mathbb{Z}_2^n$, satisfy

$$F'[x_1, \cdots, x_n] = \sum_{k=1}^{n-1} \frac{H}{2} x_k x_{k+1}$$

$$G'[x_1, \cdots, x_n] = \sum_{k=1}^{n-1} \frac{H}{2} x_k x_{k+1} + \frac{H}{2} x_1$$

So $\mathcal{F}'$ and $\mathcal{G}'$ form a Golay array pair.

The second step is to use Theorem 2.34, and take affine offsets. Thus, we know that for all $e_k, e_0' \in \mathbb{Z}_H$,

$k = 0, 1, \ldots, n,$

$$f' [x_1, \cdots, x_n] = \sum_{k=1}^{n-1} \frac{H}{2} x_k x_{k+1} + \sum_{k=1}^{n} x_k i_k + e_0$$

$$g' [x_1, \cdots, x_n] = \sum_{k=1}^{n-1} \frac{H}{2} x_k x_{k+1} + \sum_{k=1}^{n} x_k i_k + e_0' + \frac{H}{2} x_1$$

Then $\left(f' [x_1, \cdots, x_n] \right)$ and $\left(g' [x_1, \cdots, x_n] \right)$ form a Golay array pair of size $\mathbf{2}^{(n)}$ over $\mathbb{Z}_H$. The third stage requires more results given in [10], and so we won't include all the steps here. It should be fairly clear, however, that these arrays can be mapped to the Golay sequence pairs of Theorem 3.13.

Despite the advantages of the Reed Muller code, this construction does not guarantee a low code rate.

Later on, when we consider complementary sets and multiple L-shift sequences, we will discuss the ways they have been used to further the ideas in this section. For now, however, we will turn to non-Standard Golay sequences.

3.4. Non-Standard Golay Sequences

Until Li and Chu's 2005 paper [9], it was believed that standard Golay sequences were the only Golay sequences that existed for lengths 2^m over $\mathbb{Z}_{2^k}$. However, through exhaustive computer searches, Li and Chu found 1024 quaternary Golay sequences of length 16 that could not be produced by Theorem 3.31. This discovery prompted investigation into the structure of this family of Golay sequences (see [6] and [14]), and later to other families of non-standard Golay sequences (see [12]).

Lemma 3.4.1 Let $\mathcal{A} = (A[i_1, \cdots, i_r])$ be an $n_1 \times n_2 \times \cdots \times n_r$ array over $\mathbb{Z}_H$. Let $\mathcal{A}' = (A'[i_1, \cdots, i_r])$ where $A'[i_1, \cdots, i_r] = A[i_1, \cdots, i_r] + c; c \in \mathbb{Z}_H$, and let $\mathcal{A}^* = (A^*[i_1, \cdots, i_r])$ be as usual. Then

$$\mathcal{C}_\mathcal{A}(u_1, \cdots, u_r) = \mathcal{C}_{\mathcal{A}'}(u_1, \cdots, u_r) = \mathcal{C}_{\mathcal{A}^*}(u_1, \cdots, u_r),$$
$$\forall (u_1, \cdots, u_r) \neq (0, \cdots, 0)$$

Proof: Let $(u_1, \cdots, u_r) \neq (0, \cdots, 0)$

$\mathcal{C}_{\mathcal{A}'}(u_1, \cdots, u_r)$
$= \sum_{i_1} \cdots \sum_{i_r} \mathcal{E}^{A[i_1, \cdots, i_r] + c - A[i_1 + u_1, \cdots, i_r + u_r] - c}$
$= \sum_{i_1} \cdots \sum_{i_r} \mathcal{E}^{A[i_1, \cdots, i_r] - A[i_1 + u_1, \cdots, i_r + u_r]}$
$= \mathcal{C}_\mathcal{A}(u_1, \cdots, u_r)$

$\mathcal{C}_{\mathcal{A}^*}(u_1, \cdots, u_r)$
$= \sum_{i_1} \cdots \sum_{i_r} \mathcal{E}^{-A[n_1 - 1 - i_1, \cdots, n_r - 1 - i_r]}$
$\quad + A[n_1 - 1 - i_1 + u_1, \cdots, n_r - 1 - i_r + u_r]$
$= \sum_{i_1} \cdots \sum_{i_r} \mathcal{E}^{A[i_1, \cdots, i_r] - A[i_1 + u_1, \cdots, i_r + u_r]}$

$= \mathcal{C}_\mathcal{A}(u_1, \cdots, u_r)$

Let $\mathcal{A} = (A[i_1, \cdots, i_r])$ be an $n_1 \times n_2 \times \cdots \times n_r$ array over $\mathbb{Z}_H$.

$E(\mathcal{A})$

$:= \{(A[i_1, \cdots, i_r] + c \mid c \in \mathbb{Z}_H)\} \cup \{(A^*[i_1, \cdots, i_r] + c) \mid c \in \mathbb{Z}_H\}$

All arrays in $E(\mathcal{A})$ have identical autocorrelation functions. The number of arrays in $E(\mathcal{A})$ is H or $2H$, if $A[i_1, \cdots, i_r] + c = A^*[i_1, \cdots, i_r]$ for some $c \in \mathbb{Z}_H$ or not, respectively.

This construction allows us to formulate the autocorrelation crossover property, which is how the aforementioned nonstandard quaternary Golay sequences of length 16 are produced.

Suppose $\mathcal{A}$, $\mathcal{A}'$, $\mathcal{B}$, and $\mathcal{B}'$ are arrays and that $\mathcal{C}_\mathcal{A}(u_1, \cdots, u_r) = \mathcal{C}_{\mathcal{A}'}(u_1, \cdots, u_r)$ and $\mathcal{C}_\mathcal{B}(u_1, \cdots, u_r) = \mathcal{C}_{\mathcal{B}'}(u_1, \cdots, u_r)$ for all $(u_1, \cdots, u_r) \neq (0, \cdots, 0)$. Additionally, suppose that $E(\mathcal{A}) \neq E(\mathcal{A}')$, $E(\mathcal{B}) \neq E(\mathcal{B}')$, and that $(\mathcal{A}, \mathcal{B})$ and $(\mathcal{A}', \mathcal{B}')$ are Golay array pairs. Then $(\mathcal{A}, \mathcal{A}')$ are called a cross over pair. The cross over property then applies – in other words, we know that any $\mathcal{F} \in E(\mathcal{A})$ and $\mathcal{G} \in E(\mathcal{B}')$ form a Golay pair. Similarly, any $\mathcal{F} \in E\left(\mathcal{A}'\right)$ and $\mathcal{G} \in E(\mathcal{B})$ form a Golay pair.

We have seen that standard Golay sequences can be generated using the three stage process with trivial inputs in the first stage. Quaternary nonstandard Golay sequences of length 2^m can be produced using the three stage process – but their inputs are not trivial, but rather must come from the following cross-over sequences:

Example 3.4.2 (Thm 12) Let $a = \begin{pmatrix} 0 & 0 & 0 & 2 & 0 & 0 & 2 & 0 \end{pmatrix}$ and $b = \begin{pmatrix} 0 & 1 & 1 & 2 & 0 & 3 & 3 & 2 \end{pmatrix}$. Then each of the pairs in the following set is a cross over pair.

$$\{(a, b); (a^*, b); (a, b^*); (a^*, b^*); (b, a); (b, a^*); (b^*, a); (b^*, a^*)\}$$

Additionally, all quaternary Golay cross over pairs of length 8 are formed by affine offsets of these pairs (see [14]).

4. FURTHER EXTENSIONS OF GOLAY COMPLEMENTARY SEQUENCES

4.1. Complementary Sets

Definition 4.1.1 Let $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ be n $n_1 \times n_2 \times \cdots \times n_r$ arrays over $\mathbb{Z}_H$. Then they are called a Golay complementary set if

$$\sum_{i=1}^{n} \mathcal{C}_{\mathcal{A}_i}(u_1, \cdots, u_r) = 0$$

One of the earlier and more significant papers on Golay complementary sets of sequences is [3].

Theorem 4.1.2 [3] A binary complementary set of *sequences* contains an even number of sequences.

Theorem 4.1.3
Let $\{\mathcal{A}_i = A_i\,[i_1, \cdots, i_r]\,|\,i = 1, 2, \ldots, n\}$ be a Golay complementary set of n $n_1 \times n_2 \times \cdots \times n_r$ arrays over $\mathbb{Z}_H$. Let $e_k, b_j \in \mathbb{Z}_H$, $k = 0, 1, \ldots, r$, $j = 1, 2, \ldots n$.

$$A_i^{'}\,[i_1, \cdots, i_r] = A_i\,[i_1, \cdots, i_r] + \sum_{k=1}^{r} e_k i_k + b_i$$

Then replacing any $\mathcal{A}_i$ with $\mathcal{A}_i^{*} = (A_i^{*}\,[i_1, \cdots, i_r])$ still leaves a complementary set. Additionally, $\left\{\mathcal{A}_i^{'} = A_i^{'}\,[i_1, \cdots, i_r]\,|\,i = 1, 2, \ldots, n\right\}$ is a Golay complementary set where

$$A_i^{'}\,[i_1, \cdots, i_r] = A_i\,[i_1, \cdots, i_r] + \sum_{k=1}^{r} e_k i_k + b_i$$

Proof: We know that
$\mathcal{C}_{\mathcal{A}_i}(u_1, \cdots, u_r) = \mathcal{C}_{\mathcal{A}_i^{*}}(u_1, \cdots, u_r)$,
$\forall\,(u_1, \cdots, u_r) \neq (0, \cdots, 0)$

Thus, the first statement in the theorem is trivial. Let $(u_1, \cdots, u_r) \neq (0, \cdots, 0)$. Then
$\mathcal{C}_{\mathcal{A}_i^{'}}(u_1, \cdots, u_r)$
$= \sum_{i_1} \cdots \sum_{i_r} \mathcal{E}^{A_i[i_1, \cdots, i_r] + \sum_{k=1}^{r} e_k i_k + b_i - (A[i_1 + u_1, \cdots, i_r + u_r] + \sum_{k=1}^{r} e_k(i_k + u_k) + b_i)}$
$= \sum_{i_1} \cdots \sum_{i_r} \mathcal{E}^{A_i[i_1, \cdots, i_r] - A[i_1 + u_1, \cdots, i_r + u_r] - \sum_{k=1}^{r} e_k u_k}$
$= \mathcal{E}^{-\sum_{k=1}^{r} e_k u_k} \mathcal{C}_{\mathcal{A}_i}(u_1, \cdots, u_r)$

Thus,
$\sum_{i=1}^{n} \mathcal{C}_{\mathcal{A}_i^{'}}(u_1, \cdots, u_r)$
$= \sum_{i=1}^{n} \mathcal{E}^{-\sum_{k=1}^{r} e_k u_k} \mathcal{C}_{\mathcal{A}_i}(u_1, \cdots, u_r)$
$= \mathcal{E}^{-\sum_{k=1}^{r} e_k u_k} \sum_{i=1}^{n} \mathcal{C}_{\mathcal{A}_i}(u_1, \cdots, u_r)$
$= 0$

This proves the second part of the statement.

Claim 4.1.4 A sequence which is a member of Golay complementary set of size k where each sequence is of length n has PMEPR less than k.

Proof: Let $\{a_i|i = 1, 2, \ldots k\}$ and be a Golay complementary set. So we have

$$\sum_{i=1}^{k} \mathcal{C}_{a_i}(u) = 0$$

Thus,

$$\sum_{i=1}^{k} P_{a_i}(t) = kn + \sum_{u \neq 0} [\mathcal{E}^{-Hu\triangle ft} \sum_{i=1}^{k} \mathcal{C}_{a_i}(u)] = kn$$

As $P_{a_i}(t) = \big|\,s_{a_i}(t)\,\big|^2 \geq 0$, $P_{a_i}(t) \leq kn$. So the PMEPR of a_i is less than k.

Because the code rate of standard Golay sequences rapidly approaches 0 as the length of the sequences increases, it can be advantageous to use Golay complementary sets that are related to Reed-Muller Codes in

a similar way. See Theorem 1 in [18] for a family of Golay complementary sets. For more information, see [4], [15], and [18].

4.2.   Multiple L-Shift Complementary Sequences

Multiple L-shift complementary sequences were introduced as an alternative to Golay complementary sets as a way of increasing the code rate of standard sequences.

