ON THE COVER

Protected from harmful human activities, nature reserves remain critical for research opportunities and conservation efforts. On page 12, Burnett *et al.* examine the determinants of plant abundance and diversity in the Gault Nature Reserve.

The cover illustrates the white-tailed deer (*Odocoileus virginianus*), a keystone species whose population surge in recent years has greatly impacted plant biodiversity.

MCGILL SCIENCE
UNDERGRADUATE
RESEARCH JOURNAL

BOARD 2019-2020

805 Sherbrooke St. West, Room 1B22
Montreal, Quebec, H3A 2K6 Canada
Phone: (514) 398-6979
Fax: (514) 398-6766
Email: msurj.sus@mail.mcgill.ca
Website: msurj.com

TABLE OF CONTENTS

FOREWORD

Dear Reader,

Throughout our existence, the McGill Science Undergraduate Research Journal has operated with the central purpose of encouraging and showcasing scientific research at the undergraduate level. Now in our fifteenth year, we would like to celebrate not only our continued pursuit of that purpose but also the continued excellence in the work that we receive.

Each year, when we are on the cusp of bringing a new issue to publication, we are reminded of the progress that undergraduate researchers have made from the past to the present. This year, we were also inspired to look to the future. With the continued improvement in the quality of work that we receive each year, we are excited for what our authors will achieve as they continue in their careers into graduate studies or industry. We are equally excited to see the undergraduate work that we will receive for the next year and onwards.

We would like to present the 2019-2020 issue of the McGill Science Undergraduate Research Journal. This issue contains eleven research and review articles written by undergraduate students working in diverse scientific fields. The passion and effort that our authors put into their work is evident to us from the beginning. We hope that we were able to preserve that passion in the pages of our journal for all to experience.

On behalf of our entire editorial board, thank you.

Howard Li & Jacqueline Yao
Editors-in-Chief

ACKNOWLEDGEMENTS

The McGill Science Undergraduate Research Journal would like to thank its generous contributors without whom this journal would not have been possible.

We thank Dean Bruce Lennox for his continuing support as we work towards engaging undergraduate students in scientific research. His generosity and expertise has allowed MSURJ to achieve its current success. We would also like to acknowledge Mr. Victor Chisholm's dedication to the undergraduate research community and his guidance. Lastly, we thank the librarians of McGill University for helping MSURJ explore new online platforms.

We thank all the financial supporters in the McGill community for their generous support:

Faculty of Science
Faculty of Medicine
Science Undergraduate Society of McGill (SUS)
Department of Physiology
Chemistry Undergraduate Students' Society (CUSS)
McGill School of Computer Science

Finally, we are grateful to the many peer reviewers who took the time to review our submissions and the authors who submitted their work. The articles in this journal are made possible thanks to their time, expertise, and dedication.

Submitted: 01/15/20

Research Article

[1] McGill School of Environment, Montreal, QC, Canada
[2]Department of Biology, McGill University, Montreal, QC, Canada

Keywords

Plants, biodiversity, ecology, white-tailed deer, Gault Nature Reserve.

Email Correspondence

mackenzie.burnett@mail.mcgill.ca

Mackenzie Burnett[1], Imogen Hobbs[1], Alexa Ripple[2]

Seeing the Forest for the Deer: Plant Abundance and Diversity at the Gault Nature Reserve During a Spike in White-Tailed Deer

Abstract

Background: Deer populations have been rising across North America for decades. At the Gault Nature Reserve in Quebec, half of which is open to the public, the population of white-tailed deer (*Odocoileus virginianus*) has exceeded the region's carrying capacity, estimated to be 5 deer/km^2, since 1996. Given that heavy grazing profoundly impacts forests, the purpose of this paper was to investigate the potential influence of white-tailed deer on plant abundance and diversity at the Gault Nature Reserve. We hypothesized that the abundance of deer, and by extension the effect of deer on vegetation, was negatively correlated with the proximity and frequency of human visitors on pedestrian trails. Our alternative hypothesis was that the effect of deer on vegetation was positively correlated with human disturbance, which is greater on the public side of the reserve.

Methods: We recorded the abundance and diversity of vascular plants along 14 transects of increasing distance from pedestrian trails on the public and private sides of the reserve.

Results: Contrary to our hypothesis, generalized linear models indicated that overall, plant abundance and diversity declined significantly as the distance from trails increased and that the effect of distance was significantly different on the two sides of the reserve. Pearson correlation tests revealed that there was not a significant correlation between distance and plant abundance and diversity on the public side, although there was a significantly negative correlation between these variables on the private side.

Limitations: White-tailed deer were not directly studied, which limited the inferences that could be made about their influence on plant abundance and diversity.

Conclusion: The distance from trails was a strong determinant of plant abundance and diversity on the private of the reserve, but not on the public side, possibly because trail edges generally receive more sunlight and because the increased number of trails on the public side may have confounded our results. Although we did not find support for our hypothesis, the influence of trail edges on plant communities was reinforced. Researchers should continue to monitor the influence of white-tailed deer and forest managers should be mindful of edge effects when making decisions.

Introduction

In North America, white-tailed deer (*Odocoileus virginianus*) were extirpated from numerous areas and later reintroduced in the middle of the 20th century. (1-3) Their predators were also extirpated, but were not reintroduced. (1-2) In combination with restricted hunting seasons and laws, forest fragmentation, agricultural expansion, and the spread of exurban areas, deer populations have since reached an unprecedented density in many areas. Mont St. Hilaire in Quebec is one example of a region that has a high-density population of white-tailed deer. From 1996 onwards, the deer population in the area has exceeded the carrying capacity, i.e. the number of deer that a given amount of land can sustainably support (4), estimated to be 5 deer/km^2. (5) Although extensive surveys of the flora at Mont St. Hilaire have been completed (6), some questions remain about how the population of white-tailed deer impacts the Gault Nature Reserve.

The Gault Nature Reserve is located on Mont St. Hilaire, which is roughly 32 km east of Montreal and at the northern limit of the temperate deciduous forests of eastern North America. (7) The reserve is approximately 10 km^2 in size and consists of seven low peaks that form a ring around Lac Hertel. The western half of the reserve is designated for public use, while the eastern half is off limits to public visitors. Unlike the surrounding area, the forest on the mountain has experienced little human disturbance, such as tree removal for roads or properties (8). Numerous trees are more than 120 years old and sugar maples have been found that are over 400 years old. (8) Despite the reserve being well-managed and respected by local citizens, striking changes in species richness and evolutionary relationships have occurred over the past few decades. (6)

High densities of white-tailed deer dramatically affect woody and herbaceous vegetation (1, 3, 4, 9) along with insects, birds, and other mammals. (2, 3) At high densities, deer browsing can decrease vegetation density (1, 2) and lead to the extirpation of herbaceous species. (3, 4, 10, 11) Heavy deer browsing can also increase the spread of invasive species (12), alter forest succession (2), change forest regeneration (3), and modify microenvironments. (1, 9) Additionally, Takada *et al.* (2002) found that the presence of deer herbivory induced thicker spines in a spiny shrub, *Damnacanthus indicus*. (13) Since white-tailed deer negatively affect the distribution and abundance of numerous species and can influence community structure by impacting multiple trophic levels, white-tailed deer have been described as a keystone species. (3, 9) As a result, this paper aims to study how the high-density population of white-tailed deer could be impacting plant abundance and diversity at the Gault Nature Reserve. The experiment consisted of tallying the number of individual vascular plants and plant species found in circular plots (radius = 1 m) along 100 m transects at 5, 20, 40, 60, 80, and 100 m from the edges of pedestrian trails.

Fig. 1 illustrates our two hypotheses, which are based on the tendency of dense white-tailed deer populations to shrink plant abundance (1) and herbaceous species richness in the areas in which they preside. (1, 2) In our first hypothesis, depicted in Fig. 1a-b, we posited that deer were more concentrated on the private side of the reserve, possibly because they were

avoiding the abundance of visitors on the public side. If this hypothesis was correct, we expected the plant abundance and diversity to be lower on the private side of the reserve. Moreover, we expected the plant abundance and diversity on the public side to become approximately equal to the plant abundance and diversity on the private side when enough distance had passed to render the visitors' impact on the deer negligible. It was assumed that white-tailed deer would do the most grazing and thus have the greatest impact wherever they were concentrated most. Plant abundance was defined as the total number of plants observed in each plot and plant diversity was encapsulated by the Shannon diversity index, which takes both species richness and evenness into account. (14) Our second hypothesis, illustrated in Fig. 1c-d, was based on findings that herbivory had stronger effects in areas with high disturbance versus low disturbance. (15) Since the public side has more disturbance, like pedestrian trails and visitors who could be straying from the trails and trampling vegetation, in our second hypothesis we theorized that the impact of the deer population would be greater on the public side of the reserve. If this hypothesis was correct, we expected the plant abundance and diversity on the public side to be lower and to become approximately equal to the plant abundance and diversity on the private side after sufficient distance has passed to render the impact of people on deer negligible. Furthermore, in both hypotheses we proposed that plant abundance and diversity would remain the same along the transects on the private side because the private side had much less disturbance.

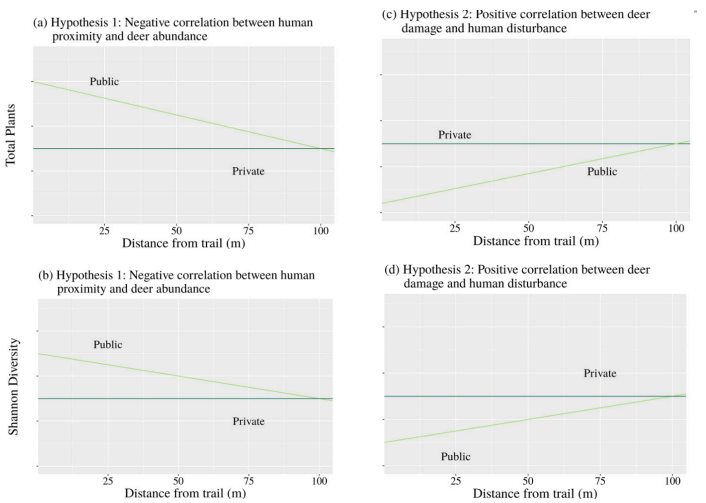

Figure 1. Hypotheses regarding the potential change in plant abundance and Shannon diversity index along transects at the Gault Nature Reserve, Quebec, Canada. *(a)* and *(b)* make up the first hypothesis, while *(c)* and *(d)* make up the second hypothesis. *Public* refers to the side of the Gault Nature Reserve where public visitors are allowed and *Private* refers to the side where public visitors are not allowed. Lastly, *Shannon Diversity* refers to an index used to assess diversity that takes both species richness and evenness into account. (14)

Methods

This study was conducted at the Gault Nature Reserve, located on Mont St. Hilaire, Quebec, Canada, in August 2019. The reserve is comprised of 10 km^2 of mature temperate forest divided into a public portion that features 25 km of pedestrian trails and a private portion that receives markedly less human traffic. (16) In between these two sides is Lac Hertel, a glacially formed depression that has a maximum depth of 9 m. (17) The climate of the region is humid continental with a mean annual temperature of 6 °C and a mean annual precipitation of 1000 mm. (4) There have been 600 species of plants and 800 species of butterflies identified at the reserve. (16) Additionally, 353 known types of minerals can be found at Mont St. Hilaire, which is considered one of the top mineralogical sites in the world. In 1978, the Gault Nature Reserve became the centre of the first Canadian reserve in the UNESCO Biosphere network. (16)

Before we started collecting data, we stratified the study area in order to reduce the types of forest and thus the number of variables being evaluated. The United States' Forest Service employs a similar phase in their sampling procedure because the species composition in plots that straddle multiple forest types may be highly variable from plot to plot. (18) Since slope can impact the distribution of plants and communities (7, 19), topographical maps of the Gault Nature Reserve were used to identify areas on the public and private side that had contour lines approximately the same distance apart. Moreover, because proximity to bodies of water can also impact plant communities (7, 20), maps were used to find areas on the public versus the private side of the reserve that were similar distances to Lac Hertel. Accordingly, seven transects were conducted on each side of the reserve, three of which started 10-50 m from the lake and four of which started more than 100 m from the lake, as shown in Fig. 2. All the transects near the lake had a similar slope and all those that were far from the lake had a similar slope as well. In the end, transects were set up in areas where American beech (*Fagus grandifolia*) and red maple (*Acer rubrum*) were consistently the dominant plants.

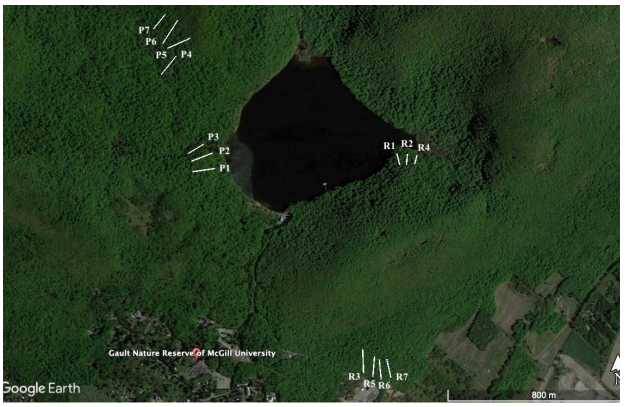

Figure 2. The locations of our surveyed transects in the mature temperature forest of the Gault Nature Reserve, Quebec, Canada. *R1-R7* refer to the seven transects on the private side of the reserve. *P1-P7* refer to the seven transects on the public side of the reserve. The labels have been placed on the sides that the transects started on. The transects were 100 m long and had circular plots (radius = 1 m) at 5, 20, 40, 60, 80, and 100 m from the edges of pedestrian trails. Image obtained from Google Earth. (21)

A detailed transect procedure was used to increase repeatability. The number of individual vascular plants and plant species in circular plots (radius = 1 m) were identified and recorded at 5, 20, 40, 60, 80, and 100 m from pedestrian trail edges. Circular plots were used because they have a smaller border per area than a square plot. As a result, circular plots have a smaller number of plants that could be along the border, which reduced the amount of variability introduced by making decisions about whether a plant was inside or outside the plot. (22) Furthermore, circular plots are utilized in the United States' Forest Service sampling procedure (22) and are one of the methods discussed in Western Australia's Department of Environment and Conservation standard operating procedure. (23) In addition, all the surveying was done between August 27th-29th, 2019 since plant diversity and abundance can vary over time. Lastly, roughly 30 minutes was spent on each transect to help standardize the surveying effort, similar to the sampling procedure used by the United States' Forest Service. (18)

The transect length was intended to capture potential edge effects in order to explore how plant composition might change with an increasing distance from pedestrian trail edges. Edge effects refer to changes in population or community structure that occur at the boundary of two or more habitats (24); however, edge effects can vary in distance and impact depending on a variety of factors. (25) Elevated forest disturbance has been evident in some cases up to 400 meters inside fragment margins (26), though the most striking changes frequently occur within 200 meters of edges. (25, 27) Other studies have stated that edge effects may only reach 50 meters into boreal and temperate forests (28), or 15 meters into managed hardwood forests. (29) Ultimately, 100 meter transects were used because 100 meters appeared to be an intermediate distance among past assessments of edge effects and because it enabled more transects to be completed.

Care was also taken to establish transects at distances from each other that would suggest independence. The most common species that was identified was American beech, which is known to have limited seed dispersal. (30) Although some small mammals may distribute seeds over short distances, and birds can potentially transport seeds several kilometers, most American beech seeds drop to the ground immediately around the tree. (30) Consequently, transects were established 50 m from each other, like in the procedure developed by Koivula *et al.* (28), and were assumed to be independent. The goal was to realize a "many and small" sampling scheme. (31) Having many, small surveyed areas yields accurate abundance estimates for the most common species, although species lists may be incomplete. Conversely, having fewer, large surveyed areas may result in a more complete species list, but the abundance of rare species could be overestimated. (31) As a result, the "many and small" scheme was implemented with transects being 50 m apart to avoid overestimating the abundance of rare species.

Plant surveying consisted of counting and identifying vascular stems that were under 1 meter in height. Although percent cover is a method of surveying plants that is quick and widely used by organizations like the United States' Forest Service, estimates can differ among observers. (22) Individual counts were used because they can be well-suited to monitoring treatment effects. (22) Despite white-tailed deer being able to remove twigs and foliage up to a height of 6 feet (32), plants under 1 meter were the focus of our survey because white-tailed deer commonly eat large amounts of vegetation that are closer to the ground. (33) When the species of a plant was not readily identifiable, it was assigned a genus or unknown code and was collected for later identification, similar to the sampling procedure utilized by the United States' Forest Service. (22) Next, when another plant was found that was not readily identifiable, it was compared to previous unidentified specimens to determine whether it constituted a new species or not. Lastly, surveying was done as a group to minimize the variation that can occur between observers. These steps facilitated the analysis of the potential effect imparted by the distance from trails and the side of the reserve.

After the plant survey was complete, the data was analyzed via generalized linear models (GLMs). GLMs are based on an assumed relationship between the mean of the response variable and the explanatory variable(s). (34) In addition, GLMs attempt to establish how the explanatory variable(s) cause the response variable to change. The interaction between explanatory variables can also be investigated via GLMs to explore whether they influence each other. In GLMs, data can be assumed to have a variety of probability distributions, like normal or Poisson. As a result, GLMs are considered more flexible and better suited for analyzing ecological relationships, which do not always fit a normal distribution. (34) The response variable in the first GLM was plant abundance, meaning the total number of plants. The explanatory variables were the distance from pedestrian trails and the side of the reserve, i.e. public versus private side. A Poisson distribution was used in this instance because plant abundance is an example of count data, which frequently follows a Poisson distribution. (34) The response variable of the second GLM was the Shannon diversity index, while the distance from pedestrian trails and the side of the reserve were once again the explanatory variables. The Shannon diversity index was calculated using the *vegan* package (35) in R (36), based on the following equation (Eq. 1). The Shannon diversity index was selected because it is a widely used diversity index in ecology that takes both species richness and evenness into account. (17) The second GLM used a normal distribution because the Shannon diversity index does not consist of count data.

$$H = \sum_{i=1}^{S} \frac{n_i}{N} \ln ln \left(\frac{n_i}{N}\right)$$

Equation 1. The formula for the Shannon diversity index (H). In this equation, *n* is the number of plants of a particular species that were observed in a plot, *N* is the total amount of plants observed in a plot, and *S* is the total number of plant species that were observed in a plot. (17)

Finally, we used Pearson correlation tests to gain further insight into the relationships at play. Correlation tests quantify the direction and strength of the linear relationship between a pair of variables. (37) While GLMs could indicate whether the side of the reserve was a significant explanatory variable of plant abundance and diversity, they could not indicate whether the relationships between the distance from trails and plant abundance and diversity were stronger on the private or public side. In order to ascertain this, Pearson correlation tests were employed. Moreover, a Pearson correlation test was used to evaluate whether there was a significant relationship between plant abundance and diversity on the two sides of the reserves.

Results

A total of 623 plants and 22 plant species were observed on 14 transects in the Gault Nature Reserve between August 27th - 29th, 2019. The three most abundant species were American beech, red maple, and beech drops (*Epifagus virginiana*), which respectively accounted for 23%, 12%, and 9% of all the surveyed plants, as shown in Fig. 3.

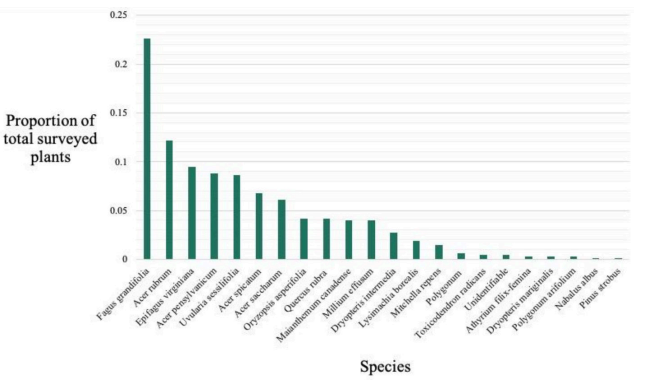

Figure 3. The abundance of the species observed on 14 transects at the Gault Nature Reserve, Quebec, Canada, expressed as a proportion of the total number of surveyed plants.

The plant abundance data, depicted in Fig. 4, was used to perform a GLM in R (36) to gain insight into the relationships between plant abundance, distance from pedestrian trails, and sides of the reserve. Plant abundance decreased as the distance from trails increased on both the public and private side of the reserve, though plant abundance dropped off to a noticeably greater degree on the private side. When the GLM analyzed the overall change in plant abundance in the plots on both sides of the reserve as the distance from trails increased, distance was found to have a significantly negative effect on plant abundance, as is summarized in Table 1.

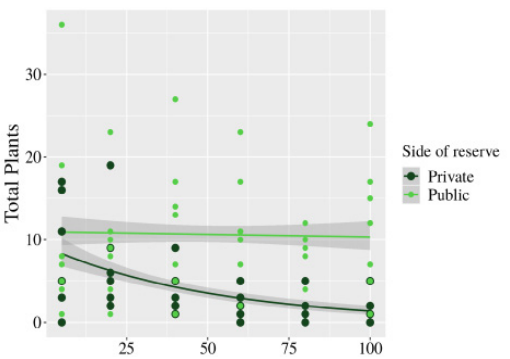

Figure 4. The relationship between the distance from pedestrian trails and the number of plants under 1 meter in six circular plots (radius = 1 m) along fourteen 100 m transects at the Gault Nature Reserve, Quebec, Canada. The figure was fitted using a Poisson distribution, and the shaded areas represent the standard error of the mean. *Private* refers to the side of the reserve where public visitors are not allowed, and *public* refers to the side of the reserve where public visitors are allowed.

However, the side of the Gault Nature Reserve did not have a significant impact on plant abundance. Lastly, the side of the reserve and the distance from trails had a significant interaction, which means that the effect of the distance from trails was significantly different on the public versus private side of the reserve. In short, the first GLM revealed that an increasing distance from trails had a significantly negative impact on the combined number of plants that were observed and that the effect of distance was significantly different on the two sides of the reserve.

GLM #1 Response variable: plant abundance	Estimate	*p*	*z*	Standard error
Explanatory variable #1: Distance from pedestrian trails	-0.018971	$6.83*10^{-13}$	-7.183	0.002641
Explanatory variable #2: Side of reserve (public)	0.181904	0.202	1.275	0.142718
Interaction between explanatory variables: distance : side (public)	0.018389	$9.55*10^{-10}$	6.117	0.003006
GLM #2 Response variable: Shannon diversity index	Estimate	*p*	*t*	Standard error
Explanatory variable #1: Distance from pedestrian trails	-0.006239	0.00758	-2.740	0.002277
Explanatory variable #2: Side of reserve (public)	0.147754	0.45113	0.757	0.195123
Interaction between explanatory variables: distance : side (public)	0.007431	0.02362	2.307	0.003221

Table 1. Results of the generalized linear models, GLMs, in which the influence of the explanatory variables on the response variable was evaluated. In the first GLM, the explanatory variables were the distance from pedestrian trails and the side of the Gault Nature Reserve, while the response variable was plant abundance. In the second GLM, the explanatory variables were the distance from pedestrian trails and the side of the Gault Nature Reserve, and the response variable was the Shannon diversity index. The interactions between the explanatory variables were also evaluated in these two GLMs.

The Shannon diversity index data, illustrated in Fig. 5, was used to perform a second GLM in R (36) to explore the relationships between plant diversity, distance from pedestrian trails, and sides of the reserve. On the public side of the reserve, the Shannon diversity index seemed to increase somewhat as the distance from trail edges increased. Conversely, on the private side of the reserve, the Shannon diversity index clearly decreased as the distance from trails increased. When the second GLM, summarized in Table 1, analyzed the overall change in the Shannon diversity index in the plots on both sides of the reserve as the distance from trails increased, distance was found to have a significantly negative effect. Once again though, the side of the reserve did not have a significant effect itself. Finally, the side of the reserve and the distance from the trails had a significant interaction. In other words, an increasing distance from the trails had an overall negative impact on the Shannon diversity indices that were observed and the effect of distance changed from one side of the reserve to the other.

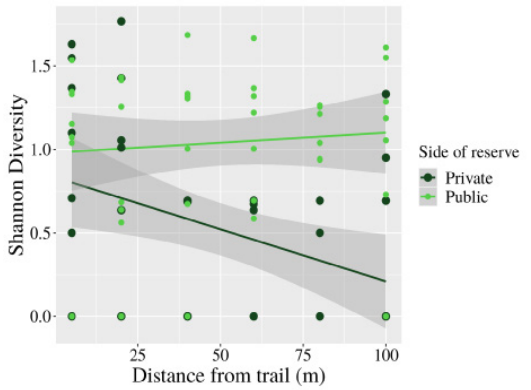

Figure 5. The relationship between the distance from pedestrian trails and the Shannon diversity index of six circular plots (radius = 1 m) along fourteen 100 m transects at the Gault Nature Reserve, Quebec, Canada. The figure was fitted using a Gaussian distribution and the shaded areas represents the standard error of the mean. *Private* refers to the side of the reserve where *public* visitors are not allowed, and public refers to the side of the reserve where public visitors are allowed.

To summarize, the relationship between each of the response variables, plant abundance and diversity, and the explanatory variable, distance from trails, was different on the public side of the reserve versus the private side. The two GLMs supported this observation because they both indicated that the effect of distance was significantly different on the two sides of the reserve. To evaluate the strength and direction of correlations between the explanatory and response variables on the private side and public side, Pearson correlation tests were used. The results of these tests are summarized in Table 2. The tests revealed that the correlation between plant abundance and distance from trails was significantly negative on the private side and insignificant on the public side. Likewise, the tests revealed that the correlation between the Shannon diversity index and distance from trails was significantly negative on the private side and insignificant on the public side. Lastly, Pearson correlation tests indicated that there was a significantly positive relationship between plant abundance and diversity on both sides of the reserve.

	p	Pearson's correlation	*t*	95% confidence interval	Degrees of freedom
Test #1: Plant abundance vs distance from trails (private side)	0.0005192	-0.51259	-3.7757	-0.706404 to -0.2471651	40
Test #2: Plant abundance vs distance from trails (public side)	0.8661	-0.02681439	-0.16965	-0.3280721 to 0.2793940	40
Test #3: Shannon diversity index vs distance from trails (private side)	0.01394	-0.3766711	-2.5717	-0.61068938 to -0.08214367	40
Test #4: Shannon diversity index vs distance from trails (public side)	0.5767	0.0886457	0.56286	-0.2212465 to 0.3822777	40
Test #5: Shannon diversity index vs plant abundance (private side)	$1.843*10^{-5}$	0.7422564	7.0054	0.5660152 to 0.8536162	40
Test #6: Shannon diversity index vs plant abundance (public side)	$3.27*10^{-6}$	0.649347	5.4015	0.4304713 to 0.7962078	40

Table 2. Pearson correlation test results. Test #1 and #2 evaluated the relationship between plant abundance and the distance from pedestrian trail edges along fourteen 100 m transects split between the public and private side of the Gault Nature Reserve, Quebec, Canada. Test #3 and #4 assessed the relationship between the Shannon diversity index and the distance from trail edges along the same transects. Lastly, test #5 and #6 analyzed the relationship between the Shannon diversity index and plant abundance.

Discussion

Initially, we hypothesized that plant abundance and diversity would be lower on the private side of the reserve because white-tailed deer may be frightened by visitors and thus concentrated on the private side. Our alternative hypothesis was that plant abundance and diversity would be lower on the public side of the reserve because herbivory has a stronger impact in areas with high disturbance. (17) In both hypotheses, we speculated that the plant abundance and diversity on the public side would be roughly equal to that of the private side after approximately 100 meters. At about 100 meters, we theorized that the influence of people on white-tailed deer and the forest would be nearly equal on both sides; yet, the results of our survey do not align with either of these hypotheses. The GLMs showed that as the distance from pedestrian trails increased, the change in plant abundance and Shannon diversity index was significantly negative overall and significantly different on the public versus the private side of the re-

serve. On the private side, Pearson correlation tests indicated that the plant abundance and Shannon diversity index were significantly greater close to the trail. Conversely, on the public side, Pearson correlation tests revealed that the plant abundance and diversity were not significantly correlated with the distance to trails.

A variety of factors might account for why our results differ from our hypotheses. The white-tailed deer in the Gault Nature Reserve may be more desensitized to the presence of people than anticipated, which could help explain why the side of the reserve was not a significant explanatory variable in the GLMs. Additionally, it is possible that people visiting the reserve could be acting as vectors for seed dispersal through seeds attaching to their clothes or equipment. This could be one of the reasons that plant abundance and diversity is generally greater near trails, where there is the highest human traffic. Elliott and Davies (2019) found that species richness on Mont St. Hilaire increased between surveys completed in 1958-1960 and 2012-2015. (6) Moreover, they found that the increased species richness could be attributed to people and other animals acting as vectors for seed dispersal and climate change impacting species distribution.

Differences in sunlight might have also contributed to our results. Researchers who investigated the environmental basis for canopy composition at Mont St. Hilaire found that slope and the amount of solar radiation, both of which influence soil moisture, were the main determinants of plot-to-plot variation. (7) Similarly, another study found that the difference in plant composition near trails is due to trail edges having lower canopy cover and therefore more sunlight reaching the understory. (29) Thus, the greater drop off in plant abundance and diversity on the private side of the reserve might be explained by the decreased disturbance it experiences, which may have caused less sunlight to reach the ground further from the trail. Correspondingly, it is possible that the increased presence of visitors could have contributed to the plant abundance and diversity not being significantly correlated with the distance from trails on the public side. If visitors were straying from paths and trampling vegetation, it could have caused more sunlight to filter through to the ground off the public paths over time. This increased amount of sunlight could have led to greater plant abundance and diversity.

Furthermore, our results could have been impacted by the numerous interconnecting pedestrian trails on the public side of the reserve, which leave smaller pockets of forest between them. At times, it was difficult to find areas on the public side where the end of the transect would be at least 100 m from every other trail while also having a similar slope and distance to Lac Hertel as the other transects. As a result, some of the public transects may have been conducted in areas where the end of the transect was too close to the edge of another pedestrian trail to reflect the actual conditions of the forest ecosystem 100 meters from a trail. Nevertheless, these three studies (29, 38, 39) and our results on the private side show that the distance to trails strongly influences plant abundance and diversity, while other research shows that the difference in sunlight may be one of the reasons why. (8, 29) Although the results on the public side may have been confounded by the increased number of pedestrian trails, the results on both sides of the reserve reinforce that edges have a powerful influence on ecosystems.

A few factors limit or impact the results of our plant survey. Species could have been overlooked because of low abundance at the time of the survey. There may have also been species that could not be identified due to their stage of development. The damaged condition of some plants rendered their identification more difficult as well. Factors such as these are inherent in essentially all plant surveys and contribute to a degree of uncertainty regarding species composition and diversity estimates. (22) Furthermore, the activity of white-tailed deer was not observed directly during this experiment, which made forming inferences about their impact more difficult. In the future, observing deer directly with camera traps might enable more causal conclusions to be drawn about their impact. Employing deer exclosures could also be informative. Finally, using percent cover as a means of surveying the Gault Nature Reserve might enable more transects to be evaluated and more plant species to be recorded. Future research could reduce the limitations that were encountered in this experiment by executing these changes.

Conclusion

Although we did not find support for our hypotheses regarding the impact of white-tailed deer on plant abundance and diversity at the Gault Nature Reserve, other sources clearly indicate that white-tailed deer can act as a keystone species. (1-3, 9-13, 17, 38) Future researchers should monitor how the overabundance of white-tailed deer impacts the ecosystem at the Gault Nature Reserve and elsewhere. If we disregard the keystone effects of deer on temperate and boreal forests, we risk losing the suite of species and processes needed to maintain ecosystem function. (9) Ultimately, our research underlined the influence that edges can have on plant communities. Individuals who manage ecosystems like the Gault Nature Reserve should be mindful of edge effects when making decisions.

Acknowledgements

We would like to thank Dr. Anna Hargreaves, Dr. Rowan Barrett, and Dr. Simon Reader for everything they taught us and for their help while this project was being designed and implemented. We would also like to thank Ph.D. student David Hunt for guiding us through the process of executing a research project and for his help during data analysis.

References

1. Shelton AL, Henning JA, SHultz P, Clay K. Effects of abundant white-tailed deer on vegetation, animals, mycorrhizal fungi, and soils. Forest Ecology and Management. 2014;320:29-49.

2. Côté SD, Rooney TP, Tremblay J-P, Dussault C, Waller DM. Ecological impacts of deer overabundance. Annual review of Ecology, Evolution, and Systematics. 2004;35:113-47.

3. Waller DM, Alverson WS. The white-tailed deer: A keystone herbivory. Wildlife Society Bulletin. 1997;25(2):217-26.

4. Beauvais M-P, Pellerin S, Lavoie C. Beta diversity declines while native species richness triples over 35 years in a suburban protected area. Biological Conservation. 2016;195:73-81.

5. Huot M, Lebel F. Plan de gestion du cerf de Virginie au Québec 2010-2017. Ministère des Ressources naturelles et de la Faune; 2012.

6. Elliott TL, Davies TJ. Phylogenetic attributes, conservation status and geographical origin of species gained and lost over 50 years in a UNESCO Biosphere Reserve. Biodiversity and Conservation. 2019;28(3):711-28.

7. Arii K, Hamel BR, Lechowicz MT. Environmental correlates of canopy composition at Mont St. Hilaire, Quebec, Canada. Torrey Botanical Society. 2005;132(1):90-102.

8. Gault Nature Reserve - Mont-Saint-Hilaire. Flora 2008 [Available from: https://web.archive.org/web/20081022110133/http:/www.mcgill.ca/gault/sainthilaire/natural/flora/].

9. Rooney TP, Waller DM. Direct and indirect effects of white-tailed deer in forest ecosystems. Forest Ecology and Management. 2003;181(1-2):165-176.

10. Augustine DJ, Frelich LE. Effects of white-tailed deer on populations of an understory forb in fragmented deciduous forests. Conservation Biology. 2001;12(5):995-1004.

11. Knight TM, Caswell H, Kalisz S. Population growth rate of a common understory herb decreases non-linearly across a gradient of deer herbivory. Forest Ecology and Management. 2009;257(3):1095-103.

12. Koh S, Watt TA, Brazely DR, Pearl DL, Tang M, Carleton TJ. Impact of herbivory of white-tailed deer (Odocoileus virginianus) on plant com-

munity composition. Aspects of Applied Biology, Vegetation management in forestry, amenity and conservation areas: Managing for multiple objectives. Wellesbourne, Warwick: Association of Applied Biologists, Horticulture Research International. 1996;44:445-450.

13. Takada M, Asada M, Miyashita T. Regional differences in the morphology of a shrub Damnacanthus indicus: An induced resistance to deer herbivory? Ecological Research. 2003;16(4):809-813.

14. Gardner M. Community Ecology: Analytical methods using R and Excel: Pelagic Publishing; 2014.

15. About the Reserve [Internet]. Available from: https://gault.mcgill.ca/en/

16. Centre de la Nature Mont Saint-Hilaire. Flora and fauna [Internet]. Le mont Saint-Hilaire. 2008. Available from: https://web.archive.org/web/20080630222323/http://www.centrenature.qc.ca/en/mountai n/faunaflora.html

17. Maron JL, Crone E. Herbivory: effects on plant abundance, distribution and population growth. Proceedings of the Royal Society B. 2006;273:2575-84.

18. Shulz BK, Bechtold WA, Zarnoch SJ. Sampling and estimation procedures for the vegetation diversity and structure indicator. United States Department of Agriculture Forest Service; 2009.

19. Spira TP. Wildflowers & plant communities of the southern Appalachian Mountains and Piedmont: The University of North Carolina Press; 2011:22.

20. National Research Council. Riparian areas: Functions and strategies for management. Washington, D.C.: National Academy Press; 2002.

21. Google Earth Pro. Mont Saint Hilaire. 7.3.2.5776 ed. 2019.

22. Shulz BK, Bechtold WA, Zarnoch SJ. Sampling and estimation procedures for the vegetation diversity and structure indicator. United States Department of Agriculture Forest Service; 2009.

23. Clarke V, Perry D. Standard operating procedure: establishing vegetation transects. Department of Environment and Conservation; 2009.

24. Levin SA, editor. The Princeton guide to ecology. Princeton, New Jersey: Princeton University Press; 2012.

25. USDA Forest Service. Three trails OHV project environmental impact statement. Klamath County, Oregon; 2011.

26. Qie L, Lewis SL, Sullivan MJP, Lopez-Gonzalez G, Pickavance GC, Sunderland T, et al. Long-term carbon sink in Borneo's forests halted by drought and vulnerable to edge effects. Nature Communications. 2017;8(1):1-11.

27. Angold PG. The impact of a road upon adjacent heathland vegetation: Effects on plant species composition. Journal of Applied Ecology. 1997;34(2):409-17.

28. Koivula M, Virta T, Kuitunen M, Valllius E. Effects of undergrowth removal and edge proximity on ground beetles and vascular plants in urban boreal forests. Journal of Urban Ecology. 2019;5. (1)
29. Watkins RZ, Chen J, Pickens J, Brosofske KD. Effects of Forest Roads on Understory Plants in a Managed Hardwood Landscape. Conservation Biology. 2003;17(2):411-419.

30. United States Department of Agriculture. Fagus grandifolia n.d. [Available from: https://www.fs.fed.us/database/feis/plants/tree/faggra/all.html#BOTANICAL%20AND%20ECOLOGICAL%20CHARACTERISTICS].

31. McCune B, Grace J. Analysis of Ecological Communities. Gleneden Beach, Oregon: Mjm Software Design; 2002.

32. Saunders DA. Adirondack mammals. New York: Adirondack Wildlife Program, State University of New York, College of Environmental Science and Forestry; 1988.

33. Dillard J, Jester S, Baccus J, Simpson R, Poor L. White-tailed deer food habits and preferences in the cross timbers and prairies region of Texas. Austin, Texas: Texas Parks and Wildlife; 2005.

34. Guisan A, Edwards TCJ, Hastie T. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling. 2002;157(2-3):89-100.

35. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. 2019.

36. R Core Team. R: A language and environment for statistical computing.
3.6.0 ed. Vienna, Austria: R Foundation for Statistical Computing; 2019.