Definition 4.2.1 Let a and b be H-phase sequences of length n, and $L \in \mathbb{Z}$. If

$$\mathcal{C}_a(u) + \mathcal{C}_b(u) = 0, \ for \ u = 0 \ mod \ L, \ u \neq 0$$

Then a and b are said to be a multiple L-shift complementary sequence pair.

I have extended this definition to arrays.

Definition 4.2.2 Let $\mathcal{A}$ and $\mathcal{B}$ be $n_1 \times n_2 \times \cdots \times n_r$ H-phase arrays and $L \in \mathbb{Z}$. If

$\mathcal{C}_\mathcal{A}(u_1, \cdots, u_r) + \mathcal{C}_\mathcal{B}(u_1, \cdots, u_r) = 0$
$for \ (u_1, \cdots, u_r) = (0, \ldots, 0)\,mod\,L, \ (u_1, \cdots, u_r) \neq (0, \ldots, 0)$

Then $\mathcal{A}$ and $\mathcal{B}$ are multiple L-shift complementary array pair.

This is a broader definition than that for Golay array pairs. As for Golay array pairs, replacing a multiple L-shift complementary array pair $\mathcal{A}$ and $\mathcal{B}$ with $\mathcal{A}^{*}$ and $\mathcal{B}^{*}$ or with affine offsets of $\mathcal{A}$ and $\mathcal{B}$ does not change the complementary property.

Claim 4.2.4 [Theorem 1, 7] A sequence which is a member of multiple L-shift complementary pair has PMEPR less than $2L$.

Theorem 4.2.5 [Theorem 2, 7] Let $L = 2^d$, ϕ a permutation of $\{1, 2, \ldots, m - d\}$, and $e_{k,l}, e_l^{'} \in \mathbb{Z}_H$. Let

$f(x_1, x_2, \ldots, x_m)$
$= \frac{H}{2} \sum_{k=1}^{m-d} x_{\phi(k)} x_{\phi(k+1)}$
$+ \sum_{k=1, \ k \neq l}^{m} \sum_{l=m-d+1}^{m} e_{k,l} x_k x_l + \sum_{l=1}^{m} e_l x_l$

Let $e, \ e^{'} \in \mathbb{Z}_H$ Then the sequences **a** and **b** whose algebraic normal forms are $a(x_1, x_2, \ldots, x_m)$ and $b(x_1, x_2, \ldots, x_m)$ respectively form a multiple L-shift complementary sequence pair, where

$$a(x_1, x_2, \ldots, x_m) = f(x_1, x_2, \ldots, x_m) + e$$

$$b(x_1, x_2, \ldots, x_m) = f(x_1, x_2, \ldots, x_m) + e^{'} + 2^{h-1} x_{\phi(1)}$$

Clearly, these are related to standard Golay sequences.

There are a number of known results on the possible lengths of multiple L-shift complementary sequence pairs. For example, they exist for length n where $L \leq n \leq 2L$, and the length must be the sum of at most 2L squares. For the proofs of these and other results, see [11].

CONCLUSION

Golay complementary sequences, from their beginning in Golay's paper, have been developed and generalized considerably. It is interesting to see the developments that have occurred over the years. Golay's ideas have been generalized from pairs of binary sequences to sets of arrays with entries over $\mathbb{Z}_H$. Questions in the area focus on possible lengths of Golay sequences, their enumeration, and methods of construction, particularly within the newer the constructions of Golay arrays and Golay sets. In this paper, we have focused on the applications within coding theory of Golay complementary sequences, because much of the development of the theory of Golay sequences has been within this context. However, Golay complementary sequences have been used in many fields, including in Ising spin systems (physics), infrared spectrometry, Hadamard matrices, and various communications systems [19]. Thus, they have both important applications, as well as involving interesting mathematics.

REFERENCES

1. W. H. Holzmann, H. Kharagani, "A Computer Search for Complex Golay Sequences", Australasian Journal of Combinatorics, pp.251-258, 10 (1994) 4,10

2. M. J. E. Golay, "Complementary Series," IRE Trans. Inform. Theory, IT-7, pp. 82-87, Apr. 1961.

3. C. C. Tseng, C. L. Liu, "Complementary sets of sequences," IEEE Trans. Inform. Theory, IT-18, no. 5, pp. 644-651, Sept. 1972

4. K. G. Paterson, "Generalized Reed-Muller codes and power control in OFDM modulation", IEEE Trans. Inform. Theory, IT-46, no. 1, 104-120, Jan. 2000

5. J. Jedwab, M. Parker, "Golay complementary array pairs", Designs, Codes, and Cryptography, Vol 44, Sept. 2007

6. F. Fiedler, J. Jedwab, "How Do More Golay Sequences Arise", IEEE Trans. Inform. Theory, Vol. 52, 4261—4266, May 2005

7. W. Chen, C. Tellambura, "Identifying a Class of Multiple Shift Complementary Sequences in the Second Order Cosets of the First Order Reed-Muller Codes", Communications, Vol. 1, 618-721, 2005

8. D. Matiae, "OFDM as a possible modulation technique for multimedia applications in the range of mm waves", Introduction to OFDM, II Edition, October 1998

9. Y. Li and W. B. Chu, "More Golay sequences", IEEE Transactions on Information Theory, vol. 51, no. 3, pp. 1141-1145, 2005.

10. F. Fiedler, J. Jedwab and M.G. Parker, A multidimensional approach to the construction and enumeration of Golay complementary sequences, J. Combinatorial Theory (Series A), 2007.

11. Y. Xin and I. Fair, Multiple L-shift Complementary Sequences, IEICE Trans. Fundamentals, Vol.E89–A, No.10 October 2006

12. F. Fiedler, J. Jedwab, and A. Wiebe, A new source of seed pairs for Golay sequences of length 2m, J. Combinatorial Theory (Series A), 117 (2010), 589–597

13. J . A. Davis and J. Jedwab, Peak-to-mean power control in OFDM, Golay complementary sequences, and Reed-Muller codes, IEEE Trans .Inform. Theory, vol. 45. pp. 2397-2417. Nov. 1999.

14. Y. Li and Y. C. Kao, "Structures of Non-GDJ Golay Sequences," Proceedings of the 2005 International Symposium on Information Theory (ISIT 2005), pp. 378-381, Adelaide, Australia, Sept. 2005.

15. Schmidt, K.-U.; , "Complementary Sets, Generalized Reed–Muller Codes, and Power Control for OFDM," Information Theory, IEEE Transactions on , vol.53, no.2, pp.808-814, Feb. 2007

16. R. Turyn and J. Storer, "On Binary Sequences", Proceedings of the American Mathematical Society , Vol. 12, No. 3 (Jun., 1961), pp. 394-399

17. Eliahou S., Kervaire M., Saffari B., On Golay polynomial pairs. Adv. Appl. Math. 12, 235–292 (1991).

18. Schmidt, K.-U., "Complementary Sets, Generalized Reed–Muller Codes, and Power Control for OFDM," Information Theory, IEEE Transactions on , vol.53, no.2, pp.808-814, Feb. 2007

19. M. Parker, K. Paterson, C. Tellambura, "Golay Complementary Sequences", Jan. 2004. http://www.isg.rhul.ac.uk/˜kp/golaysurvey.pdf

Examining Eating Habits of Undergraduate Psychology Students

Wendy Salvisberg, Peter Tom, and Sandra Ziolkowski
Department of Psychology, University of Alberta

The current study examined the relationship between parenting styles, eating and dieting behaviors, and self-esteem in undergraduate psychology students. Standardized assessments of eating and dieting behaviors, parent care and control, and self-esteem were collected from 99 undergraduate psychology students in November 2012. Expected associations of parent care and control with undergraduate students' eating and dieting habits and self-esteem were not found; however, results indicating the influence of mother care and father control on potential eating disorder diagnosis and self-esteem were discovered. Together, these results suggest that more research is needed to explore other characteristics of parenting that may be more related to unhealthy eating and dieting behaviours and self-esteem.

PARENT INFLUENCE DURING CHILDHOOD AND ADOLESCENCE

Parent interactions have a substantial impact on a child's development (Bartholomow, 1990; Bowlby, 1982; Brofenbrenner, 1979; Scaglioni, Salvioni, & Galimberti, 2008). The way parents interact with their children may influence the development of a child's physical, emotional and mental well-being, which may lead to negative repercussions in adolescence and adulthood. Specifically, family dynamics can powerfully influence the development and maintenance of unhealthy eating/dieting behaviors and lowered self-esteem and in extreme forms, initiate the development of eating disorders (Minuchin, Rosman, & Baker, 1978; Rodriguez, Novalbos Ruiz, Martinez Nieto, Escobar Jimenez, & Castro de Haro, 2004; Sira & White, 2010). Minuchin et al. (1978) proposed a familial model to explain the development of anorexia nervosa. They claimed that the interaction between enmeshment, overprotectiveness, rigidity and lack of conflict resolution within psychosomatic families ultimately contributes to the development of an eating disorder. Researchers have continued to support the link between families characterized by overprotection and excessive control, and anorexia nervosa (Cordero and Israel, 2009; Galloway, Farrow and Martz, 2010; Perry et al., 2006; Shoebridge & Gower, 2000; Sim et al., 2009; Sira & White, 2010). Research indicates that eating disturbances are often accompanied by lowered self-esteem (Button, Loan, Davies, & Sonuga-Barke, 1998).

Anorexia Nervosa, Bulimia Nervosa, and Eating Disorder Not Otherwise Specified

According to the Diagnostic and Statistical Manual, Fourth Edition Text-Revised (DSM-IV-TR) (American Psychiatric Association, 2000), an individual meeting the criteria for anorexia nervosa displays behaviours such as refusal to maintain a healthy body weight at or above a minimally normal weight for age and height, amenorrhea, an intense fear of gaining weight, and disturbances in self and body image. Individuals with anorexia may be typed as either restricting or binge-eating/purging. Individuals diagnosed with bulimia nervosa may exhibit recurrent episodes of binge eating accompanied by a sense of lack of control over eating, recurrent inappropriate compensatory behaviours to prevent weight gain, and disturbance in self and body image. The binge eating and compensatory behaviours occur, on average, twice a week for 3 months. Bulimia may be classified as either purging or non-purging type. Individuals exhibiting symptoms but not meeting the criteria for any specific eating disorder may be classified as eating disorder not otherwise specified.

Theory on Adolescent Eating and Dieting Behaviours, and Self-Esteem

The model on psychosomatic families presented by Minuchin et al. (1978) is consistent with Brofenbrenner's ecological systems theory (1979), which theorizes that child development is powerfully influenced by parental behaviors, peer relationships, and societal interactions. The environmental context in which children are raised, most importantly the home setting, helps predict developmental patterns and later emotional development in adolescents (Brofenbrenner, 1979). Parents create environments for children that may foster the development of healthy eating/dieting behaviors, or promote obesity and disordered eating (Scaglioni, Salvioni, & Galimberti, 2008).

Attachment theorists argue that children develop personal implicit attitudes and beliefs about the self on the basis of parental interactions, specifically based on the responsiveness and sensitivity of primary caregivers (Bartholomow, 1990; Bowlby, 1982). Repeated parent-child interactions thus form the basis for child mental representations of the self in relation to others. Over time and with repeated exposure, how parents respond to children leads to the development of a schema that determines how a child forms and organizes information about the self in relation to other people.

Parenting style can also have a powerful influence on the formation of both explicit and implicit self-esteem

(Baumrind, 1971; Steinberg & Morris, 2001). According to Baumrind (1971, p. 22-23), (1) authoritarian parent-child interactions characterized by a lack of nurturance and warmth or (2) permissive parent-child interactions characterized by a lack of control and guidance may lead to lower levels of self-esteem in adolescence. Authoritative parent-child interactions, in comparison, which are characterized with warmth and emotional acceptance, in addition to clearly outlined expectations for acceptable behavior, have been consistently linked with greater adolescent adjustment compared to the two former parenting styles. DeHart, Pelham, and Tennen (2006) further argue that young adults whose parents were consistently nurturing and affectionate reported higher levels of self-esteem compared to parenting styles characterized with lower levels of nurturance and cold behavior.

As healthy self-esteem and eating/dieting behaviors in childhood and adolescence have been found to be crucial indicators of later emotional well-being and mental health, understanding the ways in which parent care and control influences these areas of child development continues to be a crucial area of study. In particular, determining the extent that parents influence eating and dieting behaviours and self-esteem in adolescence is an area of particular importance. In the current study, the effect parent care and control has on undergraduate student's eating/dieting behaviours and self-esteem is investigated.

Research on Parent Care and Control, Adolescent Eating/Dieting Behaviours, and Self-Esteem

The impact of parent care and control on adolescent eating/dieting behaviours and self-esteem has been investigated in a variety of studies. For example, Renk, McKinney, Klein and Oliveros (2006) conducted a study examining the impact of parent care (i.e., emotional availability) and control (i.e., forms of discipline) on female college students' self-esteem by administering standardized questionnaires. Results of the study suggest that high levels of parent control were inversely correlated with low levels of self-esteem. However, there was no relation between parent care and self-esteem in female college students.

Studies have also investigated the influence of parental control on eating/dieting behaviors independent from self-esteem. For example, Galloway, Farrow and Martz (2010) examined the influence of parent control (i.e., controlled child feeding practices) on college students' eating/dieting behaviours. Parents completed standardized questionnaires of personal parenting styles. Significant positive correlations were found between parent control and current disordered eating patterns in college students.

Cordero and Israel (2009) examined whether perceptions of parent care (i.e., parental acceptance and emotional availability) and parent control (i.e., verbal messages relating to eating and exercise regulation) are related to risk factors pertaining to eating pathology (i.e., low self-esteem, body mass index (BMI)) in college women. Expected negative relations between parent care and daughters' eating patterns were not found; however, an element of parent control (i.e., verbal messages relating to eating and exercise regulation) was related to eating pathology. In addition, Cordero and Israel found a negative correlation between self-esteem and unhealthy eating.

Sira & White (2010) conducted a study investigating the influence of parent care and control on self-esteem, eating/dieting behaviours, and body satisfaction in college students. Based on correlational analyses, high levels of parent control and low levels of paternal care were associated directly with high disordered eating attitude scores among males. In addition, parent control was associated with eating disturbances among males, but not among females. A relationship was found between parental control and other female domains (i.e., body satisfaction), which in turn increased dieting behavior. No direct relationship was found between parent care and eating/dieting behaviours. Self-esteem was directly impacted by parent care and control for both genders.

As demonstrated by the previous studies, many researchers have not found direct relationships between parent care, eating/dieting behaviours, and self-esteem with the exception of Sira & White (2010). In addition, Perry et al. (2006) investigated the relationship between parent care (i.e., affection and warmth, or coldness and rejection) and control (i.e., encouragement of autonomy/ independence as opposed to strict control with regulation and intrusions), self-esteem and eating behaviours (i.e., eating disturbances) in university students. Unlike many of the previous studies identified, the researchers identified low levels of caring and emotional support and high levels of control to be associated with distorted eating behaviours. Low levels of self-esteem were related to unhealthy eating and dieting behaviours.

It is evident that there are a variety of similarities and differences in the literature investigating the relationship between parent care and control, adolescent eating/dieting behaviours, and self-esteem. Firstly, all of the studies collected data by administering standardized questionnaires to adolescents only; however, Galloway, Farrow and Martz (2010) obtained their results from administering questionnaires to both student and parent participants. Each of these studies used instruments examining respective measures in adolescent or college populations, including the Eating Attitudes Test (EAT-26) (Garner, Olmsted, Bohr, & Garfinkel, 1982) (Cordero and Israel, 2009; Galloway, Farrow and Martz, 2010; Perry et al., 2006; Sira & White, 2010), the Rosenberg Self-Esteem Scale (RS-E) (Rosenberg, 1965) (Cordero and Israel, 2009; Renk, McKinney, Klein and Oliveros, 2006) and the Parental Bonding Instrument

(PBI) (Hall, Peden, Rayens, & Beebe, 2004) (Perry et al., 2006; Sira & White, 2010). In addition, all of the studies found a positive relationship between parent control and either eating/dieting behaviours or self-esteem; however, only two found an association between levels of parent care and eating/dieting behaviours, and self-esteem (Perry et al., 2006; Sira & White, 2010). Overall, the findings of these studies suggest that high levels of parent control (and in two studies, low levels of parent care) are related to future outcomes of adolescents' eating/dieting behaviours and self-esteem. Few studies have examined this relationship in Canadian populations, nor taken into account participant's ethnic background. Also, none of the studies examined whether participants have received or think they could receive an eating disorder diagnosis. As few studies examining non-diagnosed college-age populations have discovered a relationship between parent care and eating/dieting behaviours, this area merits further investigation to explore what extent parent-child interactions influence the development of healthy or disturbed eating and dieting behaviours and self-esteem in university students.

Current Study

In the current study, parent care and control were assessed by behavioral interactions youth and adolescents have or had with their mother and father, for example, being spoken to in a warm and friendly voice, smiled at frequently, or feeling as if their problems and worries were understood by their parents (parent care). Parent control may be characterized by a lack of autonomy to make personal decisions, feeling babied, or overprotected. Parent care and control was assessed retrospectively, meaning that participants of this study were asked to reflect on their experience during the first 16 years of their life. The purpose of the current study was to examine whether parent care and control relate to eating/dieting behaviours and self-esteem in undergraduate university students. High levels of parent care were expected to positively relate to adolescents' self-esteem, and negatively relate to disordered eating and dieting behaviours. High levels of parent control or overprotection were expected to negatively relate to adolescents' self-esteem, and positively relate to disordered eating/dieting behaviours.

METHOD

Participants

Participants included 99 undergraduate students (69% female) in Psychology 104 and 105 from the University of Alberta in Edmonton, Alberta.

Students studied in a range of faculties with 42% from the Faculty of Science and 30% from the Faculty of Arts. Participants represented a range of ethnicities with 60% being an ethnic minority and 39% Caucasian/ Canadian/ European. Eighty-seven percent of students identified their parents as having a "married" marital status. Two students indicated that they had received a diagnosis for an eating disorder, and 4 students indicated that in their opinion they could be diagnosed with an eating disorder (compared to 16 indicating they were not sure, and 79 indicating that they could not be diagnosed).

Measures

Eating and Dieting Behaviours

Students' eating and dieting behaviors were assessed using Part B and C of the EAT-26 (Garner et al., 1982). The EAT-26 is a standardized, self-report 26-item measure that assesses a broad range of symptoms of anorexia and bulimia nervosa. Although initially validated with anorexia patients, the EAT-26 may be used in a non-clinical setting with adolescents and adults to assess risks for disordered eating. Questions on the EAT-26 assess potential concerns about dieting, food intake and body weight (i.e., "I eat diet foods," "I am aware of the calorie content of foods that I eat," and "I am occupied with the desire to be thinner"). Each was rated on a 6-point Likert-scale (1 = always to 6 = never). Respondents were asked to rate frequency of behavior associated with disordered eating and unhealthy diet behaviours (i.e., "In the past 6 months have you gone on binges where you feel that you may not be able to stop?" and "Exercised more than 60 minutes a day to lose or to control your weight?"). Each was rated on a 6-point Likert-scale (0 = never to 5 = once a day or more). Twenty-five of the items were reverse-coded to create a composite score of disordered eating and dieting behaviors with higher scores indicating higher levels of disordered eating. Multiple psychometric studies have found high internal consistency of the EAT-26 (α = .79 - .94 across studies) (Furnham & Husain, 1999).

Self-Esteem

The Rosenberg Self-Esteem Scale (RS-E) was used to assess participant's self-esteem (Rosenberg, 1965). The RS-E is a 10-item questionnaire with items answered on a 4-point Likert-scale (0 = strongly agree to 3 = strongly disagree). Examples of items include, "I feel like I have a number of good qualities," and "I wish I could have more respect for myself." The scale was originally developed with a sample of high school students; however, multiple studies demonstrated good convergent and discriminant validity when used with college students (Hall, Peden, Rayens, & Beebe, 2004). Five items

are reverse-coded to create a composite score of self-esteem, with higher scores indicating higher levels of self-esteem.

Parent Care and Control

Students' retrospective perceptions of parent care and control were measured using the Parental Bonding Instrument (PBI) (Parker, Tuplig, & Brown, 1979). The PBI is a self-report questionnaire measuring fundamental parental styles as perceived by the participant. Items are answered on a 4-point Likert-scale (0 = very unlike to 3 = very like) and twelve items are reverse-coded (3 = very unlike to 0 = very like). Twenty-five questions assessed both maternal and paternal care (i.e., "Mother/father spoke to me in a warm and friendly voice" and "Mother/father appeared to understand my problems and worries") and control (i.e., "Mother/father tended to baby me" and "Let me go out as often as I wanted"). Results were tallied for both mother and father sections to produce a score for each parent on both dimensions (i.e., mother/father care and mother/father control). Previous studies have confirmed high reliability for each of the scales (paternal care scale α = .89; paternal control scale α = .73; maternal care scale α = .86 and maternal control scale α = .74) (Sira & White, 2010).

Procedure

Students were recruited to the study via the university psychology research pool and received partial course credit for participating in the study. Participants completed 3 structured questionnaires assessing eating/dieting behaviours, levels of self-esteem, and parent care and control. In addition to other measures, the questionnaire package included demographic information of birth month/ year, faculty of study, sex, ethnicity (by continent), and parent marital status. This additional questionnaire included a self-report of whether participants had been diagnosed with an eating disorder, or whether in their opinion they could be diagnosed with an eating disorder. No research was found to support the use of self-reported measures; researchers included the questions out of interest. The researchers guaranteed participants that participation was voluntary, anonymous and confidential. The participants were given an hour to complete the 4 questionnaires.

RESULTS

Descriptive Analyses

EAT-26

As shown in Table 1.1, on average, students indicating that they had received a diagnosis for an eating disorder demonstrated higher scores on the EAT-26 ($M = 17.00$, $SD = 14.07$) in comparison to students who indicated that they were not sure whether they could receive a diagnosis ($M = 16.75$, $SD = 13.20$) and receiving no diagnosis ($M = 6.66$, $SD = 5.05$).

These results varied according to ethnic background and gender. Students indicating a European background reported the highest scores on the EAT-26 ($M = 11.75$, $SD = 12.81$), followed by students of North American background ($M = 8.23$, $SD = 9.28$), Asian background ($M = 8.70$, $SD = 7.11$) and African ($M = 8.00$, $SD = 1.41$). Females reported higher EAT-26 scores ($M = 9.43$, $SD = 9.16$) compared to males ($M = 7.13$, $SD = 6.09$).

Self-Esteem

Students indicating that they had received a diagnosis for an eating disorder reported lower self-esteem scores ($M = 17.00$, $SD = 5.72$) in comparison to students who indicated they were not sure whether they could be diagnosed ($M = 17.69$, $SD = 4.06$) and those who indicated they could not be diagnosed ($M = 20.59$, $SD = 4.44$).

These results varied according to ethnic background and gender. Students with an Asian background reported lowest levels of self-esteem ($M = 19.46$, $SD = 4.38$) followed by students of North American background ($M = 20.46$, $SD = 5.01$), European background ($M = 20.62$, $SD = 4.03$) and African background ($M = 23.00$, $SD = 1.41$). Females reported lower self-esteem scores ($M = 19.63$, $SD = 4.75$) compared to males ($M = 20.74$, $SD = 4.05$). Results are shown in Table 1.2.