37. Bewick V, Cheek L, Ball J. Statistics review 7: Correlation and regression. Critical Care. 2003;7(451):451-459.

38. Euskirchen ES, Chen J, Bi R. Effects of edges on plant communities in a managed landscape in northern Wisconsin. Forest Ecology and Management. 2001;148(1-3):93-108.

39. Shen X, Bourg NA, McShea WJ, Turner BL. Long-term effects of white-tailed deer exclusion on the invasion of exotic plants: A case study in a mid-Atlantic temperate forest. PLOS ONE. 2016;11(3): e0151825.

Submitted: 01/07/20

Research Article

[1]Department of Pharmacology and Therapeutics, McGill University, Montreal, QC, Canada

Keywords

Endochondral ossification, organophosphate esters (OPEs), developmental toxicology, skeletal staining

Email Correspondence

charlise.x.chen@mail.mcgill.ca

Charlise Xinyi Chen[1], Han Yan[1], Barbara Hales[1]

Gestational and Early Postnatal Exposure to a Mixture of Organophosphate Ester Flame Retardants Found in Canadian House Dust on Hindlimb Skeletal Development in Postnatal Day 4 Rats

Abstract

Background: Ever since organophosphate esters (OPEs) became the mainstream replacement for organobromine compounds in fire retardants (FRs), numerous studies have explored their potential as endocrine disruptors and developmental toxicants. The purpose of this study is to investigate the impact of gestational and early postnatal exposure of OPE mixtures on the ossification of hindlimb phalangeal in postnatal day 4 (PND4) rat pups, as the amount of OPEs within the diet mixture is relative to its composition in Canadian household dust.

Methods: Male and female adult Charles River Sprague-Dawley rats were exposed to OPE mixtures for 70 and 21 days, respectively. The OPE doses were determined to be 10x, 1,000x, and 30,000x the relative human exposure. The progenies were exposed to OPEs both gestationally (~21 days) and lactationally (4 days). At least 2 of each sex from each litter were sacrificed and processed at PND4 for skeletal staining using Alizarin red and Alcian blue. The samples were analyzed and compared against a reference sample to examine any abnormalities in ossification.

Results: At PND4, there is no significant effect of OPEs on the number of pups with abnormal ossification between the control and treatment groups. High doses of OPEs, at concentrations 30,000x of relative human exposures, showed a significant increase in the severity of delay of ossification at the middle phalanx of PND4 pups.

Limitations: Due to the limitation of small sample sizes (litter n=6-7) and a wide variance in data, there is no clear evidence on whether OPE exposure induces greater incidences of abnormal ossification in the digits of PND4 pups.

Conclusion: There is a delay in ossification from OPE exposure at the high dose (30,000x).

Introduction

Flame retardants (FRs) are ubiquitous within the environment; they are present in electronics, various textiles, and surface coatings of flammable household items, such as synthetic Christmas trees. A common type of organobromine compound is polybrominated diphenyl ether (PBDE) and these so-called brominated fire retardants (BFRs) dominated the market in the 1990s. However, BFRs were banned in both the European Union and United States in the early 2000s due to its proven threat to public health. Researchers have shown that BFRs induce liver toxicity and thyroid hormone dysfunction in rat models. (1) Furthermore, gestational studies have also found that exposure to BFRs in doses that approximate high human exposures (0.06 mg/kg/day) cause first-generation offspring (F1) phenotypes of fused digits and delayed ossification in the sternebrae in rats. (2) After BFRs were removed from the market, organophosphate esters (OPEs) have become the mainstream replacement chemical in fire retardants. OPEs are a class of organophosphorus chemicals that are added to flame retardants, which work by creating a charred phosphoric acid surface when encountered with a flame. The charred surface will then act like a solid shield between the fire and material, which protects the unburned portion form the fire. Since OPE flame retardants are additive flame retardants, the molecule is not chemically bonded to the polymer, which is why OPEs can be discharge to the environment, such as in house dust. (3)

However, are OPEs a responsible replacement for BFRs? Since then, many studies have investigated the effect of these chemicals on developmental toxicology in rat models. OPEs are reported to induce oxidative stress on mouse Leydig tumor cells *in vitro*, where all 7 OPEs tested (10 μM) significantly increased superoxide production. The same effect was not observed with BDEs, which is present in a majority of BFRs. (4) Previous studies used a murine limb culture model to investigate the effects on key transcription factors (TFs) in endochondral ossification, such as *Sox9*, *Runx2*, and *Sp7*. (5) These results concluded that when compared to the legacy chemical 2,2',4,4'-tetrabromodiphenyl ether (BDE-47), exposure to triphenyl phosphate (TPhP), a common OPE, caused a significant decrease of TF mRNA expression, which can alter the formation of digits. Markers for endochondral ossification were investigated using fluorescent photography in murine limb bud culture, including COL10A1-mCherry and COL1A1-YFP, which are hypertrophic chondrocyte and osteoblast markers, respectively. When the limb buds were exposed to 3 μM of TPhP, expression of both markers were greatly suppressed and the limb buds showed abnormal ossification. (5) These data suggest that OPEs cause a delay during the transition from cartilage to bone in endochondral ossification during fetal development.

This study focuses on the gestational and early lactational effect of OPEs on ossification of the hindlimbs of F1 pups. It is predicted that, as the concentration of OPEs exposure increases, there will be a greater delay in bone formation in endochondral ossification in the digits. The amount of OPEs of each treatment diet is designed to be relative to the concentration

and chemical composition humans are expected to be exposed to in Canadian house dust in mg/kg/day (Table 3). The three treatment groups are 30x, 1,000x, and 30,000x of the relative human exposure concentration. Pups are sacrificed at postnatal day 4 (PND4) for skeletal staining, as delay of ossification is evident at the middle phalanx in the hind paw during this stage. The first sign of ossification at the middle phalanx occurs at gestational day (GD) 19+, where birth of a rat fetus normally occurs on GD 20/21. Between PND0 and PND7, the middle phalanx presents a clearly defined wide band of ossification. (6) Thus, by PND4, any effect from OPEs on bone development will be evident with abnormal ossification of the middle phalanx. During endochondral ossification, mesenchymal cells are first committed to becoming cartilage cells and then condensed into compact nodules before differentiating into chondrocytes. These chondrocytes then rapidly proliferate and become hypertrophic chondrocytes, which concludes the final stages of chondrocyte differentiation before cartilage is replaced by bone. Endocardial ossification occurs in the vertebral column, the pelvis, and the limbs. (7) The combination of Alcian blue and Alizarin red dyes are optimal for staining endochondral ossifications, as they stain for cartilage and bone, respectively. Alcian blue is a cationic dye that binds to the high concentration of glycosaminoglycans (GAGs) and glycoproteins within the cartilage, while Alizarin red acts as an anionic dye that binds calcium in bones. (8) With the use of these two dyes, the stages of endochondral ossification can be clearly distinguished.

The principle results show a significant increase in the incidences of delayed digit formation in pups from the highest dose treatment group (30,000x). This abnormality in digits is more severe as the dose increases across the treatment groups. In addition, the delay of skeletal ossification is rather global. These results imply that OPEs might not be a responsible replacement for PBDE and the search for finding a better market replacement continues.

Materials and Methods

Animals and treatment

Adult Charles River Sprague-Dawley rats (St. Constant, Quebec, Canada) were separated into four groups and each animal was acclimated to either a control, low, medium, or high-dose OPE mixture treatment diet (0x, 30x, 1,000x, and 30,000x the approximate daily exposure to OPEs in Canadian house dust in mg/kg/day). The dose was approximated relative to body mass and food consumption (Table 1). Since the chemical mixture has never been tested on animals before, a broad range of doses were chosen, with the lowest dose being relative to human exposure and the highest dose being that where toxicity is expected. 30x the human exposure level is still relevant, as it is calculated relative to human exposure via dust and there are many other avenues through which humans can be exposed to OPEs.

Treatment group	Low-dose (30x)	Medium-dose (1,000x)	High-dose (30,000x)
OPEs diet mixture dose/kg of bodyweight/day	48.06 µg	1,602 µg	48,060 µg

Table 1. Dose approximation relative to body mass and food consumption.

Female rats were treated with the diet mixture for 30 days, which covers four days of the menstrual cycle, and 21 days of gestation and lactational exposure. Male rats were treated for 70 days to cover an entire cycle of spermatogenesis and epididymal transit. A male and female rat of the same treatment group were then mated when the female was in proestrus. If the female rats were confirmed to be pregnant by the presence of a vaginal plug the next day, this would be considered as gestation day 0 (GD0) and the birth of F1 pups was marked as postnatal day 0 (PD0). Two pups of each sex were sacrificed at PND4 and fixed in 95% ethanol (EtOH) until further processing.

Skeletal processing and staining

The following protocol was adapted from previous guidelines used in the Barbara Hales lab (unpublished) and the methods have been described by Rigueur and Lyons. (8) The hindlimb sections were submerged in 70°C water using forceps for five minutes to soften the tissue. The skin was incised vertically, proximally to distally, using micro-scissors along the lumbar vertebrae first (Fig. 2A) and was peeled towards the tail. Two cuts were then made from the tarsus to the phalanges on both sides of the feet (Fig. 2B, 2C). This skin was also peeled off, but the skin casing on the tip of the digits was spared. Another two additional incisions were made on the caudal and ventral side of the tail, and approximately 0.25 mm of skin casing was left at the distal end to prevent damage. Any residual organs, fat, skin, and muscle were carefully removed. The muscles were perforated using a 30 1/2 gauge needle to create channels and the skin casing on the digits was loosened with the needle tip. Then the skeleton was fixed in 95% EtOH for at least 12-24 hours before staining.

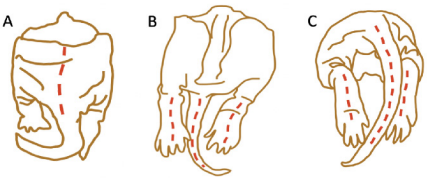

Figure 1. Sites of skin incision for skeletal preparation on the hindlegs of PND4 pup. (A) Dorsal view. (B) Ventral view. (C) Posterior view.

The staining solution contained 1.5 mL of Alcian blue, 1 mL of Alizarin red, 1 mL of glacial acetic acid, and 16.5 mL of 70% EtOH for a 20 mL vial. The skeletons were incubated at 37°C for 32 hours and washed with distilled water three times to remove the leftover dye. The stained skeletons were then left in 0.5% potassium hydroxide (KOH) for approximately five days at room temperature or until the soft tissues were mostly clear.

After the soft tissues were cleared, the skeletons were then washed three times with distilled water and placed in 2:1:1 solution (Table 2) overnight for the final stage of the clearing. For long-term storage, the skeletons were store in 1:1 storage solution (Table 2) until analysis.

1:1 solution	2:1:1 solution
1 part 70 % EtOH 1 part glycine	2 parts 70 % EtOH 1 part glycine 1 part benzyl alcohol

Table 2. Ratio of components in 1:1 and 2:1:1 solution.

Skeletal examination

Each pup was examined by comparing the ossification levels of all parts of the hindlimbs against reference samples (Fig. 2). The different ossification levels were categorized as normal, faint, or no ossification (Fig. 3A, 3B, and 3C respectively). All examinations of the pups were done by observation in the vial. In some cases, if the skeleton disintegrated during the washing or processing steps, it was poured out onto an agarose gel to be sorted and examined instead.

Data analysis

Proportions calculated relative to treatment groups (n=fetus) were analyzed using a 1-tail Fisher's exact test. Proportions calculated relative to litter size (n=litter) or paw number (n=paws) were assessed using a 1-tailed Mann-Whitney U test in GraphPad Prism 8.

Results

Middle phalanx ossification morphology

Representative samples on the degree of ossification at the middle phalanx were photographed as references. Alizarin red and Alcian blue stains for bone and cartilage, respectively. Blue-tinted areas on the paw represent cartilage and dark purple areas represent stained bones, which are considered to be positive stains by Alizarin red. An adequately developed middle phalanx at PND4 presented a thick band of ossified center that was stained positive by Alizarin red (Fig. 2A). In contrast, the ossification center at the middle phalanx can also be entirely absent (Fig. 2B), or faintly ossified (Fig. 2C).

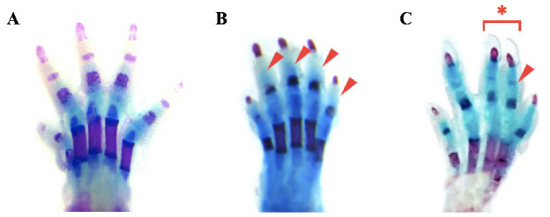

Figure 2. Effects of gestational and early postnatal exposure to OPE mixtures on the hind paw phalangeal ossification and formation. (A) Sample of normal ossification of the PND4 pup hind paw. (B) No ossification of the middle phalanx at digits 2,3,4, and 5. (C) A case of faint ossification of the middle phalanx at digits 2,3, and 4, as indicated by the *, with no ossification at digit 5.

Exposure to OPE mixtures on hind paw phalangeal ossification and formation

To determine the effect of gestational and early postnatal exposure to OPE mixtures on hind paw ossification, the proportion of affected pups with delayed ossification was calculated relative to the litter size (n=litter) or to the treatment group (n=pups). Affected pups included both individuals with either faint or no ossification of the middle phalanx in any digit. It is important to normalize proportions of abnormalities within litters, as the dams rather than the pups were acclimated to different concentrations of OPE diet mixtures.

Once the proportions were calculated relative to litter size (Fig. 3A), all treatment groups showed a large variance in the data, including the controls. All groups, except the low-dose group, had at least one litter with an abnormal ossification proportion of one, and all groups had at least one litter with an abnormal ossification proportion of zero. The data did not indicate any significant differences between the three treatment groups and the controls. Similarly, with the proportion calculated relative to the pups (Fig. 3B), the bars represent the proportion of pups that have delayed ossification of the middle phalanx in each group, and again, the data did not support a significant difference between the three treatment groups and the controls.

Effect of exposure to OPE mixtures on the ossification of the second and fifth digits in the hind-paws

In rats, the pattern of ossification of the phalanx follows the pattern of development in the metacarpals, which is in the order of 3,4,2,5. (6) Thus, reduced ossification at the middle phalanx of the second and fifth digits is a strong indication of delayed bone development.
The proportion of affected pups with delayed ossification was calculated as the number of individuals with no ossification in digits 2 and 5 relative to litter or pups. When the proportion was calculated relative to litter, the controls lost most of its variance in data compared to Fig. 3A. Only the medium and high-dose groups had litters with a proportion above 0.4, whereas in Fig. 3A, three out of four groups had a data range from 0 to 1. When the proportion was calculated relative to the number of pups (Fig. 3D), there was no significant difference between the low, medium, and high-dose groups when compared to the controls.

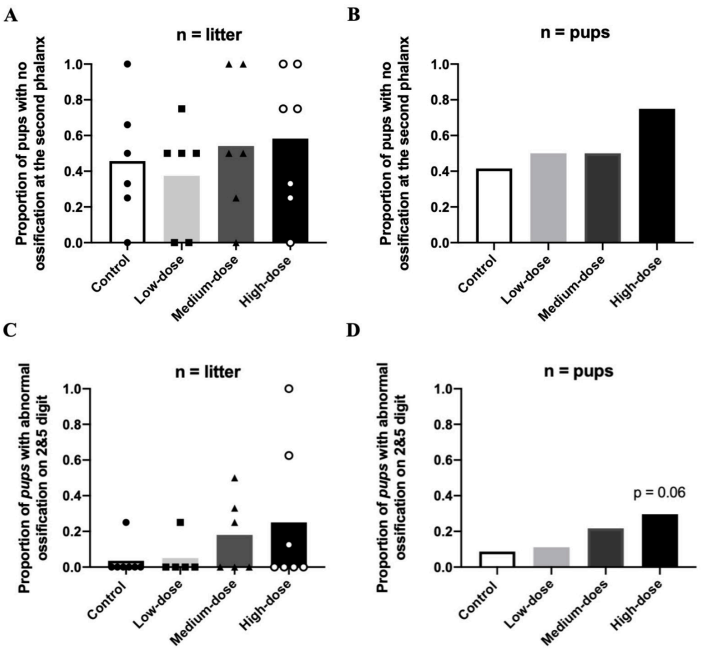

Figure 3. Proportions of pups with abnormal ossification per litter or per treatment group. (A) Proportion of pups showing abnormal ossification at the middle phalanx relative to the total number of pups in a litter (n=7,6,6, and 7 for control, low, medium, and high-doses, respectively). The bars represent the mean value. (B) Proportion of pups showing abnormal ossification at the middle phalanx relative to the total number of pups examined per treatment group (n=23,18,23, and 27 for control, low, medium, and high-doses, respectively). (C) Proportion of pups showing no ossification at the middle phalanx in both the second and fifth digit relative to the total number of pups examined in its litter (n=7,6,6, and 7 for control, low, medium, and high-doses, respectively). Bars represent the mean value of proportions. (D) Proportion of pups with no ossification of the middle phalanx in both the second and fifth digit is calculated relative to the total number of pups examined per treatment group (n=23,18,23, and 27 for control, low, medium, and high-doses, respectively).

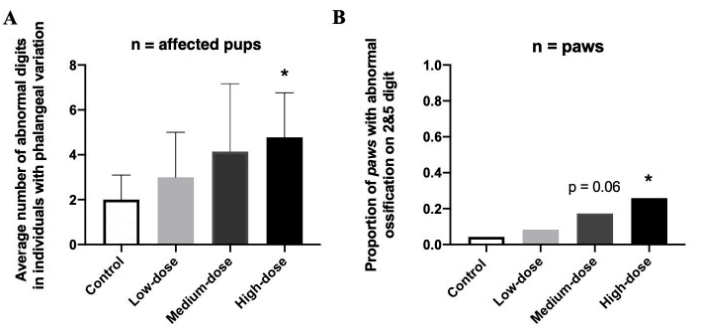

Figure 4. Number of abnormal digits per affected individual and proportion of affected paws with abnormal ossification on digits 2 and 5 in different treatment groups. (A) Average number of digits with abnormal ossification in the middle phalanx in affected PND4 pups. The error bars represent standard deviation (n=6,3,7, and 9 for control, low, medium, and high-doses, respectively). The * indicates $p<0.05$ when compared to the controls. (B) Proportion of paws showing no ossification at the middle phalanx in both the second and fifth digits. The bars represent the proportion of paws affected relative to the total number of paws examined (n=46,36,46, and 54 for controls, low, medium, and high-doses respectively). No error bars are present, as there is only one data point for proportion per group. The * indicates $p=0.009$, $p<0.05$ when compared with the controls.

High-dose treatment group causes severer delay in endochondral ossification

Within the affected individuals, the number of digits with no ossification at the middle phalanx was varied. The average number of digits affected per pup in the subgroup of affected pups was calculated to determine the dose-dependent effect of OPEs on the middle phalanx (Fig. 4A). The average number of affected digits increased as the dose of OPE mixture increased. Despite the large error bars, a significant increase was observed in the high-dose group when compared against the controls ($p=0.0114$, $p<0.05$, Mann-Whitney test). Although there is an upward trend of severity as the dose increases, the dose-dependent relationship was not significant in a 1-way ANOVA test (Fig. 4).

The proportion of pups with no ossification at the middle phalanx of digits 2 and 5 was calculated again, this time relative to the number of paws (Fig. 4B). There was a significant increase in the high-dose treatment group when compared against the controls ($p=0.009$). Although both the low and medium-dose groups had a higher proportion of affected pups than the controls, the differences were not significant. The proportions were also calculated within both sexes, however, no statistical significance was found.

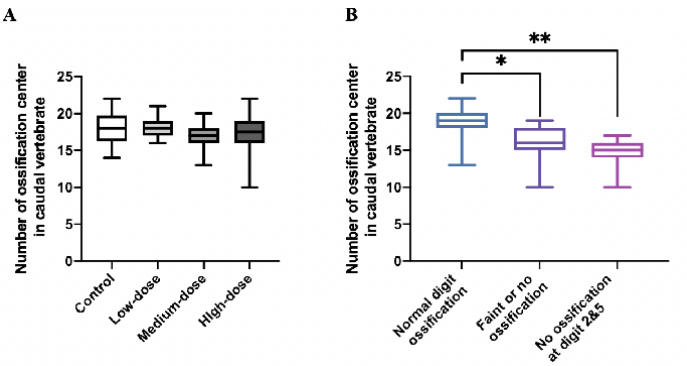

Figure 6. Correlation between middle phalanx ossification and caudal vertebrate length (* $p<0.0001$. ** $p<0.0001$). (A) Box and whisker graph of caudal vertebrae length in the control and the three treatment groups. The error bars are maximum to minimum (n=23,18,23, and 27, respectively). (B) Box and whisker graph of three different subgroups: normal digit ossification, faint or no ossification, and no ossification at digits 2 and 5. Medians are 19, 16, and 15, respectively. The error bars are maximum to minimum (n=42, 44, 15, respectively).

Abnormal ossification at the middle phalanx correlates with lower caudal vertebrate length

Other than quantifying the delay in ossification at the middle phalanx in the digits, the number of ossified centers in the caudal vertebrae was also measured to examine the over-all extent of ossification in the PND4 pups. A sample of long caudal vertebrate has around 21 ossification centers (Fig. 5A); in contrast, an abnormally short caudal vertebrae has only 14 ossification centers at PND4 (Fig. 5B). A normal tail length at PND4 has around 20 ossification centers according to past experimental results (unpublished). A single case of abnormal caudal vertebrate ossification was also observed from the high-dose group (Fig. 5C).

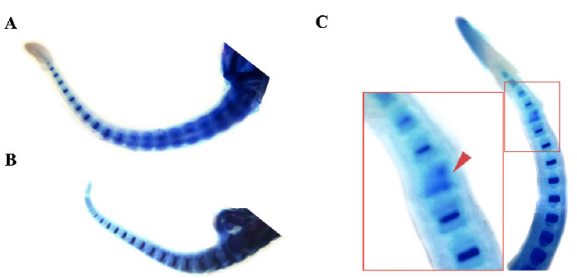

Figure 5. Images of stained caudal vertebrates comparing length and the only case of kinked tail. (A) Normal length tail with around 20 caudal vertebrae. (B) An unusually short tail of around 14 vertebrae; part of the caudal vertebrae above the pubic bone is not shown. (C) Kinked tail and abnormal ossification of the caudal vertebrae.

To investigate the relationship between caudal vertebrae length and the effect of exposure to OPE mixtures, the length was compared across the treatment groups and also between individuals with either normal or abnormally ossified digits. There was no significant difference in the average length of the caudal vertebrae between the three treatment groups and control (Fig. 6A). However, the average caudal vertebrae length in the two subgroups of individuals with affected digits was significantly lower when compared against those with normal digit ossification (*$p<0.0001$, **$p<0.0001$, Fig. 8B). No other abnormalities were observed other than in the digits and caudal vertebrae.

Log transformation of data and Pearson's correlation coefficient analysis

Log_{10} transformation was applied to the x-axis of Fig. 3B, 3C, and 4B to extract a linear correlation between the two variables: proportion and dose. Although the Pearson's R for Fig. 3D and 4B were above 0.83, the p-values were both >0.1, which is not statistically significant as p must be <0.05. Low statistical significance might be due to small sample size. Other

Discussion

In normal hind paw ossification of rats, the middle phalanx's primary ossification center appears by GD19+ in all digits except the thumb (digit 1), which lacks a middle phalanx. (6) Any middle phalanx that has less ossification than the reference sample (Fig. 2A) is considered to have delays in bone development. With these criteria, the proportion calculated in Fig. 3A and Fig. 3B for individuals with delayed ossification included pups with either faint or no ossification of the middle phalanx.

However, the data showed a large variance in all groups (Fig. 3A). This could be caused by the broad definition of "delayed ossification" that was used when calculating the proportion. Pups with faint ossification could be categorized as displaying "normal ossification" since there are natural differences in the speed of normal digit development. Large variances in data could mask out potential statistical significances. With the vast range of proportions in the controls, there was no consistent baseline data that can be used to compare with other treatment groups. To reduce the variance in data, the criteria for "delayed ossification" was refined to include only the individuals with no ossification at the middle phalanx in any digit. However, similarly to Fig. 3A and Fig. 3B, there are variances in the data with no significant difference between the groups (graph and calculation not shown). The criteria was further refined to include only the individuals with no ossification of digits 2 and 5. Since the order of rat digit development is 3,4,2, and 5, any lack of ossification at the last two digits by PND4 is a strong indication of a severe delay in the development of bones.

When the proportions of individuals with no ossification at digits 2 and 5 were calculated relative to the litter (Fig. 3C), the data was less "noisy" than Fig. 3A. All treatment groups had reduced variance, especially in the controls. A clear distinction was observed between the data points of the control and high-dose group, despite the difference not being statistically significant. When calculating the proportions relative to the pups (Fig. 3D), there was no statistical significance between the treatment groups and the controls.

The number of pups examined per treatment group were 23, 18, 23, and 27 for the control, low, medium, and high-dose groups respectively. Due to the small litter size of F0 dams and the limited time for this project, only 6 litters per treatment group were examined. To increase the sample size for data analysis, the proportion of digits 2 and 5 was calculated per paw (Fig. 4B), which increased the sample size to 46, 36, 46, and 54, respectively. By normalizing the proportion relative to the number of paws, it accounted for the severity of the effect of OPEs, which was proven to be significant in Fig. 4A. In conclusion, the number of pups that have delayed ossification in each treatment group is similar, yet abnormal ossification was present in the high-dose group.

The caudal vertebrae length is correlated with the extent of abnormal ossification at the digits (Fig. 6B). Pups with abnormal digit ossification in the middle phalanx were more likely to have a shorter tail than pups with normal ossification. This data suggests that the delay in ossification of the pups is a global effect rather than location specific. A single case of kinked tail with abnormal caudal vertebrate ossification was observed in the high-dose group (Fig. 5C). At the distal end of the caudal vertebrae, the third last caudal vertebra has an abnormal pattern of ossification: it lacks a single thick band of primary ossification center. The staining pattern is either due to the dispersion of Alizarin red or a dark stain of Alcian blue. The kink in the tail could also be an indication of neural tubal defects during fetal development, but further analysis is required to determine the exact cause of this special case of malformation. Previous studies found that OPEs such as TPhP, t-butylphenyl diphenyl phosphate (BPDP), isopropylated triphenyl phosphate (IPPP), and tricresyl phosphate (TCrP) showed adverse effects, especially on bone ossification. (5, 9) These chemicals constitute approximately 20% of the OPE mixture. Many OPEs found in house dust have not been tested regarding its effect on endochondral ossification and their effects on bone formation are still unknown.

In conclusion, there is no clear effect of exposure to OPE mixtures on hind paw development due to limitations, such as a small sample size. However, a high-dose exposure to OPEs induces a significant delay in the ossification of the middle phalanx in the hind paw. The fact that only the high-dose treatment group was statically significant could either be due to the nature of OPEs, whereby a very high dose is required to affect skeletal development, or variations in the data that masked other statistical significances. Further experiments with a larger sample size could minimize the variances in data to confirm why only the high-dose treatment groups were significant when compared to the control. Since only the hindlimb section was analyzed due to time constraints in this study, fore-limb sections could also be examined for a more extensive and complete analysis.

The ultimate goal of finding a responsible replacement is to replace the toxic legacy chemical with one that is harmless or at least, less toxic. However, the OPEs used in fire retardants that filled the market gap after BFRs were banned are not entirely safe in terms of their effects on endocrine systems. (5) Experimentally, the no-observed-adverse-effect-level (NOAEL) for variations in cervical vertebrae and phalanges of PBDEs is 0.75 mg/kg, and the NOAEL for phalangeal abnormal ossification of OPEs was 1602 mg/kg in this experiment. (1) Since this experiment has a lot of shortcomings regarding small sample size and large variance, it is difficult to fairly compare to BFR's NOAEL, which has been studied more extensively. Based on this study, there is no concrete evidence on whether OPEs are a responsible replacement for PBDEs in fire-retardants, as the data only suggests that OPEs significantly induce abnormal ossification at an extremely high concentration, which may not be particularly relevant to human exposures.

Acknowledgements

I would like to thank Dr. Hales and Aileen Yan for their support that made this project possible. Many thanks to Aimee Katen and Abi Rajkumar, who carried out the rat study and provided the PND4 samples.

% in Mixture	Symbol	CAS RN	Name	Supplier	Purity (%)
0.11%	TXP	25155-23-1	Trixylyl phosphate	ACROS	-
0.19%	BPDP	56803-37-3	t-Butylphenyl Diphenyl Phosphate	Scientific Polymer Products	-
0.48%	CDPP	26444-49-5	Cresyl diphenyl phosphate	Alfa Aesar	94
6.74%	IPPP	68937-41-7	Isopropylated Triphenyl Phosphate	NTP, NIEHS, US	97
4.27%	IDDPhP	29761-21-5	Isodecyl diphenyl phosphate	Scientific Polymer Products	-
0.79%	TnBP	126-73-8	Tri n-Butyl phosphate	Sigma	99
7.31%	TCPP	13674-84-5	tris(chloropropyl) phosphate	AKScientific	90
2.48%	TCEP	115-96-8	tri2-chloroethyl	Sigma	97
59.07%	TBOEP	78-51-3	Tributoxyphosphate	Sigma	94
1.18%	EHDPhP	1241-94-7	Ethylhexydiphenylphosphate	AKScientific	95
3.88%	TPhP	115-86-6	Triphenylphosphate	Sigma	99
5.06%	TDiCPP	13674-87-8	Tris(dichloroisopropyl) phosphate	J&K Scientific	95
8.44%	TCrP	1330-78-5	Tricresyl phosphate	Alfa Aesar	-

Table 3. OPEs mixture composition and general information of chemicals.

References

1. Tung, E. W., Yan, H. , Lefèvre, P. L., Berger, R. G., Rawn, D. F., Gaertner, D. W., Kawata, A. , Rigden, M. , Robaire, B. , Hales, B. F. and Wade, M. G. (2016), Gestational and Early Postnatal Exposure to an Environmentally Relevant Mixture of Brominated Flame Retardants: General Toxicity and Skeletal Variations. Birth Defects Res B, 107: 157-168.

2. Robert G. Berger, Pavine L.C. Lefèvre, Sheila R. Ernest, Michael G. Wade, Yi-Qian Ma, Dorothea F.K. Rawn, Dean W. Gaertner, Bernard Robaire, Barbara F. Hales,Exposure to an environmentally relevant mixture of brominated flame retardants affects fetal development in Sprague-Dawley rats,Toxicology, Volume 320, 2014, Pages 56-66,

3. van der Veen, I; de Boer, J (2012). "Phosphorus flame retardants: Properties, production, environmental occurrence, toxicity and analysis". Chemosphere. 88 (10): 1119–1153. doi:10.1016/j.chemosphere.2012.03.067.

4. Schang G., Robaire B., Hales B. F. (2016). Organophosphate flame retardants act as endocrine-disrupting chemicals in MA-10 mouse tumor Leydig cells. Toxicol. Sci . 150, 499–509.

5. Han Yan, Barbara F Hales, Effects of Organophosphate Ester Flame Retardants on Endochondral Ossification in Ex Vivo Murine Limb Bud Cultures, Toxicological Sciences, Volume 168, Issue 2, April 2019, Pages 420–429,

6. Patton, J. T., & Kaufman, M. H. (1995). The timing of ossification of the limb bones, and growth rates of various long bones of the fore and hind limbs of the prenatal and early postnatal laboratory mouse. Journal of anatomy, 186 (Pt 1)(Pt 1), 175–185.

7. Gilbert, S. (2000). Developmental biology. 6th ed. Sunderland: Sinauer Associates.

8. Rigueur, D., & Lyons, K. M. (2014). Whole-mount skeletal staining. Methods in molecular biology (Clifton, N.J.), 1130, 113–121.

9. National Center for Biotechnology Information. PubChem Database. Tri-O-cresyl phosphate, CID=6527, https://pubchem.ncbi.nlm.nih.gov/compound/Tri-O-cresyl-phosphate (accessed on Feb. 13, 2020)

Submitted: 01/13/20

Research Article

[1]Department of Psychology, McGill University, Montreal, QC, Canada

Keywords

Ethnocentrism, evolution, computer simulation, cooperation, reciprocity

Email Correspondence

ruo.feng@mail.mcgill.ca

Ruo Ying Feng[1]

Implications of Reciprocity in the Evolution of Ethnocentrism and Cooperation

Abstract

Background: Ethnocentrism is defined as an individual's tendency to favor in-group members at the expense of out-group members. Recent computer simulations have studied its evolution by modelling cooperative and defective behaviours in a Prisoner's Dilemma framework.

Methods: This paper introduces reciprocity to the study of ethnocentrism and extends Hammond and Axelrod's agent-based model by simulating the effects of five new genotypic strategies. (1)

Results: In stable-state outcomes, although ethnocentrism still dominates, moderate ethnocentrism (in-group cooperation and out-group reciprocity) is more frequent than humanitarianism and is by far the most adaptive out of all reciprocal strategies. Because it is the only reciprocal strategy that cooperates with in-group members, we conclude that it is thanks to in-group cooperation that moderate ethnocentrism is successful, which confirms previous research findings. Additionally, throughout early and late evolutionary patterns, we see that moderate ethnocentrism benefits and suffers from the characteristics of both ethnocentrism and humanitarianism, which may explain why ethnocentrism still emerges as the dominant strategy overall.

Conclusion: The strengths of the present model lie in its ability to abstractly model reciprocal behaviours in the study of ethnocentrism and may be more externally valid than Hammond and Axelrod's original agent-based model. (1) However, this model does not take in account other factors that play a role in human decision-making, such as social context, learning, or development, which could be topics of future computational simulations on ethnocentrism.

Introduction

Although biological evolution is based on the competition for resources between individuals, cooperative behaviours are prevalent in human societies. From hunter-gatherer societies to modern civilizations, cooperation is the decisive organizing principle for our survival. (2) However, this cooperation is not universal. Ethnocentrism, which is defined by the tendency to favor in-group members at the expense of out-group members, is an illustration of this selective cooperation. (3) Several studies have shown that humans have a strong predisposition towards ethnocentric behaviours. For instance, individuals favor in-group members even when group definitions are trivial and arbitrary (e.g. color preference, shirt type), as shown by research done with the minimal group paradigm. (4, 5, 6) Moreover, this in-group bias has been largely identified as a universal and implicit phenomenon, as effects have been found cross-culturally (7) and preconsciously. (8, 9, 10)

In research fields such as evolutionary biology (11) and experimental social psychology, (12) many issues may be difficult or even impossible (e.g. evolutionary patterns of different species) to investigate experimentally. As such, computer simulations have been used to understand their theoretical underpinnings and to predict the outcomes of complex processes, while abstracting away irrelevant details and focusing on essential principles. Similarly, recent studies have applied this methodology to model the evolution of ethnocentrism. (1, 3, 13) In these simulations, the Prisoner's Dilemma game is often used to embody cooperative and defective behaviours between individuals abstractly. Before presenting our simulation and that of Hammond and Axelrod, (1) in addition to Schultz *et al.*'s simulations (13) upon which we will be extending, we will briefly review the Prisoner's Dilemma framework.

The Prisoner's Dilemma framework

Widely considered as a classic framework to study mutual cooperation, (13) The Prisoner's Dilemma was originally designed in the 1950s by nuclear strategists. In that context, researchers were interested in how humans make decisions of cooperation, which involved not sending a missile, or defection (i.e., not cooperating), which involved launching a missile, when they did not know how the opposing party would respond. When experimental psychologists became interested in this paradigm, it was reframed to be about two prisoners who have been arrested for a crime. Defection involves confessing and cooperation involves keeping silent. If they confess while their partner does not, they are freed from all charges. However, the dilemma that the prisoners face is that while each of them is better off if they confess, the outcome of both confessing is worse than if they both kept silent. (14)

In the classic version of the game, two autonomous agents A and B each make a decision to cooperate or defect against one another. The cost to cooperate is c=0.01 while the benefit of receiving cooperation is b=0.03. As such, the outcome of an interaction for an agent is defined as $O=b-c$. If both agents cooperate, they will each receive an outcome of O=0.02. If only agent A cooperates but agent B defects, agent A will receive a negative outcome of O=-0.01, while agent B will receive a large positive outcome of O=0.03. If both agents defect, there is neither cost nor benefit and the outcome will be null. A summary of the basic outcomes for agent A can be found in Table 1.

	B Cooperation	B Defection	Mean Outcome
A Cooperation	b-c	-c	0.005
	0.02	-0.01	
A Defection	b	Null	0.015
	0.03	0.00	

Table 1. Basic outcomes of the Prisoner's Dilemma Game for agent A.

Simulating ethnocentrism

As we observe from Table 1, the mean outcome of defecting, O=0.015, is higher than the mean outcome of cooperating, O=0.005. Therefore, the most adaptive behavior for rational agents should be to always defect and the outcome of two defecting agents will be null. This mutual defection is called the *Nash equilibrium*. When running the Prisoner's Dilemma game with species ranging from bacteria to humans, however, participants largely choose to cooperate. (15) From this intriguing observation, Hammond and Axelrod ran their first agent-based computer simulation. (1)

Although the *Nash equilibrium* predicts that rational agents should always choose defection, the original simulation found cooperation to dominate defection in 74% of interactions. The prevalence of ethnocentric strategies explains this finding, meaning in-group cooperation and out-group defection, which appears in 76% of agents. (1) Additionally, in Schultz *et al.*'s study, we see an early stage of humanitarianism dominance, in which an agent always cooperates. (13) Recent studies have also simulated reciprocity in an iterated Prisoner's Dilemma Game with tit-for-tat agents, who replicate the previous response of an opposing agent. (16, 17) In Baek *et al.*'s study, proportions of cooperation and defection were found to remain invariant despite changing proportions of agents in the population. (16) That said, reciprocity has not yet been studied in the context of in-group and out-group cooperation and defection.