Variables	N	Mean	SD	Range
Diagnosed Eating Disorder Opinion				
No	79	6.66	5.05	0.00 – 22.00
Yes	4	17	7.04	9.00 – 38.00
Not sure	16	16.75	3.3	4.00 – 45.00
Ethnicity by Continent				
North American	35	8.23	9.28	1.00 – 45.00
European	8	11.75	12.81	2.00 – 39.00
Asian	54	8.7	7.11	0.00 – 39.00
African	2	8	1.41	7.00 – 9.00

TABLE 1.1. Descriptive Statistics of EAT-26 Scores, Diagnosed Eating Disorder Opinion, and Ethnicity by Continent

Variables	N	Mean	SD	Range
Diagnosed Eating Disorder Opinion				
No	79	20.59	4.44	3.00 – 29.00
Yes	4	17	5.72	10.00 – 24.00
Not sure	16	17.69	4.06	8.00 – 26.00
Ethnicity by Continent				
North American	35	20.46	5.01	3.00 – 28.00
European	8	20.62	4.03	14.00 – 27.00
Asian	54	19.46	4.38	22.00 – 24.00
African	2	23	1.41	3.00 – 29.00

TABLE 1.2. Descriptive Statistics of Self-Esteem, Diagnosed Eating Disorder Opinion, and Ethnicity by Continent, and Gender

Parent Care and Control

Students indicating that they had received a diagnosis for an eating disorder reported lower mother care ($M = 19.75$, $SD = 5.52$) and higher mother control ($M = 19.00$, $SD = 9.06$). Students diagnosed with an eating disorder reported lower levels of father care ($M = 18.25$, $SD = 7.37$) as well as higher levels of father control ($M = 17.75$, $SD = 6.24$). Participants indicating no eating disorder diagnosis reported highest levels of mother care ($M = 29.33$, $SD = 5.52$) and lowest levels of mother control ($M = 14.08$, $SD = 6.49$). Levels of father control were lowest for participants indicating that they had not been diagnosed with an eating disorder ($M = 10.17$, $SD = 5.77$). Levels of father care were similar for participants indicating that they were unsure whether they could receive a diagnosis ($M = 26.56$, $SD = 5.07$) or no diagnosis ($M = 26.32$, $SD = 6.99$). Results are indicated in Table 1.3. Males reported higher levels of mother care ($M = 29.03$, $SD = 5.40$) than females ($M = 28.72$, $SD = 5.48$); however, females reported higher levels of mother control ($M = 14.75$, $SD = 6.61$), father care ($M = 26.18$, $SD = 7.10$) and father control ($M = 11.61$, $SD = 6.81$). Variation in scores by gender is outlined in Table 1.4.

ANOVA

EAT-26 and Eating Disorder Diagnosis Opinion

A 3x1 (opinion of eating disorder diagnosis x EAT-26 score) between-groups ANOVA design was used to analyze the interaction between whether students felt they could receive a diagnosis of an eating disorder and their reported EAT-26 score. Significant effects were further analyzed using Tukey's HSD ($p < .05$). A between-subjects factor analysis revealed a statistically significant interaction between opinion of eating disorder diagnosis and EAT-26 scores, $F(2, 96) = 15.15$, $p < .001$.

Participants believing they could be diagnosed with an eating disorder ($M = 17.00$, $SEM = 7.04$, $n = 4$) and participants unsure as to whether they could be diag-

Variable I	Variable II	N	Mean	SD	Range
Parental Bonding	*Diagnosed Eating Disorder Opinion*				
Mother Care	No	78	29.33	5.52	15.00 – 36.00
	Yes	4	19.75	3.3	16.00 – 23.00
	Not sure	16	28.56	2.99	26.00 – 35.00
Mother Control	No	78	14.08	6.49	3.00 – 28.00
	Yes	4	19	9.06	12.00 – 31.00
	Not sure	16	15.75	5.52	3.00 – 27.00
Father Care	No	78	26.32	6.99	8.00 – 36.00
	Yes	4	18.25	7.37	9.00 – 27.00
	Not sure	16	26.56	5.07	18.00 – 36.00
Father Control	No	78	10.17	5.78	0.00 – 30.00
	Yes	4	17.25	6.24	9.00 – 24.00
	Not sure	16	12.44	7.96	3.00 – 33.00

TABLE 1.3. Descriptive Statistics of Parent Care and Control and Diagnosed Eating Disorder Opinion

Variable I	Variable II	N	Mean	SD
	Gender			
EAT-26	Male	31	7.13	6.09
	Female	68	9.43	9.16
Self-Esteem	Male	31	20.74	4.05
	Female	68	19.63	4.75
Mother Care	Male	31	29.03	5.4
	Female	68	28.72	5.48
Mother Control	Male	31	14.13	6.26
	Female	68	14.75	6.61
Father Care	Male	31	25.68	6.41
	Female	68	25.18	7.1
Father Control	Male	31	9.13	4.77
	Female	68	11.61	6.81

TABLE 1.4. Descriptive Statistics of EAT-26, Self-Esteem, Parental Care and Control and Gender

nosed with an eating disorder ($M = 16.75$, $SEM = 3.30$, $n = 16$) had significantly higher EAT-26 scores than participants not receiving a diagnosis ($M = 6.66$, $SEM = .57$, $n = 79$).

A between-subjects factor analysis of EAT-26 scores and ethnicity indicated no significant differences between ethnicity and EAT-26 score.

Self-Esteem and Eating Disorder Diagnosis Opinion

A 3x1 (opinion of eating disorder diagnosis x self-esteem) between-groups ANOVA design was used to analyze the interaction between whether students felt they could receive a diagnosis of an eating disorder and their reported self-esteem score. Significant effects were further analyzed using Tukey's HSD ($p < .05$).

A between-subjects factor analysis revealed opinion of eating disorder diagnosis interacted to affect self-esteem scores, $F(2, 96) = 3.82$, $p < .05$. Participants believing they could not be diagnosed with an eating disorder ($M = 20.59$, $SEM = .50$, $n = 79$) had significantly higher self-esteem scores than participants unsure as to whether they could receive a diagnosis ($M = 17.69$, $SEM = 1.02$, $n = 16$). However, no significant differences were found between participants believing they could be diagnosed with an eating disorder and self-esteem scores ($M = 17.00$, $SEM = 2.86$, $n = 4$).

A between-subjects factor analysis of self-esteem scores and ethnicity indicated no significant differences between the two variables, $F(3, 95) = .70$, $p > .05$.

Parent Care and Control and Eating Disorder Diagnosis Opinion

A 3x4 (opinion of eating disorder diagnosis x mother/father care and control) between-groups ANOVA design was used to analyze the interaction between whether students felt they could receive a diagnosis of an eating disorder and their reported parent care and control score. Significant effects were further analyzed using Tukey's HSD ($p < .05$).

Analysis revealed opinion of eating disorder diagnosis interacted to affect mother care scores, $F(2, 95) = 6.63$, $p < .01$ and father control scores, $F(2, 95) = 3.14$, $p < .05$. Participants believing they could be diagnosed with an eating disorder ($M = 19.75$, $SEM = 1.65$, $n = 4$) reported significantly lower mother care scores than participants not believing they could be diagnosed ($M = 29.33$, $SEM = .63$, $n = 78$) and students unsure as to whether they could receive a diagnosis ($M = 28.56$, $SEM = .75$, $n = 16$). Further analyses found no significant differences in means of reported father control scores and opinions of eating disorder diagnosis.

Independent Sample T-Test

Gender Differences in EAT-26, Self-Esteem, and Parent Care and Control Scores

A two-tailed independent samples t-test was used to compare EAT-26, self-esteem and reported parent care and control scores in male and female participants. Analyses indicated no significant differences between any of these factors and gender. Marginally significant differences, $p = .07$ were found between female ($M = 11.61$, $SEM = .83$, $n = 67$) and male ($M = 9.13$, $SEM = .86$, $n = 31$) ratings of paternal control, with females reporting higher levels of paternal control than males. Results indicated in Table 2.1.

DISCUSSION

Inconsistent with previous research (Cordero and Israel, 2009; Galloway, Farrow and Martz, 2010; Perry et al., 2006; Renk, McKinney, Klein and Oliveros, 2006; Sira & White, 2010), the expected positive relations between parent care and self-esteem, and parent control and disordered eating and dieting behaviours were not found in this study. In addition, expected negative relations between parent control and self-esteem, and between parent care and eating/dieting behaviors were also not found. Rather, it was observed that a significantly higher level of maternal care is associated with students believing they could not be diagnosed with an eating disorder, or not sure whether they could receive a diagnosis. Results also indicate students believing they could be diagnosed or were unsure as to whether they could receive a diagnosis reported significantly higher unhealthy eating/dieting behaviors. Students reporting higher levels of self-esteem reported lower levels of unhealthy eating and dieting behavior.

Association between Eating and Dieting Behaviours, Self-Esteem and Parent Care and Control

Relations between parent care and control and self-esteem, and eating/dieting behaviors were not found as expected in research done by Perry et al. (2006), and Sira and White (2010). Previous research found these associations between high levels of control (i.e., encouragement of autonomy/ independence, controlled feeding practices during childhood, discipline and verbal messages relating to eating and exercise regulation) and irregular eating/dieting behaviors, and diminished self-esteem (Renk, McKinney, Klein and Oliveros, 2006; Galloway, Farrow and Martz, 2010; Cordero and Israel, 2009; Sira & White, 2010; Perry et al., 2006). This challenges Minuchin et al.'s. psychosomatic familial model (1978) that families characterized by overprotectiveness contributes to the development of unhealthy eating and dieting patterns symptomatic of eating disorders. It seems plausible to expect heightened parental control may contribute to the development of unhealthy eating and dieting patterns, and diminished self-esteem, as meticulous behavior surrounding exercise and food are characteristic of developing eating disorders. A possible explanation for the lack of association found may be that participants may have had variable interpretations of parental care and control, and depending on the familial context, the parent care and control variables outlined in the questionnaire may not be applicable to participants. However, females reported marginally higher levels of paternal control than males. This may indicate that excessive paternal control disables the development of autonomy and independence in adolescent females, in turn triggering unhealthy eating and dieting behavior (Sira & White, 2010). Also, as suggested by Perry et

	Male			Female			Levene's Test (Eq. variances assumed)		Samples T-Test (2-tailed)	
Variable	n	M	SD	n	M	SD	F	Sig.	t	Sig.
Eat-26	31	7.13	6.09	68	9.43	9.15	0.87	0.35	-1.27	0.21
Self-Esteem	31	20.74	4.05	68	19.63	4.75	0.38	0.54	1.13	0.26
Mother Care	30	29.03	5.4	68	28.72	5.48	0.01	0.91	0.26	0.79
Mather Control	31	14.13	6.26	67	14.75	6.61	0.31	0.58	-0.44	0.66
Father Care	31	25.68	6.41	65	26.18	7.1	0.61	0.44	-0.34	0.74
Father Control	31	9.13	4.77	67	11.61	6.81	4.48	.04*	-0.18	0.07

TABLE 2.1. Independent Samples T-Test of EAT-26, Self-Esteem and Parent Care and Control by Gender

al. (2006), a mediation model may be more effective at explaining the relationship between parent control, eating/dieting behaviours, and self-esteem. For example, parenting variables influence self-image, which in turn impacts eating/dieting behaviours and self-esteem.

Eating Disorder Diagnosis Opinion and Mother Care

Although not supported by the literature, students reporting that they believe they could not or were not sure whether they could be diagnosed with an eating disorder also reported higher levels of mother care. These results indicate that a caring mother may act as a protective factor from developing unhealthy eating and dieting behaviors. Adolescents who view their mother as a source of emotional support and warmth may in turn adopt a healthy lifestyle characterized by regular food intake and exercise. The lack of findings between paternal warmth and healthy eating patterns may be attributed to the authoritative role that a father plays within a family. A father may take it upon himself to act as a leader and primary provider of the family, and as a result not exhibit as much warmth as a mother. Children, in turn, may not feel comfortable turning to their father as a source of emotional support. These speculations, however, may vary by individual families.

Eating Disorder Diagnosis Opinion, Unhealthy Eating and Dieting Behaviours, and Self-Esteem

Results indicate that students thinking they could be diagnosed with an eating disorder reported higher levels of unhealthy eating/dieting behavior than individuals thinking they could not be diagnosed. What is interesting to note is that individuals reporting being unsure as to whether they could be diagnosed also reported high levels of unhealthy eating/dieting behavior. This indicates that individuals unsure of deserving diagnosis are equivalent to those sure of deserving diagnosis, and exhibit behaviors symptomatic of a potential eating disorder. Students indicating that they could not receive a diagnosis reported higher levels of self-esteem than those who were not sure. This indicates that these students engaging in normal eating/dieting behaviors

feel positively about themselves; however, once unsure as to whether a diagnosis could be received, this in turn may lower an individual's perception of oneself.

Limitations

It is important to acknowledge limitations of this investigation. First, the sample size in this study included only 99 students, of which 69% were female. These implications may impact our ability to generalize our findings to a population other than an adolescent post-secondary population. Other means of collecting parent-child relationship data could have been used, for example, by observing the parent's relationship with their child which would lead all researchers to accurately gauge the dynamic of this relationship and decrease self-report and retrospective reporting biases. As well, a survey reporting parents' relationships with their children could be used in the future.

Conclusion and Future Implications

The purpose of this current study was to examine whether parent care and control relate to eating/dieting behaviours and self-esteem in undergraduate university students. The expected relations between parent care and control, eating/dieting behaviours, and self-esteem were not found. However, results indicate that higher levels of maternal care are associated with students believing they could not or were not sure whether they could be diagnosed with an eating disorder. Results also indicate students believing they could or were unsure as to whether they could receive a diagnosis reported significantly higher unhealthy eating and dieting behaviours. Finally, students reporting higher levels of self-esteem reported lower levels of unhealthy eating/dieting behavior. These findings provide further evidence that although parent care and control were not associated with eating/dieting behaviors, other parenting variables and attitudes more closely related to the consumption of food and exercise behaviors, may merit attention in research that explores the role of parenting in the developing of eating disorders. This research will contribute to the advancement of knowledge by a) help-

ing to explore behaviors that parents should or should not be engaging in to help prevent the development of unhealthy eating/dieting behaviors; b) sparking the interest in the research of familial implications in the etiology and course of eating disorders, and c) reinforcing the importance of a family-based therapy program for eating disorders or revealing a need for change in the treatment methods. Parenting behaviors should be considered along with other factors that may be contributing to the development of unhealthy eating and dieting behaviors to ensure that adolescents are presented with the best opportunity to engage in a healthy lifestyle.

REFERENCES

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed., text rev.). Washington, DC: Author.

Bartholomew, K. (1990). Avoidance of intimacy: An attachment perspective. Journal of Social & Personal Relationships, 7, 147-178.

Baumrind, D. (1971). Current patterns of parental authority. Developmental Psychology Monographs, 4(1, Pt. 2), 1-103.

Bowlby, J. (1982). Attachment and loss (Volume 1: Attachment). London: Hogarth Press

Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design. Cambridge, Massachusetts, USA: Harvard University Press.

Button, E. J., Loan, P., Davies, J., & Sonuga-Barke, E. J. S. (1998). Self-esteem, eating problems, and psychological well-being in a cohort of schoolgirls aged 15-16: A questionnaire and interview study. International Journal of Eating Disorders, 21(1), 39-47.

Cordero, E. D., & Israel, T. (2009). Parents as protective factors in eating problems of college women. Eating Disorders, 17, 146-161. 10.1080/10640260802714639

DeHart, T., Pelham, B.W., & Tennen, H. (2006). What lies beneath: Parenting style and implicit self-esteem. Journal of Experimental Social Psychology, 42, 1-17.

Furnham, A., & Husain, K. (1999). The role of conflict with parents in disordered eating among British Asian females. Social Psychiatry & Psychiatric Epidemiology, 34, 498505.

Galloway, A. T., Farrow, C. V., & Martz, D. M. (2010). Retrospective reports of child feeding practices, current eating behaviours, and BMI in college students. Obesity, 18, 1330-1335. doi:10.1038/oby.2009.393

Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The eating attitude test: psychometric features and clinical correlates. Psychological Medicine, 12, 871878.

Hall, L. A., Peden, A. R., Rayens, M. K., & Beebe, L. H. (2004). Parental bonding: A key factor for mental-health of college women. Issues in Mental Health Nursing, 25, 277291.

Minuchin, S., Rosman, B. L., & Baker, L. (1978). Psychosomatic families: Anorexia nervosa in context. Cambridge, Massachusetts, USA: Harvard University Press.

Parker, G., Tuplig, H., & Brown, L. B. (1979). A parental bonding instrument. British Journal of Medical Psychology, 52, 110.

Perry, J. A., Silvera, D. H., Neilands, T. B., Rosenvinge, J. H., & Hanssen, T. (2007). A study of the relationship between parental bonding, self-concept and eating disturbances in Norwegian and American college populations. Eating Behaviours, 9, 13-24.

Renk, K., McKinney, C., Klein, J., & Oliveros, A. (2005). Childhood discipline, perceptions of parents, and current functioning in female college students. Journal of Adolescence, 29, 73-88. doi:10.1016/j.adolescence.2005.01.006

Rodriguez, M. A., Novalbos Ruiz, J. P., Martinez Nieto, J. M., Escobar Jimenez, L., & Castro de Haro, A. L. (2004). Epidemiological study of the influence of family and socioeconomic status in disorders of eating behaviour. European Journal of Clinical Nutrition, 58(6), 846-852.

Rosenberg, M. (1965). Society and the adolescent self-image. Princeton, NJ: Princeton, University Press.

Scaglioni, S., Salvioni, M., & Galimberti, C. (2008). Influence of parental attitudes in the development of children eating behavior. British Journal of Nutrition, 99(1), S22- 25. doi: 10.1017/S0007114508892471

Shoebridge, P. J., & Gower, S. G. (2000). Parental high concern and adolescent-onset anorexia nervosa: A case-control study to investigate direction of causality. British Journal of Psychiatry. 176, 132-137. doi: 10.1192/bjp.176.2.132

Sim, L. A., Homme, J. H., Lteif, A. N., Vande Voort, J. L., Schak, K. M., & Elingson, J. (2009). Family functioning and maternal distress in adolescent girls with anorexia nervosa. International Journal of Eating Disorders, 42, 531-539.

Sira, N. & White, C. P. (2010). Individual and familial correlates of body satisfaction in male and female college students. Journal of American College Health, 58(6), 507-514.

Steinberg, L. & Morris, A.S. (2001). Adolescent development. Annual Review of Psychology, 52, 83-110.

A Science of Evil: An Exploration into Terror Management Theory and a Psychoanalytic Theory of Religious Extremism

Sayeed Devraj-Kizuk

Department of Psychology, University of Alberta

On September 11th, 2001, the United States was profoundly shaken when they were the victim of a tragic terrorist attack, in which 19 hijackers crashed two commercial airliner planes into New York's most treasured symbol of American prosperity, the towering world trade centres. Another plane hit the pentagon, bringing the total cost in human lives to approximately 3000 civilians. This event rocketed the Islamic terrorist organization Al Qaída into infamy, and has caused countless people to question how people could carry out these actions that were starting to dominate world news. Al Qaída is not a structured organization of terrorists. There is no base of operations or capitol for Al Qaída, no cultural obligation to be a part of the extremism, nor can people be born into Al Qaída, family obligations notwithstanding. People willingly join the extremists according to their own beliefs and conduct their operations in isolated small groups, without much contact with the leaders of the cause (The 9/11 commission, 2004). How could people willingly join something that causes so much destruction and pain? Even more confusing, how could people willingly die for this cause? Intentional suicide is a very different thing from joining a war with the knowledge that you may die in the fight. The idea of an individual deliberately choosing to end their own life, to accept the terror of death, in service to their cause is a shocking and troubling one, and it goes against everything that is intuitive about the behaviour of life-loving human beings. In this paper, I will attempt to describe the history and central themes of modern psychoanalytic theory, and in doing so, provide some insight into how the ideas of Terror management theory can explain how individuals can be capable of such extreme and perverse actions.

INTRO TO TMT:
THE QUEST FOR KNOWLEDGE

Man is an inquisitive animal. For as long as we have been able to consider intelligently the world around us, we have been asking questions. The earliest records of religious behaviour date to at least the Upper Paleolithic era, or Late Stone Age, approximately 10-40 thousand years ago (Rossano, 2006). At the very birth of our species we began to question the nature of our place on this earth and the meaning of our existence. Mankind's quest for philosophical enlightenment begins with Socrates around 400 B.C.E. Socrates was a questioner. He sought knowledge through analysis of

others, rather than postulating his own personal viewpoints. He would seek to understand courage, for example, by continually interrogating courageous men until he found a contradiction (Internet Encyclopedia of Philosophy, 2010). In this way, he can indeed be regarded as the first psychoanalyst. The history of philosophy, from this point on, continually builds on itself; the Greeks, the Christian philosophers of the medieval period, and the scholars of the Renaissance, all drew upon those that came before them, and all were influenced by Socrates' psychoanalytic questioning. Eventually, in the late nineteenth century, philosophy became incorporated into science in the work of an extremely influential German in the field of psychology named Sigmund Freud (Latham, n.d.; Becker, 1973) Figure 1 charts a few milestones in the understanding of the self.

Freud was the creator of psychoanalysis, a technique for understanding the structure of the mind based on logical deduction of the unconscious pressures and stresses on the psyche. Since the theory is based on logic and clinical observation rather than experiment, psychoanalysis has an element of the philosophical, which is clear in the parallels between Freud and Nietzsche (Latham, n.d.). This process produced theories that greatly improved comprehension of the human psyche and in recent years has allowed psychologists to study the mind through scientific experimentation. Although Freud himself believed that most of the behaviours that can be observed in human beings are due to repressed sexual and aggressive urges and instincts, this hypothesis has become increasingly less satisfactory to newer generations of researchers. Freud's train of thought, however, was not wrong, it was simply not going deep
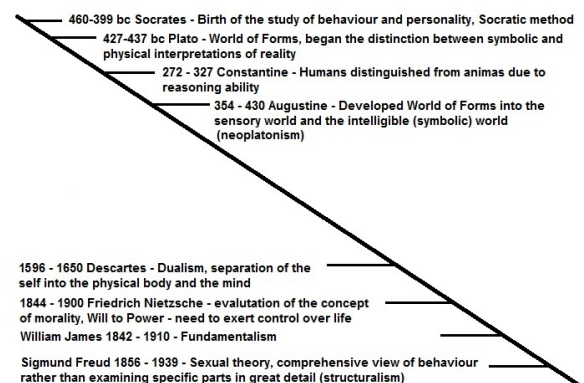

FIG. 1. A timeline of the understanding of the self.

enough. Sexual and aggressive instincts undeniably exist, however they aren't convincing as the primary, intrinsic motivators for behaviour. A deeper, more fundamental instinct was needed, namely, the instinct to survive. Through the ingenious insights of psychoanalysts such as Jung, Adler, and Rank, and as clarified through the work of cultural anthropologist Ernest Becker (1973), a new theory has begun to emerge, one that places fear of death as the primary motivating factor, and has been named terror management theory.

Psychology, like all of science, has been building upon itself since the emergence of empirical investigation. In the study of the universal unconscious factors that influence human behaviour, through humanity's insistence to question the nature of our existence, two and a half millennia of thought have developed this one theory. The theory shows all-inclusive comprehension of the human condition, and can be used to understand and predict a myriad of different complex behaviours. Unlike its predecessors, it is supported by scientific experiments, which have profound implications for how we should proceed in the field of social psychology. It is the science of all that we are, of the entire human experience. terror management theory has the capability to accomplish at long last the fundamental goals of psychoanalysis. It provides insight into how to live happily, and provides an explanation for all of humanity's capacity for good and for evil.

THE INTERNAL CONTRADICTION OF MAN

A. The Terror of Death

Everything must die. From the smallest microbacterium to the highest towering redwoods, everything that lives is caught in a battle for survival, driven by the instinct to survive as long as possible, and produce some offspring to carry on the lineage. These are the basic animal instincts. But this is a fight that we all eventually lose. Every plant, animal, and microbe on this planet eventually runs out of time, and is reincorporated into the cycle of life. Even the sun itself will one day die, and in its catastrophic death-scene, vaporize the earth and the planets, ending this life cycle once and for all. The universe has a central theme of creation, and destruction.

Ernest Becker argues, in *The Denial of Death* (1973), that of all the living breathing things on this planet, only humanity has the misfortune of knowing that it is doomed. We have just as much of that animal instinct to survive as any other species, and what's more, we have the intelligence to rejoice in life, to understand how fantastic it really is to be alive, to see the beauty in the world, in ourselves, and in each other. To us, every day is a gift, every experience a wonder, every second spent alive an affirmation to the greatness of our kind. We are also able to visualize ourselves as separate, indi-

vidual beings. Furthermore, we have the ability that no other species has to remove that sense of self from our present location in position and time. Through our infinite imagination, we can project ourselves onto the surface of the moon and look down at our own world, or we can shrink ourselves down and contemplate the structure of matter while sitting on the nucleus of an atom (p.26). We can travel back in time years to our childhoods, and project ourselves forward in time to our old age, or some distant future. Or to the day of our death. The same intelligence that allows us to fully appreciate life also forces us to acknowledge the inevitability of our biology: that we are impermanent, if superlative, beings on this planet, and that for all of the creativity and divinity of our minds, our uniqueness in all of nature, we still cannot escape the finitude of our bodies. It is utterly both inacceptable and inescapable for us to die. This is the eternal confliction within man's psyche. It is the cause of tremendous anxiety, that, if not dissipated in some way, would be so overwhelming that not a single one of us would be able to function, to do anything without cowering in fear of accidentally slipping and cracking our heads open like eggshells on the sidewalk. This is the terror of death that terror management theory postulates as the central driving force behind our actions, and all that we do in life is part of our desperate attempt to repress our fear of death, and to conquer it (Becker, 1973; Shen, 2006).