Accordingly, I will be modelling reciprocity in simple abstract agents with a Prisoner's Dilemma game. Following Hammond and Axelrod's agent-based model, group membership is defined by a single arbitrary color and agents will have one of nine types of strategies. These strategies include the four original ones from Hammond and Axelrod's study, (1) as well as five new ones that will simulate reciprocal behaviours towards in-group and/or out-group members. I will be observing the impact of these new reciprocal strategies on stable-state evolutionary outcomes of cooperative and defective behaviours. Additionally, similar to Schultz *et al.*, I will be examining the effects of reciprocal behaviour in earlier stages of ethnocentrism. (13)

Materials and Methods

The Hammond and Axelrod Model

The model used for this paper largely follows Hammond and Axelrod's original simulation. (1) In their simulation, each agent possesses four traits: a tag representing one of four abstract groups, a strategy towards agents with the same group tag (in-group members), a strategy towards agents with a different group tag (out-group members), and a reproductive rate of 0.12. These agents live in a world represented by a two-dimensional 50 by 50 lattice, in which every position in the lattice can be occupied by one agent. When an agent encounters other agents directly adjacent to it in any of four cardinal positions (i.e., north, south, east or west), an interaction occurs in the form of a Prisoner's Dilemma game. The outcome O of these interactions (up to four, one with each potential neighbor) has a direct effect on the agent's reproductive potential. For example, if agent A decides to cooperate with its neighbor agent B while agent B defects, agent A's reproductive potential will be reduced by 0.01, leaving agent A with a reproductive potential of 0.12-0.01=0.11. Moreover, the 50 by 50 world is folded from north to south and from east to west to ensure that all agents have an equal number of potential neighbors.

As mentioned earlier, each agent possesses a strategy (cooperate or defect) toward other agents with the same group tag (in-group members) and another strategy toward agents with a different group tag (out-group members). As such, four genotypic strategies emerge: selfish, traitorous, ethnocentric and humanitarian. A summary of these four strategies can be found in Table 2.

Strategy	In-group	Out-group
Selfish	Defect	Defect
Traitorous	Defect	Cooperate
Ethnocentric	Cooperate	Defect
Humanitarian	Cooperate	Cooperate

Table 2. Four genotypic strategies in Hammond and Axelrod's simulation.

To assess which strategy dominated at stable-state outcomes, Hammond and Axelrod simulated 2000 evolutionary cycles. At each cycle, four stages occur:

1. Immigration: new agents are created according to the immigration rate of 1 (one new agent per cycle). All agent characteristics are randomized, including group tag, genotypic strategy and lattice placement. Reproductive rate is set to 0.12.

2. Interaction: each agent has a chance to play a game of Prisoner's Dilemma with its four possible neighbors, and the outcome affects its reproductive potential. As noted previously, cooperating with an agent will reduce reproductive potential by 0.01, while receiving cooperation increases reproductive potential by 0.03. If all agents decide to defect, reproductive potential will remain unchanged.

3. Reproduction: after being placed in a random order, each agent has a chance to reproduce according to its reproductive potential. This entails creating an offspring in an adjacent empty position in the lattice who will inherit all traits of the parent, with a mutation rate of 0.005 for each trait (i.e. group tag, in-group strategy and out-group strategy).

4. Death: each agent has a death rate of 0.1, which would lead to its removal from the lattice.

The Reciprocal Model

For the purposes of the present study, all parameters (lattice size, number of group tags, cost and benefit of cooperation, reproduction rate, mutation rate and death rate) are kept the same except for the addition of reciprocal behavior, which increases the total number of genotypic strategies (code is available in Appendix B). In the context of the model, reciprocity is defined as an agent replicating the opposing agent's behaviour. Specifically, if the opposing agent cooperates, the reciprocal agent will also cooperate. Similarly, if the opposing agent defects, the reciprocal agent will also defect. If two reciprocal agents interact, the interaction ends early and results in no outcome. As such, agents will have one of three behaviours against in-group or out-group members: defect, cooperate or reciprocate. Thus, nine genotypic strategies are formed. The first four are repeated from Hammond and Axelrod's simulation: selfish (always defect), traitorous (in-group defection and out-group cooperation), ethnocentric (in-group cooperation and out-group defection) and humanitarian (always cooperate). (1) With the addition of reciprocity, five new strategies emerge: moderate selfish I (in-group defection and out-group reciprocity), moderate selfish II (in-group reciprocity and out-group defection), moderate traitorous (in-group reciprocity and out-group cooperation), moderate ethnocentric (in-group cooperation and out-group reciprocity), and universal reciprocal (reciprocating regardless of group). The term "moderate" for these new strategies reflects that reciprocity can be seen as a milder way to defect opposing agents. The behavioral outcomes of all nine strategies are detailed in Table 3.

To study the effects of reciprocity on both stable-state outcomes and earlier stages of evolution, we ran 23 simulations of different worlds with 2000 evolutionary cycles each. We first analyzed the stable-state results of mean strategy frequencies averaged across the 23 worlds after the 2000th cycle. Then, similar to Shultz *et al.*'s method, (13) which also extends Hammond

and Axelrod's original study, the number of agents in each of the nine strategies were tabulated after each cycle, and two separate chi-square tests were performed with critical values at the p=0.01 level. The first chi-square test assessed whether one strategy dominated over the others overall. If the result was significant, a second chi-square test was performed on the two most frequent strategies to assess which one of them was dominant.

Strategy	In-group	Out-group
Selfish	Defect	Defect
Traitorous	Defect	Cooperate
Ethnocentric	Cooperate	Defect
Humanitarian	Cooperate	Cooperate
Moderate selfish I	Defect	Reciprocate
Moderate selfish II	Reciprocate	Defect
Moderate traitorous	Reciprocate	Cooperate
Moderate ethnocentric	Cooperate	Reciprocate
Universal reciprocal	Reciprocate	Reciprocate

Table 3. Nine genotypic strategies in the Reciprocal Model.

	Mean frequency	SD
Selfish	44.56522	21.090338
Traitorous	22.73913	13.705398
Ethnocentric	781.34783	139.478121
Humanitarian	261.60870	79.285869
Moderate selfish I	21.00000	12.336200
Moderate selfish II	26.21739	8.284468
Moderate traitorous	18.69565	13.596239
Moderate ethnocentric	402.26087	122.326767
Universal reciprocal	19.13043	11.442760
	Df	8
	F-value	347
	p-value	<0.001

Table 4. Mean genotypic strategy frequencies and standard deviation after 2000th cycle.

Results

Stable-state outcomes

Mean genotype strategy frequencies across the 23 simulated worlds after the 2000th evolutionary cycle are presented in Figure 1, with error bars indicating plus or minus 1 standard deviation of the mean. Table 4 reports these means and standard deviations in more detail, as well as the results of a simple linear regression summarized with an ANOVA. A large main effect of strategy was revealed, $F(8)=347$, $p<0.001$. Then, to distinguish which genotypic strategies were significantly more frequent than others at stable-state outcomes, Tukey Honest Significant Differences were computed between each pair-wise comparison of strategy. The results of these tests can be seen in Appendix A. Similar to Hammond and Axelrod's findings, (1) ethnocentrism (in-group cooperation and out-group defection) emerged as the dominant strategy and was significantly more frequent than all other strategies (Mean=781.35, SD=139.48). The second significantly most frequent strategy was moderate reciprocity (in-group cooperation and out-group reciprocity) with a mean of 402.26 and standard deviation of 122.33. No other reciprocal strategies were significant. Lastly, humanitarian strategy was the third significantly more frequent strategy (Mean=261.61, SD=79.29), while selfish and traitorous strategies were both not significant.

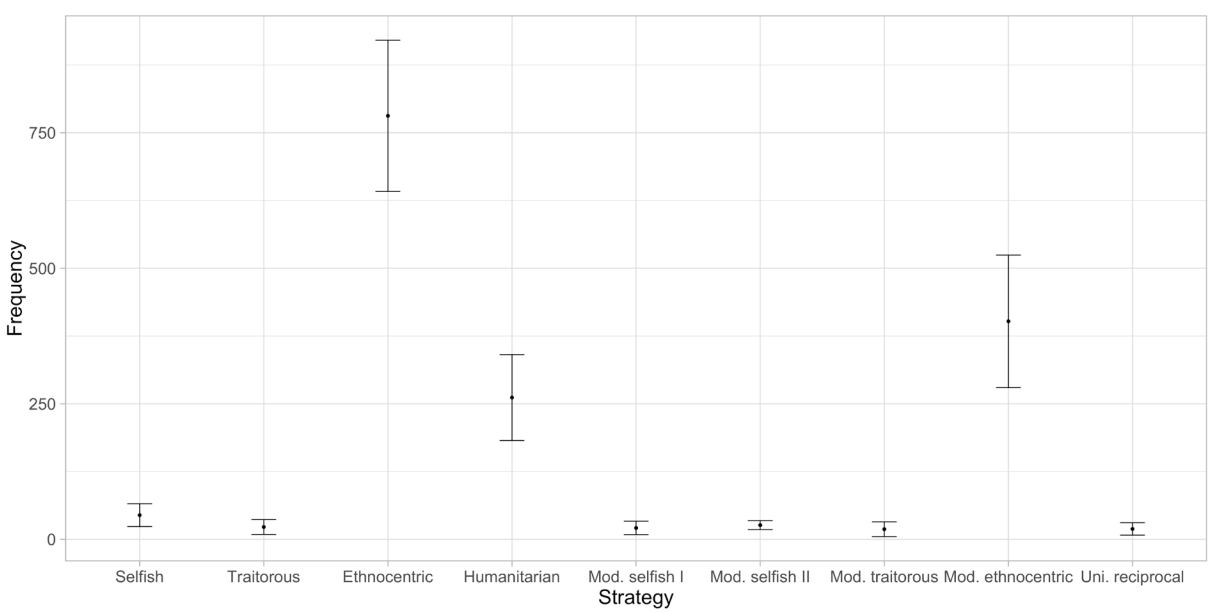

Figure 1. Mean genotypic strategy frequencies after 2000th cycle.

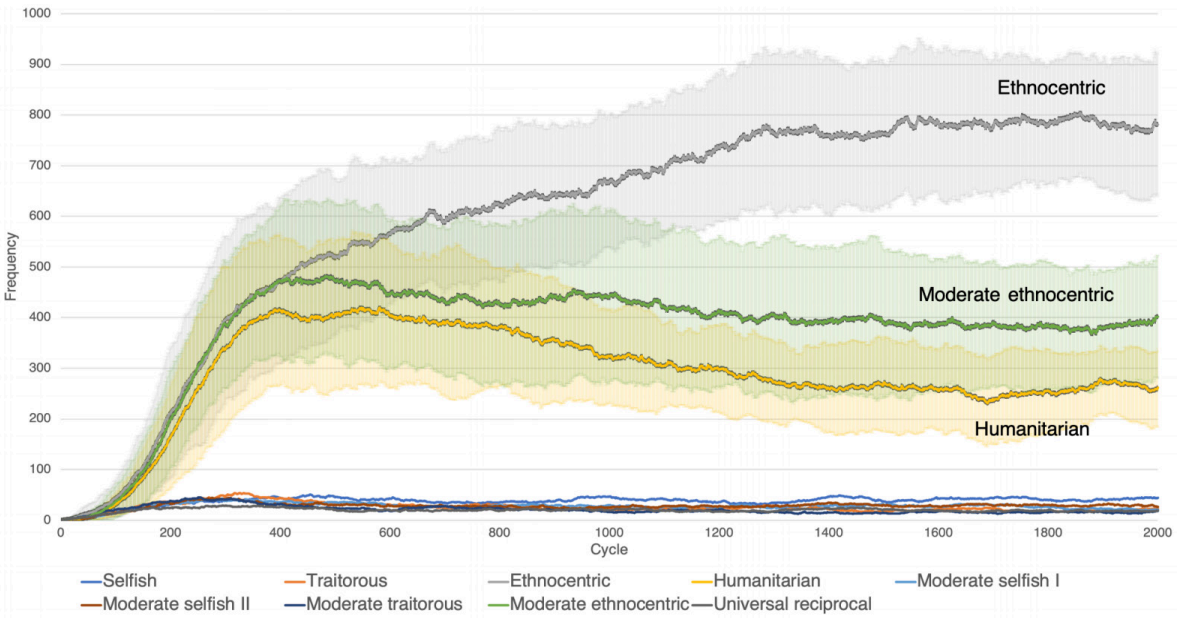

Figure 2. Mean genotypic strategy frequencies throughout 2000 cycles.

Stages of evolution

In Figure 2, mean strategy frequencies across the 23 simulated worlds are plotted for each evolutionary cycle, with error bars indicating plus or minus 1 standard deviation of the mean. As mentioned earlier, strategy frequencies were tabulated after each cycle and two separate chi-square tests were performed with critical values at the p=0.01 level. The critical values for overall strategy dominance and dominance between the top two strategies were $\chi^2_{crit}(8)=20.09$ and $\chi^2_{crit}(1)=6.64$ respectively. Many strategies' evolutionary patterns resemble the findings of Schultz *et al.* (13) However, the inclusion of reciprocal behaviors adds interesting findings to these patterns. Indeed, there is early competition between ethnocentric, humanitarian and moderate ethnocentric (in-group cooperation with out-group reciprocity) strategies. Although overall chi-square tests reach significance from cycle 85 ($\chi^2(8)=20.25$, $p<0.01$) until the end of the simulations, ethnocentric strategy only statistically dominates over the second most frequent strategy (moderate ethnocentrism) from cycle 573 ($\chi^2(1)=7.11$, $p<0.01$). Moderate ethnocentrism and humanitarianism continue to compete until the latter half of the simulations, when moderate ethnocentrism eventually significantly dominates over the latter from cycle 865, ($\chi^2(1)=6.64$, $p<0.01$). Moreover, moderate ethnocentrism even statistically dominated ethnocentrism over multiple cycles in 13 out of the 23 simulated worlds. Lastly, similar to the previous results of stable-state outcomes, no other reciprocal strategies were notable.

Discussion

The dominance of in-group cooperation

These simulations confirm many findings from previous ethnocentrism studies and strengthen Hammond and Axelrod's original finding that evolution favors in-group cooperation due to its positive effect on reproductive potential. (1) As seen in stable-state mean strategy frequencies in Fig. 1 and Table 4, the three most frequent genotypic strategies are ethnocentrism, moderate ethnocentrism and humanitarianism. Interestingly, they are the only three strategies that involve in-group cooperation, as seen in Table 3. Similarly, moderate ethnocentrism (in-group cooperation and out-group reciprocity) is the only strategy that achieves notable evolutionary outcomes, while also being the only reciprocal strategy that involves cooperation with in-group members. As such, this indicates that the reason for its dominance over other reciprocal strategies is not due to its reciprocal behaviour towards out-group members, but its cooperative behaviour towards in-group members, which has positive effects on reproductive potential. Much like ethnocentrism, this strategy allows for selective cooperation, which increases in-group members' fitness without extending it to free riders who only defect. As such, with mechanisms such as in-group favoritism, keeping offspring close, and environmental viscosity, strategies that involve in-group cooperation dominate. (13)

The poor outcomes of other reciprocal strategies further confirm that in-group cooperation is by far the most adaptive evolutionary behaviour. As explained by Schultz *et al.*, selfish and traitorous strategies do poorly due to their inability to cooperate with each other, resulting in lower chances of reproduction. (13) Similarly, all other reciprocal strategies either defect in-group members or reciprocate, except for moderate ethnocentrism. We could also hypothesize that strategies involving in-group reciprocity do better than in-group defection, but there were no significant results found. Therefore, this strengthens the finding that agents need to always cooperate with in-group members to thrive evolutionarily.

Hammond and Axelrod have also argued that ethnocentric agents were vulnerable to in-group selfish agents, or "free riders", who benefit from in-group cooperation without contributing. (1) These free riders are controlled by nearby ethnocentric agents from other groups. By introducing reciprocal behaviors, agents who reciprocate with in-group members like moderate selfish II (in-group reciprocity and out-group defection), moderate traitorous (in-group reciprocity and out-group cooperation) and universal reciprocal agents might have an advantage over ethnocentric or moderate ethnocentrics, because they can suppress their own free riders without reliance on other group members. However, no advantages of in-group reciprocity were found in stable-state outcomes, nor in earlier stages of evolution. This is most likely due to the same reasons why selfish and traitorous strategies fail to dominate: the absence of in-group cooperation. Indeed, even if free riders are suppressed, the lack of cooperation between in-group members results in lower reproductive potential than ethnocentric, moderate ethnocentric or humanitarian strategies. As such, we can infer that in-group cooperation is a more adaptive factor for evolutionary success. Another question we might ask, then, is what distinguishes ethnocentrism from moderate ethnocentrism and humanitarianism, and why does it dominate in stable-state outcomes?

The dominance of ethnocentrism

In Fig. 2, mean frequencies of genotypic strategies throughout all 2000 cycles are represented. From these evolutionary outcomes, two patterns of

competition for dominance emerge: ethnocentrism versus moderate ethnocentrism (with ethnocentrism achieving dominance at cycle 573) and moderate ethnocentrism versus humanitarianism (with moderate ethnocentrism achieving dominance at cycle 865). Starting with the second pattern of competition, the reason why moderate ethnocentrism is able to surpass humanitarianism in later stages may be due to its greater resistance to ethnocentrism. As stated by the direct hypothesis, humanitarians lose to ethnocentrics in later stages of evolution because they cooperate, while ethnocentrics do not. (3) Thus, since moderate ethnocentrics only cooperate with ethnocentrics when reciprocity is present, they are less affected by ethnocentrics in the long run.

As for the competition between ethnocentrism and moderate ethnocentrism, an interesting question is why the latter statistically dominated over ethnocentrism itself in certain cycles in half of the simulated worlds (13 out of 23). An explanation may be that like humanitarianism, moderate ethnocentrism benefits from cooperation early on and is not hurt by encounters with defectors. However, in the same realm of thought, this strategy doesn't benefit from cooperation as much as humanitarianism does due to its reciprocal nature towards out-group members, which might explain why ethnocentrism still dominates in stable-state outcomes. Moreover, the difference in interactions of ethnocentric versus moderate ethnocentric agents with humanitarian agents may also explain this dominance. Indeed, moderate ethnocentric agents always cooperate with both in-group and out-group humanitarians, because it replicates the humanitarians' universally cooperative behavior. On the other hand, ethnocentric agents always defect with out-group members, thus taking advantage of these out-group humanitarians and increasing their reproductive potential at a faster rate than moderate ethnocentric agents. All in all, by analyzing evolutionary patterns, we see that moderate ethnocentrism benefits from the advantages of both ethnocentrism and humanitarianism. Consequently, it also suffers from the disadvantages of both strategies, which might explain why it is not the dominant strategy in stable-state simulations. As such, ethnocentrism still prevails as the most adaptive evolutionary strategy, replicating the findings of Hammond and Axelrod, and Schultz *et al.* (1, 13)

Conclusion

By adding reciprocal behaviours to Hammond and Axelrod's agent-based model of cooperation and defection, (1) we were able to study the implications of reciprocity in evolutionary patterns of ethnocentrism and humanitarianism. By observing both stable-state outcomes and earlier evolutionary patterns, the finding that in-group cooperation is the most adaptive strategy is strengthened by the prevalence of moderate ethnocentrism (in-group cooperation and out-group reciprocity) in overall strategies. As such, the strengths of this model lie in its ability to abstractly model reciprocal behaviours in the study of ethnocentrism. While Hammond and Axelrod's original model did not include reciprocity in order to maximize abstraction, the present model may be more externally valid. In face of a decision of cooperation or defection, humans are not black or white; multiple factors including reciprocity, context, and learning all play in role in our behaviour. Similarly, this model does not consider other factors like context or development, which could be topics of future computational simulations. Future research could also simulate a world in which only reciprocal behaviours exist, which may resemble the world we live in today. Additionally, the implications of reciprocity could be studied in other decisional games such as the iterated Prisoner's Dilemma game, the Closed-bag exchange, or the Friend or Foe game.

Acknowledgements

I would like to thank Dr. Thomas Shultz and the teaching assistants for their valuable feedback and guidance throughout the writing of this paper. Additionally, I would like to express my gratitude to the anonymous reviewers who provided insightful comments on my manuscript.

Appendix A

	Tukey HSD	Adjusted *p*-value
Humanitarian-Ethnocentric	-519.73913	0.00000
Mod. ethnocentric-Ethnocentric	-379.08696	0.00000
Mod. Selfish I-Ethnocentric	-760.34783	0.00000
Mod. selfish II-Ethnocentric	-755.13043	0.00000
Mod. traitorous-Ethnocentric	-762.65217	0.00000
Selfish-Ethnocentric	-736.78261	0.00000
Traitorous-Ethnocentric	-758.60870	0.00000
Uni. reciprocal-Ethnocentric	-762.21739	0.00000
Mod. ethnocentric-Humanitarian	140.65217	0.00000
Mod. selfish I-Humanitarian	-240.60870	0.00000
Mod. selfish II-Humanitarian	-235.39130	0.00000
Mod. traitorous-Humanitarian	-242.91304	0.00000
Selfish-Humanitarian	-217.04348	0.00000
Traitorous-Humanitarian	-238.86957	0.00000
Uni. reciprocal-Humanitarian	-242.47826	0.00000
Mod. selfish I-Mod. ethnocentric	-381.26087	0.00000
Mod. selfish II-Mod. ethnocentric	-376.04348	0.00000
Mod. traitorous-Mod. ethnocentric	-383.56522	0.00000
Selfish-Mod. ethnocentric	-357.69565	0.00000
Traitorous-Mod. ethnocentric	-379.52174	0.00000
Uni. reciprocal-Mod. ethnocentric	-383.13043	0.00000
Mod. selfish II-Mod. selfish I	5.21739	0.99999
Mod. traitorous-Mod. selfish I	-2.30435	1.00000
Selfish-Mod. selfish I	23.56522	0.96153
Traitorous-Mod. selfish I	1.73913	1.00000
Uni. reciprocal-Mod. selfish I	-1.86957	1.00000
Mod. traitorous-Mod. selfish II	-7.52174	0.99999
Selfish-Mod. selfish II	18.34783	0.99208
Traitorous-Mod. selfish II	-3.47826	1.00000
Uni. reciprocal-Mod. selfish II	-7.08696	0.99999
Selfish-Mod. traitorous	25.86957	0.93438
Traitorous-Mod. traitorous	4.04348	0.99999
Uni. reciprocal-Mod. traitorous	0.43478	1.00000
Traitorous-Selfish	-21.82609	0.97580
Uni. reciprocal-Selfish	-25.43478	0.94029
Uni. reciprocal-Traitorous	-3.60870	1.00000

Table 5. Tukey HSD pair-wise comparisons of strategies.

Appendix B

The full code is available at https://osf.io/ah5gm/.

References

1. Hammond RA, Axelrod R. The Evolution of Ethnocentrism. Journal of Conflict Resolution. 2006 Dec 1;50(6):926–36.

2. Nowak MA. Five Rules for the Evolution of Cooperation. Science. 2006 Dec 8;314(5805):1560–3.

3. Hartshorn M, Kaznatcheev A, Shultz T. The Evolutionary Dominance of Ethnocentric Cooperation. JASSS. 2013;16(3):7.

4. Tajfel H. Experiments in Intergroup Discrimination. Sci Am. 1970 Nov;223(5):96–102.

5. Tajfel H, Billig MG, Bundy RP, Flament C. Social categorization and intergroup behaviour. European Journal of Social Psychology. 1971;1(2):149–78.

6. Diehl M. The Minimal Group Paradigm: Theoretical Explanations and Empirical Findings. European Review of Social Psychology. 1990 Jan;1(1):263–92.

7. Fischer R, Derham C. Is in-group bias culture-dependent? A meta-analysis across 18 societies. SpringerPlus. 2016 Jan 22;5(1):70.

8. Dovidio JF, Gaertner SL. Chapter 8 - Stereotypes and Evaluative Intergroup Bias. In: Mackie DM, Hamilton DL, editors. Affect, Cognition and Stereotyping [Internet]. San Diego: Academic Press; 1993 [cited 2020 Feb 27]. p. 167–93.

9. Lamont M, Molnár V. The Study of Boundaries in the Social Sciences. Annual Review of Sociology. 2002;28(1):167–95.

10. Otten S, Wentura D. About the impact of automaticity in the minimal group paradigm: evidence from affective priming tasks. European Journal of Social Psychology. 1999 Dec;29(8):1049–71.

11. Hoban S, Bertorelle G, Gaggiotti OE. Computer simulations: tools for population and evolutionary genetics. Nat Rev Genet. 2012 Feb;13(2):110–22.

12. Ostrom TM. Computer simulation: The third symbol system. Journal of Experimental Social Psychology. 1988 Sep;24(5):381–92.

13. Shultz TR, Hartshorn M, Hammond RA. Stages in the Evolution of Ethnocentrism. Proceedings of the 30th Annual Conference of the Cognitive Science Society. 2008;1244–9.

14. Kuhn S. Prisoner's Dilemma [Internet]. Winter 2019. Metaphysics Research Lab, Stanford University; 2019 [cited 2020 Feb 28]. (The Stanford Encyclopedia of Philosophy). Available from: https://plato.stanford.edu/entries/prisoner-dilemma/#Bib

15. Axelrod R, Hamilton WD. The evolution of cooperation. Science. 1981 Mar 27;211(4489):1390–6.

16. Baek SK, Yi SD, Jeong H-C. Duality between cooperation and defection in the presence of tit-for-tat in replicator dynamics. Journal of Theoretical Biology. 2017 Oct 7;430:215–20.

17. Zheng X-D, Li C, Yu J-R, Wang S-C, Fan S-J, Zhang B-Y, et al. A simple rule of direct reciprocity leads to the stable coexistence of cooperation and defection in the Prisoner's Dilemma game. Journal of Theoretical Biology. 2017 May 7;420:12–7.

Submitted: 01/14/20

Research Article

[1]Department of Psychology, McGill University, Montreal, QC, Canada

Keywords

psychosocial stress; cognitive effort; motivation; avoidance; sex differences

Email Correspondence

noa.givon@mail.mcgill.ca

Noa Givon[1]

The Effects of Psychosocial Stress and Sex Differences on Cognitive Effort Avoidance

Abstract

Background: Recent research suggests stress may affect cognitive performance including memory, executive functioning, decision-making, and task-switching. However, it is unknown whether these effects are aversive or advantageous for effort exertion. This experiment aimed to evaluate the effects of acute psychosocial stress on willingness to exert cognitive control processes in a cognitive-effort-based decision-making task.

Methods: To test this, 40 participants (20 female) in a within-subject, fully crossed, randomized design, were exposed to both a psychosocial stress induction condition (the Trier Social Stress Test; TSST) and a control condition. Subsequently, they underwent the Demand Selection Task (DST) that tests for participants' effort aversion by manipulating switch probabilities in a task-switching paradigm.

Results and Conclusion: The induction of stress did not lead to significant error or accuracy rates, or significant differences in cognitive effort avoidance. Previous research indicated sex differences in response to stress. However, there is a lack of data on sex differences in the avoidance of demanding cognitive processes. Therefore, we assessed sex differences in the DST and found that women were more likely to avoid cognitive effort, choosing the less cognitively demanding cue more often than men.

Limitations: A limitation of this study is the small sample size. Future research should increase the sample size and take individual differences in stress responders, type of stressor, and biases on effort exertion into account.

Introduction

Generally, people prefer taking shortcuts, such as taking the escalator instead of the stairs. Hull's "Law of Less Work" explains this tendency for humans to select actions that minimize effort, stating that when organisms are presented with two or more behavioural options, there is a preference towards the less demanding course of action. (1) While the "Law of Less Work" was intended for physical effort, it was suggested that this could be applied to mental tasks. (2) Kool *et al.* posit that individuals typically avoid situations that carry a high cognitive load. (2) They developed an experimental paradigm confirming this preference for less cognitively strenuous tasks, called the Demand Selection Task (DST) to test their hypothesis. (2) The DST requires participants to choose between two cues associated with higher or lower switch frequencies in a task-switching paradigm. Task-switching requires cognitive control, which is thought to be effortful. (3,4) In the DST, it is assumed that higher switch frequencies require more effort expenditure by the participant. Since the publication by Kool *et al.* (2), further studies have shown there is a general tendency for effort avoidance and that people require large incentives to exert effort. (5-8)

While studies have shed light on cognitive effort avoidance, when and why people exert effort is rarely investigated. One potential mechanism that may influence motivation is stress. Evidence suggests that stress affects cognitive performance, which impacts memory, decision-making, and cognitive control. (9-12) Stress elicits a physiological response that activates the autonomic nervous system (ANS) and the hypothalamic-pituitary-adrenal (HPA) axis. (13) The ANS activity results in elevated blood pressure (BP) and heart rate (HR), whereas the HPA axis regulates the release of cortisol, a major stress mediator in the brain. (14) Kirschbaum and Hellhammer (14) demonstrated that stressful situations, such as public speaking, cause increases in salivary cortisol levels. Cortisol is important in the mediation of how stress affects higher cognitive control processes, presumably through its effects on the prefrontal cortex (PFC). The PFC is responsible for mediating complex cognitive processes that are assumed to be effortful, however, it is highly vulnerable to stress. (15,16) While physiological stressors directly affect the paraventricular nucleus (PVN) of the hypothalamus, psychological stressors pass through the limbic system (PFC, hippocampus, amygdala) before activating the HPA axis, and therefore affect higher levels of processing. (17) Because the PFC possesses a high density of cells with receptors sensitive to neurotransmitters released in response to stress, it is reasonable to hypothesize that decision-making and task-switching processes that rely on the PFC can be directly affected by stress. Under stress, executive functions can impair cognitive performance, altering decision-making and task-switching processes. (18,19) Despite the large body of literature demonstrating a link between cognitive functioning and acute stressors, to our knowledge the effects of stress on effort avoidance have not been examined to date. In this study, our main interest was to examine DST performance in response to stress from the TSST.

Beyond the overarching motivations of this study, we further examined sex differences in the DST in response to stress. Research has shown sex differences in cognitive processes such as spatial skills, verbal skills, and memory. (20-22) However, little is known about individual differences in motivational effort. Recent studies have examined individual differences in response to stress due to the high prevalence of somatic and pathological disorders related to stress, such as anxiety, depression, and pain. (23-26) Since the HPA axis is implicated in many of these disorders that have sex differences, scholars have posited that stress responses may be different in men and women. (27,28) Studies have found sex differences in cortisol levels, BP, and HR in response to stress, showing that men have greater overall physiological responses under acute stressors. (17,29,30-33) These studies state that a potential factor causing these results may be due to hormonal fluctuations related to the menstrual cycle and/or the use of hormonal contraceptives. (31,34,35) Finally, evidence regarding sexually divergent reactions to psychological stressors reveals that women report higher levels of negative affect than men. (31,36-38)

This experiment sought to answer the following questions: First, how does acute psychosocial stress affect cognitive effort-based decision-making, and how do these changes relate to the participants' cognitive performance on the DST? Second, do men and women differ in their response to stress and does this divergence mediate behaviour in the DST? In order to test DST performance in response to stress, we conducted a within-subject, fully crossed, randomized study. We hypothesized that stress would de-

crease participants' willingness to exert effort and impair task-switching performance in the DST by disrupting executive functioning and working memory, causing more effort avoidance. With respect to sex, we speculated that women would have lower accuracy rates and would be more effort averse than men, as psychological stressors are more often activated in women. To induce stress, we employed the TSST. (39) Over a two-day testing period, 40 healthy participants (20 female) underwent both a stress-inducing condition (TSST) and a control condition before completing the DST. To ensure effective physiological and psychological stress induction, we collected subjective mood reports, salivary cortisol samples, BP, and HR.

Methods

Participants

A total of 40 healthy participants, between the ages of 18 and 30, were recruited from a participant pool of students and the community (20 female; age, mean±SEM: 23.43±0.45 years; body mass index, 22.4±0.45 kg/m2). Participants were free of neurological or psychiatric conditions and did not have any current or acute illnesses (GAD, mean±SEM: 1.7±0.36; PHQ, 2.1±0.37). Exclusion criteria included medication intake, smoking, substance abuse, pregnancy, and use of hormonal contraceptives in women. Women were not tested during their menses. In addition, participants were instructed to abstain from exercise, food, and caffeine during the 2 h period before the testing session. Test protocols were approved by the Research Ethics Board of the Psychology Department at McGill University. Participants provided informed consent and received a financial reward of $50 and a $3 bonus in return for their participation.

Demand selection task

In the DST, we presented participants with two abstract patterned choice cues, with a small cue centered between them marking the midpoint. (2) In each run, participants selected one of the two patterns by rolling a mouse cursor over the desired pattern, which revealed a coloured number within a magnitude/parity task-switching protocol (Fig. 1). In this task, if participants saw a purple number, they were asked to judge the magnitude of the numbers by clicking the left mouse button in response to numbers less than 5 and the right button for numbers above 5. If a yellow number was presented, the task switched to a parity protocol, where participants had to click the left or right mouse button for even and odd numbers, respectively. After each run, participants were required to roll the mouse over the midpoint to reset the choice cues. Participants were unaware that the choice cues were associated with either "high demand" or "low demand" trials. For low demand trials, the colour of the number matched that of the preceding trial on 90% of trials. On high demand trials, the colour from the preceding trial was only repeated 10% of the time. The high-demand choice cue required more frequent task-switching and thus carried a greater cognitive load. The DST had a total of 300 trials and was divided into four blocks, with 75 in each block. The individual's accuracy and low demand choice percentages were recorded.

Experimental stress induction

The TSST is a standardized psychosocial laboratory stressor used to induce stress in humans. (39) The test involves an anticipation period, an oral presentation and a complicated mental arithmetic task in front of a panel of two "experts". In this study, participants were told to think of their dream job and convince the panelists why he/she is the perfect candidate. The panel was comprised of two unresponsive experimenters (one male, one female), who were introduced to participants as experts trained in analyzing behaviour and took notes during the speech. Participants were given time to prepare before delivering their 5-minute uninterrupted speech to the experts and were told when to stop. If the speech was less than 5 minutes, the panel responded with standardized comments reminding the subject there was time left, and that they should continue. Following the speech, the judges asked the participant to perform a difficult mental arithmetic task in which they serially subtracted the number 17 from 2043 as fast and as accurately as possible. If an incorrect number was stated, they were asked to restart from 2043, and when participants were too slow, they were asked to increase their speed. In the control condition, participants completed a neutral public speaking task in which the panel of judges was removed. With the room to themselves, participants were asked to speak for 5 minutes on a topic of their choice followed by a simpler arithmetic task of counting upwards in increments of five. To assure participants complied with the instructions, the experimenter listened through the door.

Procedures

Shortly after the participants' arrival, they were exposed to either a stress-inducing condition (TSST) or a control condition before completing the DST. Participants were tested on two days, separated by a one-week interval, lasting for 2.5 h each session. All testing occurred between 1:00 P.M. and 6:00 P.M to control for circadian fluctuations in salivary cortisol levels. We obtained informed consent on the first day of experimentation. Afterwards, baseline measurements of subjective feelings were collected as measured by the Positive and Negative Affect Schedule (PANAS; 40) Questionnaire as well as BP and HR measurements. Saliva samples were taken to control for cortisol level variation. The BP and HR measurements were collected seven times throughout day one (baseline, during TSST, +15, +25, +45, + 60, and +75 after stressor onset). The salivary cortisol and PANAS were collected six times (baseline, +15, +25, +45, + 60, and +75 after stressor onset). The Demand Questionnaire was given to the participants following the TSST. Finally, participants completed the DST approximately 45 minutes after the stress onset, in addition to three other effort and decision-making tasks which will not be further discussed. On the second day, participants underwent the same procedures as day one,

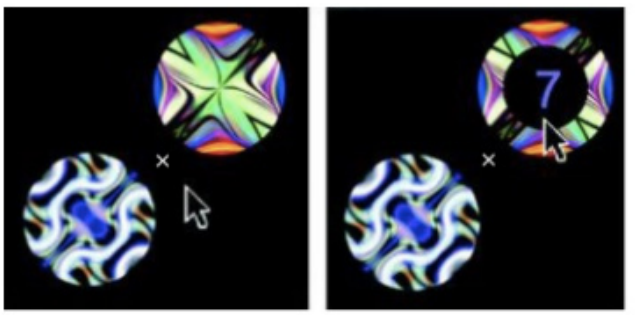

Figure 1. Example of cues in the DST. Participants used the mouse to reveal a coloured number then clicked the left or right mouse button to indicate magnitude (purple) or parity (yellow). One stimulus (high-demand) changed colours more often, demanding frequent task-switching. The other stimulus (low-demand) repeated colours more often, demanding less effort. Cues reset by returning the mouse to midpoint. (2)

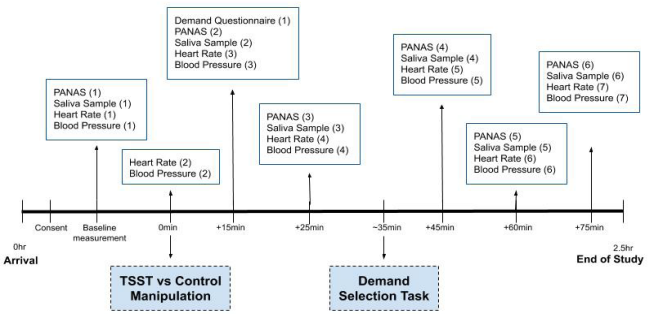

Figure 2. Timeline of the procedure for both days. Participants were exposed to the TSST on one day and the control manipulation on the other day before completing the DST. Time points reflect the number of minutes after stressor onset. PANAS is the Positive and Negative Affect Schedule. Numbers in brackets reflect the number of measurements for the respective stress variables.

but with the opposite TSST condition than they experienced on the day prior (Fig. 2).

Psychological questionnaires

Throughout both days of the experiment, participants reported their subjective affective states. The Positive and Negative Affect Schedule (PANAS; 40) is a self-report questionnaire that consists of two 10-item scales that measure positive and negative affect. The positive affect (PA) scale includes attentive, active, alert, excited, enthusiastic, determined, inspired, proud, interested, and strong. The negative affect (NA) scale includes distressed, upset, guilty, scared, hostile, irritable, ashamed, nervous, jittery, and afraid. Participants rated the degree to which they felt these items on a 5-point scale of 1 (not at all) to 5 (very much). PANAS data were collected six times simultaneously with the physiological measurements (Fig. 2). Using the Demand Questionnaire, ratings of difficulty, unpleasantness, and stressfulness were collected following the TSST. These ratings range from 0 (not at all) to 100 (extremely) in intervals of 10, which resembles the widely validated Visual Analogue Scales (VAS). (41)

Physiological measures

To evaluate the effectiveness of the stress induction, in addition to subjective reports, endocrine and autonomic responses were measured. Six saliva samples were collected per day using Salivette® collection devices (Sarstedt Inc., Rommelsdorf, Germany). (42) The first sample was taken before the TSST, another following the stress or control exposure, and the rest were collected throughout the remainder of the study, after the cognitive tasks. Saliva samples were stored at -18 °C and analyzed for cortisol concentrations using a chemiluminescence immunoassay (IBL Hamburg, Germany). BP and HR data were collected along with the self-report questionnaires and saliva samples, with one additional measurement taken during the TSST/control manipulation, for a total of seven measurements. A LotFancy BP-103H arm BP monitor assessed the participants' diastolic and systolic BP as well as their HR. We recorded the mean of the two consecutive measurements.