B. A Symbolic Life

If humanity has been burdened by its intelligence, thrown into an unwinnable struggle for life and be forced to see it as unwinnable, then at least it is given by that same intelligence the means, through complex psychological structures, to deal with this burden. In order to understand these structures fully, we must talk briefly about Humanity's paradoxical nature. The human being, according to Becker (1973) is a combination of two entities, half animal and half symbolic. The animal nature of humanity is simple and undeniable, we have instincts and urges, and our bodies require food and drink and air or they will expire. And most importantly, sooner or later, they will expire. The symbolic nature of humanity, what separates us from the rest of nature, is more complex, and has more profound implications. We have imagination, creativity, and individuality. We have an identity, a name, a history, and a future; non-concrete things that only have meaning in the minds and words of human beings. These traits are what allow us to shake off our fear of death, to deny the reality of our physical condition by living fully in the symbolic, endowing us victory over physical mortality through what Becker calls *symbolic immortality*.

As human beings grow and make their way through the world, they develop a unique set of values and beliefs, as well as a set of aspirations, hopes and dreams.

Put simply, people generally figure out which things in the world are most important to them – their religion, art, work, lovers, family, country, a cause, or any combination of these ideals. Together, they come together to form a personal world-view, a mental picture of what is good and right in the world. terror management theory postulates that personal worldviews develop because they serve to shed the existential dilemma, and allow people to occupy their minds with goals that are actually attainable, and will result in some measure of symbolic immortality. (Becker, 1973, Shen, 2006) This makes those values no less meaningful or significant, it simply explains why people cling so tightly to their worldviews and ways of life, and why they are so devastated if their aspirations are not realised.

The success in realizing this personal worldview becomes tantamount to their success in conquering the anxiety of death, in a symbolic and personally meaningful way. Naturally, it should follow that people should feel more anxious and unstable the more the reality of their lives deviate from the ideal of their personal worldview, and people should feel more confident and happy the closer they are to fully realising that set of ideals. Indeed, what we know as self-esteem is defined in terror management theory as how successful a person is in realizing, in themselves, their personal worldview (Shen 2006).

SCIENTIFIC SIGNIFICANCE

While the theory certainly presents some interesting ideas, one must enquire on the scientific significance of these various and far-reaching claims. How can we be convinced that death is really the central anxiety, and not mere speculation? The science of functional psychology has reached a stage in which it is highly theoretical, formulated not from concrete experimental data, but from clinical observations and logical deduction. Experimental data in support of TMT has been widely conducted a propos of the theories, as proof and evidence. In a science such as this, the significance and truth of a theory can be evaluated by showing how all-inclusive the theory is. Much like the difficulty faced by the physical sciences in describing the nature of the very small, we may not be able to directly observe these processes as they are happening, but if the theory encompasses and accurately predicts all behaviour that we can observe, then it can be taken as an accurate model of the subject in consideration. In the next few sections, I am going to show how multiple behavioural patterns that we see in our modern society are convincingly explained by terror management theory, and then use those explanations to examine the behaviour of some people which we might otherwise see as absolutely baffling.

TRANSFERENCE

Transference is a complicated but integral idea in the study of human behaviour. In order to overcome the fear of death, and the impotence associated with it, people will often make use of some type of meaningful object in order to project their feelings of insecurity onto someone or something. This can happen in many ways, and is an explanation for a variety of human behaviours.

A.   The Hero and Anti-hero

Throughout the chronicles of history, we see time and again the emergence of heroic individuals. Wherever there is a great event in human history, a great war or great empire, or an influential cultural movement or revolution, there is an unusually powerful person in the centre, shaping the history of the world by their actions as though tied to the very fabric of fate. Examples of such heroes could be Alexander the Great, Winston Churchill, Vladimir Lenin, Martin Luther King Jr., the Prophet Mohammed, or Jesus Christ. According to TMT (Becker, 1973), heroism is transference of death-anxiety from a group of people onto a single relatively virtuous person, which has the effect of elevating their perceived virtuosity and importance.

Transference, Becker (1973) states, is ubiquitously agreed on by psychoanalysts as a problem with courage. When people are in a vulnerable position, and doubt their own abilities to achieve symbolic immortality, they choose another person to do it for them. They see the person, not for who he or she really is, but as the embodiment of all the characteristics that they wish they could embody in themselves. Those who see themselves as weak, and foolish, and impotent, will see the hero as infinitely strong and wise and powerful. They elevate the hero to a position where it would be impossible for him or her not to achieve symbolic immortality, to change the fate of the world. Followers achieve their own immortality through these great and powerful people, in a diminished way, by convincing themselves that they are part of some great cause championed by a great hero. This is why great people in history almost always are in command of, or worshipped by, huge numbers of people. Their greatness is not due, in most part, to their own inherent properties, but those imposed upon them by their followers.

This process also explains why people can do things that would normally be completely contrary to their personal worldview, such as murder, for their cause. People place their entire life-project, their only chance to evade the terror of death and find peace, in one individual, and in their associated cause. The cause begins to define the individual, to become a new personal worldview, with the hero as a focusing point, and the success of the cause becomes the success of the individual. This creates a moral dilemma for the individual. What happens

when the beloved hero commands a peaceful person to destroy? The individual has committed himself or herself to the cause, and in effect abandoned his or her own personal worldview, his or her own individual quest for immortality as dictated by personal beliefs, values, and morals, and allowed it to be subjugated by a stronger defense, the hero/cause transference object. The individual will then commit the act because to defy the hero would be to lose that bid for immortality. However, this does not occur without some consequences for the individual. If the individual is completely immersed in the hero and the cause, he or she can no longer return to the original, personal worldview once the ideals of that worldview have been violated. Within the cause, the individual is a holy warrior, achieving his or her self-esteem by conforming to the wishes of the transference object, even if they contradict personal values. If the individual were to leave the group, he or she would be a pathetic and miserable failure in the terms of the old, personal worldview. The cause has become his or her new shield against the terror of death. It is close to impossible for a person to exchange a strong defense for a weaker one, even if he or she at some point realizes that the strong defense is fundamentally wrong, because that would mean facing the terror of death, braving the fear that defines a person at the very core, and accepting it as a part of life. If the essence of a person's life is an attempt to run away from the reality of death, how can anyone be expected to act in a way that will only bring them closer to this truth? (Becker, 1973).

As a final point on hero-based transference, Becker (1973) states that transference of the kind discussed above can occur positively, as in the hero, or negatively, as in the anti-hero. In this case, the process is exactly the same but with reversed emotions. The anti-hero is seen as the embodiment of everything that the person despises within his personal worldview, and is elevated to a level of ultimate evil rather than supreme virtuosity. The individual then finds his self-esteem in the quest to destroy the negative transference object. The individual will stop at nothing to realise that goal, and in doing so, earn his or her symbolic immortality by cleansing the Earth of some great evil. During this quest, individuals may have to resort to means that contradict their personal worldviews in order to proceed. After the quest is over, after the evil is destroyed or these people are forced to give up, they find it difficult to return to the old worldview for the same reason. The self-esteem gained from the personal worldview was tarnished by the act of pursuing the evil. This idea of negative transference can help explain why people will so self-destructively hold onto the past, as well as some obsessive behaviour. A couple archetypical examples of such cases could be police officers who become obsessed with a criminal they were unable to catch (Mikulenka, 2007), or soldiers that experience difficulty returning to their regular lives after a war (Helzer, 1987).

B.   The Failure of Transference

Heroism is characterized by a group of people projecting their anxiety onto a single person. However, there's quite a variety of projections that we human beings are capable of. Two people can project their anxieties onto each other, which occurs during what we know as love. Sometimes, only one is projecting onto the other, which is characteristic of a platonic relationship, or in more extreme circumstances of obsession. People can project onto inanimate objects or meaningless rituals, as seen in fetishism, superstition, and other kinds of magical thinking. People can even project onto the products of their own minds, which gives the world our greatest artists, musicians, and authors. Indeed, anything that has contributed to the development of culture can be explained through the need of all human beings to distract their minds from the anxiety of death. All of these transference objects, however, are doomed to fail, because they all are based off other human beings or the products of human beings. This makes the objects unstable, and the process has the potential to cause deep anguish for the individual (Becker, 1973).

Becker (1973) argues that the fundamental flaw with human beings as transference objects is that in order for the object to be an effective distraction against the individual's imperfection and impermanence, he or she must elevate it to a divine status, or the transference has no weight. This entire psychological structure is thus weakened by the fact no person can ever be perfect or immortal. Inevitably, all individuals will be forced to see that some of the things that they despise are in some way present in that person. If one has deluded his or her hero, or beloved, to a state of immortal perfection, and one's own bid for immortality is dependent on the other's immortality, then what does it mean when one finds imperfection in his or her object of transference? The individual's own bid for immortality is of course diminished. When people become "disillusioned" in their love or hero, their self-esteem gradually drops along with their opinion of the transference object, and feelings of hurt and frustration towards the object are the result. Worse yet, if the object is ripped away, if the hero dies or abandons the cause, or if the beloved dies or leaves them, then the self-esteem gained from that relationship is brought crashing down. Of course, some objects are better choices than others, and it is perfectly possible to live your entire life, if you are lucky, entirely fulfilled by a finite transference object. However, it is these possibilities for failure that ultimately make transference onto persons intrinsically weak and unstable. Figure 2 shows an example of one of the stronger human transference objects in history.

This idea of the failures of transference provides insights into many behaviours of people in their social relationships. Why can people who were once passionate lovers show so much hatred towards each other? Why is domestic violence often associated with rela-

FIG. 2.   The Pope is one of the strongest human transference objects in the world today.   How can it be possible that so many people show such devotion to a person with whom they have had no personal interaction? (http://mediagallery.usatoday.com/Pope+Benedict+XVI)

tionships that those involved could not bear to leave? Why, througout history, were authoritarian leaders so often thrown from power and executed, only to be replaced by another who is rarely any improvement? Why are celebrities made the focus of a whole culture one day and demonized the next? These violent fluctuations in public opinion towards influential people can be convincingly explained by the fact that transference towards a person simply is not viable as a strong, lasting solution to the problem of death. People are only just that, people. No one can last as the image of divine perfection for another. It causes too much strain, puts too much weight on the shoulders of the hero or love object, and gives them an impossible role to fulfill. Eventually, individuals will lose faith in the immortality of the transference object, either indirectly, by rejection or disillusionment, or directly, by the actual loss of the person. They find themselves having forsaken their personal worldview to focus on another individual, yet the other is bringing more pain than happiness. They cannot succeed in this new immortality project, nor can they return to the old one, the original personal worldview. This is the eventual fate of any form of transference to another human being. In truth, there is only one entity in existence that can serve as a strong, viable transference object, the only thing that can forever act as an immortal, perfect being in the mind of the individual. This being, as Becker (1973) explains, is of course what we know in our culture as God.

C.   Religion

The concept of religion is perhaps one of the most complex and interesting topics that one can study using the approach of psychoanalytic theory. Religion is a staple of all cultures across the world and has been

around as a solution to the problem of death since the Stone Age (Rossano, 2006).   An explanation for its remarkable prevalence is that as such a solution, it provides the most stable and most fulfilling answer that man has yet achieved. From a psychoanalytic standpoint, religion is simply transference, but one that uses a spiritual and all powerful, yet intangible, entity as the transference object. The presence of a deity takes the transference object from the physical realm and places it in the spiritual, or symbolic realm (Becker, 1973). The individual no longer has to elevate a physical being to a level of divine perfection in which to place his trust and faith, he instead relies directly on a pre-existing, all-powerful, all-knowledgeable, all virtuous being. This does not have any of the complications that arise from transference to persons. The individual cannot become disillusioned in the imperfection of God, because God by definition has no imperfection. Because God is always perfect, no fault can be due to God. If there is evil in the world, it is due to humanity, or to himself, but never to God. He cannot be abandoned by his transference object because God cannot die, or leave him, or be lost. God is always there, an eternal presence. Because God never interacts directly with the individual, God can never be blamed for his failures. Neither can He feel the pressure of being the source of happiness for the individual, and so reject him. It is obvious. If transference fails when we try to make imperfect beings perfect and immortal, then the difficulties vanish once a transference object is found that already contains these qualities. God is the only strong, viable transference object that a person can safely trust with his or her psychological security without the possibility of having it crushed, sooner or later, by the object's infallibility (Becker, 1973).

What God really is to humanity, is the very essence of the symbolic immortality that humanity wants so desperately. Human beings, in religion, create the finest of illusions to veil from themselves the reality of their death. It is similar to the illusion of the loved one or the hero, but assigned, not to a physical object which would weaken the illusion, but to the very idea that the individual is trying to achieve. In order to achieve symbolic immortality, humanity transfers their death-anxiety onto a being that represents the concept of symbolic immortality itself. We also give ourselves the added pleasure of making the fulfillment of this new personal worldview unbelievably easy to fulfill. Essentially, the only thing that must be done to be accepted into a religion, and to be a good follower of religion, is to believe in that religion. Humanity has created for itself a cultural worldview, wherein, simply believing in that worldview fulfills it, securely warding off the omnipresent terror of death.

A CASE STUDY: EXTREMIST THEORY AND THE PSYCHOLOGICAL MOTIVATIONS BEHIND AL QAÍDA

In the past century, there is no lack of instances of perverse and inexplicable group human behavior. Nazism, Genocides in Rwanda and across Africa, and on a smaller scale the many examples of regular, moral people acting atrociously out of prejudice towards ethnicity or sexual orientation, all can attest to the propensity of considerable numbers of people to engage in objectively immoral lifestyles for prolonged periods of time. In order for a large group of people to live in a fashion that would normally be so far from their regular personal worldviews, there has to be opportunities for extremely effective cause/hero transference. For an extreme and perverse cause to take hold, there are multiple factors that can strongly contribute to the strength of the transference. These factors include:

1. An environment that contains strong and consistent death-reminders

2. A strong unifying motive that draws the people of the group together

3. A charismatic and heroic individual

4. An easily demonized and targeted scapegoat or Anti-hero

The first two factors deal with the environmental hardship that creates the need for the extremist cause to come about. This need for an escape must be strong enough that it pushes people into something they might not otherwise believe in. The second two factors deal with the appeal of the cause itself, which must be strong enough to pull people in when they are troubled by the environmental hardship. When both of these factor groups are strong, it creates a situation that forces the people into the extremist cause, because it is the easiest and often the only way available to them to achieve their symbolic immortality.

It is somewhat of a non-consequential observation that people most often react in extreme ways when they are confronted with extreme situations. Al Qaída was created in Afghanistan and has grown into an extremist movement throughout the Middle East (The 9/11 Commission, 2004). Many countries in this area have weak and disparate economies, extreme poverty, and are war-torn and dangerous. These types of environments cause extremely high levels of existential anxiety for the inhabitants because there is a constant reminder of death and the possibility of dying. In places like this, it is inescapable. The regular sight of malnutrition and sickness, the hidden explosives and remnants of wars long past, and the armed troops and warfare that is still being seen by many in these regions, all present the perception that death is everywhere. If human beings seek symbolic immortality through their personal passions as a reaction to anxiety about death, then it should follow that if death is more prevalent in an individual's mind, the result will be a heightening of anxiety and the need to escape death using some kind of immortality symbol. This is the idea of death-reminders or mortality salience. People living in stressful situations are reminded more acutely about their own death, and so they will go to more extreme measures for some way to transcend their fear and achieve their immortality.

The second aspect of an environment that provides the ideal substrate for extremism is a strong unifying factor that brings people together, and provides the emotion that the extremist cause draws on. In the Middle East, this has to do with Islamic, particularly Sunni political history. This is addressed in detail in the 9/11 Commission report, (2004) and I will give a very brief description. In the history of Islam, there was no separation of the church and the state. Muslim communities had governments that enforced law in accordance with Islam's teaching only, and the state was a completely religiously directed entity. This idea that the country's religion and politics are inseparable has been present since Islam's inception, and it retained a measure of that idea for many hundreds of years. Over time, the Muslim empire weakened, and became loosely represented by the Ottoman empire, which was slowly defeated by European powers, the remnants of which is now Turkey. In the following years, Europe and the Americas became more profitable and the Muslim world became generally poorer. The extremists view this trend as the defeat of Islamic nations by Judeo-Christian nations. In the modern world, some Islamic countries do not enforce religious obligations, and do not act as defenders of Islamic ideology as they once did. The extremists view this embracement of western practices as tantamount to Christian and Jewish influence destroying aspects of Islamic culture. This is the unifying concept that the extremist cause uses to foster a spirit of unity within Al Qaída, that the members of the cause are fighting to restore Muslim nations to their once former glory and to regain aspects of their culture that have been lost. In this way, the members of the cause are allowed to see themselves as champions of their world, fighting for a truly noble cause.

Now we move on to the idea of the hero and the anti-hero. The extremist cause in this case is one that is very strong because it combines aspects of religious and heroic transference. The member has strong, charismatic leaders, the most prominent being Osama Bin Laden, until his death in 2011. Osama Bin Laden was an extremely virtuous figure when seen from the extremist Islamic perspective. He was university educated, and spoke eloquently and fluently on religious scripture. His first efforts in Al Qaída, at which time the movement is generally thought to have begun, were in Afghanistan, where he used his considerable family fortune to train local militants to combat and free themselves from a brutal and oppressive Soviet occupation. Long before his involvement in any terrorist ac-

tivity, many Arabs idolized Osama Bin Laden for his intellectualism, ambition to protect Muslim rights and freedoms, the position of power given to him by his fortune, and ultimately the successes that came from those qualities (The 9/11 Commission, 2004). However, there is a strong difference between him and the regular hero, because he is a spiritual leader of the cause. Osama Bin Laden does not direct all of Al Qaída like his own private army. Al Qaída is a disjointed group of extremists that idolize him and follow his teachings and instructions, but do not act directly out of submission to him. The real drive for their actions is the ultimate transference object, religion. The members are then self-directed, acting in the style dictated by their hero but ultimately not to please the hero, but to please God. In this way, they have the strong appeal and love for the hero, who is divinity embodied. But, they also have the strength and stability of the religious cause. Indeed, almost all religions have a lineage of human representatives associated with their religion, be it the Pope, Imam, or divinely appointed royalty, because it strengthens the religious transference by giving it a more grounded, physical nature. The desire for extremist Muslims to follow the wishes of Osama Bin Laden is identical, in nature, for the desire for astute Catholics to adhere to the wishes of the Pope.

And finally, we come to the anti-hero. When the historical background is considered, the United States is clearly a strong candidate. If the extremist cause is fighting in opposition to the defeat of Islamic culture by western culture (Judeo-Christian world) and the unjust reduction of the Muslim empire to poverty while the western world has become wealthy and prosperous, then there is no greater example of western influence than the United States. The extremist cause is strongly religious in nature and the United States has become a non-religious state, at least from the perspective of the government. The American-dominated popular culture, media, and music of the west often promote promiscuity, rebellion, and self-promotion, the exact opposite of Islamic religious ideology. The United States has the reputation of being the most prosperous and wealthy western nation, and it is the most successful superpower that has grown out of the Judeo-Christian culture of colonial Western Europe. In addition, the United States supports Israel in their foreign policy, and had a strong military presence in the Middle East before the War on Terror. All of these factors made the United States an extremely tempting scapegoat. All of the problems of the Muslim World can be seen inversely reflected in this one country, and because the extremist cause is primarily a holy war, then it is no surprise that they call the United States the "Great Satan" (The 9/11 Commission, 2004).

Now that we have considered all these factors, we can see how the members of Al Qaída can perform such terrible actions, and even give their own lives willingly and knowingly, for their cause. They live in an extremely stressful environment, which burdens their minds with an unbearable amount of death related anxiety. This anxiety is manifested in religious devotion, and indeed a huge majority of the Middle East are strong believers in Islam (The Pew Forum on Religion & Public Life, 2011). But for some in these situations simply having religion is not enough, and during the 1970's there rose a strong, reputable, and heroic figure who they could follow. This hero gave them a purpose, a way to achieve further symbolic immortality by being part of a great and noble cause that would shape the fate of the world. In addition to this, their cause was intricately engrained in their religion albeit twisted from its true values and so their transference object was the eternal God, and their self-esteem was bolstered on two levels, by service to their religion and service to their cause. If they began to see that the cause was in actuality acting against the principles of their religion, or they began to be asked to perform actions that they knew were not within Islam or their personal values, they could not turn back. They had pledged allegiance to the cause, and to leave it would be to sacrifice their symbolic immortality, and return to a life where they can only be branded as criminals. For those who still did believe, to die in service to the cause becomes desirable, because they cannot fail in their mission. If a suicide bomber dies with the belief that his act will bring his cause ever closer to its goals, and will strike a blow at the heart of the enemy, then by his act of death he is achieving his symbolic immortality, and so he is capable of dying without fear. Instead, he is guaranteed his immortality, because once he completes his act, he finishes his life-struggle in the highest possible state of death denial.

The people who compose Al Qaída are often depicted in the West as monsters, religious fanatics with no goals but to destroy those who do not believe what they believe. But to truly believe that view is ignorance. They are human beings, with thoughts and emotions, and have just as much capacity as the rest of us to feel love and happiness, anger and indignation, sorrow and guilt. They are put in a difficult situation and given an easy solution. This solution may cause pain for others, it's true, but that pain comes as the price of their own self-esteem and peace of mind. Without it, they run the risk of living a life that has no meaning, one that gives them no feeling of self-worth, and no hope for immortality. A life where their life could end at any moment and it would have meant nothing at all. It is too easy for us here in the western world to judge those whose behaviour would be inacceptable in our safe, comfortable, and supportive society, but allow yourself, for a moment, to imagine living in their circumstances. Can you honestly say you would act any differently?

EXPERIMENTAL EVIDENCE

While Terror management theory is based on sound and extensive clinical work by some of the most renowned psychologists in the field (Becker, 1973), it cannot be taken as scientifically significant without being supported by specific experimental study. In past few decades, social psychologists such as Jeff Greenberg, Sheldon Solomon, Tom Pyszczynski , Jeff Schimel, and Joseph Hayes have been conducting lab research in support of terror management theory. Some of their experiments yield results that are convincing and indeed quite striking.

Two main hypotheses are focused on in TMT studies. The first has already been mentioned, that is the hypothesis of mortality salience, which dictates that if knowledge of mortality is made salient to a subject (if they are reminded of death) they will demonstrate stronger support for their cultural world-view and weaker support for an opposing worldview. Rosenblatt, Greenberg, Solomon, Pyszczynski, and Lyon (1989) asked judges whose occupation demands a strong sense of cultural values to recommend a bond for a prostitute, whose occupation infringes upon cultural values. The average bond recommended by the control was $50. If mortality was made salient, however, the average bond recommended rose to $455. This was repeated with students, and found that the same result occurs, however, even when mortality was made salient, an increased negative reaction towards prostitution was not observed if the student did not beforehand report a negative attitude towards prostitution. This suggests a direct relationship between mortality salience and cultural values. The study was also reversed for a hero who upholds cultural values (a woman who turns in a violent burglar in spite of possibility of harm to herself). The control recommended a reward of $1,112. When mortality was made salient, the reward rose to $3,476. Further research (Rosenblatt et al, 1990) showed that when mortality was made salient, subjects with a Christian religious background, when asked to form impressions of otherwise identical Christian and Jewish target persons, showed especially positive evaluations of the in-group member (Christian) and especially negative evaluations of the out-group member (Jew) when MS was induced. This research was further applied to attitudinally dissimilar persons, and persons who either praised or criticized the subject's cultural worldview, with supportive results.

These findings with Mortality Salience were used to study self-esteem (Greenberg et al., 1997). Subjects were first primed with either positive or neutral feedback on a personality test. The positive feedback served the purpose of bolstering the subject's self-esteem, and a self-test was administered to record this effect. The effects of MS were then tested. Subjects were given two essays, pro U.S. and anti U.S., and asked to rate the authors. In the control (neutral feedback) MS induced a large difference between the subject's opinion in the two authors.