Experimental design and statistical analyses

For both the stress condition and the DST condition, the study utilized a within-subject randomized design, in which the two factors, stress induction (stress vs control condition), and demand task (high vs low demand), were fully crossed. Subjective, physiological, and sex parameters were analyzed using a mixed model ANOVA, with the within-subjects factors "condition" (stress and control), "time" of sample collection (systolic and diastolic BP, and HR: seven samples; PANAS and salivary cortisol: six samples), and the between-subject factor "sex" (male and female). Significant main and interaction effects were followed up by post hoc tests to analyze the differences in single time point evaluations. Paired samples t-test were applied to analyze responses in the Demand Questionnaire as well as accuracy and low demand effort choices in the DST of the TSST and control conditions. Data were analyzed using SPSS, Version 24. All analyses were two-tailed, with the significance level set at $p<0.05$, and adjusted by Bonferroni correction for multiple comparisons. Greenhouse-Geisser degrees of freedom adjustment was applied to correct for violations of sphericity.

Results

Cortisol responses to stress

Consistent with previous research, salivary cortisol levels increased significantly following the TSST (main effect of time, $F(1.688,65.844)=12.161$, $p<0.001$, $\eta p2=0.238$; main effect of condition, $F(1,39)=18.99$, $p<0.001$, $\eta p2=0.328$; time x condition interaction, $F(2.32,90.489)=9.022$, $p<0.001$, $\eta p2=0.199$). Cortisol concentration from the samples taken at +15 minutes and +45 minutes of the study were significantly higher than at +60 and +75 time points (both $p<0.001$). Cortisol levels peaked +25 minutes after the stressor onset (third time point) and were significantly higher than all time points after the TSST (all $p<0.021$) (Fig. 3a). Post hoc analyses confirmed that, overall, participants had significantly higher concentrations of cortisol in the stress condition following stress onset (all $p<0.001$). These results indicate stress was successfully induced and was increased at the time participants performed the DST.

Heart rate and blood pressure responses to stress

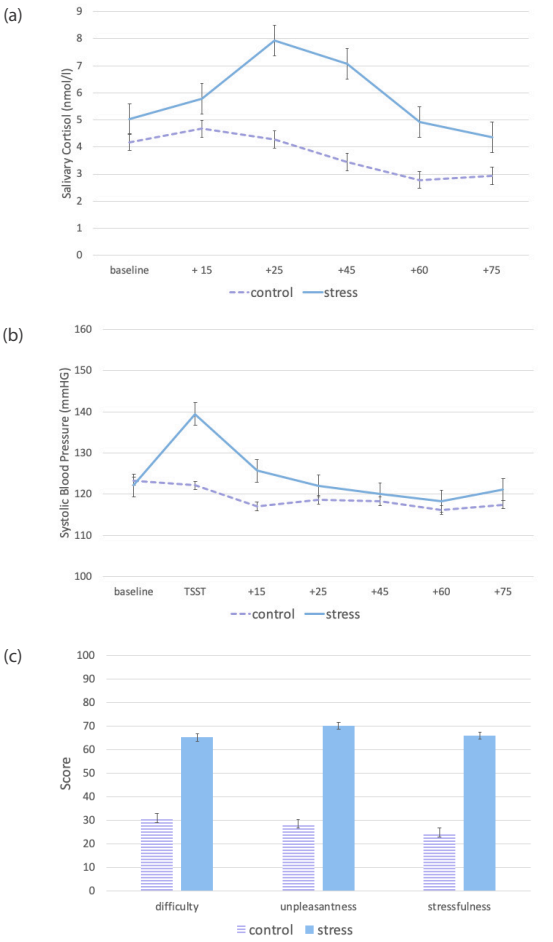

Figure 3. Successful stress induction in response to the TSST. (a) Mean salivary cortisol levels during stress induction and control manipulation. (b) Mean systolic blood pressure for control vs stress condition. (c) Mean scores on ratings of difficulty, unpleasantness, and stressfulness immediately after the TSST vs control manipulation. Error bars represent the standard error of the mean (SEM).

There was a significant increase in blood pressure (BP) and heart rate (HR) in response to the exposure of the experimental stress induction (main effect of time, systolic: $F(3.799,132.965)=14.829$, $p<0.001$, $\eta p2=0.298$; diastolic: $F(3.236,113.262)=9.426$, $p<0.001$, $\eta p2=0.212$; HR: $F(3.236,113.262)=9.426$, $p<0.001$, $\eta p2=0.212$; main effect of condition, systolic: $F(1,35)=22.409$, $p<0.001$, $\eta p2=0.390$; diastolic: $F(1,35)=17.753$, $p<0.001$, $\eta p2=0.337$; time x condition interaction, systolic: $F(6,210)=10.954$, $p<0.001$, $\eta p2=0.238$; diastolic: $F(3.83,134.050)=5.978$, $p<0.001$, $\eta p2=0.146$; HR: $F(3.83,134.050)=5.978$, $p<0.001$, $\eta p2=0.146$). Differences in HR across the two conditions were not significantly different (main effect of condition, $F(1,35)=3.476$, $p=0.071$, $\eta p2=0.090$). Post hoc tests revealed that individuals in the stress condition had significantly higher systolic and diastolic BP and HR during the TSST (all $p<0.001$) and at the +15 min time point for systolic BP only ($p<0.001$) (Fig. 3b). There was no significant change in BP at other time points (systolic: all $p>0.168$, diastolic: all $p>0.189$, HR: all $p>0.125$). Overall, BP and HR peaked during the TSST manipulation, then returned to baseline for the remainder of the experiment.

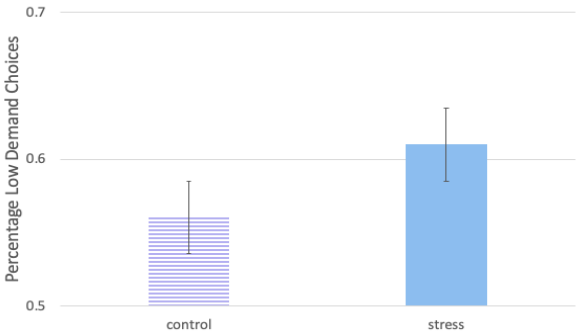

Figure 4. Percentage of participants' choice of the low demand option on the DST in both conditions. Results did not significantly differ by the TSST. Error bars represent the standard error of the mean (SEM).

Subjective feeling responses to stress

Participants in the stress induction showed the expected increase in scores of psychological stress. A paired samples t-test evaluating responses from the Demand Questionnaire demonstrated that participants reported the stress induction as significantly more difficult, unpleasant, and stressful following the TSST compared to the control task (difficulty: t(38)=-7.857, p<0.001, d=1.258; unpleasantness: t(38)=-8.257, p<0.001, d=1.322; stressfulness: t(38)=-9.170, p<0.001, d=1.468) (Fig. 3c). PANAS results revealed greater NA directly after the TSST (main effect of time, F(3.274,121.145)=9.054, p<0.001, ηp2=0.196; time x condition interaction, F(2.963,109.615)=11.256, p<0.001, ηp2=0.233). However, reported feelings of NA did not differ across conditions (main effect of condition, F(1,37)=9.844, p=0.576, ηp2=0.009). Feelings of PA decreased from baseline consistently throughout the study (main effect of time, F(3.568,132.027)=15.687, p<0.001, ηp2=0.298). Participants' PA was not affected by the TSST (main effect of condition, F(1,37)=0.704, p=0.407, ηp2=0.19; time x condition interaction, F(3.235,120.362)=0.364, p=0.795, ηp2=0.1). Post hoc tests for PA and NA were performed to assess differences between the conditions at each time point. While no significance was found for PA throughout the experiment (p>0.235), individuals reported feelings of greater NA at baseline when experiencing the stress condition (p<0.004) and at the time point following the TSST (p<0.001).

DST results

Mean accuracy rates on the DST in both stress and control conditions were 88%, yielding insignificant results, F(1,39)=0.054, p=0.817, ηp2=0.001. Additionally, the rate in which participants chose the low versus the high demand choice did not significantly differ across the two conditions, F(1,39)=3.714, p=0.061, ηp2=0.087, in which the mean of choosing the low demand cue was 56% in the control condition and 61% in the stress condition (Fig. 4).

Sex differences in stress responses

To compare the sex differences in stress responses, we subtracted the baseline from the peak level in each modality. In physiological measurements of salivary cortisol concentration (Fig. 5a) and BP (Fig. 5b), we found that men had significantly greater increases than women, following the TSST (salivary cortisol: t(38)=2.791, p<0.008; systolic: t(38)=2.071, p<0.046; diastolic: t(38)=2.358, p<0.024). For self-reported measures, women reported significantly higher feelings of unpleasantness after the TSST compared to men (t(37)=-2.023, p<0.05) (Fig. 5c). There were no significant sex differences in HR, or reports of NA, difficulty or stressfulness following the TSST (HR: t(38)=0.791, p=0.434; NA: t(38)=-0.057, p=0.955; difficult: (37)=-0.604, p=0.550; stress: t(37)=-0.402, p=0.69).

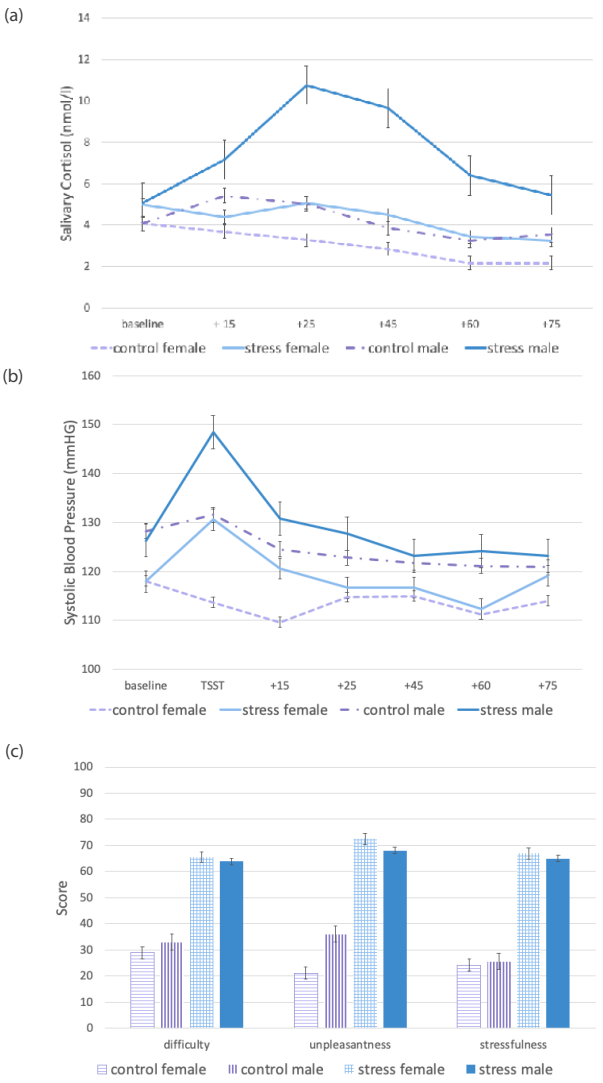

Figure 5. Sex differences in response to the TSST. (a) Mean cortisol concentration for male and female participants across the control and stress conditions. (b) Mean systolic blood pressure between males and females on both conditions. (c) Mean difficulty, unpleasantness, and stressfulness scores in males and females on both conditions. Error bars represent the standard error of the mean (SEM).

Sex Differences in the DST

For the DST, we investigated sex differences in high versus low demand choices under TSST and control conditions. A repeated-measures ANOVA assessing the between-subjects factor of sex revealed that in both conditions, women more often chose the low demand choice (Fig. 6), with a mean rate of 60% in the control condition and 66% in the stress condition, while men chose the low demand choice 52% of the time in the control condition and 56% in the stress condition (main effect of sex, F(1,38)=4.83, p<0.034, ηp2=0.113). There were no significant differences between conditions nor a sex x condition interaction (main effect of condition, F(1,38)=3.626, p=0.064, ηp2=0.087; sex x condition interaction, F(1,38)=0.077, p=0.783, ηp2=0.002). A post hoc test confirmed that across both conditions, women chose the low demand choice more often, 63% of the time, compared to men who chose the low demand choice 54% (p<0.034).

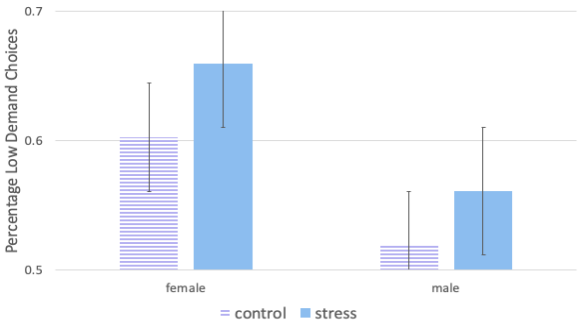

Figure 6. Sex differences in low demand choices in the DST. Error bars represent the standard error of the mean (SEM).

Discussion

Stress is often present during cognitively effortful tasks in our everyday lives. However, stress has been shown to jeopardize executive functions and working memory. (18,19) The purpose of this study was to assess whether the induction of acute psychosocial stress promotes or hinders motivation to exert effort. Participants' efforts were assessed using the DST after experiencing a standardized laboratory psychosocial stressor (TSST) or a control condition. We hypothesized that motivation to exert effort and task-switching performance during the DST would be impaired by stress. Specifically, we predicted that the activation of the HPA axis would affect the PFC, diminishing cognitive performance such that it causes individuals to make errors and avoid using effort. Further, we predicted that stress from the TSST would have greater effects on women, causing more errors and effort avoidance on the DST than in men. The present experiment demonstrated women were more likely to avoid cognitive effort, but stress induced by the TSST did not significantly influence cognitive performance.

Our study confirmed that the TSST elicited significant physiological changes. This was in line with previous research that demonstrated that public speaking tasks enhance stress and NA. (31,36-38) Our results uphold that feelings of NA were significantly greater following the TSST manipulation than at other time points of the experiment and compared to the control condition. Participants regarded the stress induction as more difficult, unpleasant, and stressful. Salivary cortisol levels reached its highest point approximately 25 minutes after participants were exposed to the TSST, which is consistent with previous research stating that cortisol in saliva peaks 20-30 minutes after stressor onset. (43) These physiological and psychological stress responses reflect prolonged HPA-axis activity, indicating successful stress induction.

The results of this study reveal sexually divergent responses to stress, as male participants exhibited greater increases in both cortisol levels and BP following the TSST. These findings are in line with previous research that suggests men exhibit higher cortisol outputs under acute stress. (17,31-33) Additionally, women reported higher feelings of unpleasantness after the TSST compared to men, consistent with previous findings that women subjectively report more NA. (31,36-38) However, the results of this study do not confirm previous research demonstrating impaired task switching performance (44) and decision-making (19,45) while under stress. Nonetheless, we showed that women chose the low demand choice more often than men. Because we controlled for potential hormonal confounds such as pregnancy, hormonal contraceptive use and menses, this finding cannot be attributed to these factors. Biases for effort demand were examined, revealing that there are individual differences in cognitive effort. (46) People may be Demand Avoiders, who avoid more frequent task-switching tasks, or Demand Seekers, who routinely choose them. (46) Future research should look more carefully at these effort biases to allow for a better understanding of why these preferences develop and how these biases affect stress responses.

Our experiment revealed that performance on the DST (accuracy rates and low versus high demand choices) did not vary significantly whether participants were exposed to stress or not. A possible explanation could be due to the differences in the way people respond to the TSST. Studies have shown there are habitually high and low responders to stress, as well as sex differences in responders. A study by Preston *et al.* (19) investigated how stress from the anticipation of public speaking affects performance on the Iowa Gambling Task (IGT), a paradigm to assess risky decision-making. (47) Women showed a higher stress response yet performed better under anticipatory stress, while men performed worse. Similarly, Van den Bos *et al.* (45) examined the effects of the TSST on decision-making performance on the IGT. They concluded that there are sex differences in response to the TSST and that cortisol reactivity is a decisive factor in behavioural performance. In women, low responders had slightly elevated cortisol levels after the TSST, which improved their IGT performance, whereas high responders had highly elevated cortisol levels, which hindered performance. In men, overall increases in cortisol levels impaired their IGT performance. It is reasonable to consider these responder differences may have differently affected our participants' cognitive performance. Moreover, our sample size was divided in half in order to observe sex differences, which presented as a limitation. Further studies should increase sample size to aid in better controlling for individual differences when assessing the effects of stress on cognitive performance tasks.

As previously mentioned, differences in cognitive performance as a result of differences in stress response may be due to the nature of a stressor. Preston *et al.* (19), Plessow *et al.* (44) and Van den Bos *et al.* (45) used public speaking stressors, which may produce a phenomenon called social-evaluative threat. (13) This occurs when important aspects of one's identity have the potential to be negatively judged by others. Because participants in our study were required to speak on a topic of their choice, this implies a social-evaluative threat and may be stress eliciting. It has been shown that women are more sensitive to social rejection, as they tend to respond to stress in more socially oriented ways (38,48), so it could be expected that women may be more susceptible to the social-evaluative components of the TSST. However, women seem to be less sensitive to the TSST than men. (17,35) Previous findings suggest that men are more sensitive to the achievement components of the TSST (48) and exert more effort in response to the challenges. (39) Accordingly, men may interpret stressors differently than women and depending on the type of stressor used, men or women may be more sensitive to its effects. Future research should take the type of stressor into account.

Our results showed high effort avoidance in women but not in men. This study did not reveal results consistent with our hypothesis which predicted that stress induction would impair cognitive performance such that participants would make more errors and avoid effort. Potential factors that led to these results include individual biases in effort exertion (Demand Avoiders versus Seekers), differences in stress response, and type of stressor. It may be beneficial for future studies to investigate these confounding factors and control for their effects. Nevertheless, given the increasing evidence of prefrontal cortex involvement in the regulation of the HPA axis, it is important to find reliable methods to reduce or prevent stress-induced cognitive impairments.

Acknowledgements

The author wishes to thank Ross Otto and Mario Bogdenov of The McGill Otto Lab, as well as the many individuals who have collaborated and assisted with the present research study.

References

1. Hull CL. Principles of behavior: An introduction to behavior theory. New York: Appleton-Century-Crofts; 1943. 422 p.

2. Kool W, McGuire JT, Rosen ZB, Botvinick MM. Decision making and the avoidance of cognitive demand. Journal of Experimental Psychology:

General. 2010 Nov;139(4):665.

3. Yeung N, Monsell S. Switching between tasks of unequal familiarity: The role of stimulus-attribute and response-set selection. Journal of Experimental Psychology: Human Perception and Performance. 2003 Apr;29(2):455.

4. Arrington CM, Logan GD. The cost of a voluntary task switch. Psychological science. 2004 Sep;15(9):610-5.

5. Kool W, Botvinick M. A labor/leisure tradeoff in cognitive control. Motivation Science. 2014;1(S):3-18.

6. Shenhav A, Musslick S, Lieder F, Kool W, Griffiths TL, Cohen JD, Botvinick MM. Toward a rational and mechanistic account of mental effort. Annual review of neuroscience. 2017 Jul 25;40:99-124.

7. Westbrook A, Braver TS. Cognitive effort: A neuroeconomic approach. Cognitive, Affective, & Behavioral Neuroscience. 2015 Jun 1;15(2):395-415.

8. Westbrook A, Kester D, Braver TS. What is the subjective cost of cognitive effort? Load, trait, and aging effects revealed by economic preference. PloS one. 2013 Jul 22;8(7):e68210.

9. Shields GS, Sazma MA, Yonelinas AP. The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. Neuroscience & Biobehavioral Reviews. 2016 Sep 1;68:651-68.

10. Bogdanov M, Schwabe L. Transcranial stimulation of the dorsolateral prefrontal cortex prevents stress-induced working memory deficits. Journal of Neuroscience. 2016 Jan 27;36(4):1429-37.

11. Lupien SJ, Maheu F, Tu M, Fiocco A, Schramek TE. The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. Brain and cognition. 2007 Dec 1;65(3):209-37.

12. Lupien SJ, McEwen BS, Gunnar MR, Heim C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. Nature reviews neuroscience. 2009 Jun;10(6):434.

13. Dickerson SS, Kemeny ME. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychological bulletin. 2004 May;130(3):355.

14. Kirschbaum C, Hellhammer DH. Salivary cortisol in psychobiological research: an overview. Neuropsychobiology. 1989;22(3):150-69.

15. McEwen BS, Gianaros PJ. Stress-and allostasis-induced brain plasticity. Annual review of medicine. 2011 Feb 18;62:431-45.

16. McEwen BS, Morrison JH. The brain on stress: vulnerability and plasticity of the prefrontal cortex over the life course. Neuron. 2013 Jul 10;79(1):16-29.

17. Kudielka BM, Kirschbaum C. Sex differences in HPA axis responses to stress: a review. Biological psychology. 2005 Apr 1;69(1):113-32.

18. Al'Absi M, Hugdahl K, Lovallo WR. Adrenocortical stress responses and altered working memory performance. Psychophysiology. 2002 Jan;39(1):95-9.

19. Preston SD, Buchanan TW, Stansfield RB, Bechara A. Effects of anticipatory stress on decision making in a gambling task. Behavioral neuroscience. 2007 Apr;121(2):257.

20. Halari R, Hines M, Kumari V, Mehrotra R, Wheeler M, Ng V, Sharma T. Sex differences and individual differences in cognitive performance and their relationship to endogenous gonadal hormones and gonadotropins. Behavioral neuroscience. 2005 Feb;119(1):104.

21. Hyde JS. Sex and cognition: gender and cognitive functions. Current opinion in neurobiology. 2016 Jun 1;38:53-6.

22. Marrocco J, McEwen BS. Sex in the brain: hormones and sex differences. Dialogues in clinical neuroscience. 2016 Dec;18(4):373.

23. Charmandari E, Tsigos C, Chrousos GP. Neuroendocrinology of stress. Ann Rev Physiol. 2005;67:259-84.

24. Chrousos GP, Gold PW. The concepts of stress and stress system disorders: overview of physical and behavioral homeostasis. Jama. 1992 Mar 4;267(9):1244-52.

25. Chrousos GP, Torpy DJ, Gold PW. Interactions between the hypothalamic-pituitary-adrenal axis and the female reproductive system: clinical implications. Annals of internal medicine. 1998 Aug 1;129(3):229-40.

26. Vgontzas AN, Mastorakos G, Bixler EO, Kales A, Gold PW, Chrousos GP. Sleep deprivation effects on the activity of the hypothalamic–pituitary–adrenal and growth axes: potential clinical implications. Clinical endocrinology. 1999 Aug;51(2):205-15.

27. Cahill L. Why sex matters for neuroscience. Nature Reviews Neuroscience. 2006 Jun;7(6):477.

28. Kudielka BM, Hellhammer DH, Wüst S. Why do we respond so differently? Reviewing determinants of human salivary cortisol responses to challenge. Psychoneuroendocrinology. 2009 Jan 1;34(1):2-18.

29. Matthews KA, Gump BB, Owens JF. Chronic stress influences cardiovascular and neuroendocrine responses during acute stress and recovery, especially in men. Health Psychology. 2001 Nov;20(6):403.

30. Steiner H, Ryst E, Berkowitz J, Gschwendt MA, Koopman C. Boys' and girls' responses to stress: affect and heart rate during a speech task. Journal of Adolescent Health. 2002 Apr 1;30(4):14-21.

31. Childs E, Dlugos A, De Wit H. Cardiovascular, hormonal, and emotional responses to the TSST in relation to sex and menstrual cycle phase. Psychophysiology. 2010 May;47(3):550-9.

32. Cornelisse S, van Stegeren AH, Joëls M. Implications of psychosocial stress on memory formation in a typical male versus female student sample. Psychoneuroendocrinology. 2011 May 1;36(4):569-78.

33. Kudielka BM, Buske-Kirschbaum A, Hellhammer DH, Kirschbaum C. Differential heart rate reactivity and recovery after psychosocial stress (TSST) in healthy children, younger adults, and elderly adults: the impact of age and gender. International journal of behavioral medicine. 2004 Jun 1;11(2):116-21.

34. Guillermo CJ, Manlove HA, Gray PB, Zava DT, Marrs CR. Female social and sexual interest across the menstrual cycle: the roles of pain, sleep and hormones. BMC women's health. 2010 Dec;10(1):19.

35. Kirschbaum C, Kudielka BM, Gaab J, Schommer NC, Hellhammer DH. Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. Psychosomatic medicine. 1999 Mar 1;61(2):154-62.

36. Kelly MM, Tyrka AR, Anderson GM, Price LH, Carpenter LL. Sex differences in emotional and physiological responses to the Trier Social Stress Test. Journal of behavior therapy and experimental psychiatry. 2008 Mar 1;39(1):87-98.

37. Kudielka BM, Hellhammer J, Hellhammer DH, Wolf OT, Pirke KM, Varadi E, Pilz J, Kirschbaum C. Sex differences in endocrine and psychological responses to psychosocial stress in healthy elderly subjects and the impact of a 2-week dehydroepiandrosterone treatment. The Journal of Clinical Endocrinology & Metabolism. 1998 May 1;83(5):1756-61.

38. Taylor MK, Larson GE, Hiller Lauby MD, Padilla GA, Wilson IE, Schmied EA, Highfill-McRoy RM, Morgan CA. Sex differences in cardio-

vascular and subjective stress reactions: prospective evidence in a realistic military setting. Stress. 2014 Jan 1;17(1):70-8.

39. Kirschbaum C, Pirke KM, Hellhammer DH. The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology. 1993;28(1-2):76-81.

40. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. Journal of personality and social psychology. 1988 Jun;54(6):1063.

41. Folstein MF, Luria R. Reliability, validity, and clinical application of the Visual Analogue Mood Scale. Psychological medicine. 1973 Nov;3(4):479-86.

42. Kirschbaum C, Hellhammer DH. Salivary cortisol in psychoneuroendocrine research: recent developments and applications. Psychoneuroendocrinology. 1994 Jan 1;19(4):313-33.

43. Lopez-Duran NL, Mayer SE, Abelson JL. Modeling neuroendocrine stress reactivity in salivary cortisol: adjusting for peak latency variability. Stress. 2014 Jul 1;17(4):285-95.

44. Plessow F, Kiesel A, Kirschbaum C. The stressed prefrontal cortex and goal-directed behaviour: acute psychosocial stress impairs the flexible implementation of task goals. Experimental brain research. 2012 Feb 1;216(3):397-408.

45. Van den Bos R, Harteveld M, Stoop H. Stress and decision-making in humans: performance is related to cortisol reactivity, albeit differently in men and women. Psychoneuroendocrinology. 2009 Nov 1;34(10):1449-58.

46. Sayalı C, Badre D. Neural systems of cognitive demand avoidance. Neuropsychologia. 2019 Feb 4;123:41-54.

47. Bechara A, Damasio AR, Damasio H, Anderson SW. Insensitivity to future consequences following damage to human prefrontal cortex. Cognition. 1994 Apr 1;50(1-3):7-15.

48. Stroud LR, Salovey P, Epel ES. Sex differences in stress responses: social rejection versus achievement stress. Biological psychiatry. 2002 Aug 15;52(4):318-27.

Submitted: 01/26/20

Research Article

[1]McGill University, Montreal, QC, Canada

Keywords

Microplastic, Gault Nature Reserve, limnology, pollution, freshwater

Email Correspondence

yael.lewis@mail.mcgill.ca

Emily Brown[1], Laura Mackey[1], Libby Rothberg[1], Mackenzie Burnett[1], Noelle Bergeron[1], Yael Lewis[1]

An Evaluation of Microplastics in Lac Hertel Sediment Over Time

Abstract

Background: Microplastics, defined as plastic smaller than 5 mm, are pervasive in both marine and freshwater ecosystems. Humans, zooplankton, and fish have been shown to ingest microplastics, which could have detrimental health impacts. Consequently, this project investigated the question: are there microplastics in the sediment of Lac Hertel, located in the Mont Saint Hilaire Biosphere Reserve in Quebec, and if so, how has the amount of microplastics changed over time?

Methods: One sediment core was obtained from the centre of the lake and one was obtained from the edge near the mesocosm dock. Next, one section from the top, middle, and bottom of each core was collected. Afterwards, the microplastics were extracted from the sediments, counted with a dissecting microscope under regular light, and a subset of fragments were tested with a hot needle to confirm that they were plastic.

Results: A generalized linear model indicated that the number of microplastics in our samples increased significantly over time and that the sediment samples from the mesocosm dock had significantly fewer microplastics than the lake's centre. Similarly, a Pearson correlation test revealed that an increasing sediment depth had a significantly negative relationship with the number of microplastics at the lake's centre. However, another Pearson correlation test determined that this trend was not reflected at the mesocosm dock, potentially because of sediment focusing.

Limitations: Due to resource and time constraints, we had a small sample size, only analyzed microplastics larger than 250 µm, and counted microplastics instead of weighing them.

Conclusion: Our results suggest that there has been a significant increase in microplastics in Lac Hertel sediment over time. Ultimately, our results emphasize the need to mitigate plastic pollution.

Introduction

Plastic is ubiquitous in modern society. The cheap cost, light weight, and durable structure of plastic are just some of the factors that created a global annual demand of about 245 million tonnes in 2011. (1) Aquatic biologists have only recently discovered that tiny particles of plastic are also consumed by organisms of a much wider size range than initially thought, allowing high concentrations of plastics to circulate food webs much more easily. (1) Studies have now shown that microplastics, defined as plastic particles smaller than 5 mm, can be ingested by and pose risks to the health of organisms such as zooplankton, fish, and potentially humans via the consumption of seafood. (2–4) Yet there is little evidence that the problem is slowing; since the 1960s, demand for plastics has been increasing exponentially, likely leading to an increased accumulation of microplastics in aquatic ecosystems. (2)

The Gault Nature Reserve is located on Mont Saint Hilaire and has been designated as an International Biosphere Reserve by the United Nations. (5) Situated roughly 32 km east of Montreal, the reserve consists of seven low peaks that form a ring around Lac Hertel (5), a glacially formed depression that has a maximum depth of 9 m. (6) The lake is fed by three streams and a fourth flows toward the Richelieu River. Swimming, fishing, and boating are not permitted because Lac Hertel is a secondary reservoir of drinking water in the region. (6) Although the reserve has experienced little known human disturbance relative to the surrounding area (7), it is vital that the lake is monitored for pollution that could negatively affect the community of species within it and its use as a research site. As a result, our research question was the following: are there microplastics in Lac Hertel's sediment and if so, how has the amount of microplastics changed over time?

Despite the regulations in place at the Gault Nature Reserve to preserve the ecosystem, like prohibiting fishing, swimming, and boating at Lac Hertel, the accumulation of plastic pollution remains a potential threat. One possible source of plastic pollution could be atmospheric deposition of microplastics. (8–10) In a study quantifying atmospheric deposition in the remote French Pyrenees, Allen *et al.* found that microplastic fibers up to ~750 µm long and fragments smaller than 300 µm were being deposited over distances of up to 95 km. (8) Similarly, Dris *et al.* recorded a rate of atmospheric fiber fallout between 2 and 355 particles/m^2/day, 29% of which was estimated to be synthetic. (9) Moreover, the UV photo-degradation (11), wave deterioration (1,11), and weather degradation (1) of larger plastic items originating from litter in or around Lac Hertel could be another source. Research has been conducted at the reserve since the 1950s, potentially also contributing to the accumulation of microplastics in the area. (12) Since the creation of the Gault Nature Reserve, the number of annual visitors to the lake and use of plastics by society has been increasing. (13) Specifically, the number of annual visitors surpassed 63 000 in 2017 (14) and plastic production has exponentially increased over the last ~65 years, with 8,300 million metric tons having been produced by 2017. (15) Such an increase in visitation and plastic consumption has likely tested the efficacy of the reserve's visitor use regulations, potentially still allowing for plastic litter from visitors and researchers to be degraded into microplastics and contaminate the sediment of Lac Hertel.

Understanding the scope of microplastic pollution in this protected area is crucial to recognizing the extent to which its protection has actually been effective. Past research has highlighted that many protected areas are protected in name only. For example, though the establishment of Marine Protected Areas (MPAs) theoretically aimed to limit anthropogenic pollution, pollution status is unknown "in most MPAs worldwide". (16) The lack of current pollution data is itself alarming, but without long term records it can be even more difficult to ascertain the causes of pollution. Historical records may also assist in predicting future trends in pollution (17), which can inform what kind of restrictions are necessary for protected areas to limit future risks. Unfortunately, such long-term records are even more

scarce than current data on pollution, partially because certain inconspicuous pollutants like microplastics have only been recognized as problematic in recent years. (1,2) Consequently, paleoenvironmental methods provide an extremely valuable mechanism of tracking how changes to ecosystems, such as pollution, have accrued over time. (17)

In this vein, this study aimed to discover: to what extent is Lac Hertel's sediment polluted with microplastics and how has this pollution changed over the last several decades? We hypothesized that there has been a significant increase in microplastics in Lac Hertel sediment over time. To test this hypothesis, we analyzed sediment cores from two areas of the lake for microplastic pollution. We then compared the microplastic abundance of the top, middle, and bottom of the cores, representing sediment from recent years and from the past several decades, in order to evaluate these historical trends.

Methods

The first step of our methodology was completed in November 2019 and involved retrieving columns of sediment using a piston corer, a widely-used instrument that results in minimal sediment disturbance. (18) We took one sediment core at the mesocosm dock, depicted in Fig. 1, which was installed in 2011. We analyzed sediment from the mesocosm dock, the site of the reserve's partially enclosed outdoor limnological experiments (19), because large plastic bags are used as part of the set-up. Since larger pieces of plastic debris are known to fragment into microplastics over time (11,20), we wanted to test whether this was potentially contributing microplastics to the sediment underneath it. Next, we obtained another sediment core from the centre, i.e. deepest part, of Lac Hertel. Sampling from the centre of a lake is standard practice in limnology (21), since it has the highest probability of representing the whole-lake assemblage. (22,23) The centre of a lake is more representative due to sediment focusing, which refers to how sediments redistribute from shallow to deep areas of a lake because of waves and currents. (23–25)

Figure 1. Lac Hertel, Quebec, Canada, surrounded by the mature temperate forest of the Gault Nature Reserve. We retrieved our sediment cores from the centre, i.e. deepest part, of the lake and the mesocosm dock. (26)

After retrieving the cores, we used an extruding device and spatula to take incremental 3 cm samples of the sediment. In each sediment core, we obtained the sediment that was 0-3 cm and 15-18 cm from the top. We also obtained the sediment that was 3-0 cm from the bottom of each core, although the total depth of each core was different. Accordingly, the bottom sediment obtained from the mesocosm dock was 30-33 cm from the top of the core, while the bottom sediment from the lake's centre was 35-38 cm from the top. Gélinas, Lucotte & Schmit (2000) concluded that the sediment accumulation rate in Lac Hertel is 0.55 cm/year. (27) Based on this rate, the top sections contained sediment that was ~0 to 5.5 years old (2019-2013) and the middle sections contained sediment that was ~27.3 to 32.7 years old (1991-1986). Finally, the bottom section from the mesocosm dock contained sediment that was ~54.5 to 60 years old (1964-1959) and the bottom section from the centre of the lake contained sediment that was 63.6 to 69.1 years (1955-1950). As a result, we were able to determine whether microplastics in Lac Hertel's sediment have increased over at least the past 60-69 years. However, the bottom sections are likely older than 60-69 years because sediment tends to compact as it accumulates. (16,17,28)

We observed several of the precautions outlined in Crew *et al.* (2020) to minimize the potential contamination of the sediment samples. (29) To avoid contamination in the field, we rinsed all our tools with water between sample extrusions. In addition, the samples were stored in glass containers and covered with aluminum foil. Both in the field and in the lab, we avoided wearing clothes made out of synthetic material and wore cotton lab coats while handling the samples. Moreover, lab surfaces were regularly wiped down with 70% ethanol and glassware was washed between samples with 4% non-foaming detergent and distilled water. Next, petri dishes with wet borosilicate filters were placed beside all our workstations to serve as procedural blanks. The filters were later analyzed under a microscope to assess whether airborne particles had contaminated our samples. Lastly, we implemented a positive control by spiking an extra sediment sample with a known number of easily identifiable microplastics and subjecting it to the same lab procedure as the rest of our samples. We also analyzed this sample under a microscope to evaluate the proportion of microplastics that were recovered. This provided us with a recovery rate that we used to account for the microplastics we lost throughout our lab procedure.

Our lab procedure was adapted from Crew *et al.* (2020) and consisted of the following steps. (27) Firstly, the samples were transferred to 250 mL Erlenmeyer flasks and 10 mL of 10% KOH was added to each flask. Next, the samples were deflocculated for 25 minutes in a hot water bath set to 60 °C in order to digest the organic material in the samples. Afterwards, each sample was placed on a 250 µm sieve and rinsed with distilled water to remove the digested organic matter and KOH. What remained on the mesh sieve was then transferred to a petri dish lined with aluminum foil and dried in an oven set to 60 °C for 12 hours. Since the melting points of most plastics are above, or just below, 100 °C (30–32), the oven and water bath would not have melted any of the potential plastic in the samples. Once the samples were dried, they were placed in clean 250 mL Erlenmeyer flasks and an oil extraction procedure adapted from Crichton *et al.* (2017) was performed to isolate the microplastics. (33) This method was chosen because it is quick and cost-effective. The extraction procedure involved adding 2.5 mL of canola oil and 50 mL of distilled water to each Erlenmeyer flask, swirling the flasks for 30 seconds, and leaving the mixtures to settle for approximately two minutes. Since plastic is oleophilic, this traps microplastics in the oil layer. (33) Next, the contents of a flask were poured into a 250 mL separatory funnel and several rinsing, swirling, and settling steps were carried out, with the water layers being discarded. Finally, the oil mixture was dispensed onto a funnel equipped with a polycarbonate membrane filter and dried using vacuum filtration. In between each sample, 13 mL of 4% non-foaming detergent was added to the separatory funnel and swirled vigorously, then poured onto the polycarbonate filter. Following this, the funnel was rinsed 3 times with distilled water.