(M = 3.06, higher statistics mean larger disparity) The control without MS showed relatively little difference (M = 0.92). In the positive feedback group, however, both the control and MS showed the same difference in opinion as the neutral feedback control. (M = 0.89, 0.90). Only the group that had induced MS and received neutral feedback showed especially positive and negative responses towards the authors that supported and criticized their culture, respectively. This suggests that increasing a subject's self-esteem has a buffering effect on mortality salience and death anxiety. This is important research in TMT, because it shows that if a subject has a high opinion of themselves, it removes the need for death suppressive constructs even after MS. This trend was replicated (Greenberg, 2008) using threatening videos and electric shock as the MS, and anxiety state tests and skin conductivity in order to measure anxiety.

Mortality salience has also shown subjects to be more supportive of charismatic leaders and less supportive of practical leaders (Arndt et al., 2004), more distasteful of sex unless reminded of romantic love first more fair, and an increased striving for uniqueness or similarity to others, depending on which was threatened (Greenberg, 2008).

The second hypothesis is the death-thought accessibility hypothesis. DTA is basically reversed MS; it states that if a particular aspect of a subject's cultural world-view is attacked or threatened, it will cause thoughts of death to be unconsciously more prevalent in the mind of the subject. In one study performed at the University of Alberta, Schimel, Hayes, Williams, and Jahrig (2007) led subjects to a webpage that either derogated Canadian or Australian values. DTA was then measured using a word fragment completion task and a lexical decision task as described in detail by Schimel et al., (2007). After reading the webpage that threatened the Canadian values, the subjects showed a greater accessibility for death-related words. The researchers also tested to see if the same results were found if generally negative words, rather than specifically death-related words, were tested for. Interestingly, there was no difference in the results for the negative words control and the neutral words control. This shows that attacking the subject's cultural worldview increases the accessibility of death thoughts only, and doesn't just put the subject in a negative state of mind.

As a final point of interest, another study conducted at the University of Alberta (Schimel, Hayes, Williams, 2007) tested whether the annihilation of those who threaten a subject's cultural worldview could have a buffering effect on DTA much like the bolstering of self-esteem has a buffering effect on MS. A sample of Christian subjects read either a non-worldview-threatening article about the aurora borealis, or worldview-threatening article on how the childhood home of Jesus Christ, Nazareth, is falling increasingly under Muslim control and influence. Half of the

subjects in the worldview-threatening group, however, had an additional paragraph at the bottom of the article that informed them that a plane full of 117 Muslims on their way to Nazareth had crashed and that there were no survivors. The subjects then completed a word fragment completion task to assess the accessibility of death thoughts. The results showed that the threat group showed increased DTA, whereas the control and threat/annihilation groups showed little difference. In addition, the threat group rated their liking for Muslims and others (Jews, Hindus, Atheists) much lower than the control or annihilation groups. The study shows that the annihilation of those who threaten one's worldview can have a buffering effect on DTA, much like bolstered self-esteem can buffer the effects of MS. The lower rating for other religious groups in non-annihilation subjects shows a decreased opinion of those that differ from their cultural worldview when that worldview is threatened. This was not seen in the annihilation group because they no longer needed to defend their own worldview; learning of the death of members of an opposing worldview serves this function for them. While some may argue that the annihilation groups rated their liking for Muslims higher out of sympathy, this explanation does not address why they also increased their opinion of Jews, Atheists, and Hindus. This can only be explained by the buffering of death-related anxiety leading to the loss of especially negative reactions to culturally dissimilar persons.

This is by no means a comprehensive account of all the studies that have been done on this topic. Over 200 studies have been done in North America and overseas on Mortality Salience alone. This is simply a small account of some of the most convincing research that has been done, some of which at our own university. The principles of TMT have been tested extensively and are supported by a large volume of psychological study. Other cognitive-behavioural and evolutionary ways to describe the behaviours observed are addressed in many of the studies, and in many cases are not capable of explaining the results with the same certainty and clarity of TMT. These concrete studies as well as the profound theoretical insight based on sound clinical observation both serve to establish TMT as a comprehensive and reliable model for predicting social behaviour in human beings.

CONCLUSION

Terror management theory, while both refreshing and convincing, is fundamentally just that, a theory. Many in the field of psychoanalysis retain the belief that social factors such as development of the personality during childhood, repressed experiences and emotions, and sexual and aggressive instincts have a stronger role in development. This bivalency can be described in terms of the familiar debate of nature vs. nurture. In this per-

spective, TMT argues in the favour of the importance of a person's existential nature, stating that aspects of behaviour originate from a common intrinsic dilemma for all people, and childhood development serves to differentiate how people deal with this dilemma using the complex social structures available to them.

Terror management theory is a look into the human experience that looks beyond the specifics of neuroscience and the simplicity of previous behavioural studies, and attempts to form an intelligent general theory for the shared life experience of all human beings. It looks into the deepest and most fundamental essence of humanity and attempts to explain with a holistic perspective the central motivations behind our instinctive actions, our most profound thoughts, and our innermost emotions. Through the examples of transference, heroism and religion, I have attempted to explain some of these behaviours, only a small subset of the emotions, behaviours and disorders that TMT can elaborate on, and the complexities that lead to their development.

TMT is also the first psychoanalytic theory to be supported by scientific experiment. Hundreds of studies in the past few decades have provided convincing evidence, and research has only but begun in this promising new psychological perspective. Further research in this field will allow psychologists to answer not just how behaviours develop, as is the focus in psychology at the present, but why they develop. This perspective is crucial because it will allow humanity to understand from a common viewpoint our most unsettling behaviours, and thereby allow us to address of the vast quantities of negative emotion and hurtful behaviour that is present in the world. Once we, as a species, can realise that evil, as well as good is simply a necessary product of our nature as human beings, only then can we intelligently self-regulate our hateful and aggressive emotions, sober, and without shame. Only then can we choose worldviews and paths in life that are not so insecure that they will eventually lead to psychological distress. Only if we can understand the science of evil, do we stand the chance of abolishing it.

REFERENCES

Arndt J, Cook A, Greenberg J, Landau MJ, Miller D, Ogilvie D, Solomon S. (2004). Deliver us from evil: The effects of mortality salience and reminders of 9/11 on support for president George W. Bush. *Personality and Social Psychology Bulletin, 30*(9), 1136-1150.

Becker E. (1973). *The Denial of Death*. New York, NY: The Free Press.

Greenberg J. (2008) Understanding the vital human quest for self-esteem. *Perspectives on Psychological Science, 3*(1).

Greenberg J, Harmon-Jones E, McGregor H, Pyszczynski T, Simon L, Solomon S. (1997). Terror management theory and self-esteem: Evidence that increased self-

esteem reduces mortality salience effects. *Journal of Personality and Social Psychology, 72*(1), 24-36.

Greenberg J, Kirkland S, Lyon D, Pyszczynski T, Rosenblatt A, Solomon S, Veeder M. (1990). Evidence for terror management theory II: The effects of mortality salience on reactions to those who threaten or bolster the cultural worldview. *Journal of Personality and Social Psychology, 58*(2), 308-318.

Greenberg J, Lyon D, Pyszczynski T, Rosenblatt A, Solomon S. (1989). Evidence for terror management theory: I. the effects of mortality salience on reactions to those who violate or uphold cultural values. *Journal of Personality and Social Psychology, 57*(4), 681-690.

Hayes J, Jahrig. J., Schimel J, Williams T. (2007). Is death really the worm at the core? Converging evidence that worldview threat increases death-thought accessibility. *Journal of Personality and Social Psychology, 92*(5), 789-803.

Hayes J, Schimel J, Williams T. (2007). Fighting death with death. The buffering effects of learning that worldview violators have died. *Psychological Science, 19*(5), 501-507.

Helzer JE, Robins LN, McEvoy L. (1987). Post-traumatic stress disorder in the general population. *England Journal of Medicine, 317*(26), 1630-1634.

Internet Encyclopedia of Philosophy. (2010). *Ancient Greek Philosophy*. Retrieved from: http://www.iep.utm.edu/greekphi/

Latham T. (n.d) The birth of psychoanalysis: Nietzsche's legacy and influence on Freudian thought document Research Paper. *Psychology Today*. Retrieved from http://www.psychologytoday.com/files/attachments/58144/the-birth-psychoanalysis.pdf

Mikulenka J. (Director). (2007). *Hunting the zodiac* [Documentary]. United States: John Mikulenka

Rossano JM. (2006) The religious mind and the evolution of religion. *Review of General Psychology, 10*(4), 346-364.

Shen P. (Director). (2006). *Flight from death: The quest for immortality*. [Documentary]. United States: Patrick Shen, Greg Bennick

The 9/11 Commission. (2004). *The 9/11 commission report*. (United States National Commission on Terrorist Attacks Upon the United States). Washington, DC

The Pew Forum on Religion & Public Life. (2011). *World Muslim Population Data Tables*. Retrieved from: http://features.pewforum.org/muslim-population/.
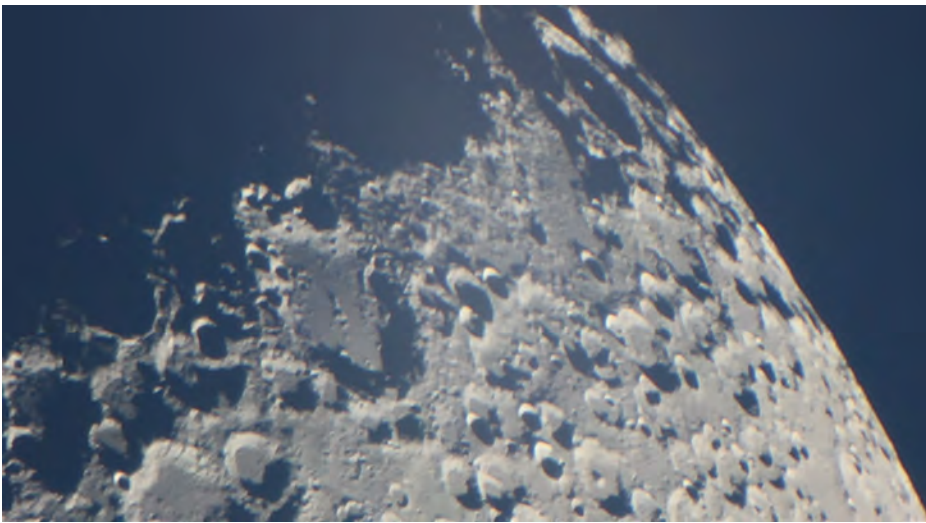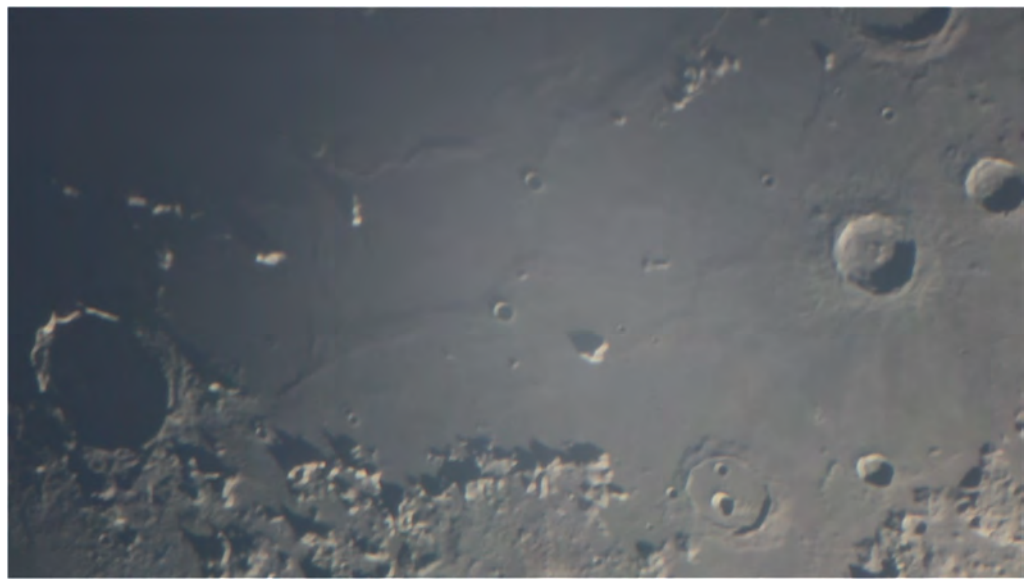
Eureka

Volume 4, Number 1

Summer 2014

University of Alberta Science Undergraduate Research Journal

http://www.eurekajournal.com