Our next step involved identifying and counting the microplastics under a dissecting microscope at 40X magnification. Microscopy is a widely used identification method for microplastics in the hundreds of micrometer range, meaning 100-999 µm (29), like the particles we were investigating. Two individuals counted each sample and then the average of the two counts was rounded to the nearest whole number. We used the identification trees that were created by Crew *et al.* (29) to differente between microplastics and other oleophilic particles under regular light and then applied the hot needle test to a subset of what we observed. The hot needle test has been used to verify particles as plastic in a number of studies (3,34–36) and consists of touching the suspected microplastic particles with a hot needle. The hot needle test can help distinguish plastics due to their melting points; many plastics have relatively wide temperature ranges as melting points, creating a zone between the solid and liquid states where their physical properties shift. (37) When heated to temperatures in this

intermediate zone, plastic is semi-solid and flexible. Thus, if the particles are plastic, the hot needle will be able to make the particle sticky and leave a mark, or the particle will curl. If a particle is non-plastic, it will either fragment into smaller pieces or not react at all.

We analyzed our data using a Generalized Linear Model (GLM) in RStudio Cloud version 3.6.0. (38) GLMs are based on an assumed relationship between the mean of the response variable and the explanatory variables. In GLMs, data can be assumed to have a variety of probability distributions, like normal or Poisson. (39) As a result, GLMs are considered more flexible and better suited for analyzing ecological relationships, which do not always fit a normal distribution We employed a GLM to help us determine: 1- whether sediment from the centre of the lake or the mesocosm dock had more plastic, 2- whether the number of microplastics changed over time, i.e. sediment depth, and 3- whether the relationship between sediment depth and number of microplastics was different at the centre of the lake versus the mesocosm dock. Accordingly, the response variable in the GLM was the average number of microplastics in each sample, adjusted by our recovery rate, which we determined to be 65%, and rounded to the nearest whole number. The explanatory variables were area, i.e. centre of the lake versus mesocosm dock, and sediment depth. Furthermore, the interaction between the explanatory variables was also evaluated in order to answer the third aforementioned question. Our GLM equation was thus "average number of microplastics ~ depth*area". Microplastic abundance is an example of count data, which frequently follows a Poisson distribution. (39) To check if our data fit the Poisson distribution, we applied a chi-square goodness-of-fit test. The resulting p value was 0.4050313, which indicated that our data was not significantly different from the Poisson model. Consequently, the GLM was implemented with a Poisson distribution.

Although the GLM could help answer the above-mentioned questions 1, 2, and 3, it could not answer our fourth question: was the relationship between sediment depth and number of microplastics stronger at the centre of the lake or the mesocosm dock? As a result, we also analyzed our data using Pearson correlation tests in RStudio Cloud version 3.6.0. (38) The response variable in the two correlation tests was the average number of microplastics found in the sediment at the centre of the lake and the mesocosm dock, respectively, rounded to the nearest whole number and adjusted to incorporate our recovery rate. The explanatory variable of the two tests was the depth of the sediment. Although Pearson's correlation coefficient is not as robust in the presence of non-normality, it can still be used if data is approximately normal. (40) To determine if this was the case, we applied a Shapiro-Wilk test to the average number of microplastics found at the centre of the lake and at the mesocosm. The resulting p values were respectively 0.9449 and 1, suggesting that the distribution of our data was not significantly different from the normal distribution. This seems contradictory to the results of the chi-square goodness-of-fit test but may be explained by our small sample size. Nevertheless, given that the variables in question are examples of ratio data, that the relationships appeared to be linear when graphed (see Fig. 2), and that the Shapiro-Wilk test implied that the normal distribution fit our data, the Pearson correlation test was used.

Results

We found between 23-82 microplastics in every sample we retrieved, as is depicted in Table 1. Examples of the microplastics we found include fibers, clear fragments, and microbeads. When we touched a hot needle to a subset of the pieces that we suspected to be plastic, the pieces curled or melted, confirming that they were plastic. We also confirmed that the debris we excluded from our counts were not plastic, as these items broke when they came into contact with the hot needle. Fig. 3 demonstrates some examples of microplastics and non-plastic debris that we tested with a hot needle. When we looked at our samples under a microscope, we noticed that there was an abundance of synthetic-looking fibers. In addition, we found an abundance of very similar fibers on the procedural blanks that we left beside our workstations. Consequently, we suspected that these fibers may have largely been the result of contamination and we deducted all the fibers from the average number of microplastics to be conservative. After counting the number of microplastics we retrieved in the spiked sediment sample, we calculated that our recovery rate was 65%. In comparison, Crew *et al.* (2020) reported their recovery rate for fibers was 67% ± 2.3 (SE), 63% ± 3.5 (SE) for microbeads, and 61% ± 2.2 (SE) for fragments. Since our recovery rate was similar to those of Crew *et al.* (29), we adjusted our average number of microplastics to reflect how 33% of the microplastics had been lost throughout the lab procedure.

Sample	Average number of fibers	Average number of clear fragments	Average number of other microplastics
Top/Centre	34	43	5
Middle/Centre	28	33	3
Bottom/Centre	15	22	0
Top/Mesocosm	20	13	12
Middle/Mesocosm	10	15	2
Bottom/Mesocosm	22	20	1

Table 1. Summary of the types of microplastics that were found in the sediment samples obtained from Lac Hertel. Each column reflects the average of two counts, rounded to the nearest whole number but not adjusted by our 65% recovery rate. *Centre* denotes the samples from the centre of the lake and *mesocosm* denotes the samples from the mesocosm dock. Moreover, *top*, *middle*, and *bottom* refer to the section of the sediment core that the samples were from. Examples of microplastics that fell under *other* include microbeads and coloured fragments.

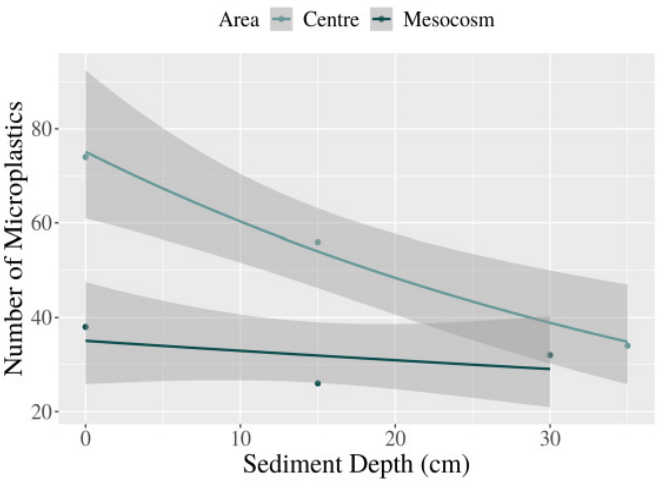

Figure 2. The average number of microplastics that we found in our sediment samples from the centre of Lac Hertel and the mesocosm dock. The average number of microplastics was minus the fibers we observed, adjusted to account for our 65% recovery rate, and rounded to the nearest whole number. The figure was fitted using a Poisson distribution and the shaded areas represent the standard error of the mean.

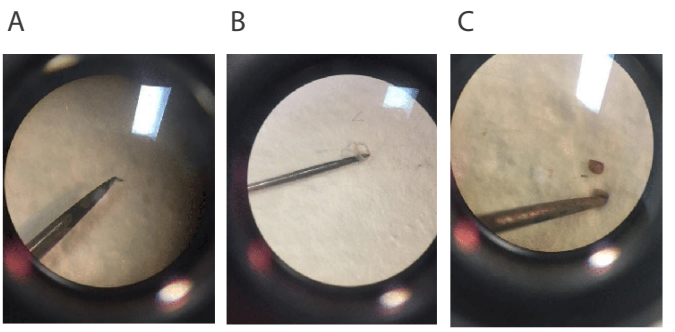

Figure 3. Images of items recovered from the oil extraction procedure. (A)-(B) Microplastics, confirmed as such since they did not break during the hot needle test. (C) Not a microplastic, confirmed as such since it broke during the hot needle test.

Analyzing the data using a GLM revealed three things, summarized in Table 2. Firstly, when the GLM analyzed the overall change in microplastics over sediment depth, i.e. time, meaning without distinguishing whether samples came from the centre or mesocosm, the GLM indicated that the abundance of microplastics increased significantly over time. Secondly, when the GLM analyzed the number of microplastics at the centre of the lake versus the mesocosm dock, it determined that the mesocosm dock had significantly fewer microplastics than the centre of the lake. Third, the GLM showed there was not a significant interaction between the two explanatory variables, depth and area, meaning that the effect of depth on the number of microplastics was not significantly different at the centre of the lake versus the mesocosm dock. Next, the Pearson correlation tests, summarized in Table 3, revealed one more thing: the first test indicated that at the centre of the lake, there was a significantly negative relationship between the depth of sediment and the number of microplastics in our samples. Conversely, a second Pearson correlation test indicated that at the mesocosm dock, there was not a significant relationship between the sediment depth and the number of microplastics in our samples.

	Estimate	*p*	Standard error	*z*
Explanatory Variable #1: Depth (cm)	-0.021952	0.000154	0.005802	-3.784
Explanatory variable #2: Area (mesocosm)	-0.762377	$5.16*10^{-5}$	0.188321	-4.048
Interaction between variable #1 and #2:	-0.015693	0.122774	0.010169	1.543

Table 2. Results of the generalized linear model (GLM) in which the explanatory variables were the depth of the sediment samples in cm and the area that the samples came from, i.e. centre of Lac Hertel versus the mesocosm dock. The response variable was the average number of microplastics that we identified in the samples minus the fibers we observed, adjusted to account for our 65% recovery rate, and rounded to the nearest whole number. The interaction between the explanatory variables was also evaluated.

	Pearson Correlation Coefficient	*p*	*t*	Degrees of freedom
Test #1: Number of microplastics at the lake's centre vs sediment depth	-0.9985728	0.0340 2	-18.697	1
Test #2: Number of microplastics at the mesocosm dock vs sediment depth	-0.5	0.6667	-0.57735	1

Table 3. The results of two Pearson correlation tests. Test #1 and #2 analyzed the relationship between sediment depth and the average number of microplastics in samples from the centre of Lac Hertel and the mesocosm dock, respectively. Fibers were removed from the average number of microplastics, which was then adjusted by our 65% recovery rate and rounded to the nearest whole number.

Discussion

The results of the GLM and Pearson correlation tests revealed novel information about Lac Hertel. Based on the lake's sedimentation rate (27), the oldest samples from the mesocosm dock and lake centre were at least 54.4-60 years old and 63.6-69.1, respectively. The GLM demonstrated that overall, the number of microplastics increased significantly with time. This was supported by the first Pearson correlation test, which indicated that there was a significantly negative correlation between sediment depth and abundance of microplastics in the samples from the centre of Lac Hertel. However, the second Pearson correlation test revealed that this trend was not present at the mesocosm dock. Since the mesocosm dock is located on a steep slope and sediment is particularly susceptible to redistribution along steep slopes (25), it is likely that the sediment and microplastics are being redistributed towards the centre of the lake. This may partially account for the lack of a significant correlation between sediment depth and number of microplastics at the mesocosm dock. The GLM also determined that the samples from the mesocosm dock had significantly fewer microplastics. To be clear, the *area* row in Table 2 depicts the effect of area on the number of microplastics. The GLM output specified that these values refer to the mesocosm dock, and since the p-value was significant and the estimate was negative, we can conclude that the sediment taken from the mesocosm dock had significantly fewer microplastics than the centre of the lake. In short, our results indicate that there are microplastics in Lac Hertel sediment and the amount of microplastic pollution has increased significantly over at least the last 60-69 years.

Our findings are notable because the literature documenting microplastic contamination in protected lakes is scarce; yet, our results are consistent with the limited literature concerning microplastic pollution in freshwater. For instance, microplastics have been found in Lake Tahoe, a freshwater lake that does not experience wastewater dumping and that is surrounded by National Forest, a classification of protected lands in the United States. (41) However, Lac Hertel has the additional protection of being closed-off to visitor usage, like boating and swimming, as well as having a catchment that is encapsulated by the biosphere reserve. Microplastics have also been found to be abundant in other freshwater environments (2,41–44), similar to how microplastics have been found in Lac Hertel. Furthermore, plastic debris has been found along the beach of the remote Biosphere Reserve of Lanzarote in the Canary Islands (45), near the Pelagos Sanctuary in the Mediterranean Sea (46), as well as in a Marine Protected Area in Croatia. (47) Microplastic pollution has even reached the Arctic. (48,49) Our results support the findings that microplastic contamination is an increasing environmental problem that has spread to even the most protected habitats, and caution against viewing protected areas as pristine havens from anthropogenic pollution.

Our results add to the growing collection of research documenting microplastics' presence and increase over time using sediment coring. Matsuguma *et al.* found that the abundance of microplastics in sediment cores from Japan, Thailand, Malaysia, and South Africa increased toward the top layer of sediment, demonstrating an increase of microplastic pollution over time on a global scale. (50) Specifically, the cores from Japan demonstrated that microplastic pollution started in the 1950s and then increased considerably as the sediment depth decreased. These results are consistent with our data, as the bottom sections of our cores reflect the presence of microplastics in Lac Hertel as early as 1950 and microplastics were far less abundant in this section. Research by Turner *et al.* that examined microplastics in the sediment record of an urban lake in London similarly found an increase in microplastic accumulation over time, starting in the 1950s. (51) Furthermore, sediment cores taken from an estuary in Tasmania were found by Willis *et al.* to have microplastics present in every layer sampled, dating from 1744 to 2004, and increasing in abundance closer to the surface. (52) The authors found that deeper sediment layers pre-dating the proliferation of plastic in society mostly contained fibers, leading them to suspect that fibers were a major form of contamination, which is consistent with what we discovered in our results. Still, there is a relative dearth of studies that utilize sediment coring in areas less obviously affected by anthropogenic influence, further underlying the relevance of our results to knowledge in the field.

Given that Lac Hertel has a small catchment and is not open to the public, our results lead us to ask: where are the microplastics coming from? One possible culprit could be atmospheric deposition, which is an important pathway for microplastic pollution. (8–10) Atmospheric deposition, however, is likely to deposit microplastics too small to identify under 40x magnification; one study found that the majority of deposited microplastics were ≤50 μm. (8) Typically, atmospheric deposition also mainly deposits microplastic fibers. (10,11) Thus, atmospheric deposition may not be as substantial of a contributing source to the microplastics we observed, which were greater than 250 μm and mostly fragments after we accounted for the suspected fiber contamination. Instead, microplastics could stem from littering on Lac Hertel's beach, which is open to the public, as well as the use of the lake for research purposes.

Researchers have been doing limnological experiments at the mesocosm dock on Lac Hertel since it was built in 2011, which often uses large plastic bags that can become torn. We observed that this had occurred when we were carrying out our experiments. These activities could result in large plastic fragments polluting the environment that could then be further broken down (1,11), contributing to the microplastics found in our top sediment samples. Though we controlled for plastic contamination in the

petri dishes and glass jars that the sediment was stored and transported in, the one relatively unavoidable source of potential microplastic contamination was the plastic coring tube itself. Though concerns have been raised about the potential for coring tubes to be scratched by sediment particles and contaminate samples (53), there is limited research to show that such contamination is substantial in practice. However, the presence of plastic in sediment coring is not limited to this one study; sediment coring with plastic corers is employed regularly by researchers at Lac Hertel. In addition, there has been long-term use of other plastic equipment like sampling bottles, buoys, Secchi disks, life vests, and ropes. Even the simple presence of researchers on the lake could enable their clothing to pollute the lake with microplastics if it is made of synthetic fibers, which could eventually settle into the sediment. Little to no data exists to document the microplastic footprint of limnological research, which should be a focus for future investigation. The Gault Nature Reserve has been used for research, teaching, and public recreation dating back to the 1950s and 1960s. (12) As such, it is probable that pollution sources like atmospheric deposition, littering, and research activities may have contributed to the microplastics that we detected in all of our samples.

A few factors limit the scope of our results. Due to time constraints, we were not able to weigh the amount of microplastics that we found in each sample. Resource constraints prevented us from using Nile Red dye and fluorescent microscopy, thus we could only evaluate microplastics larger than 250 μm. Since our cores were not radiometrically dated, the dates were approximated from the lake's sedimentation rate, which does not account for compaction. Consequently, our bottom layers are likely older than our calculations, resulting in some ambiguity about exactly how far back microplastic pollution extends in Lac Hertel's sediment. Another notable limitation was our small sample size, since we only had three data points per core. We did not finely partition our core as this would have multiplied the number of oil extractions we had to perform, which was not feasible within our timeframe. Still, a more finely partitioned core would allow for a better understanding of the relationship between microplastic abundance and time. Additionally, we chose to leave the procedural blanks out for the entire time that we were conducting experiments to get a better idea of the kind of contamination that could be coming from our surroundings. However, this meant that the amount of contamination on the blanks was likely greater than what any single sample would have been exposed to, since they were only exposed for brief periods of time. As a result, we could not draw direct conclusions about the quantity of contamination in the samples from the procedural blanks, but the blanks still enabled us to observe what kinds of microplastics may have been the result of contamination.

Future research would benefit from the following adaptations to our methodology: weighing microplastics, using additional methods to analyze even smaller microplastics, obtaining more sections from each sediment core, dating the sediment cores, and only setting procedural blanks out when samples are exposed. The primary focus of future research could be to determine the origins of the microplastics at Lac Hertel, specifically focusing on two potential sources: the beach, which is used by the public, and the mesocosm dock, which is used by researchers. The beach and dock areas should additionally be monitored for macroplastic pollution, e.g. torn mesocosm bags and incidents of littering, so that original sources can be tracked prior to degradation into microplastics. Further research could also explore whether these microplastics are present in tissues of biota of the lake, as well as the biological impacts of microplastics on these organisms. This is especially important given the use of this lake as an experimental system for researchers, where unaccounted microplastic pollution may confound experiment results. For example, studies interested in the diets or life history of the lake's biota may not account for the consumption of microplastic and its possible effects on development and reproduction. (54,55) More protected lakes could be studied to see if our hypothesis is supported in other protected areas as well. Most importantly, further regulations to limit non-essential plastic use by researchers at the Gault Nature Reserve should be considered where possible, at least until the source(s) of the plastic are identified.

Conclusion

Microplastics have become a major pollutant that poses a risk to the health of numerous species. Today, microplastics can be found in even the most remote corners of the world. Despite an increasing amount of research on freshwater microplastic pollution, there is still very little research about how microplastics impact protected areas. As a result, we evaluated whether there are microplastics in the sediment at Lac Hertel, which is part of an UNESCO biosphere reserve, and if yes, whether microplastics have increased over time. We hypothesized that microplastic pollution is present and has significantly increased in abundance over time. To be conservative, all fibers were excluded from our counts, as we suspected they were largely the result of contamination. We also factored in our 65% microplastic recovery rate to account for the plastic lost during our lab procedure. Despite our conservative microplastic counts, we found microplastics in all our samples and determined that microplastics significantly increased in abundance over time. It is important to note that we only analyzed microplastics that were larger than 250 μm; further analysis of smaller particles would likely reveal a greater extent of contamination. Nevertheless, our results support our hypothesis. Ultimately, our findings are consistent with other research, which concludes that microplastics are a pervasive, growing issue that warrants more research and action.

Acknowledgements

We would like to thank Professor Gregor Fussmann (McGill University, Department of Biology) for his guidance and support of this work. We are also grateful for the insight and ongoing assistance provided to us throughout this project by the Fall 2019 BIOL 432 teaching assistants: Alexandre Baud, Marie-Pier Hébert and Egor Katkov, as well as the Tomlinson Engagement Awardee for Mentoring, Alexa Schubak. Additionally, we appreciate the work of all the staff at the Gault House and Gault Nature Reserve who maintain the facility that allowed us to conduct this research. We would also like to extend our sincere thanks to Genevieve D'Avignon and Alex Crew for going out of their way to contribute equipment, microplastic analysis and identification protocols, and general guidance throughout this project. Finally, we are grateful to members of the Gregory-Eaves Lab for allowing us use of their lab space to complete our microplastic count and identification.

References

1. Andrady AL. Microplastics in the marine environment. Marine Pollution Bulletin. 2011 Aug 1;62(8):1596–605.

2. Eerkes-Medrano D, Thompson RC, Aldridge DC. Microplastics in freshwater systems: A review of the emerging threats, identification of knowledge gaps and prioritisation of research needs. Water Research. 2015 May 15;75:63–82.

3. Vandermeersch G, Van Cauwenberghe L, Janssen CR, Marques A, Granby K, Fait G, et al. A critical view on microplastic quantification in aquatic organisms. Environmental Research. 2015 Nov 1;143:46–55.

4. Cole M, Lindeque P, Fileman E, Halsband C, Goodhead R, Moger J, et al. Microplastic Ingestion by Zooplankton. Environmental Science & Technology 2013 Jun 18;47(12):6646–55.

5. Arii K, Hamel BR, Lechowicz MJ. Environmental correlates of canopy composition at Mont St. Hilaire, Quebec, Canada. Torrey Botanical Society. 2005;132(1):90–102.

6. Centre de la Nature Mont Saint-Hilaire. Flora and fauna [Internet]. Le Mont Saint-Hilaire. 2008. Available from: https://web.archive.org/web/20080630222323/http://www.centrenature.qc.ca/en/mountai n/faunaflora.html

7. Gault Nature Reserve. Flora [Internet]. 2008. Available from: https://

web.archive.org/web/20081022110133/http:/www.mcgill.ca/gault/saint-hilaire/natural/flora/

8. Allen S, Allen D, Phoenix VR, Le Roux G, Durántez Jiménez P, Simonneau A, et al. Atmospheric transport and deposition of microplastics in a remote mountain catchment. Nature Geoscience. 2019 May;12(5):339–44.

9. Dris R, Gasperi J, Saad M, Mirande C, Tassin B. Synthetic fibers in atmospheric fallout: A source of microplastics in the environment? Marine Pollution Bulletin. 2016 Mar 15;104(1):290–3.

10. Zhang Y, Gao T, Kang S, Sillanpää M. Importance of atmospheric transport for microplastics deposited in remote areas. Environmental Pollution. 2019 Nov 1;254:112953.

11. Boucher J, Friot D. Primary microplastics in the oceans: A global evaluation of sources [Internet]. Gland, Switzerland: IUCN; 2017. Available from: https://portals.iucn.org/library/sites/library/files/documents/2017-002.pdf

12. Reserve's History - Gault Nature Reserve [Internet]. [cited 2020 Feb 29]. Available from: https://gault.mcgill.ca/en/the-reserve/detail/reserves-history/

13. Fondation Hydro-Quebec pour l'environnment. Annual Report 2017 [Internet]. 2017. Available from: http://www.hydroquebec.com/data/fondation-environnement/pdf/English_Annual_Report_2017.pdf

14. Duval, Martin, Maneli, David, Malka, Eric, Poirier, Genevieve. SPF Application Form: Provide a transit oriented and community access to Gault Nature Reserve (GNR) [Internet]. McGill University; 2017. Available from: https://mcgill.ca/sustainability/files/sustainability/17-320_gault-communityaccess_aug2017.pdf

15. Geyer R, Jambeck JR, Law KL. Production, use, and fate of all plastics ever made. Science Advances. 2017 Jul 1;3(7):e1700782.

16. Corcoran PL, Norris T, Ceccanese T, Walzak MJ, Helm PA, Marvin CH. Hidden plastics of Lake Ontario, Canada and their potential preservation in the sediment record. Environmental Pollution. 2015 Sep 1;204:17–25.

17. Liu Y, Ma T, Du Y. Compaction of Muddy Sediment and Its Significance to Groundwater Chemistry. Procedia Earth and Planetary Science. 2017 Jan 1;17:392–5.

18. Committee on National Needs for Coastal Mapping and Charting. A Geospatial Framework for the Coastal Zone: National Needs for Coastal Mapping and Charting [Internet]. Washington, D.C.: The National Academies Press; 2004 [cited 2020 Jan 23]. Available from: https://www.nap.edu/read/10947/chapter/4

19. Kwak JI, An Y-J. The current state of the art in research on engineered nanomaterials and terrestrial environments: Different-scale approaches. Environmental Research. 2016 Nov 1;151:368–82.

20. Kalogerakis N, Karkanorachaki K, Kalogerakis GC, Triantafyllidi EI, Gotsis AD, Partsinevelos P, et al. Microplastics Generation: Onset of Fragmentation of Polyethylene Films in Marine Environment Mesocosms. Frontiers in Marine Science [Internet]. 2017 [cited 2020 Jan 23];4. Available from: https://www.frontiersin.org/articles/10.3389/fmars.2017.00084/full

21. Ohio Environmental Protection Agency. Inland Lakes Sampling Procedure Manual [Internet]. Ohio; 2010 p. 1. (Manual of Ohio EPA Surveillance Methods and Quality Assurance Practices). Available from: https://epa.ohio.gov/portals/35/inland_lakes/Lake%20Sampling%20Procedures-Final42910.pdf

22. Heggen MP, Birks HH, Heiri O, Grytnes J-A, Birks HJB. Are fossil assemblages in a single sediment core from a small lake representative of total deposition of mite, chironomid, and plant macrofossil remains? Journal of Paleolimnology. 2012 Dec;48(4):669–91.

23. Dearing JA. Lake sediment records of erosional processes. Hydrobiologia. 1991 May 1;214(1):99–106.

24. Drevnick PE, Cooke CA, Barraza D, Blais JM, Coale KH, Cumming BF, et al. Spatiotemporal patterns of mercury accumulation in lake sediments of western North America. Science of The Total Environment. 2016 Oct 15;568:1157–70.

25. Blais JM, Kalff J. The influence of lake morphometry on sediment focusing. Limnology and Oceanography. 1995;40(3):582–8.

26. Google Earth Pro [Internet]. Gault Nature Reserve, Quebec; 2017. Available from: https://www.google.com/maps/@45.5447659,-73.1510673,1303m/data=!3m1!1e3

27. Gélinas Y, Lucotte M, Schmit J-P. History of the atmospheric deposition of major and trace elements in the industrialized St. Lawrence Valley, Quebec, Canada. Atmospheric Environment. 2000 Jan 1;34(11):1797–810.

28. Holzbecher E. Advective flow in sediments under the influence of compaction. Hydrological Sciences Journal. 2002 Aug 1;47(4):641–9.

29. Crew A, Gregory-Eaves I, Ricciardi A. Distribution, abundance, and diversity of microplastics in the upper St. Lawrence River. Environmental Pollution. 2020 Jan 14;113994.

30. Klein, Rolf. Material properties of plastics. In: Laser Welding of Plastics [Internet]. 1st ed. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2011. p. 17–22. Available from: https://application.wiley-vch.de/books/sample/3527409726_c01.pdf

31. Carr SA, Liu J, Tesoro AG. Transport and fate of microplastic particles in wastewater treatment plants. Water Research. 2016 Mar 15;91:174–82.

32. Tunçer S, Artüz OB, Demirkol M, Artüz ML. First report of occurrence, distribution, and composition of microplastics in surface waters of the Sea of Marmara, Turkey. Marine Pollution Bulletin. 2018 Oct 1;135:283–9.

33. Crichton EM, Noël M, Gies EA, Ross PS. A novel, density-independent and FTIR-compatible approach for the rapid extraction of microplastics from aquatic sediments. Royal Society of Chemistry. 2017;9(9):1315–528.

34. De Witte B, Devriese L, Bekaert K, Hoffman S, Vandermeersch G, Cooreman K, et al. Quality assessment of the blue mussel (Mytilus edulis): Comparison between commercial and wild types. Marine Pollution Bulletin. 2014 Aug 15;85(1):146–55.

35. ICES. OSPAR request on development of a common monitoring protocol for plastic particles in fish stomachs and selected shellfish on the basis of existing fish disease surveys [Internet]. 2015. Available from: http://www.ices.dk/sites/pub/Publication%20Reports/Advice/2015/Special_Requests/OSPAR_PLAST_advice.pdf

36. Devriese LI, van der Meulen MD, Maes T, Bekaert K, Paul-Pont I, Frère L, et al. Microplastic contamination in brown shrimp (Crangon crangon, Linnaeus 1758) from coastal waters of the Southern North Sea and Channel area. Marine Pollution Bulletin. 2015 Sep 15;98(1):179–87.

37. Verschoor AJ. Towards a definition of microplastics [Internet]. National Institute for Public Health and the Environment; 2015. Available from: https://www.rivm.nl/bibliotheek/rapporten/2015-0116.pdf

38. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.

39. Guisan A, Edwards TC, Hastie T. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modeling. 2002 Nov 30;157(2):89–100.

40. Kowalski CJ. On the Effects of Non-Normality on the Distribution of the Sample Product-Moment Correlation Coefficient. Journal of the Royal Statistical Society, Series C (Applied Statistics). 1972;21(1):1–12.

41. Kerlin K. Microplastics: Not Just an Ocean Problem [Internet]. Science and Climate. 2019 [cited 2020 Jan 26]. Available from: https://climatechange.ucdavis.edu/news/microplastics-not-just-an-ocean-problem/

42. Toumi H, Abidli S, Bejaoui M. Microplastics in freshwater environment: the first evaluation in sediments from seven water streams surrounding the lagoon of Bizerte (Northern Tunisia). Environmental Science and Pollution Research. 2019 May 1;26(14):14673–82.

43. Fu Z, Wang J. Current practices and future perspectives of microplastic pollution in freshwater ecosystems in China. Science of The Total Environment. 2019 Nov 15;691:697–712.

44. Dikareva N, Simon KS. Microplastic pollution in streams spanning an urbanisation gradient. Environmental Pollution. 2019 Jul 1;250:292–9.

45. Edo C, Tamayo-Belda M, Martínez-Campos S, Martín-Betancor K, González-Pleiter M, Pulido-Reyes G, et al. Occurrence and identification of microplastics along a beach in the Biosphere Reserve of Lanzarote. Marine Pollution Bulletin. 2019 Jun 1;143:220–7.

46. Panti C, Giannetti M, Baini M, Rubegni F, Minutoli R, Fossi MC. Occurrence, relative abundance and spatial distribution of microplastics and zooplankton NW of Sardinia in the Pelagos Sanctuary Protected Area, Mediterranean Sea. Environmental Chemistry. 2015 Sep 1;12(5):618–26.

47.Blašković A, Fastelli P, Čižmek H, Guerranti C, Renzi M. Plastic litter in sediments from the Croatian marine protected area of the natural park of Telaščica bay (Adriatic Sea). Marine Pollution Bulletin. 2017 Jan 15;114(1):583–6.

48. Lusher AL, Tirelli V, O'Connor I, Officer R. Microplastics in Arctic polar waters: the first reported values of particles in surface and sub-surface samples. Scientific Reports. 2015 Oct 8;5(1):1–9.

49. Zarfl C, Matthies M. Are marine plastic particles transport vectors for organic pollutants to the Arctic? Marine Pollution Bulletin. 2010 Oct 1;60(10):1810–4.

50. Matsuguma Y, Takada H, Kumata H, Kanke H, Sakurai S, Suzuki T, et al. Microplastics in Sediment Cores from Asia and Africa as Indicators of Temporal Trends in Plastic Pollution. Archives of Environmental Contamination and Toxicology. 2017 Aug 1;73(2):230–9.

51. Turner S, Horton AA, Rose NL, Hall C. A temporal sediment record of microplastics in an urban lake, London, UK. Journal of Paleolimnology. 2019 Apr 1;61(4):449–62.

52.Willis KA, Eriksen R, Wilcox C, Hardesty BD. Microplastic Distribution at Different Sediment Depths in an Urban Estuary. Frontiers in Marine Science [Internet]. 2017 [cited 2020 Mar 1];4. Available from: https://www.frontiersin.org/articles/10.3389/fmars.2017.00419/full

53. Tsuchiya M, Nomaki H, Kitahashi T, Nakajima R, Fujikura K. Sediment sampling with a core sampler equipped with aluminum tubes and an onboard processing protocol to avoid plastic contamination. MethodsX. 2019 Jan 1;6:2662–8.

54. Paul-Pont I, Lacroix C, González Fernández C, Hégaret H, Lambert C, Le Goïc N, et al. Exposure of marine mussels Mytilus spp. to polystyrene microplastics: Toxicity and influence on fluoranthene bioaccumulation. Environmental Pollution. 2016 Sep 1;216:724–37.

55. Gandara e Silva PP, Nobre CR, Resaffe P, Pereira CDS, Gusmão F. Leachate from microplastics impairs larval development in brown mussels. Water Research. 2016 Dec 1;106:364–70.

Submitted: 01/15/20

Research Article

[1]Department of Psychology, McGill University, Montreal, QC, Canada

Keywords

Bat-and-ball problem, attribute substitution, dual process theory, replication study

Email Correspondence

alexandra.machalani@mail.mcgill.ca

Alexandra Machalani[1], Amanda Gallant[1], Victoria Orha[1], Nora Ostbo[1], Eric Hehman[1]

Substitution Sensitivity and the Bat-and-Ball Problem: A Direct Replication of De Neys *et al.* (2013)

Abstract

Background: Cognitive misers are no happy fools. Earlier findings (1) came to this conclusion by assessing people's sensitivity to attribute substitution, which they defined as the situation that occurs when we are confronted with a problem that demands greater cognitive effort, for which we rely on automatic and intuitive processes that substitute the complex situation for an easier one.

Methods: Through the exploration of the "bat-and-ball" problem, (2) De Neys, Rossi, and Houdé (1) found that participants were indeed sensitive to the substitution bias. Specifically, participants who incorrectly answered the question that gave rise to the substitution bias were significantly less confident in their answer relative to their answer on a control problem that did not give rise to the substitution. Using the same methods, we conducted a direct replication study on a sample of 264 undergraduate psychology students.

Results and Conclusion: Our results suggest that we successfully replicated the original conclusions; participants who answered by substituting the difficult question for an easier one significantly ($p<.0001$) decreased their confidence ratings on the version of the problem that gave rise to the substitution bias, relative to the problem that did not.

Limitations: Though there may have been limitations, it seems that we are sensitive to attribute substitution.

Introduction

As social creatures, humans have always been keen on finding the simplest, most quick and effective solutions. One could even say that we are cognitive misers; we not only engage in but rely on fast and intuitive processing rather than more effortful and deliberate thinking. First conceptualized by Peter Wason and Jonathan Evans, (3) this dual process theory proposes that we engage in two distinct types of processes: heuristic, in which an individual chooses which information is relevant and filters out the irrelevant information, and analytic, in which they analyze the information relevant for further processing. (4) This first theoretical proposition allowed for further research in the field of human cognition with the emergence of impactful theories such as Daniel Kahneman's (5) dual systems processes, which distinguishes processes as either intuition or reasoning based. A year later, Fritz Strack and Roland Deutsch (6) proposed their own dual process theory which puts forth the reflective and impulsive systems of the human mind. The model relies on separate systems they call the reflective and impulsive systems, where decisions are made using knowledge that comes from the situation or by using existing schemas without conscious thought. While faster cognitive processes might seem advantageous when we desire quick and effective problem-solving, they sometimes fall short when we are faced with complex situations that demand deeper reasoning. These processes thus lead us to biases in our judgment that we fail to recognize… or so it seems.

It has been theorized that the underlying problem of most cognitive biases and perceptual illusions is one of attribute substitution, a psychological process also known as the substitution bias. (1,2) Specifically, this concept is different from the decision-making strategy known as satisficing, as the latter entails "searching through the available options just long enough to find one that reaches a preset threshold of acceptability". (7p670) The substitution bias was first argued by Shane Frederick and Daniel Kahneman, (2) who based their theory on early research on the representativeness and availability heuristics. They relied on the hypothesis that "when confronted with a difficult question people often answer an easier one instead, usually without being aware of the substitution". (2p53) De Neys, Rossi, and Houdé (1) similarly define substitution bias, and revisit the 'bat-and-ball' problem first posed by Frederick and Kahneman (2):

> "A bat and a ball together cost $1.10. The bat costs $1 more than the ball. How much does the ball cost?" (1p269)

The reasons for the high rate of errors in this easy problem lie in the fact that people are not accustomed to thinking hard, and instead trust judgment that comes to mind quickly, especially in situations in which they are not deeply committed. (2p58-59) Further explanations rely on the substitution of the critical relational statement 'more than' with an easier problem without the relational statement. Academics (2,8) who have explored the attribution bias concluded that the ultimate problem is that substitution is not detected. De Neys *et al.* however have challenged this assumption and believed that not deliberately reflecting upon one's responses does not necessarily imply that the substitution process is undetected. (1)

The original study

The authors hypothesized that people who engage in attribute substitution must be somewhat sensitive to their error process. (1) More specifically, people should have some unconscious awareness that the substituted '10-cent' answer they gave must not be completely accurate. To study this, De Neys *et al.* (1) relied on the assumption that when people engage in attribute substitution on the bat-and-ball problem, they will most commonly incorrectly respond in a specific, pre-determined way ('10 cents'). Because they were interested in one's sensitivity, they designed an isomorphic control problem (without the "more-than" relational statement) in order to obtain base rates of people's confidence levels in a similar problem that would not give rise to substitution:

> "A magazine and a banana together cost $2.90. The magazine costs $2. How much does the banana cost?" (1p260)

A total of 248 undergraduate students who took an introductory course

in psychology were recruited to take part in the study. All participants completed both the control and standard versions of the problem in randomly assigned orders where half would start with the control version. After each problem was solved, they were asked to rate how confident they were that their answer was correct on a scale from 0% to 100%. To support their hypothesis, the authors expected their data to show a significant decrease in confidence levels when biased reasoners rated the accuracy of their answer to the standard problem, relative to that of the control problem. However, if people are completely unsensitive to the substitution, as suggested by previous research, their confidence levels following the standard problem should not differ from that of the control problem.

De Neys *et al.* (1) found that 21% (SE=2.3%) of participants successfully solved the standard bat-and-ball problem (i.e. 5-cents answer) while 99.5% of those who answered incorrectly responded in line with the substitution bias (i.e. 10-cents answer). On the control version of the problem, 98% (SE=1%) of participants gave the correct answer. This indeed supports the theory that people do not just randomly guess when they are faced with complex questions, but rather solve the problem in an automatic and intuitive way. In line with their hypothesis, the data showed that those who engaged in substitution on the standard problem also rated their confidence levels as being significantly lower ($p<.0001$, $\eta 2p=.23$) than on the control problem. These findings thus support their hypothesis in that biased reasoners are sensitive to the substitution rather than being completely blind, relative to those who resisted the substitution and gave greater confidence ratings.

These results are interpreted by the authors as a clear example of how we tend to "minimize cognitive effort and stick to mere intuitive processing" and that "cognitive misers might have more accurate intuitions about the substitution process than hitherto believed". (1p271) Additionally, they argue that this process of attribute substitution might be an explanation for other cognitive biases, such as the base-rate neglect or the conjunction fallacy. (1)

The current study

The current era of the 'replication crisis' has led to a widely felt sense of uncertainty in the scientific community, especially in experimental psychology. In his manifesto, Chris Chambers (9) offers seven sins of psychological inquiry, the third of which is the sin of unreliability. He critiques the field's reluctance to replicate and the tendency to easily dismiss different outcomes as untrue replications when replications do occur. Therefore, as part of the Collaborative Replications and Education Project (CREP), the present paper will explore and discuss a direct replication of De Neys *et al.*'s substitution sensitivity study. (1) Along with the original hypothesis, an additional hypothesis was analyzed. Ethics approval for the study was received on October 22, 2019 by McGill University's Research Ethics Board.

Additional hypothesis

To add on to De Neys *et al.*'s original hypothesis, (1) we were interested in assessing participants' self-esteem. Specifically, we looked at optimism, a correlate of self-esteem, defined as "a belief that desirable outcomes are attainable". (10p517) In fact, Hale *et al.* explain that optimists have "a favorable outlook on life, expect things to go their way, and believe that good rather than bad things will happen to them". (10p271) Because it has been found to be related to psychological well-being, such as lowering levels of depression (11) and neuroticism, (12) we were intrigued by the idea of capturing a possible correlation between optimism and substitution sensitivity. More specifically, we wanted to investigate whether optimism can influence someone's perception of their own performance on tasks, and whether it has an important effect on their confidence. As such, we have hypothesized that participants who fall victim to the substitution bias – and therefore have lower confidence ratings in the standard condition of the De Neys (1) problem – will be more inclined to rate their expectancy for success lower, relative to those who are not affected by attribute substitution. Interestingly, if a correlation were to emerge and original findings were to be replicated, we might be able to conclude that these preliminary findings suggest that only a specific population is more sensitive to the substitution bias, such as those who are less optimistic, and more neurotic.

Methods

Participants

A total of two-hundred and sixty-four (N=264) undergraduate students from McGill University completed our study, or sixteen more than the original study. Participants were recruited through SONA, an online portal that manages voluntary study participation for the McGill Psychology Human Participation Pool in exchange for course credit(s). Inclusion criteria for our study were (i) participation for psychology course credit, (ii) undergraduate student status, and (iii) being 18 years or older. In line with the De Neys *et al.* (1) study, all participants were undergraduate students who have taken psychology courses, which is important to consider when inferring the generalizability of the present and other experiments investigating substitution sensitivity.

All participants read through the consent form explaining the purpose and contents of the study and indicated that they were informed and were voluntarily participating in our study. It was also made clear that participants could end the study at any time and would still be receiving compensation. would fit the background models, fit the source parameters and obtain their resulting "test statistics". The test statistic value is a way to quantify the quality of the maximum likelihood fit, and it roughly represents $\sigma 2$ significance for a normal ("Gaussian") distribution - so a larger value implies a higher likelihood of a gamma-ray signal. The formula for the test statistic is

Procedure

As this was a direct replication study, the detailed procedure that was made available in the original paper was closely followed. However, due to time constraints, an important difference should be noted: consistently across all testing sessions, participants had to complete our study following another replication study by Griskevicious, Tybur, and Bergh. (13) The experiment took the form of a computer-based survey through Qualtrics, completed in-person and in the presence of other participants. Sessions ranged from testing 5 participants to 20 participants at a time. Because the original study had not specified data collection methods, this was done to maximize sample size.

All testing sessions were completed in the same computer laboratory at McGill University's downtown campus and were in the presence of two experimenters (although a few sessions were conducted by one experimenter).

After students had completed the Griskevicious, Tybur, and Bergh (13) study and had read and agreed to our consent form, Qualtrics automatically randomly assigned them to one of four conditions. Following the original De Neys repeated-measures design, (1) all participants completed the control version of the bat-and-ball problem as well as a variation of the standard problem. Here, the superficial item content of the original bat-and-ball problem was modified to minimize surface similarity. The standard problem that was presented to participants was as follows:

> "A pencil and an eraser together cost $1.10.
> The pencil costs $1 more than the eraser.
> How much does the eraser cost?"

The four possible conditions included a control and a standard version using either the pencil/eraser combination or the magazine/banana combination. Consequently, about half of the participants completed the control version first, and the other half completed the standard question first. After answering each problem, participants were asked to type in how confident they were that their answer was correct on a scale of 0% (totally not sure) to 100% (totally sure). After the replication portion of the study was

completed, participants were then redirected to our additional measure, the GESS-R (see section 2.3), followed by basic demographic questions. Finally, participants were debriefed and were informed of the right answer to the standard problem.

Materials and measures

As this was a computer-based study, participants were not provided with any additional materials other than the computers available to them in the laboratory. A URL was presented to them on a whiteboard and students were asked to access the survey by themselves.

In addition to the original study's materials, the Revised Generalized Expectancy for Success Scale (GESS-R) (10) was used to measure optimism. Originally developed in 1978 by Fibel & Hale, (11) we used the 1992 revised scale published in the Journal of Clinical Psychology. The GESS-R is designed to measure optimism specifically in one's expectations for successful outcomes. Psychometric measures from the 1992 report suggests an acceptable level of reliability over time as well as a high level of internal consistency. Importantly, the GESS-R was not found to be correlated to neuroticism, giving way for a better interpretation of the results as attributable to optimism. Permission for the use and reproduction of the GESS-R was granted on October 27, 2019 by Dr. Daniel Hale.

To recall our main hypothesis, we expect biased reasoners on the standard problem to rate their confidence levels as significantly lower than their confidence levels on the control problem. In other words, the confidence rating in the correctness of one's answer to the standard problem was the main dependent variable that was being measured. Indirectly, confidence levels are meant to measure the latent construct of substitution sensitivity, where those who engage in substitution are shown to be sensitive by feeling less confident in their answer to a complex problem, relative to an easier problem. Additionally, we are also assessing optimism as a potential moderator of the relationship between confidence ratings and sensitivity to substitution. If data were to support our hypothesis that those who are less optimistic would perceive their answers as less accurate – thus being more sensitive to substitution – we would expect to see a greater decrease in confidence ratings in participants who score lower than normal on the GESS-R.

Data processing and analysis

In line with the original study and the CREP-provided step-by-step methods, separate analyses were conducted. To assess the possible differences in confidence levels, participants were grouped according to accuracy in their response on the standard problem. Specifically, accuracy was coded as "biased" (i.e. '10-cents' answer) and "correct" (i.e. '5-cents' answer). From this, the data was analyzed using a repeated-measures ANOVA, as this experiment was a two by two within-subjects design, in order to assess the differences in confidence levels of biased and unbiased reasoners on both control and standard problems.
For our additional hypothesis, a bivariate Pearson correlation analysis was performed in order to determine if there is a correlation between confidence levels in the standard problem and GESS-R score. Alpha values to determine statistical significance were set at 5% for both the original and additional hypotheses. Analyses were carried out by EH on IBM SPSS Statistics software version 26.0. (14)

Results

Participant characteristics

The majority of participants (84.47%) were cisgender females and 12.88% were cisgender males, where 'cisgender' was defined as identifying to one's biological sex assigned at birth. It is important to note that upon request, students were provided with a brief explanation of the term 'cisgender'. Among the sample, 50.8% were Caucasian, 20.5% were Asian, 10. 6% were of mixed ethnicities, and, the remaining 18.1% were Arab or Middle Eastern, Hispanic or Latin American, or Black or African American (Table 1).

Accuracy of responses

In agreement with the original study's results, 97.72% of participants answered the control problem correctly (Fig. 1). The other seven people who answered the control problem incorrectly were subsequently excluded from further analyses.

In the standard problem, results were not as dramatically distinct: 50.38% of participants answered correctly, while 43.94% answered incorrectly by engaging in a substitution (Fig. 2). The other fifteen people who answered incorrectly to the standard problem in a way that was not in line with the substitution bias were subsequently excluded from further analyses. Although the data suggests that it is not the majority who fall into the trap of attribute substitution, it does however show that the standard question is much trickier than the control. If the standard question would not give rise to some sort of confusion, results would be much more heterogeneous in which there would be an unequal split between correct and incorrect answers among participants.

	Number of participants *(N)*	Percent *(%)*	Mean *(M)*
Age			20.42
Ethnicity			
Asian	20	7.6	
Arab or Middle-Eastern	54	20.5	
Black or African American	8	3.0	
Caucasian or White	134	50.8	
Hispanic or Latin American	12	4.5	
Mixed	28	10.6	
Other	8	3.0	
Gender			
Cisgender male	34	12.88	
Academic Year			
U0	10	3.79	
U1	83	31.44	
U2	80	30.30	
U3	75	28.41	
U4	16	6.06	
First Language			
Arabic	12	4.55	
English	122	46.21	
French	50	18.94	
Mandarin	25	9.47	
Spanish	11	4.17	

Table 1. Demographics of participants.

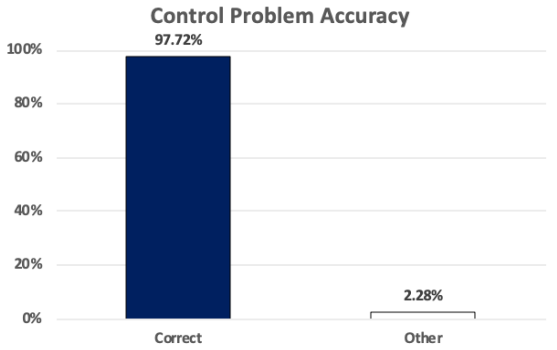

Figure 1. Accuracy of participants' answers on the control problem.

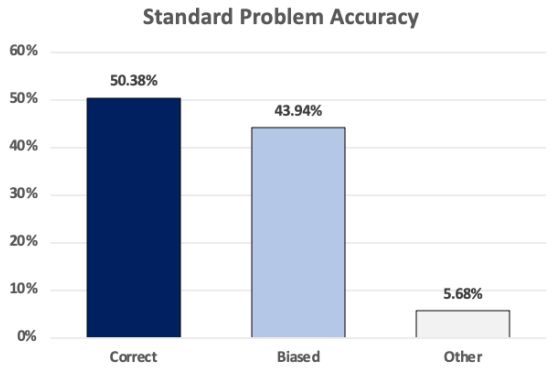

Figure 2. Accuracy of participants' answers on the standard problem.

Confidence ratings

Consistent with the central hypothesis, results show that biased reasoners (participants who engaged in the substitution bias) rated their confidence levels as significantly lower on the standard problem compared to the control problem, $F(1, 114)=31.62$, $p<0.0001$. In participants who correctly answered the standard problem, there was a negligible confidence discrepancy between the two conditions (Fig. 3).

Analyses of optimism

Following a bivariate correlation analysis between GESS-R score and confidence levels, it appeared that there was no significant correlation between variables ($r=0.006$, $p=0.929$). Hence, optimism as we have defined it has not been found to be correlated with confidence levels on either the standard or control problem.

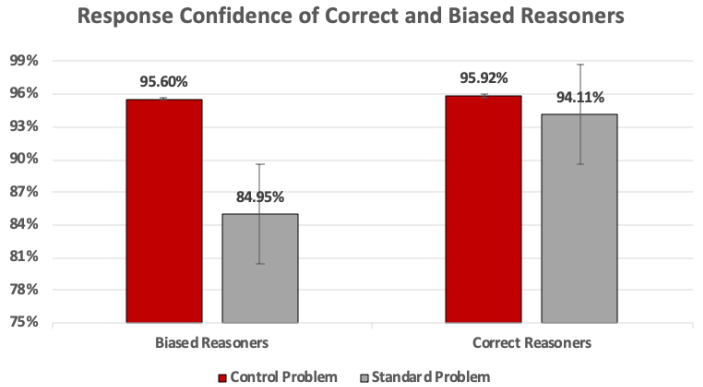

Figure 3. Response confidence on the standard and control versions of the problem for participants who answered the standard problem incorrectly (biased) and correctly (correct). Error bars are standard errors.

Discussion

The present data suggests that we have successfully replicated the results of De Neys *et al.*'s original paper. (1) First, we have replicated the way in which we obtained significant results in the same direction by using the same confidence levels as the original study ($\alpha=95\%$). Second, we have replicated the observed significant decrease in biased reasoners' confidence ratings on the standard problem relative to the control. In line with this, we have also replicated the narrow gap in confidence ratings of correct reasoners in both conditions. Finally, we have replicated the magnitude of answers in the control problem, where the majority of participants answered correctly.

What we have not replicated, however, is the percentage of people who incorrectly answered the standard problem. De Neys *et al.* (1) found that only 21% got the standard question correct, while we found that 50.38% got the standard question correct. Possible confounds to this could be explored, such as the possibility that students who took part in the study were from different cultural backgrounds, as well as in different academic levels (Table 1). It may be that some participants had already been exposed to this type of question through preparing for the Graduate Record Examination (GRE), for instance.

It is therefore safe to conclude that we have successfully replicated De Neys *et al.*'s original Bats, balls, and substitution sensitivity study, (1) but with certain conditions. Our results support the theory that we are not completely blind to questionable cognitive processes, and that we might actually be sensitive to the fact that we are giving unreliable answers to complex questions. What this could suggest is that when we are faced with difficult situations, we tend to prefer less effortful ways of reasoning. In other words, we prefer minimizing our cognitive effort and resort to intuitive processing instead.

The results for our additional hypothesis demonstrated that optimism did not have a significant effect on the confidence ratings. The negative findings regarding the construct imply that it does not have a significant impact on the way people perceive their confidence regarding their performance. In the current literature, however, optimism has been known to impact future performance positively. (15) It is possible that our results may therefore indicate that when individuals assess their performance immediately after completing a task, they tend to be more realistic than optimistic. Future research should further investigate the relationship between optimism and both immediate and future performance on tasks.

Limitations

Although most of our data supports our hypothesis, there are a few limitations to consider. First, the way confidence was measured might not be entirely appropriate. Because confidence is a latent construct, we cannot be sure that it is confidence levels in participants' answers that we are measuring. Uncertainty in a measure's construct validity poses a threat to the interpretation of our results, especially since this is a self-reported measure. Further, confidence levels in one's answer are meant to measure substitution sensitivity. It may be that being more or less confident in one's answer does not represent sensitivity to attribute substitution, but rather confidence in one's answer due to mathematical skills, complex wording of the problem, or any other confound not measured. Specifically, it is interesting to think about how the wording of the question might impact a participant's confidence ratings, since the only difference between the standard and the control problem is the use of the relational statement "more than". The former is used to elicit a specific incorrect response from participants, which in turn should decrease their confidence in their answer. A second limitation is that external validity in this study is not substantial. Because this sample was a convenience sample, it may be that these results are only of interest for young psychology undergraduates. In addition, the time at which the study was offered poses a threat to external validity. In fact, the Fall semester is often very stressful when undergraduates can be overwhelmed with work, study, and applications to graduate school.

As for our additional hypothesis using the GESS-R to assess optimism, we

expected a pattern of results showing that higher scores on the GESS-R would have influenced confidence ratings in both conditions in a positive direction, as well as a high confidence levels in biased reasoners. Although we could easily conclude that optimism is unrelated to confidence levels and substitution sensitivity, it is important to note that we had no base rates to compare participants' scores to. It would have been more powerful to assess their optimism score before exposing them to potentially threatening situations, which might decrease confidence in one's abilities for a short period of time, causing optimism levels to also decrease.

Implications and future directions

Future research wanting to explore possible covariates affecting confidence levels and substitution bias may want to include certain personality characteristics such as impulsivity or neuroticism. It may be that those who are more impulsive or neurotic tend to rate themselves as more or less confident respectively regardless of the type of condition they are in. Other possible variables that could be tested is simple test anxiety, where examiners can send a scale assessing people's anxiety when taking tests, or specifically arithmetic examinations to see if these people are more likely to get answers wrong or rate themselves as less confident.

The implications of this type of research for the field of human cognition is important to consider. Recent work has shown that biased reasoners on the bat-and-ball problem do not lazily monitor their intuitive reasoning. (16) Rather, they do evaluate quite automatically their substituted answers by answering the mathematical and the confidence questions much more slowly than on the control question. The authors suggest that it all boils down to the relational term "more than", which causes conflict between our automatic, linguistic operations and our quick, intuitive reasoning processes, thus providing this semantic awareness that we have not fully complied with the relational terms of the sentence. (16) Thus, future research in this interdisciplinary area of cognitive awareness and decision-making will give deeper insight into the ways we are not entirely "happy fools".

Acknowledgements

Dr. Eric Hehman carried out statistical analyses, along with Psychology graduate student Eugene Ofosu. We would like to thank the Collaborative Replications and Education Project and the McGill Research Ethics Board for approving our replication project. We would also like to thank anonymous reviewers for their helpful and constructive feedback.

References

1. De Neys W, Rossi S, Houdé O. Bats, balls, and substitution sensitivity: cognitive misers are no happy fools. Psychonomic Bulletin & Review. 2013;20(2):269–73.

2. Kahneman D, Frederick S. Representativeness Revisited: Attribute Substitution in Intuitive Judgment. Heuristics and Biases. 2002, Aug;49–81.

3. Wason P, Evans J. Dual processes in reasoning? Cognition. 1974;3(2):141–54.

4. Evans JSBT. Heuristic and analytic processes in reasoning*. British Journal of Psychology. 1984;75(4):451–68.

5. Kahneman D. A perspective on judgment and choice: Mapping bounded rationality. American Psychologist. 2003Sep;58(9):697–720.

6. Strack F, Deutsch R. Reflective and Impulsive Determinants of Social Behavior. Personality and Social Psychology Review. 2004Aug1;8(3):220–47.

7. Colman AM. A dictionary of psychology [Internet]. 2nd ed. Oxford: Oxford University Press; 2006 [cited 2020Mar2]. Available from: archive.org

8. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. Science. 1974;185(4157):1124–31.

9. Chambers C. The seven deadly sins of psychology: a manifesto for reforming the culture of scientific practice. Princeton, NJ: Princeton University Press; 2017.

10. Hale WD, Fiedler LR, Cochran CD. The revised Generalized Expectancy for Success Scale: A validity and reliability study. Journal of Clinical Psychology. 1992;48(4):517–21.

11. Fibel B, Hale WD. The Generalized Expectancy for Success Scale: A new measure. Journal of Consulting and Clinical Psychology. 1978;46(5):924–31.

12. Smith TW, Pope MK, Rhodewalt F, Poulton JL. Optimism, neuroticism, coping, and symptom reports: An alternative interpretation of the Life Orientation Test. Journal of Personality and Social Psychology. 1989;56(4):640–8.

13. Griskevicious V, Tybur JM, Bergh BVD. Going green to be seen: Status, reputation, and conspicuous conservation. Journal of Personality and Social Psychology. 2010;98(3):392–404.

14. IBM. SPSS statistics for Windows. Version 26.0. Armonk, NY: IBM; 2019.

15. Loftus TJ, Filiberto AC, Rosenthal MD, Arnaoutakis GJ, Sarosi GA, Dimick JB, et al. Performance advantages for grit and optimism. The American Journal of Surgery. 2020.

16. Johnson ED, Tubau E, Neys WD. The Doubting System 1: Evidence for automatic substitution sensitivity. Acta Psychologica. 2016Dec23;164:56–64.

Submitted: 01/10/20

Research Article

[1]Department of Psychology, McGill University, Montreal, QC, Canada

Keywords

Motivation, life history theory, individual differences

Email Correspondence

mana.moshkforoush@mail.mcgill.ca

Mana Moshkforoush[1], Eliane Roy[1]

Correlations of Fundamental Social Motives with Personality Measures and Life History Variables

Abstract

Background: In response to the replication crisis in the field of psychology, the authors conduct a replication of the Neel *et al.* (2016) (1) study examining individual differences in fundamental social motives.

Methods: Using the Fundamental Social Motives Inventory, we explore the relationships of the fundamental social motives to other individual differences and personality measures and the extent to which life history variables (e.g., age, sex, childhood environment) predict individual differences in the fundamental social motives. In addition to the replication study, the authors also incorporate the Behavioral Inhibition/Activation Scale (BIS/BAS) as a new variable to determine this measure of personality's correlation with all seven fundamental social motives of Self-Protection, Disease Avoidance, Affiliation, Status, Mate Seeking, Mate Retention, and Kin Care. A total of 34 participants are recruited from Amazon Mechanical Turk to complete the measures of personality under question. The replication criteria are set at ±0.15 r/β-units from the original study results and effect sizes greater than or equal to r/β=0.5 have to demonstrate statistical significance at the $p<0.05$ level.

Results: Results demonstrate that between a third and a half of all effect sizes replicate Neel *et al.*'s (1) findings.

Limitations: These results should be considered carefully with respect to the low sample size of our study.

Conclusion: The BIS/BAS variable proves to be most informative, indicating that the seven motives cluster under either the BIS or BAS factors with medium to large strengths of correlation. These findings contribute to discussions on considering the most accurate measures of social motivation and the implications of individual differences in psychology's understanding of such motivational systems.

Introduction

Social motivation is a defining feature of the cognitively gifted human; what differentiates us from other animal species confined to biological motives of a Darwinian nature. Beyond the primitive goals of obtaining food and passing on our genes to viable offspring, human motivation extends to higher levels in Maslow's hierarchy of needs and is often tied to our social nature. (2) Although an over-arching aspect of what makes us all human, the fundamental social motives are also key to understanding individual differences that lend to people's cognitive uniqueness. These individual differences in motivation – and by extension, the fundamental social motives that drive the differences – are worthy of extensive research as key predictors of perception and behavior in human psychology. (3) In order to have a descriptive/explanatory value in the study of personality, the biologically-informed fundamental social motives approach is built upon a multidisciplinary perspective of how humans have adapted to their social nature and are thus defined by Neel *et al.* (1) as "systems shaped by our evolutionary history to energize, organize and select behavior to manage recurrent social threats and opportunities to reproductive fitness". From existing literature and established theory, fundamental social motives include Self-Protection, Disease Avoidance, Affiliation, Status Seeking, Mate Seeking, Mate Retention, and Kin Care. (4) This seven-motives approach is a middle-ground between few aggregative, broad motives and many non-aggregative, specific goals. (5,6,7) The balance achieved is aimed at reflecting both distinctive and overlapping motivational inclinations in response to adaptive problems in social interactions. As part of their analytical predictability of human cognition and behavior, empirical findings suggest that activating the fundamental social motives attunes social phenomena such as stereotyping, conformity, intergroup prejudice, economic decision-making, political beliefs, self-presentation, aggression. (1) With such a wide range of functional applicability in the field of social psychology, research on the fundamental social motives approach and its links to other individual differences and personality measures is justified both in terms of scientific interest and importance.

It is intuitive that mere individual differences, as measured by various scales such as the Big Five inventory (8), would manifest as between-person variability in social motives. The social situation in which people find themselves also plays a key role in the motivational inclinations that then drive our behaviour and responses to the adaptive challenges and opportunities that social group living affords. For example, a situation containing a sexually attractive neighbor is likely to activate the fundamental social motive of Mate Seeking. However, individual characteristics and social situations alone cannot be credited for eliciting social motives. The prominence of motives in social situations is also a function of life history variables such as age, sex, relationship status, and parent status that calibrate the trade-offs faced by investing effort in particular social goals. (9) Applied to the previously-mentioned example of a sexually attractive neighbor, prominence of the Mate Seeking motive will likely vary between a 28-year-old single individual and a 58-year-old married parent. Thus, motivational inclinations are accounted for by the social situation, as well as the biological framework that describes how individuals' resource allocation changes over the course of a life-time. Life history theory (10,11) addresses trajectories and timing of shifts in the prominence of social motives, which account for significant between-person variability in motives.

In 2016, Neel *et al.* (1) published a paper on the relations among the different fundamental social motives, the relationships of the motives to other individual difference, and personality measures including the Big Five personality traits, the extent to which the motives are linked to recent life experiences, and the extent to which life history variables predict individual differences in the fundamental social motives. Their hypothesis that such relations exist addressed the prediction of individual differences in social motives based on factors that shape life history variables, thereby

providing a framework for understanding changes in social motives over the life span. Our team at McGill University sought to determine whether a direct replication of the Neel *et al.* (1) study – including analyses of fundamental social motives, conceptually related scales, and life history variables – would bear similar enough results to constitute reproducibility of the original effects found under the scientific method. The study's purpose is to provide a unifying, approach for examining individual differences at the level of fundamental social motives. By testing the hypotheses with this purpose in mind, we seek to further extend understanding of human motivation and personality. Our decision to conduct a replication as well as to add an original variable – BIS/BAS – is a response to the decade's replication crisis and serves as an effort to actively reorganize the disciplinary social structure that discourages reproducibility in the field of social psychology. (12)

Additional Variable Hypothesis

The Neel *et al.* (1) study had initially aimed for the Fundamental Social Motives Inventory to reflect both promotion and prevention of each of the seven motives. However, the distinctions between approach and avoidance behavior were not borne out in the scale development process. As defined by Ellen Crowe and E. Higgins (13), the promotion focused motivation framework is concerned with advancement, growth, and accomplishment. Promotion-focused goals are about doing something one would ideally do, theorized to ensure hits and minimize false negatives. In contrast, the avoidance focused motivation framework is concerned with security, safety, and responsibility. Prevention-focused goals are about fulfilling responsibilities and doing the things that you feel you ought to do, theorized to ensure correct rejections and minimize false positives.

We hypothesized that the Behavioral Inhibition/Activation Scale variable would correlate with the fundamental social motives in some way. Since no previous research had compared the two measures, it was not yet known which motives would correlate with either the inhibition or the activation systems, nor which direction these correlations would take. Furthermore, the addition of the BIS/BAS variable to this study is meaningful in that it adds validity to the fundamental social motives' theory since certain aspects of the two measures are inherently linked. As examples, we would expect that a participant who scores high on the Self-Protection motive would also score high on the Behavioral Inhibition Scale, and that a participant who scores high on the Mate Seeking motive would also score high on the Behavioral Activation Scale. Additionally, the comparison of the BIS/BAS measure and fundamental social motives inventory provides information about how the seven motives are similar and/or different from one another. Given that the current theory encompasses the largest set of motives ever proposed, significant clustering in the BIS and BAS dimensions could provide an argument that a more restricted set of motives might be just as good in conceptualizing human motivation. The specific and over-arching hypothesis involving the additional BIS/BAS variable: is promotion of achievement of a fundamental goal versus avoidance of failure to achieve that goal dependent on the fundamental social motive that drives that goal?

Methods

Design and participant demographics

This study is a correlational study, a type of research design where the kind of relationships naturally occurring variables have with one another is sought to be understood. Naturally occurring variables are those that have not undergone any manipulation by the researcher; in this case, all fundamental social motives, individual differences, life history variables, and behavioral inhibition/activation orientations.

We recruited participants from the Amazon Mechanical Turk (MTurk) platform, a crowdsourcing website where participants receive monetary compensation for answering surveys and/or participating in studies. For this replication study, the participants were compensated with roughly $0.12/min. Given the fact that the participants volunteered for the study and that we did not control to get a representative sample of a certain population, the sample collected is considered a convenience sample. However, since the original study used the same sampling method and the same platform, it does not constitute a limitation regarding the replication potential of the study. Demographic data was collected amongst our participants and it was found that our sample consisted of roughly equal male and female participants, mostly Christian Caucasians between the ages of 20 and 49 years old (see Table 1). In terms of level of education, household income, and political beliefs, our sample showed great diversity.

Gender					
	n	%		*n*	%
Male	19	56%	Prefer not to specify	1	3%
Female	14	41%	TOTAL	34	100%
Age			**Ethnicity**		
	n	%		*n*	%
20-29	6	18%	Caucasian	24	71%
30-39	14	41%	Asian	5	15%
40-49	6	18%	Hispanic/Latino	2	6%
50-59	3	9%	African	1	3%
60-69	3	9%	Black (I'm not African)	1	3%
70-79	2	6%	Prefer not to specify	1	3%
TOTAL	34	100%	TOTAL	34	100%
Highest level of education attained			**Religious affiliations**		
	n	%		*n*	%
High school completed or GED	4	12%	Christian	18	53%
Some college	9	26%	Atheist	10	29%
Associate's degree (2 years)	2	6%	Agnostic	3	9%
Bachelor's degree (4 years)	12	35%	Noahide	1	3%
Graduate or professional degree	7	21%	Hindu	2	6%
TOTAL	34	100%	TOTAL	34	100%
Household income			**Political beliefs**		
	n	%		*n*	%
Under $20,000	5	15%	1 very conservative	4	12%
$20,000-$29,999	6	18%	2 conservative	4	12%
$30,000-$39,999	7	21%	3 somewhat conservative	1	3%
$40,000-$49,999	2	6%	4 neutral	4	12%
$50,000-$59,999	4	12%	5 somewhat liberal	6	18%
$60,000-$79,999	5	15%	6 liberal	8	24%
$80,000-$99,999	4	12%	7 very liberal	7	21%
Decline to answer	1	3%			
TOTAL	34	100%	TOTAL	34	100%

Table 1. Demographics from the particpants of the replication study.

Procedure

Participants first responded to items assessing their relationship status and parent status, so that the Mate Retention and Kin Care (Child) scales could be presented only to those in relationships and those with children, respectively. Participants completed the Fundamental Social Motives Inventory (66-item set retained for analyses reported in original paper), the Big Five Inventory, and questions about their life experiences. By random assignment, participants then completed one of two possible sets of measures of individual differences in constructs often used to measure fundamental social motives or motive-relevant vulnerabilities and strategies: one set consisted of the Sociosexual Orientation Inventory (14), Perceived Vulnerability to Disease Scale (15), and the Dominance and Prestige Scales (16); the other set consisted of the Belief in a Dangerous World Scale (17), the Need to Belong Scale (18), and the Experiences in Close Relationships–Revised Scale. (19)

All participants then provided information on a number of life history variables, beginning with their age (continuous) and sex (male coded -1, female coded 1). Participants used the following response options to indicate their relationship status: married, in a committed relationship, dating one person, separated, divorced, and single. Only those who indicated that they were either married, in a committed relationship, or dating one person were considered "in a relationship" (coded 1), and only those who responded as single, divorced, or separated were considered "not in a relationship" (coded -1). Participants indicated whether they had children with a "yes" (coded 1) or "no" (coded -1). The childhood stability scale consisted of three items (e.g., "Compared to the average person, how [sta-

ble, predictable, hard] was your home life when you were growing up?" 1=very [stable/predictable/easy], 7=very [unstable/unpredictable/hard], reverse-coded so that higher scores reflect greater stability. The childhood resources scale consisted of four items (e.g., "My family usually had enough money for things when I was growing up," 1=strongly disagree, 7=strongly agree). Although people may not have veridical memories of childhood experiences (20), and thus their self-reported memories of childhood likely contain some error, Neel *et al.* (1) drew items used from past research that has successfully used these items to assess the influence of childhood environments on life history strategies. (21) The current resources scale consisted of two items (e.g., "I don't currently need to worry much about paying bills," 1=strongly disagree, 7=strongly agree). Finally, all participants completed the BIS/BAS scales. Results of this measure would be mapped onto three BAS-related scales and only one BIS-related scale. The instrument author does not encourage combining the BAS scales because they focus on different aspects of incentive sensitivity. (22)

Analytical strategy

The sample size n=34 was obtained after the elimination of a) any participant who did not complete the study, b) two participants who "completed" the study within 5 and 8 minutes – well below the average time of 26.7 minutes – and c) two suspected internet bots – one whose answer was tremendously irrelevant to the survey question, and another whose reported age of conception was earlier than puberty. While the researcher initially intended to recruit around 100 participants, the lack of funding did not allow them to achieve this objective. The subjective decision to eliminate these participants were made according to the principle of eliminating data that we are absolutely certain is defective. This practice is not considered p-hacking because the effects of removing versus retaining the data were never compared or considered in the decision-making process. Despite increasing our certainty regarding the remaining data, the elimination of some data contributed to the low number of our sample size, which has negative effects on the certainty of our replication conclusions.

For the fundamental social motives scale, selected items were averaged into eleven scores, one for each motive and sub-motive. The conceptually related scales – Sociosexual Orientation Inventory, Perceived Vulnerability to Disease, Dominance and Prestige, Belief in a Dangerous World, Need to Belong, and Experiences in Close Relationships (Revised) Scales – each gave a single score corresponding to an average of all scale items. As for the Life History data, each of the seven questions yielded a single item score. The scores for age, childhood stability, childhood resources, and current resources were on a continuum. The scores for sex, relationship status, and parent status were coded in a binary fashion, where men=-1, women=1; single=-1, in relationship=1; non-parent=-1, parent=1.

The researchers computed correlation coefficients between each fundamental social motive score and each additional scale, but only the ones with meaningful relationships were further analyzed and compared with the correlation coefficients found in the Neel *et al.* (1) study. Correlation coefficients were also computed between the fundamental social motives scores and the single item scores for the Life History questions. Furthermore, three scales were presented to only a subset of the participants. As such, correlation coefficients were calculated for the Mate Retention (general) and Mate Retention (breakup concern) scales only amongst participants who indicated being in a relationship (n=23). The same went for the Kin Care (child) scale, which was only presented to participants who had a child (n=23).

Results

Since the correlations of fundamental social motives with the behavioral inhibition/ activation scales were not part of the Neel *et al.* (1) study, there are no replications to be assessed. The data obtained can give us valuable preliminary information on the relationships between the motives and this social motivation framework. Beyond assessment of the correlations themselves, two intriguing patterns can be discerned. First, that the statistically significant correlations highlighted are directionally consistent, where significant Behavioral Inhibition Scale (BIS) correlation signs are in the opposite direction of signs of significant Behavioral Activation Scale (BAS) correlations (see Table 2). The second pattern is the emergence of two clusters of motives: the first group, which includes motives that correlated either negatively with BIS or positively with BAS, can be labelled as the Behavioral Activation group. Within this category, Affiliation (Group) motive significantly correlated with all BAS indicators (r=0.40) and with BIS (r=-0.39). Status motive also correlated with BAS (drive) (r=0.30) and BAS (fun seeking) while Mate Retention (General) motive correlated with BAS (reward response) (r=0.50). The second group, which includes motives that correlated either positively with BIS or negatively with BAS, can be labelled as the Behavioral Inhibition group. Within this category, Self-Protection motive (r=-0.42) and Disease Avoidance motive (r=-0.34) significantly correlated with BAS (fun seeking). Disease Avoidance motive was also associated to BIS (r=0.30). Affiliation (Independence) motive was negatively correlated with BAS (drive) (r=-0.41) and BAS (fun seeking) (r=-0.50)

Fundamental Social Motive	BAS (Drive)	BAS (Fun Seeking)	BAS (Reward Response)	BIS
	r =	*r* =	*r* =	*r* =
Self-Protection	-0.30	-0.42*	-0.21	0.30
Disease Avoidance	-0.20	-0.34*	0.00	0.30*
Affiliation (Group)	0.40*	0.40*	0.40*	-0.39*
Affiliation (Exclusion Concern)	0.05	0.16	-0.45*	0.25
Affiliation (Independence)	-0.41*	-0.50*	-0.20	0.50
Status	0.30*	0.30*	-0.20	-0.34
Mate Seeking	0.05	0.31	-0.19	0.02
Mate Retention (General)	0.24	0.02	0.50*	0.07
Mate Ret. (Breakup Concern)	0.32	0.24	0.00	0.04
Kin Care (Family)	-0.16	-0.11	-0.60*	-0.09
Kin Care (Child)	0.14	0.37	0.00	-0.40

Note. $^*p < 0.05$.

Table 2. Fundamental Social Motives as Predictors of Behavioral Inhibition and Behavioral Activation Systems.

Replication criteria

Using the Fundamental Social Motives Inventory, our team explored the relationships of the fundamental social motives to other individual difference and personality measures; the extent to which fundamental social motives are linked to recent life experiences; and the extent to which life history variables predict individual differences in the fundamental social motives. As the study being replicated committed to the statistical significance requirements of p-value and effect size requirements of r- and β-values, the 2019 replication team committed to the following requirements for what is to be considered replication: 1) correlations found must be within 0.15 r/β-units of the original effect size reported. 2) correlations found must be in the same direction as the original results; with the exception of correlations within 0.15 r/β-units of one another around r/β = 0, in which case "no effect" is replicated. 3) r/β-values above 0.5 must be significant in order to ensure validity of the correlation found in order to be considered "replicated." The first two requirements are relatively intuitive in terms of justification. If two correlations are in opposite directions, the fundamental relationships that they describe are diametrically different and therefore clearly not replications of one another. Similarly, if two correlations are far apart in effect size, they are effectively descriptions of two different correlations, just like two differently angled diagonal lines on

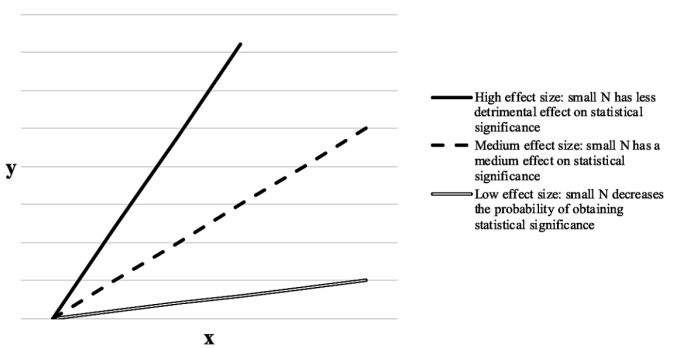

Figure 1. Theoretical representation of effect sizes in relation to statistical significance.

an x-y correlation graph (see Figure 1).

The third requirement of replication is related to the statistical effects of a small sample size. When the effect size of a correlation is large – defined here as r/β≥0.5 – a small sample size theoretically should not have a detrimental affect on the certainty of the p-value. Therefore, it is expected that a truly high strength of correlation would be significant, despite our replication study's small sample size. Whenever the effect size of a correlation is large but not significant, we may thus question the validity of the strength of correlation and exclude the finding from what we consider to be replication. Contrastingly, when the effect size of a correlation is small – defined as r/β<0.5 – a small sample size does indeed decrease the certainty of the p-value. Therefore, it is not expected that a truly low strength of correlation be significant and whenever that is the case, we may attribute the non-significance of the strength of correlation to the small sample size and consider the finding as valid anyway.

Correlations of fundamental social motives with their conceptually related scales

When comparing the fundamental social motives and the related scales, we found that four correlation coefficients out of twelve replicated: the Self-Protection motive which correlates with Belief in a Dangerous world (r=0.51), the Affiliation (Exclusion Concern) which correlate with the Need to Belong scale (r=0.77), the Mate Retention (General) which correlates with the Avoidance scale (r=-0.57), and the Mate Retention (Breakup Concern) which correlates with the Anxiety Attachment scale (r=0.96) (see Table 3). Correlation of the Status motive with the Prestige scale is the only correlation found in the opposite direction of that reported by the original study. The remaining non-replications did not meet the requirement of falling within 0.15 r/β-units of the original effect size reported.

Fundamental Social Motives	Conceptually Related Scale	r (Neel et al.)	r (CREP)
Self-Protection	Belief in a Dangerous World	0.38*	0.51*
Disease Avoidance	Perceived Vulnerability to Disease	0.64*	0.20
Affiliation (Group)	Need to Belong	0.37*	0.19
Affiliation (Exclusion Concern)	Need to Belong	0.75*	0.77*
Affiliation (Independence)	Need to Belong	-0.46*	-0.08
Status	Dominance	0.52*	0.78*
Status	Prestige	0.33*	-0.10
Mate Seeking	Sociosexual Orientation Inventory	0.44*	0.72*
Mate Retention (General)	Avoidance	-0.45*	-0.57
Mate Retention (Breakup Concern)	Anxiety	0.84*	0.96*
Kin Care (Family)	Avoidance	-0.28*	-0.71*
Kin Care (Child)	Dominance Scale	-0.28*	-0.50

Note. *$p < 0.05$.

Table 3. Zero-Order Correlations of Fundamental Social Motives with their Conceptually Related Scales

Life history predictors of fundamental social motives

Our team replicated more than half of the correlations of life history predictors of the fundamental social motives (see Table 4). Out of the 74 correlations, 40 replicated and only 11 were in the opposite direction as the original findings. The remaining 23 correlations did not meet the requirement of falling within 0.15 r/β-units of the original effect size reported by Neel *et al.* (2016). (1) The most successfully replicated life history variable correlations are with the fundamental life variables of Status and Mate Seeking – both only one correlation away from perfect replication. Interestingly enough, the most poorly replicated life history variable correlations are with the fundamental life variables of Mate Retention – both General and Breakup Concern replicating in only one correlation out of the seven life history variables. Relationship Status in Mate Retention was not analysed since, by definition, the motive to retain a mate would assume relationship status to be "in a relationship" with a β-value of 1.00. Similarly, Parent Status in Kin Care (Child) was not analysed since the motive to care for one's child would assume parent status to be "parent" with a β-value of 1.00.

	Self-Protection		Disease Avoidance		Affiliation (Group)	
	β (Neel)	β (CREP)	β (Neel)	β (CREP)	β (Neel)	β (CREP)
Age	-0.12*	-0.09	0.04	-0.20	0.00	-0.07
Sex	0.17*	0.30	0.08*	0.21	-0.01	-0.11
Relationship status	0.02	-0.24	0.00	-0.01	0.04	0.54*
Parent status	0.15*	-0.38	0.03	-0.15	0.10*	0.45*
Childhood stability	-0.04	-0.28	-0.05	-0.04	0.03	0.35
Childhood resources	0.07	0.51	0.05	0.14	0.05	-0.14
Current resources	-0.08*	-0.34	-0.02	-0.42	0.11*	0.10
	Affiliation (Exclusion Concern)		**Affiliation (Independence)**		**Status**	
	β (Neel)	β (CREP)	β (Neel)	β (CREP)	β (Neel)	β (CREP)
Age	-0.26*	-0.25	0.05	0.05	-0.20*	-0.17
Sex	0.06	0.08	0.06	0.11	-0.04	-0.05
Relationship status	-0.04	0.07	-0.10*	-0.39*	0.00	0.08
Parent status	-0.04	0.03	-0.02	-0.52*	0.04	0.09
Childhood stability	0.02	-0.33	-0.01	-0.35	-0.02	-0.36
Childhood resources	0.03	0.25	-0.01	0.08	0.11*	0.19
Current resources	-0.05	0.05	-0.05	-0.08	0.05	0.20
	Mate Seeking		**Mate Retenton (General)**		**Mate Retention (Breakup Concern)**	
	β (Neel)	β (CREP)	β (Neel)	β (CREP)	β (Neel)	β (CREP)
Age	-0.14*	-0.21	0.09*	0.26	-0.23*	-0.09
Sex	-0.17*	-0.06	0.18*	-0.09	-0.06	0.32
Relationship status	-0.52*	-0.25	N/A	N/A	N/A	N/A
Parent status	-0.04	0.08	-0.07	0.15	-0.02	-0.19
Childhood stability	-0.06*	-0.05	0.12*	0.33	-0.12*	-0.40
Childhood resources	0.04	0.11	-0.05	-0.21	0.04	0.35
Current resources	0.01	0.09	0.01	-0.06	-0.14*	0.36
	Kin Care (Family)		**Kin Care (Child)**			
	β (Neel)	β (CREP)	β (Neel)	β (CREP)		
Age	0.04	0.22	-0.22*	0.19		
Sex	0.17*	0.03	0.14*	0.36		
Relationship status	0.09*	0.09	-0.04	-0.12		
Parent status	0.16*	0.08	N/A	N/A		
Childhood stability	0.21*	0.47	0.22*	0.10		
Childhood resources	-0.02	-0.41	-0.11*	-0.15		
Current resources	0.03	-0.06	-0.19*	-0.16		

Note. *$p < 0.05$.

Table 4. Life History Predictors of Fundamental Social Motives

Conclusion

In Neel *et al.* (1), a large sample of participants showed that individual differences in the fundamental social motives relate meaningfully to other individual differences. In our replication study, a small sample of participants replicated the original findings in only one third of correlations between conceptually related fundamental social motives and scales of individual differences. Furthermore, Neel *et al.* (1) found that individual differences in the fundamental social motives can be partially accounted for by life history variables. In the replication of the study, more than half of the correlations between life history variables and each fundamental social motive were replicated. The addition of Behavioral Inhibition/Activation Scales demonstrated that promotion of achievement of a goal versus avoidance of failure to achieve that goal depends on the fundamental social motive that drives the goal.

Limitations

Overall, our replication team achieved partial replication of the Neel *et al.* (1) results. However, this is based on a researcher-determined definition of what constitutes a replication and the requirements that were put into place for analysis of the effects found. For example, if the requirement that correlations found must be within 0.15 r/β-units of the original effect size reported were narrowed to 0.10 r/β-units, many of the correlations considered replications would no longer meet requirements. This trivial change in definition would lower the 40/74 replication ratio to below 50%. In that case, we most likely would not consider the replication attempt even partially successful, rather concluding non-replication overall. Beyond the definition of replication, it is noteworthy to mention that many of the data in the original study was not collected, and many that were collected were not analysed. For example, a major aim of the Neel *et al.* (1) study was to build construct validity for the fundamental social motive scale. Therefore, not having performed any exploratory analysis on the items of the scale, nor having compared the fundamental social motives to the Big Five, can place limitations on our conclusions since we may not be certain that the fundamental social motive scale is working the same way as it was in the original paper. If these concerns are most central to this replication project, many other factors could have influenced these results and they shall be discussed below.

The most evident limitation of the replication study is the small sample

size and lack of statistical power. While the replication had a sample size of 34, the original Neel *et al.* (1) study had a sample size of 220–770. With such a clear discrepancy in sample size, external validity concerns make us doubt the extent to which the results can be generalized to other populations, other environments, other times, etc. Moreover, in terms of internal validity, history effects of the 3-year duration between the original study and the replication must be considered. Since 2016, the world has changed, society has changed, people have changed, and importantly, we believe MTurk might have also changed. Past research showed a shift in participants' motivation to join MTurk in the past years, approaching it as a full-time job rather than hobby-like. (23) Since then, message boards have appeared with discussions of payments and study features such as deception, etc. If the participants or tools of research themselves have changed between original and replication, then the data collected may reflect these changes rather than report on the true variables being targeted.

Beyond financial and technological limitations, some features of the correlational survey design also pose concerns about trusting the data obtained. One concern is the length of the study and the large number of items that participants are expected to commit their undivided attention to. It may be the case that as participants work their way through scale after scale, the quality and accuracy of the responses obtained diminishes progressively. In addition to this design limitation, order effects may be very prominent in a lengthy study. This is especially the case for the additional BIS/BAS variable placed at the end of the list of scales in order to stay true to the original study design. Furthermore, testing effects such as polarization may threaten the internal validity of the results. After completing scale after scale, repetition may lead to more extreme and polarized responses merely due to the structural aspects of the study. Future directions in this field of research must dedicate resources to limiting the many threats to validity endured by this study. However, on a more optimistic note, an unresolved question to explore is how to incorporate a parsimony objective in explaining the complex relationship between the fundamental social motives and all other variables addressed in this study.

Acknowledgements

The authors wish to express their deep gratitude to fellow students Elisa Migliara and Yihong Bao for their immense contributions to this study. Special thanks should be given to Dr. Eric Hehman of the McGill Psychology Department for his supervision of this research work.

References

1. Neel R, Kenrick DT, White AE, Neuberg SL. Individual differences in fundamental social motives. Journal of Personality and Social Psychology. 2016;110(6):887–907.

2. Maslow AH. A theory of human motivation. Psychological review. 1943;50(4):370.

3. McAdams DP, Pals JL. A new Big Five: Fundamental principles for an integrative science of personality. American Psychologist. 2006;61(3):204–17.

4. Kenrick DT, Neuberg SL, Griskevicius V, Becker DV, Schaller M. Goal-Driven Cognition and Functional Behavior. Current Directions in Psychological Science. 2010;19(1):63–7.

5. Bakan D. The duality of human existence: Isolation and communion in Western man: Beacon Press (MA); 1966.

6. McClelland DC. How motives, skills, and values determine what people do. American psychologist. 1985;40(7):812.

7. Deci EL, Ryan RM. The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. Psychological inquiry. 2000 Oct 1;11(4):227-68.

8. Goldberg LR. The development of markers for the Big-Five factor structure. Psychological assessment. 1992;4(1):26.

9. Sacco DF, Young SG, Hugenberg K. Balancing Competing Motives. Personality and Social Psychology Bulletin. 2014Feb;40(12):1611–23.

10. Ellis BJ, Figueredo AJ, Brumbach BH, Schlomer GL. Fundamental Dimensions of Environmental Risk. Human Nature. 2009;20(2):204–68.

11. Stearns SC. The Role of Development in the Evolution of Life Histories. Evolution and Development. 1982;:237–58.

12. Maxwell SE, Lau MY, Howard GS. Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? American Psychologist. 2015;70(6):487.

13. Crowe E, Higgins E. Regulatory Focus and Strategic Inclinations: Promotion and Prevention in Decision-Making. Organizational Behavior and Human Decision Processes. 1997;69(2):117–32.

14. Jackson JJ, Kirkpatrick LA. Sociosexual Behavior Measure. PsycTESTS Dataset. 2007;

15. Duncan LA, Schaller M, Park JH. Perceived vulnerability to disease: Development and validation of a 15-item self-report instrument. Personality and Individual Differences. 2009;47(6):541–6.

16. Cheng JT, Tracy JL, Henrich J. Pride, personality, and the evolutionary foundations of human social status. Evolution and Human Behavior. 2010;31(5):334–47.

17. Altemeyer B. Marching In Step. The Sciences. 1988Apr;28(2):30–8.

18. Leary MR, Kelly KM, Cottrell CA, Schreindorfer LS. Construct Validity of the Need to Belong Scale: Mapping the Nomological Network. Journal of Personality Assessment. 2013;95(6):610–24.

19. Fraley RC, Waller NG, Brennan KA. An item response theory analysis of self-report measures of adult attachment. Journal of Personality and Social Psychology. 2000;78(2):350–65.

20. Conway MA, Pleydell-Pearce CW. The construction of autobiographical memories in the self-memory system. Psychological Review. 2000;107(2):261–88.

21. Griskevicius V, Delton AW, Robertson TE, Tybur JM. Environmental contingency in life history strategies: The influence of mortality and socioeconomic status on reproductive timing. Journal of Personality and Social Psychology. 2011;100(2):241–54.

22. Carver CS, White TL. Measurement Instrument Database for the Social Sciences. Available from: https://www.midss.org/sites/default/files/bis.pdf.

23. Ross J, Irani L, Silberman MS, Zaldivar A, Tomlinson B. Who are the crowdworkers? shifting demographics in mechanical turk. CHI '10 Extended Abstracts on Human Factors in Computing Systems; Atlanta, Georgia, USA: Association for Computing Machinery; 2010. p. 2863–72.

Submitted: 01/15/20

Research Article

[1]Department of Civil Engineering and Applied Mechanics, McGill University, Montreal, QC, Canada

Keywords

Barbados, rainwater harvesting, affordable, water tank, natural disaster, water scarcity

Email Correspondence

laura.vanderweyen@mail.mcgill.ca

Laura Vanderweyen[1], Xiao Yi Zhang[1], Vladislav Zasmolin[1]

Emergency Rainwater Harvesting, Water Storage, and Distribution System for an Affordable Housing Development in Barbados

Abstract

Background: To sustain both permanent residents and an intense tourism industry, Barbados overpumps its sole source of natural freshwater—the aquifer. Climate change is projected to increase both storm intensity and drought, further hampering groundwater recharge. These intense rainfalls quickly saturate topsoil and result in extensive surface run-off that causes flooding, erosion, sedimentation, and eutrophication. By providing more water for households and reducing aquifer withdrawal, rainwater harvesting has the potential to both mitigate water scarcity and reduce the amount of harmful run-off. However, rainwater harvesting is not currently practiced in Barbados. This paper proposes a hurricane-resistant rainwater harvesting, storage, and distribution system to be implemented in an affordable housing community in St. Thomas, Barbados.

Methods: In the creation of the distribution system, social, economic and environmental concepts need to be considered. A field survey was conducted within the neighbourhood to understand what residents felt the local water supply lacked. Afterwards, a detailed rainfall analysis allowed for determination of the amount of rainwater that can realistically be captured. Finally, after consulting with various academic experts, local industry members, and supply stores, this allowed for the determination of an affordable design.

Results: The findings suggest that an initial household investment of $2,790.90 BBD appears adequate to provide a system for rainwater harvesting, dual plumbing, and communal distribution which can withstand and utilize 1/50 years storms.

Limitations: Given that rainwater harvesting is not currently practiced in Barbados, government initiatives are needed to encourage its development.

Conclusion: Affordable rainwater harvesting, dual plumbing, and community distribution systems can be implemented to possibly reshape life in countries facing water scarcity.

Introduction

As a tropical Island, Barbados is exposed to climate change-powered natural disasters, droughts, and extreme rainfalls. Since 1960, temperatures on the island have risen by 0.6 °C; rainfall during the wet season has increased by 6.2% per decade, and rainfall during the dry season has decreased by 2.9%. (1) Combining this with rising sea levels and aging infrastructure makes Barbados vulnerable to floods, storms, and hurricanes, and it is uncertain whether the country is prepared for a potential disaster. (2-3)

In Barbados, 90% of potable water is pumped from the aquifer; the remaining is produced by desalination plants. (4) Whereas desalination is extremely expensive, the aquifer provides cheap, high-quality water treated through limestone percolation and disinfected by chlorination. Rainfall is the sole source of water that contributes to recharging the aquifer. As population, urbanization, and droughts increase, this indispensable source of water is progressively being depleted. Overpumping leads to a lowering of the water table which puts the reservoirs at risk of salt-water intrusion. As a result of a proposal by the Barbados Water Authority (BWA) to impose water rationing in some neighbourhoods, Barbados residents are left with less than 281.5 m^3 of renewable water per capita per year, making it the 15th most water scarce country in the world. (5-6) In parishes like St. Thomas, residents claim to experience water shutoffs from 9 a.m. to 4 p.m. twice a week to provide for water-demanding agricultural areas to the north of the island. Meanwhile, countries like Bermuda, Nigeria, and Guyana have developed sustainable programs to move towards rainwater harvesting methods for greywater, irrigation, and potable purposes. (7-9) Chennai, India has made it mandatory to include rainwater systems in all establishments. (10) Furthermore, the city created a sustainable design that directs the overflow of water harvested into wells that recharge aquifers. As a result, Chennai has seen a replenishment in the top aquifer and have retarded their water scarcity issues. (11) Efficiently harnessing an abundantly available resource like rainwater can provide a viable route to mitigating water scarcity. Unfortunately, widescale rainwater harvesting efforts have not been implemented in Barbados.

As the frequency, intensity, and duration of hurricanes is expected to increase, the ability for communities to maintain themselves can be compromised. Hurricane Dorian recently devastated areas in the Bahamas and left some residents without water and power, resulting in more than 56 deaths, 600 missing peoples, and over $7 billion USD in damages. (12-13) Barbados has not endured this kind of disaster since Hurricane Janet in 1955 and is both overdue and unprepared for a disaster. Tropical storms regularly cause flooding and erosion in Barbados. Storm run-off also leaches through fertilized agricultural land where it picks up nutrients. These sediment-rich and nutrient-rich waters eventually reach the coral reef causing eutrophication and bleaching. Systems that have the capability to harness energy and water from tropical storms and hurricanes could assuage the destructiveness of these events. (14)

In Barbados, owners of houses with a surface area greater than or equal to 1,500 ft^2 are legally bound to purchase water tanks that are supplied by water from the mains. (15) Residents with smaller houses also invest in water storage tanks to better deal with water shut offs. Some of them rely on an

electric pump to provide the house with the previously stored water in times of shut-off. Others choose to elevate their tank on a tall foundation to use gravity head to supply their house with water. However, these structures are usually built without consultation with engineers and can potentially be dangerous, especially during a hurricane with high wind speeds. Safe circulation of water during power outages has not been extensively researched and may be essential to developing a water resilient country.

Based on stormwater's potential for reuse, this paper proposes a water storage and distribution system centered around an assessment survey conducted in the affordable housing neighbourhood of Content, St. Thomas. The design's intended purpose is to provide relief to residents by having a reliable source of water when main water is unavailable. It is categorized by three major components. Firstly, a rainwater harvesting system gathers the abundance of fresh water into individual water storage tanks that are already common throughout Barbados. Secondly, a dual plumbing system assures that rainwater and main water are being separated and allocated to specific utilities within the house. Lastly, a community distribution system allows the excess water harvested to be shared between neighbours instead of being released as run-off.

Survey and Needs Assessment

On September 28, 2019, 25 households scattered randomly around the existing development in Content were surveyed using a questionnaire relating to water use and quality, frequency of shutoffs, willingness to invest in a tank, and disaster preparedness. A mix of gender, age, and backgrounds were observed, but all households were within the same financial conditions as they cohabited the same neighbourhood. Although the community needs are coherent within Content, it might not necessarily be the case elsewhere on the island, which forms a limitation. Based on this survey, it was found that rainwater harvesting is not usually practiced due to a lack of awareness.

Residents were asked, amongst other things, about their experience with water management and their willingness to invest in a tank. The questions gauged the resident's value of the following eight parameters: maintainability, workability, affordability, security, practicability, water quality, durability, and aesthetics. The most important decision parameter for Barbadians seemed to be affordability, while aesthetics did not matter as much. Accordingly, a Tuff Tank will be used for the design. Tuff Tanks are readily available and affordable, black coloured, above-ground water storage tanks.

Rainwater Harvesting System

Rainfall analysis

The study site was localized on qGIS along with 24 rain gauges across the Barbadian territory for which monthly volume recordings from 2000 to 2015 were provided by Tara Mackey, a University of the West Indies Ph.D. candidate. Thiessen polygons were subsequently drawn to attribute an area to each gauge and the proposed site fell on two different polygons (Fig. 1). To accurately represent the neighbourhood, data were averaged over these two gauges and compared with rainfall data (1887-1986) from monthly rainfall volume maps as shown in Fig. 1. (15)

When designing for extreme events, high-intensity-short-duration and long return period data are expected. Due to the lack of public short-duration rainfall data for Barbados, local IDF curves for 1/50 years floods were not readily available. (16) The longest return period available for Barbados is currently 1/25 years. Hence, it was decided to use the Bahamas' IDF curve for 1/50 years storms. (17) This decision will result in a conservative design since, from a comparison of the 1/25 years storm curves for the two countries, it is clear that the rainfall intensity of the Bahamas is much larger than that of Barbados. (17) Indeed, in a 10-minute duration, the 1/25 years rainfall intensity was 250 mm/hour in the Bahamas while it was only 150 mm/hour in Barbados. Furthermore, the shortest duration represented in the Bahamas' IDF curve for a 1/50 years rainfall was 10 minutes. To account for the most extreme event, the rainfall intensity for a 5-minute duration was extrapolated and found to be approximately equal to 350 mm/hour or 13.8 inch/hour.

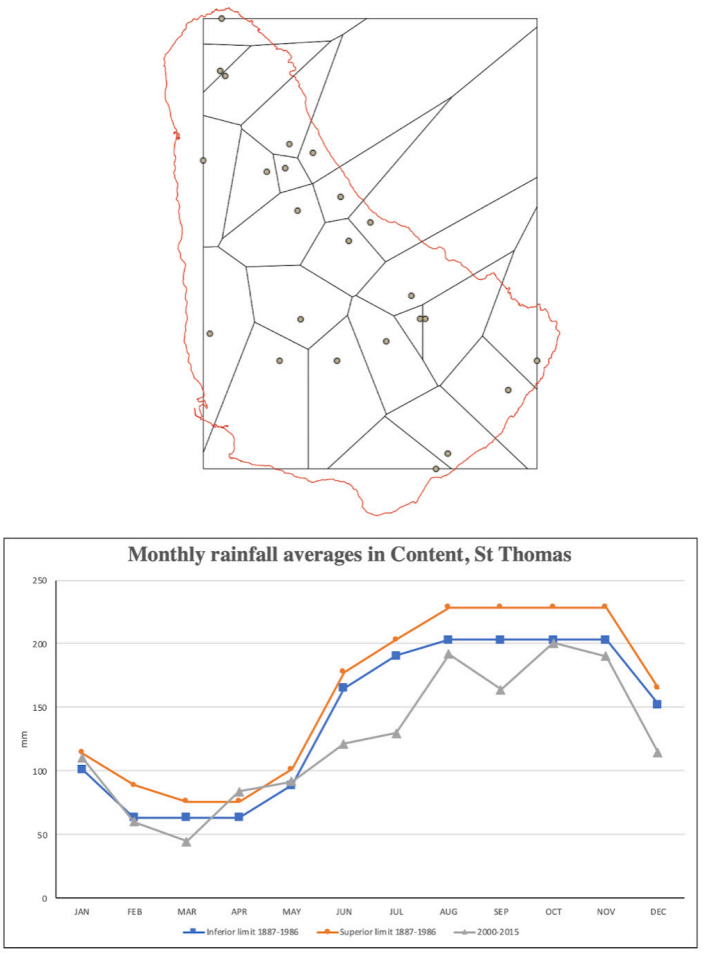

Figure 1. Thiessen polygon for rainfall gauges in Barbados (top) and plotted rainfall data for each month in millimetres (bottom).

Water tank sizing

With an average family of 5 using 823 L/week/person, a 1000 gallon Tuff Tank should supply six days of water to the family. (18) The tank must hold enough water in a hurricane situation both to provide head towards the house and not to tip over. Using the 1/50 years wind speed of 41.12 m/s for Barbados results in a 2.614 kN force acting on the centre of the water tank which has a 1.65 m diameter and 2.457 m height corresponding to the European Code. (19) Therefore, using force summation $\Sigma F=0$ and moment summation $\Sigma M=0$ at the corner of the tank, the critical amount of water needed in the tank to prevent tip over was calculated at 162 gallons. In a power outage, the head required to transfer water from the tank to the least elevated appliances must be provided by a water level lower than 187 gallons (see *Rainwater Plumbing*). Thus, as long as the tank contains at least 187 gallons of water, it will both prevent tip-over and supply head.

Removing particulate matter

Not only do corrugated iron roofs reduce pathogens through the dry-heat effect, but their high run-off coefficient and smooth surface allow for optimal surface cleansing. (20-21) Additionally, gutter guards, or sheets of pierced metal placed on top of the gutter, will filter out solids larger than

1 mm. The first flush diverters, placed on each side of the roof, will then capture the initial volume of rainfall that swept sediments and small particles off the roof into separate storage. As the first flush compartment fills up, a spherical float will rise to close the storage and direct cleaner water into the tank (Fig. 2).

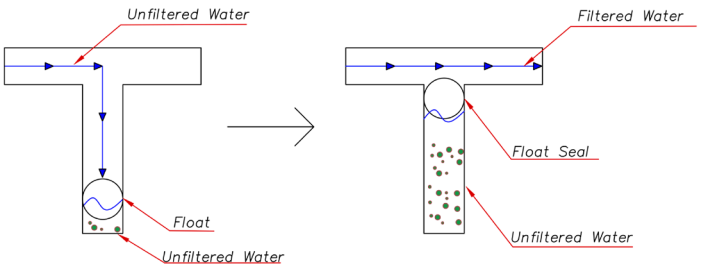

Figure 2. First flush diverters showing the flow of filtered and unfiltered water depending on the volume of water in the first flush compartment.

The first flush compartments were sized according to Blue Mountain Co rainwater harvesting regulations; for every square foot of roof, there should be a litre of first flush diverter volume. (22) To contain the calculated volumes, two diverters (10 gallons each) for each side of the roof will be 2361 mm long, 6 inch diameter pipes that will be connected to the 3 inch horizontal pipe above through a 3 inch corner tee and two consecutive expanders, a 3 to 4 inch and a 4 to 6 inch from Barbados Steel Works (Fig. 3). An end cap that can be manually unscrewed after a rainfall to empty the system will be placed at the bottom of the diverter.

Gutter sizing and pipe sizing

The proposed development contains three different types of precast houses that have different surface areas (from smallest to largest: Evergreen, Flamboyant, and Tulip). The houses contain 2 to 3 bedrooms and 1 to 2 bathrooms. They are made of prestressed concrete designed for earthquake loads, and the roofs are made of corrugated steel, ensuring a large run-coefficient. (23) To withstand the massive influx of water during a hurricane, gutters and pipes were sized based on the largest roof catchment area (Tulip) and the Bahamas's 1/50 years rainfall intensity. Calculations followed Architectural Graphic Standards for Residential Construction and resulted in 6 by 6 inch gutters and 3 inch diameter downpipes on each side of the roof (Fig. 3). (24)

In order to maintain a consistent water velocity at a given flow rate, the pipes to the tank have the same cross-sectional area as the downpipe given that Q=AV, where Q is the volume flow rate, A is the cross-sectional area, and V is the mean velocity of the fluid. Hence, PVC pipes are 3 inches in diameter.

Cost

Based on quotes from Akeem Nurse at Alan Armstrong and the BWA, the cost of installing a 1,000 gallon water tank in Content was estimated to be $9,024.3 BBD by adding the cost of gutters, pipes, connectors, and the first flush system (Table 1).

The monthly volume of water captured by the harvesting system was estimated by multiplying historical rainfall intensity recordings by the catchment area of the roof for each house type. The economic savings on the water bill was then estimated by subtracting the volume captured from household consumption for a family of five and calculating the BWA water bill that would be incurred by such a volume (Table 2). (25) Household consumption was based on per capita estimates for St. James, Barbados. (18)

The initial investment was then divided by the average monthly savings to calculate the time to break even. As shown in Table 2, when taking into account the total cost (tank installation with connection to the main, construction, and harvesting system), the time to break even based on savings from harvested water is high (about 20 years depending on the house type). However, people already do invest in water tanks and pay for installation to store main water and avoid experiencing frequent shutoffs. The only addition to current practices is the harvesting system which allows for main water savings. The initial investments for the harvesting system (gutters, meshes, etc.) should be recovered in less than three and a half years. The installation cost of a harvesting system is only about 2% of the average local yearly salary of $62,000 BBD. (26)

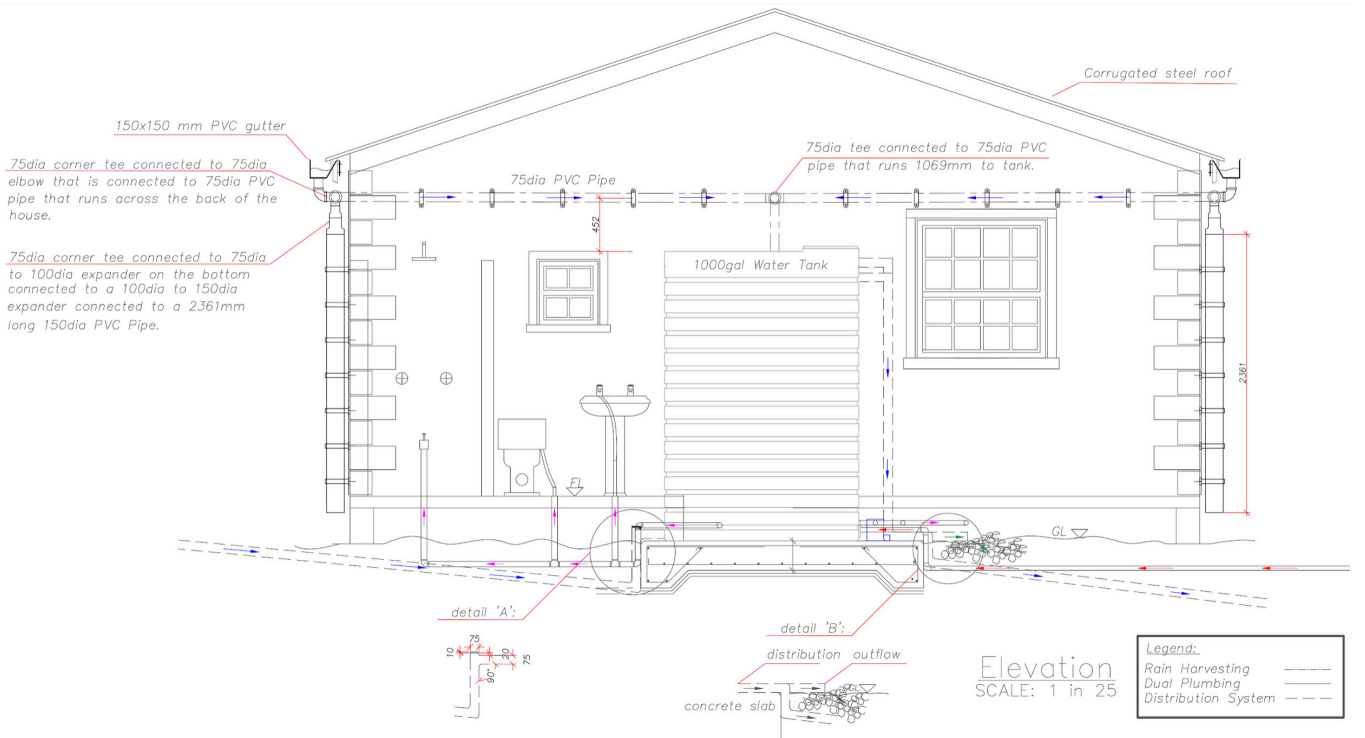

Figure 3. Elevation View of Rain Harvesting, Dual Plumbing, and Distribution System

	Unit Price/rate	Quantity	Total price
MATERIALS	2,384.78	1	2,384.78
3VM06		1	
19G Pressure Tank		1	
Pressure Switch		1	
Pressure Gauge		1	
Float Switch		1	
1000 GALLON TANK	1,677.87	1	1,677.87
Pipework including fill line, discharge, 2" supply with 1/2" Return		1	
ELECTRICAL INSTALLATION	1,105	1	1,105
EARTHWORKS			
excavation of foundations	120	1	120
Filling and compaction			
General filling		1	0
Excavated marl fill under water tank	120	1	120
Excavated marl fill to trench at connection	12	1	12
Excavated 20 mm single size stone	55.8	1	55.8
INSTITU CONCRETE			
Concrete slab X-section 0.01-0.02m2	874.5	1	874.5
binding to slab	142.5	1	142.5
CONCRETE ANCILLARIES			
Formwork with 25*25mm chamfer	211.2	1	211.2
mild steel bars 10mm diameter	32	1	32
High yield bar to BS 4449 12mm	340	1	340
steel trowel finish	34	1	34
OVERHEADS AND PROFITS	570.6	1	570.6
TOTAL before tax			7,680.25
Tax @ 17.5%			1,344.04
TOTAL after tax			9024.29

Table 1. Construction and water tank installation costs based on DM Simpson quotes sent by Akeem Nurse for a system in Trents, St. Lucy. Monetary values in BBD.

House	Evergreen	Flamboyant	Tulip
Tanks Installation BDD$	9,024.29	9,024.29	9,024.29
Harvesting system BDD$	931.52	1,057.60	1,266.58
Total Cost BDD$	9,955.82	10,081.90	10,290.87
Average monthly saving BDD$	24.28	29.45	31.59
Yearly savings on water bill %	48%	59%	63%
Months to break even	410.10	342.36	325.75
Years to break even	34.18	28.53	27.15
Years to Break Even for harvesting only	3.20	2.99	3.34

Table 2. Estimated costs, savings based on water not withdrawn from the mains, and time to break even on the initial investment. Monetary values in BBD.

Dual Plumbing System

Concept

The BWA provides potable water that is mainly sourced from aquifers and treated with chlorine. Amenities like the toilet, sinks, faucets, and showers are already connected to the mains as part of the design for a residential development. Rainwater harvesting provides an additional source of water that can be installed during construction or added to an existing house. The storage tank contains primarily rainwater collected from the harvesting system but is also branched to the main for back up when rainwater is scarce and pressure head becomes too low to feed the house.

A dual plumbing system consists of a network of two separate pipes: the typical main-to-house for potable water and the tank-to-house for tank water. The treated main water is hence physically separated from the non-treated rainwater, preventing contamination. The main plumbing feeds the amenities that require potable water like the kitchen sink, or high-water heads such as the shower, while the secondary plumbing feeds amenities that require neither potable water nor a high-pressure head, such as the toilet. For such utilities, non-return valves and knobs allow users to switch between sources.

Main water plumbing

The main water pathway going to the kitchen faucets remains unaltered as potable water is indispensable not only for drinking, but also for cooking and washing food (Fig. 4).

2 inch polyblue pipes, which are flexible yet sturdy, will be used to account for unideal angles (30 degrees, 45 degrees, or 90 degrees). Although polyblue can be used above ground, it is strongly suggested to be run underground to limit risks of damage in hurricane circumstances. The elevation view in Fig. 3 illustrates how the main water pipe runs below ground until it reaches the concrete slab.

Rainwater plumbing

Rainwater is distributed to the house using 2 inch pipes as shown on Fig. 4. A hole punctured in the concrete floor slab will allow the pipe to enter the house. This hole must be properly sealed to prevent faulty waterproofing and cracking in the floor slab.

A faucet located 0.525 m above floor level in the shower cabin (below the showerhead) can provide rainwater. Various fittings (T-fittings, Y-fittings, elbows, etc.) were used to direct the water flow (Fig. 3).

Bernoulli's equation of energy conservation with minor losses

$$H_1 = H_2 + \sum S_{fi} L_i + \sum K_i \frac{V_i^2}{2g}$$

was used to determine the head needed to allow for gravity flow from the tank to the house. It states that the head at the free surface in the tank equals the head level of the utilities plus the friction losses along the pipe and minor losses due to parts (elbows and valves). The S_{fi} is given by the Hazen-Williams equation for circular conduits

$$S_f = \left(\frac{Q}{0.278 * C * D^{2.63}}\right)^{1.65}$$

where C, the roughness coefficient, is 145 for a PVC pipe. (27) Knowing all the manufactured parts' head loss coefficients allows for the calculation of minor head losses in the pipes. (27) It was found that a 1000 gallon tank will supply water using only gravity flow until the water level in the tank reaches 187 gallons (0.399 m) for the toilet, 426 gallons (0.912 m) for the sink, and 246 gallons (0.526 m) for the shower faucet. This is ideal because water will always stay above 187 gallons when there is no electricity (only gravity flow) meaning it will never reach the critical tipping point of 162 gallons (see *Water Tank Sizing*) in emergencies. A standard 1 horsepower pump, according to Alan Armstrong Associates engineers, is placed to supply water into the house when electricity is running.

Maintenance and repairs

If electricity is running and the water volume in the tank falls below 187 gallons, the pump and pressure tank activate to supply main water into the tank to provide enough head. The pressure tank acts as a sensor to detect water level and activate the pump only when required. All components of the dual plumbing system are protected from backflow by non-return valves (Fig. 4).

For cleaning or repairs, the tank must be disconnected from the plumbing network and emptied through an outflow valve. A 38 mm pipe (Fig. 4), bypasses the tank, pump, and pressure tank and redirects main water into the building.

Emergency scenario

During an emergency such as a storm, or hurricane, damages to the electrical grid and water mains will likely jeopardize power and potable water availability. As a result, the residents must rely on stored water for their daily necessities. A 1,000 gallon tank at full capacity will require 187 gal-

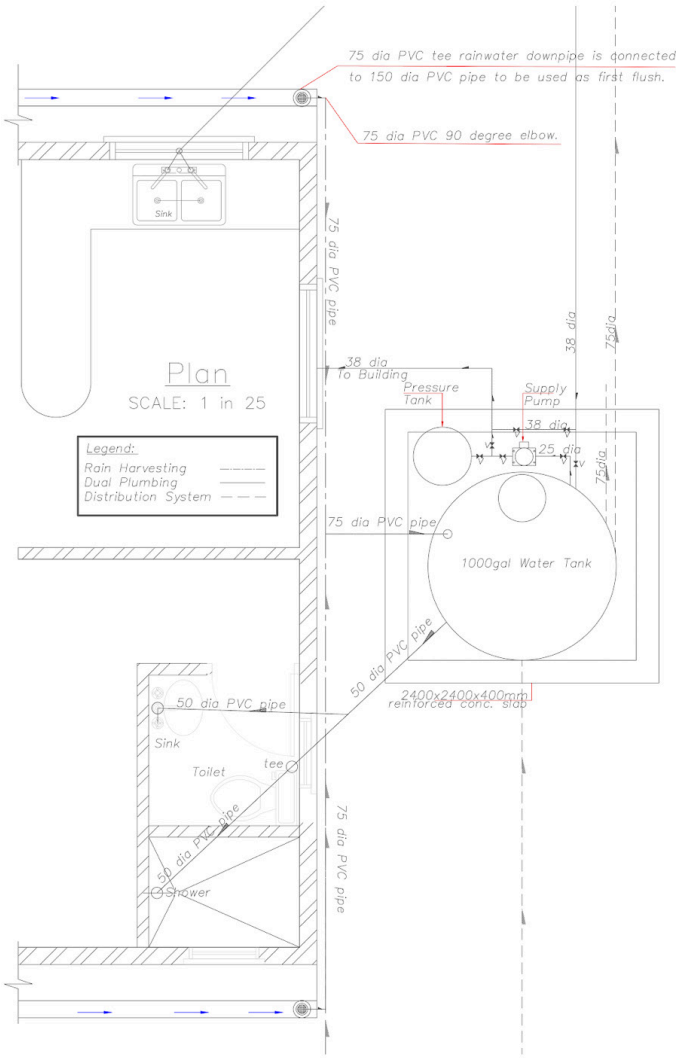

Figure 4. Plan View of Rain Harvesting, Dual Plumbing, and Distribution System. The solid lines represent potable water sourced from the main and the storage tank water distributed to the house. Non-return valves are represented as two triangles pointing at each other, and denoted by "v."

Figure 5. Distribution System across 4 houses in Content, St. Thomas, Barbados.

lons to pressurize the water distribution (see *Rainwater plumbing*) to the toilet, leaving a remaining 813 gallons for usage. However, the collected rainwater is not potable since it did not go through disinfection. The addition of an emergency kit to render this rainwater potable is thus the most realistic and affordable solution. The proposed kit contains chlorine pills and a chart that gives chlorine to water ratios for optimal disinfection. Chlorine is chosen as it is readily available and cost-effective. Alternatively, boiling the water is another option as Barbados relies heavily on gas stoves which do not require electricity.

For other usages of water such as washing clothes, cleaning, and flushing the toilet, the rainwater can be used directly. Rainwater will go directly to the toilet bowl as long as the appropriate valve is turned on and the source is above 187 gallons. Similarly, the lower shower faucet can be used. The dual system supplies water directly from the tank to indoors preventing people from venturing outdoors during a storm, where they are putting their life at risk with heavy rainfall, slippery grounds, strong winds, and flying objects.

Distribution System

Rationale

The creation of a connected residential distribution system is based on the premise that in a situation of natural disaster, mutual help and fraternity can increase resilience of the whole community. In a hurricane situation, wind speeds and rainfall levels may display high local variations. Hence, houses that are located or oriented differently on the land's topography could be subjected to different levels of damage, disruption, and rainfall. The proposed distribution system would allow overflow from one tank to feed the adjacent tank. That way, if the gutters of any house break, and the rainwater harvesting system is disabled, the tank would be replenished from the neighbours' tank.

Design

A series of four houses within the development were chosen as a proof of concept for design of the distribution system. Since it is impossible to push the water back uphill without providing head through a pump, the distribution network flows from high to low elevations (Fig. 5). Consequently, the uphill house is a provider only, while the downhill house is a receiver only.

The tanks were placed on the CAD drawing, and their elevations were found using contour lines. These elevations set the gradients of the distribution pipes that are placed underground for protection. A set of angled elbows were then chosen to drive a 3 inch pipe from the distribution overflow to the bottom of the next tank (Fig. 3). To prevent any backflow, a non-return valve is also placed on the incoming distribution pipe.

Head differences were calculated both in hurricane and in normal circumstances according to Bernoulli's equation of energy conservation with minor losses. The friction slope, S_{fi}, was calculated through the Hazen-Williams equation. In hurricane circumstances, when the tank is full to the distribution overflow level, head goes from 5.41 m in the highest tank to 0.92 m in the lowest; in normal circumstances, it goes from 5.41 m to 2.50 m. These differences in head assure a continued flow.

The potential risk of local debris accumulation and pressure build up at the 135° elbow (Fig. 3 detail 'A') led to a design modification. Since debris would accumulate in the 135° angle elbow, a cap was simply added to the 90° angle elbow to allow physical access to the elbow for cleaning (Fig. 3 detail 'A').

Cost

The cost of the distribution system, including the three pipes connecting

the four houses, the second overflow pipe for the tanks, excavation, filling, and labour was estimated at $2,939.3 BBD which is equal to $734.8 BBD per house. Such a cost represents only 1.1% of the average yearly salary (22). However, this investment is made for emergency situations only and does not yield any return.

Conclusion

Water storage and distribution is a prominent issue within water-scarce Barbados. With constant water rationing in communities, power shut offs, and aging infrastructure, Barbados needs immediate changes to their water policies and technologies. Furthermore, susceptibility to natural disasters and unpredictability of rainfall amplifies the dangers of water scarcity, causing malaise within residents. Indeed, climate change continues to exacerbate scarcity of a resource that is both indispensable and economically inelastic. Therefore, following this trajectory, the BWA will be forced to increase water prices. The design proposed above for Content, St. Thomas would potentially provide residents with a consistent viable source of water through rainwater harvesting, dual plumbing, and water distribution for $2,790 BBD per household. This could be economically viable for residents who are interested in purchasing a water tank to palliate water rationing. Harnessing the abundance of rainfall for household use appears to be a sustainable approach to mitigate water scarcity. Furthermore, using this as a framework for further projects globally can revolutionize the lives of people living under similar circumstances. With these innovative solutions, not only would residents be equipped with a sustainable solution to provide water, but run-off would be reduced, slowing down coral eutrophication.

Future Work

In designing the distribution system, the issue of bringing water uphill for the houses located at the highest elevations arises. The only way to accomplish this task is to pump the water back up. In the case of a power outage, which is likely during category 3, 4, or 5 hurricanes, any electrical pump would be non-functional. Solutions to this constraint could possibly be found in investigating mechanical pumps and ram pumps. Moreover, since this distribution system is shared, the cost to build and maintain it needs to be fairly allocated among residents.

In order to implement rainwater harvesting systems, awareness of its potential to solve environmental issues need to be raised within the Barbadian community. Building and testing the proposed design is left for future work.

Acknowledgements

The authors would like to gratefully acknowledge Akeem Nurse and Tara Mackey for their constant support and guidance throughout the research. Akeem Nurse is an engineer affiliated with Alan Armstrong Associates in Barbados, and Tara Mackey is a Ph.D. fellow working closely with the main water network in Barbados.

References

1. World Bank Climate Change Knowledge Portal [Internet]. [cited 2020 Mar 4]. Available from: https://climateknowledgeportal.worldbank.org/

2. Isaacs W. Opportunities to Mainstream Gender in Water and Wastewater Infrastructure Projects: A Case Study in Barbados. Graduate Thesis and Dissertation, University of South Florida. 2017 Mar 17; 157 pp.

3. Belle N, Bramwell B. Climate Change and Small Island Tourism: Policy Maker and Industry Perspectives in Barbados. Journal of Travel Research. 2005 Aug 1;44(1):32–41.

4. Cashman A. Water Security and Services in the Caribbean. Water. 2014 May 5; 6(5): 1187-1203.

5. FAO Aquastat. Country Profile – Barbados. 2015. Available from: http://www.fao.org/aquastat/en/countries-and-basins/country-profiles/country/BRB

6. Singh A. Barbados National Assessment Report of Progress made in addressing Vulnerabilities of SIDS through Implementation of the Mauritius Strategy for Further Implementation (MSI) of the Barbados Programme of Action. Government of Barbados; 2010.

7. Rowe MP. Rain Water Harvesting in Bermuda. Journal of the American Water Resources Association. 2011;47(6):1219–1227.

8. Efe S.I. Quality of rainwater harvesting for rural communities of Delta State, Nigeria. Environmentalist. 2006; 26:175-181.

9. Peters E.J., Mohamed Z. Willingness to pay for water on the Guyana east coast. Proceeding of the Institution of Civil Engineers - Water Management. 2015; May 25;163(6): 315-323.

10. Raghavan D. Rainwater Harvesting – The Success Story of Chennai. IEEE India Info. 2018 June;13(2):5.

11. Elangovan A, Ravichandran S, Kumar LM. Groundwater responses to artificial recharge of rainwater in Chennai, India: A case study in an educational institution campus. Indian Journal of Science and Technology. 2010 Jan;3(2);124-130.

12. International Medical Corps. Hurricane Dorian Situation Report #12 - Bahamas. Available from: https://reliefweb.int/report/bahamas/hurricane-dorian-situation-report-12-september-30-2019

13. International Medical Corps. WFP Support to the NEMA/CDEMA-led humanitarian response in the Bahamas Situation Report #02 – Bahamas. Available from: https://cdn1.internationalmedicalcorps.org/wp-content/uploads/2019/08/Int-Med-Corps-Hurricane-Dorian_SitRep2.pdf

14. White G. Human Adjustment to Floods. Research Paper 29. Department of Geography, University of Chicago. 1945; 225 pp.

15. Hutchinson A.P. "Rain Water Harvesting" The Barbados Experience. Stantec Caribbean. 2009 May 28; Presented at St Lucia Water Week.

16. Lumbroso DM, Boyce S, Bast H, Walmsley N. The challenges of developing rainfall intensity–duration–frequency curves and national flood hazard maps for the Caribbean. Journal of Flood Risk Management. 2011;4(1):42–52.

17. Boyce S. Intensity Duration Frequency Curves for Barbados– A Review. Caribbean Institute for Meteorology and Hydrology. 2009 Jul 20; 8 pp.

18. Suchorski A. Socio-Economic and Physical development Influences on Water use in Barbados. Graduate Thesis, McGill University. 2009 May; 106 pp.

19. Vickery P.J., Wadhera D. Wind Speed Maps for the Caribbean for Application with the Wind Load Provisions of ASCE. Applied Research Associates. 2003; 12 pp. Report No.: 18108-1.

20. Meera V, Ahammed M.M. Water quality of rooftop rainwater harvesting systems: a review. Journal of Water Supply: Research and Technology-Aqua. 2006 Jun 1;55(4):257–68.

21. Biswas B.K., Mandal H.R. Construction and Evaluation of Rainwater Harvesting System for Domestic Use in a Remote and Rural Area of Khulna, Bangladesh. Hindawi. 2014 September 14; 2014(751952); 6 pp. Available from: https://doi.org/10.1155/2014/751952

22. Rain Harvesting by Blue Mountain Co. Calculate your diversion needs. Available from: https://rainharvesting.com.au/products/first-flush-diverters/first-flush-downpipe/

23. Eastern Land Development. House & Land Developments. Available from: http://www.eldbarbados.com/house_land.html

24. Rumbarger J. Architectural Graphic Standards for Residential Construction: The Architect's and Builder's Guide to Design, Planning, and Construction Details. John Wiley & Sons; 2003

25. Barbados Water Authority. Calculate your Bill. Available from: http://barbadoswaterauthority.com/?p=393

26. Average Salary in Barbados 2020 - The Complete Guide [Internet]. [cited 2020 Mar 4]. Available from: http://www.salaryexplorer.com/salary-survey.php?loc=19&loctype=1

27. Hughes C.T., Jeppson W.R. Hydraulic Friction Loss in Small Diameter Plastic Pipelines. Journal of the American Water Resources Association. 1978 October; 14(5); 1159-1166.

Submitted: 01/15/20

Research Article

[1]Department of Math, McGill University, Montreal, QC, Canada
[2]Department of Physiology, McGill University, Montreal, QC, Canada

Keywords

Cellular automaton, FitzHugh-Nagumo model, excitable media, cardiac dynamics, convolutional neural network

Email Correspondence

yujing.zou@mail.mcgill.ca

Yujing Zou[1,2], Gil Bub[2]

Comparison of Complexity and Predictability of a Cellular Automaton Model in Excitable Media Cardiac Wave Propagation Compared with a FitzHugh-Nagumo Model

Abstract

Background: Excitable media are spatially distributed systems that propagate signals without damping. Examples include fire propagating through a forest, the Belousov-Zhabotinsky reaction, and cardiac tissue. (1) Excitable media generate waves which synchronize cardiac muscle contraction with each heartbeat. Spatiotemporal patterns formed by excitation waves distinguish healthy heart tissues from diseased ones. (3) Discrete Greenberg-Hastings Cellular- Automaton (CA) (1) and the continuous FitzHugh- Nagumo (FHN) model (7 are two methods used to simulate cardiac wave propagation. However, previous observations have shown that these models are not accurately predictive of experimental results as a function of time. We hypothesize that cardiac simulations deviate from the experimental data at a rate that depends on the complexity of the experimental data's initial conditions (I.C.).

Methods: To test this hypothesis, we investigated two types of propagating waves with different complexities: a planar (i.e. simple) and a spiral wave (i.e. complex). With the same I.C., we first compared simulation results of a Greenberg-Hastings Cellular Automaton (GH-CA) model to that a FitzHugh-Nagumo (FHN) continuous model which we used as a surrogate for experimental data. We then used median-filtered real-time cardiac tissue experimental data to initialize the GH-CA model and observe the divergence of wave propagation in the simulation and the experiment.

Results and Conclusion: The alignment between the CA model of a planar wave and the FHN model remains constant, while the degree of overlap between the CA and FHN models decreases for a spiral wave as a function of time. CA simulation initialized by a planar wave real-time cardiac tissue data propagates like the experimental data, however, this is not the case for the spiral wave experimental data. We were able to confirm our hypothesis that the divergence between the two models are due to initial condition (I.C.) complexity.

Discussion: We discuss a promising strategy to represent a GH-CA model as a Convolutional Neural Network (CNN) to enhance predictability of the model when an initial condition is given by the experimental data with higher level of complexity.

Introduction

An excitable media system can be viewed as a group of coupled individual elements where each element can pass information to its neighbors with various neighborhood size. A signal over a certain threshold initiates a wave of activity moving across the excitable medium. (2) They are spatially distributed systems that propagate signals without damping. An excitable media is characterized by its threshold of excitability, which is a certain level of excitation to be reached before the system can generate travelling waves whose shape and speed remain unchanged through the medium. Examples of travelling waves in excitable media include fire propagating through a forest, the Belousov-Zhabotinsky (BZ) reaction, and propagating waves for means of communication within and across nerves as well as generating contraction (8) in cardiac tissue. (1) More specifically, the heart supports propagating waves in a variety of different geometrical patterns including plane waves, spirals, and multiple spirals. As a physical system passing signals by diffusion, this travelling wave is a result of propagating electrical activity in cardiac muscle involving sodium and potassium ions moving to neighboring cells. (2)

The wave dynamics and the resultant spatiotemporal patterns are essential to the heart's function. (3) Changes in the spatial patterns of these waves can cause potentially deadly arrhythmias, therefore, spatiotemporal patterns formed by excitation waves can distinguish healthy heart tissues from diseased ones. In the context of this paper, we define planar waves as travelling waves emanating from a central pacemaking source that acts to synchronize contraction during a healthy heartbeat. In contrast, aberrant re-entrant waves, which have a characteristic spiral geometry, re-excite the tissue rapidly and underlie potentially deadly tachycardias and fibrillation. (3) We define them as spiral waves. Therefore, a planar wave's initial condition is deemed to have less complexity than that of a spiral wave. We use cardiac monolayers which are thin sheets of heart muscle tissue grown in a petri dish to examine the dynamics of these propagating waves. These cultured cardiac cells can form connections and generate excitation waves which propagate out from an initiating point (a target) or a re-entrant circuit (a circuit of electricity where an impulse re-enters and a region of the heart is repetitively excited) with a spiral shape. Thus, studying these waves from the cardiac monolayers can help improve our understanding of cardiac arrhythmias. (2)

In this paper, we focus on computational methods to study these excitable waves. *Discrete cellular-automaton (CA)* and *continuous FitzHugh-Nagumo (FHN) models* are both well-known systems for simulating cardiac wave propagation. In a cellular automaton model, each cell has a finite number of states. These states are updated based on the states of their neighbors and their own previous state. It is an extremely powerful method for studying the dynamics of an excitable media due to its simple rules which underpin the nature of connected cells, even while the processes driving the physical system can be rather complicated (Fig. 1, first row). However, previous results have demonstrated that neither of the CA and FHN models show consistently accurate predictions of excited wave behaviors that align with experimental results (Fig. 2, second row) after a

few seconds. Consequently, it is crucial to determine the roots of the discrepancies seen between the CA model and experimental result in order to generate an accurate model of the dynamics. Therefore, we hypothesized that cardiac simulations deviate from the experimental data as a function of the complexity of initial conditions (I.C.) of the experimental data.

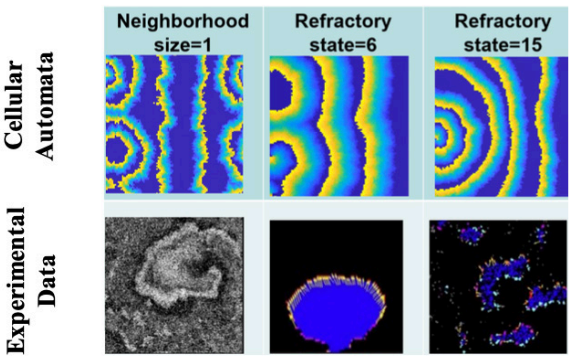

Figure 1. Greenberg-Hastings Cellular Automaton (GH-CA) wave (first row) propagation simulations generated at various parameter values can accurately represent experimental data. The GH-CA algorithm is detailed in section 2.1. The first row shows Cellular Automaton simulations of wave propagation: where the first subfigure shows a planar wave and the second and the third subfigure show a spiral wave with various refractory states where the spiral source is to the very left of the panel. In the second row: the first subfigure is a snapshot of a spiral wave during tachycardia, the second and the third pictures show an example output of wave segmentation and tracking using an automated wave tracking software in cultured cardiac monolayers called 'Ccoffinn'. (4)

The cardiac membrane potential function of the FHN model is continuous. It is relatively simple and not computationally expensive so it was used to produce surrogate data as our "ground truth" for wave propagations. Here, we provide a direct comparison between Greenberg-Hastings Cellular Automaton and FHN model simulations for two types of excited waves with differential complexity, a planar wave I.C. for the simple type and a spiral wave I.C. for the complex type. Our results were able to confirm our hypothesis that when the same I.C. was given, the two models diverge earlier in simulation time steps when the I.C. is more complex.

Methods

To study what gives rise to the divergence between a cellular automaton model and experimental results over time, we investigated two types of propagating waves with different complexity: a planar and a spiral wave. We first built a discrete GH-CA model of size 100 by 100 cells. Then we used diffusively coupled FHN equations to generate surrogate data sets as our "ground truth experimental data" since its (fast and slow) variables together reflect the cardiac membrane potential, which is continuous. We then compared the CA and FHN models with the same initial condition using algorithms to quantify wave behavior.

Greenberg-Hastings Cellular Automaton Model

The GH-CA model follows a simple set of rules to represent the complex physiological processes that result in electrical impulse generation, conduction, and propagation. It does so by representing electrical activity propagation by cardiac action potentials on a discrete lattice of points in space (i.e. representing the volume of the myocardium) as a form of information transmission. The cellular automaton model is made up of discrete integer numbers where each number represents its own state. States in a CA model are categorized as being at rest, excited, or refractory. Important parameters used in the CA rules for governing wave propagation are the following: the threshold (T) where $0<T<1$, the excitatory state (E), the refractory state (R), the resting state (defined as 0), size of the cardiac tissue (N) which is the number of cells in a row of a 2D CA square array, and the neighborhood size (r) which determines the number of neighboring cells that affect the current cell state in next time step. We define a 'cell' state as a discrete integer that represents the state of the cell at position (i,j) in the 2D square matrix. For the wave to progress, the state of each cell must be updated during each generation based on the simple rules we define in our GH-CA model. Let $u_j(t)$ be a cell state at a certain time step (or generation). If $1 \leq u_j(t) \leq E+R$, then $u_j(t+1) = u_j(t) + 1$. In simulations using a neighborhood with square boundaries (a Moore neighborhood), we saw unrealistic sharp edges (Fig.2) in our CA simulations. This is due to the condition for a resting cell (state=0) to become the first state of an excited cell (state=1). For square boundaries, when

$$\frac{\#\ of\ excited\ neighboring\ cells}{total\ \#\ of\ neighboring\ cells} > threshold \text{ (Eq. 1)},$$

a resting state becomes 1. We therefore adapted a method where we created a new coordinate system initially developed by Bub *et al.* (1) into our GH-CA model. This algorithm made the edges of our waves significantly smoother (Fig. 2). Specifically, we re-defined the original coordinate of any cell from (x_0, y_0), where x_0 and y_0 are integer numbers, to $(x_0 + \varepsilon x, y_0 + \varepsilon y)$, where εx and εy are uniformly distributed decimal values between -0.5 and 0.5. We also assigned a random weight S_j to each cell in a 2D GH-CA matrix where S_j is uniformly distributed between 0.5 and 1.5. Here, new coordinates and random weights are assigned to each cell every new time step. Then we compute the distance between a resting cell to all its Moore neighboring cells $D[j,i]$. A resting cell becomes excited if

$$\frac{\sum D[j,i]<r; 0<u_j(t)\leq E,\ S_j}{\sum D[j,i]<r; u_j(t)=0, or\ u_j(t)>E, S_j} > threshold \text{ (Eq. 2) [1]}.$$

This randomization process of the new coordinate system successfully eliminated unwanted edges in our GH-CA simulation with the Moore neighborhood counting method is used (Fig. 2).

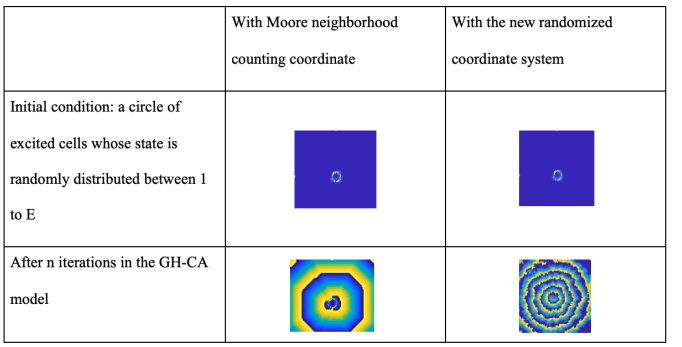

Figure 2. The effect of integrating the new coordinate system into our GH-CA model. We could see in the left column; edge wave front appears in the CA simulation whereas the wavefronts become much smoother when our new coordinate algorithm is adopted.

FitzHugh-Nagumo Equations Model

The FHN model (5) is popular for simulating excitable media because of its analytical tractability (7), relative simplicity, and ease of geometrical analysis. The basic form of the FHN model has two coupled, non-linear ordinary differential equations. One of these depicts the fast evolution of the neuronal membrane voltage while the other equation represents the slower recovery (refractory) action of sodium channel de-inactivation and potassium channel deactivation. For simulating a travelling wave, a spatial diffusion term (i.e. a second derivative in spatial coordinates) is needed for the first equation to model an action potential propagation process, which turns the FHN model into a coupled-diffusive partial differential equation. (7) The electrical propagation properties in an excitable media like nerve fibers are analogous to that of myocardium. Since the model tracks

the membrane voltage continuously and is easily controllable compared to real-time experimental cardiac tissue data, we chose to use the FHN model to create our surrogate cardiac tissue wave propagation dataset as our 'ground truth' to be compared with our GH-CA simulation.

$$\frac{\partial v}{\partial t} = f(v) - r + \frac{D\partial^2 v}{\partial x^2}$$

Where we used $f(v)=-v(a-v)(1-v)$

$$\frac{\partial r}{\partial t} = bv - gr \ \textbf{(Eq. 3)} \ (7)$$

Where we used a=0.25, b=0.001, g=0.003, and D=0.05. The function $f(v)$ is a third order polynomial that describes the fast evolution of the cardiac membrane voltage, whereas the slower recovery variable r provides negative feedback. We used the ode45 solver from Matlab 2018b to solve the above equations (Fig. 3).

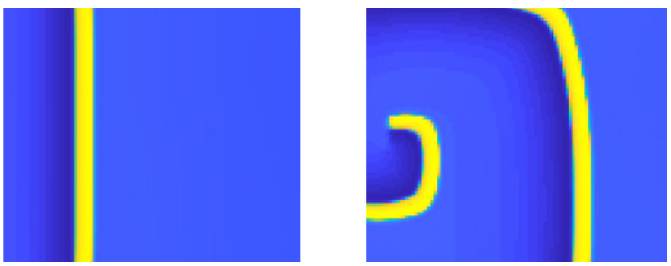

Figure 3. Simulation results of FHN model of a planar wave (left) and a spiral wave (right)

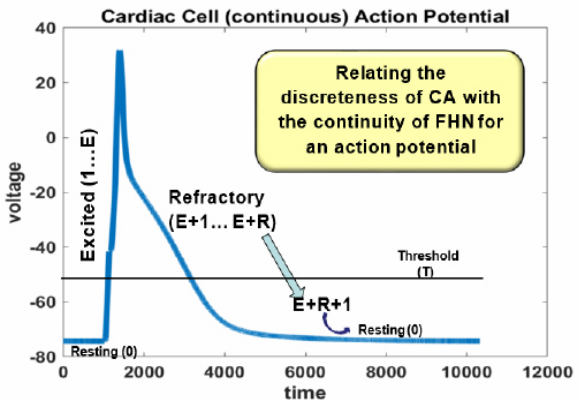

Figure 4. relating the discrete states and parameters of the GH-CA model (i.e. E, R, resting, and T) to an action potential simulated by the FHN model. The GH-CA model simulates cardiac wave propagation whose underlying process is an action potential in discrete states, while the FHN model does so with continuous membrane voltage values. This figure was generated by the Simulink toolbox of Matlab 2018b to simulate a cardiac cell action potential.

Fraction Method

To compare our GH-CA simulation with our FHN surrogate data, we define a measure called the Fraction Method which we used as a first pass to test our hypothesis that GH-CA simulation deviates from our surrogate FHN data at a rate depending on the initial condition complexity. The same initial condition representing a simple (planar) and a complex (spiral) wave were given to both the CA-GH and the FHN model. We computed a ratio of $\frac{\#ofexcitedcells}{\#ofunexcitedcells}$ (Eq. 4) for each iteration of the CA and FHN simulations for both the planar and spiral waves. This algorithm can track wave propagation dynamics regardless of wave travelling direction. A limitation of the Fraction Method has to do with the observation that it does not track the direction of the wave. In other words, the result of the same wave travelling from the left and the right of the 2D CA matrix is the same. However, this method still provides a general sense of wave dynamics regardless of the travelling direction of the wave.

To give the same IC to both models, we initially created a spiral wave from the FHN model. At a certain iteration step of interest in the FHN simulation, we first identified which CA-equivalent state this cell is in, and then directly converted the FHN values into discrete values that corresponded to the state values in a CA model based on the parameter value of a from the FHN model. Let the FHN value be γ. If $0 \leq \gamma < a$, then this cell received an equivalent resting state of 0. If $\gamma > a$ and $\gamma < 1$, then we assumed this cell to be at an excited state; since we would not be sure which discrete state the cell is in, we gave the all FHN values belonging to this range a value of 1 as its excitatory CA state. The rest of the values from the FHN model is assumed to be in a refractory state; similarly, as we do not know the specific state the cell is in, we gave all cells in this FHN value range a state of E+1, the first possible refractory state. This was the initial condition retrieved from our FHN simulation converted into discrete values that our GH-CA model could accept (Fig. 5).

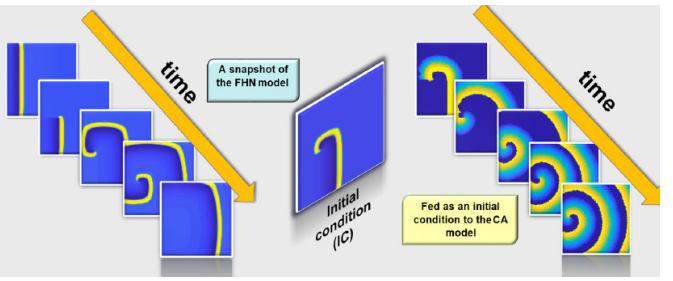

Figure 5. FHN simulation spiral wave propagation (left) where 2D values at a certain instance (middle) were retrieved as an I.C. for the CA, which were then converted into discrete values a GH-CA model can accept (right). This 2D discrete IC was then fed into the GH-CA model which then was let run and created a spiral wave propagation in the GH-CA simulation (right).

Overlap Method

Due to the difference in travelling wave velocity between both models given the same initial condition, the CA simulation always travels seemingly faster than that of the FHN model. Therefore, though the Fraction Method portrays the dynamics for both models, it cannot track how much two simulation with the same initial condition align as a function of time. We then quantify how much the two models overlap in order to test whether differences in initial conditions lead to discrepancies between the GH-CA model and experimental data. Thus, to directly compare the two models, we designed a technique to quantify the degree of overlap between the CA simulation and our 'surrogate' FHN data when both models were given the same initial condition (I.C.) by calculating a ratio of $\frac{\#ofoverlappingcells}{\#ofnon-overlappingcells}$ (Eq. 5) at each time step. We call this ratio the given I.C.'s complexity score. An "overlapping cell" is defined as when a cell is at the same state in both models (i.e. excited, refractory or at rest). This algorithm generates either a spiral or a planar wave in the FHN model, then initializes the GH-CA model with the same initial condition. Due to FHN model's slower wave propagation velocity, we let it iterate 200 steps before starting the GH-CA simulation. We then found the time point of maximal overlap between the two models by determining the ratio in Eq.5 while excluding resting cells with a state of 0. We then iterated the FHN model for another 200 steps and waited for the GH-CA model to achieve the largest match between the two models, then compute the ratio. The same process was repeated for 4 times at five different threshold values of the GH-CA model (Fig. 8).

Using Experimental data as an Initial Condition for the Cellular Automaton Model

We first obtain experimental real-time cardiac activity data (i.e. movies capturing cell's activity frame by frame) that depict a planar wave and a spiral wave from our microscope built by Bub *et al.* published in their *Nature Photonics* paper. (3) With an algorithm written in Python, we con-

verted the experimental data into readable 128 x 128 sized matrix whose further manipulation was performed in Matlab. We then removed the noise from a chosen arbitrary frame using a median filter. After this filtering, this frame's pixel values become integers. For the planar wave (Fig. 5), this chosen frame was fed into the GH-CA model directly as an initial condition. We arbitrarily let the excitatory state level (E) be 7, refractory state level (R) be 1 and the threshold value be 0.3. For the spiral wave (Fig. 6), instead of median-filtering one frame as the initial condition, all frames from the recording were filtered.

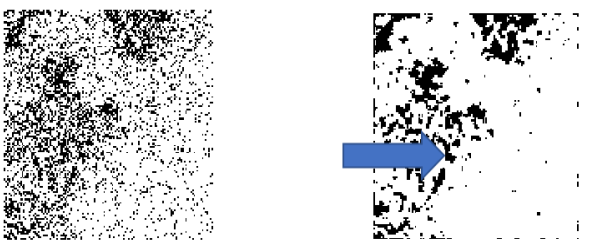

Figure 5. For the planar wave, the left image is the raw data from our experiment converted to be a matrix appearing to be noisy; the right image is the left image that has been median filtered, we can see (at the arrow) a very clear front wave compared to the left image.

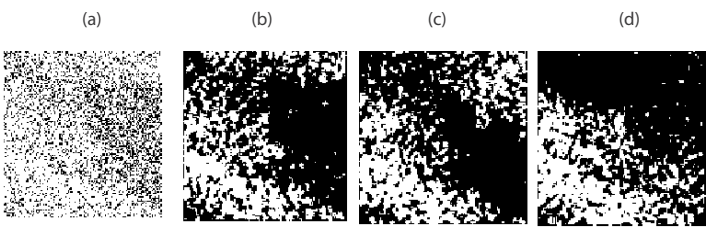

Figure 6. For a spiral wave, (a) a chosen starting frame of the raw data from our experiment converted to be a matrix appearing to be noisy; (b) is image (a) after it has been median filtered; since all frames of the raw data have been filtered, (c) and (d) represent the spiral wave at two later time points from (b), we can clearly see a wave spiraling into the center of the plate in the experiment.

Results

Following the Fraction Method, for a simple planar wave, we observed a rather constant $\frac{\#ofexcitedcells}{\#ofunexcitedcells}$ ratio after the FHN model is caught up with the GH-CA model at about the 60th GH-CA iteration (Fig. 7). A planar wave (i.e. simple wave) in the cellular automaton model is always better aligned with its FitzHugh-Nagumo model of the same initial condition than a spiral wave. We show in Fig.7 that the ratio from the FHN and CA simulations becomes the same at 0.05 as iteration step continues to 100, which indicates the two models are behaving the same with a planar wave initial condition. In contrast, when applying the Fraction Method with a spiral wave I.C.to both the FHN and GH-CA models, we saw a strong deviation between the ratio between excited and unexcited cells as iteration steps increase for both models. Since a CA wave travels faster than an FHN wave because of its state's discreteness, the CA was iterated for 100 steps and the FHN was iterated for 872 steps, which achieve the same 'cell distance'. Evidently in Fig.8, the fraction method traced the FHN simulation dynamics of a spiral wave until the wave disappears at the end of its simulation at the edges of the 2D FHN matrix, which was why its excited to unexcited cells ratio dropped dramatically to 0 at its 780 iteration step. However, it the FHN spiral wave behaviors is still captured by the fraction method precisely. In contrast, the GH-CA simulation whose IC was given by CA at a certain instant never completely left the edges of the 2D GH-CA matrix. Therefore, its excited to unexcited cells ratio was significantly higher than that of the FHN simulation around 50 GH-CA iterations. However, regardless of whether the spiral wave exits the 2D array or not, a noticeable discrepancy between the excited to unexcited cell count ratio can be seen between the GH-CA (red) and the FHN (blue) models.

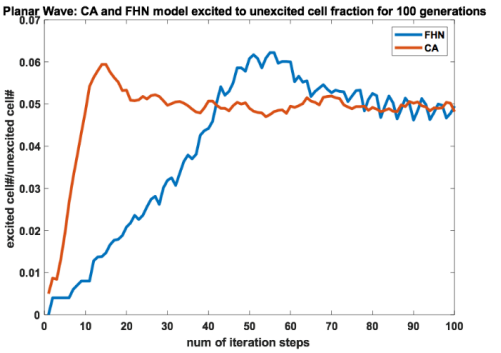

Figure 7. Fraction method performed to a simple planar wave initial condition given to the FHN (blue) and GH-CA (orange) models for 100 GH-CA and FHN iterations.

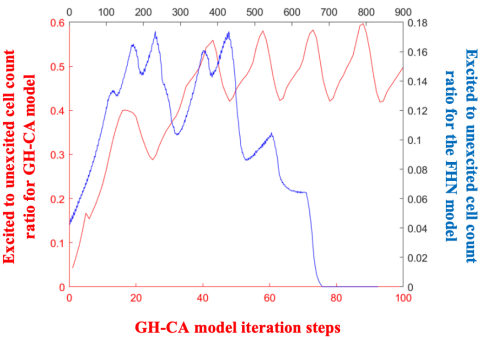

Figure 8. Fraction method performed to a complex spiral wave initial condition given to the FHN (blue, bottom line) and GH-CA (red, top line) models for 100 GH-CA iterations and 872 FHN iterations which travel the same cell equivalent distance.

We quantified the degree of overlap between the CA model simulation using the **Overlap method**. We were able to generate the degree of overlap of a planar and a spiral wave compared between the FHN and CA models. We directly compared the CA and FHN models for five different CA thresholds (color-coded) shown in Fig. 9. We can see 1) the cellular automaton model's planar waves are better aligned with the FHN models initiated by the same I.C. than the spiral waves, and 2) the fit with the spirals decreases as the number of iterations increase, while the plane waves do not decrease as much.

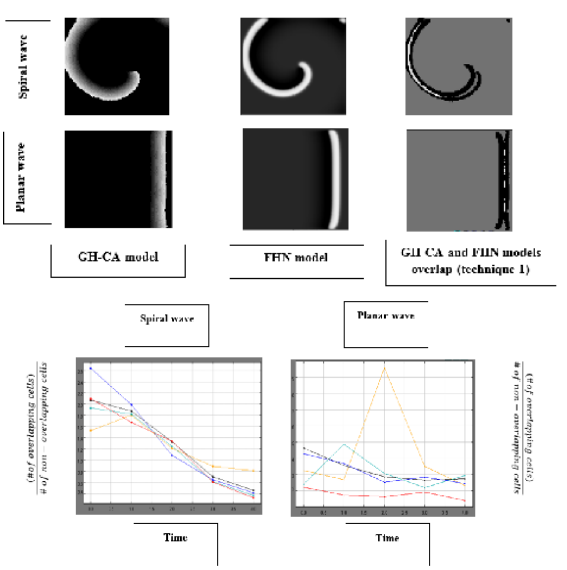

Figure 9. Overlapping to non-overlapping cells ratio between GH-CA and FHN models by capturing the maximal overlap after allowing the FHN simulation to run 200 iterations ahead of the GH-CA one and let the latter to catch up.

Furthermore, our method of initializing the CA with one frame of median-filtered real-time experimental cardiac tissue data was successful in reproducing the behavior of the raw data for the planar wave (Fig. 10), but not for the spiral wave. As for the spiral wave, its initial condition evolves to be a planar wave rather than spiraling into the center.

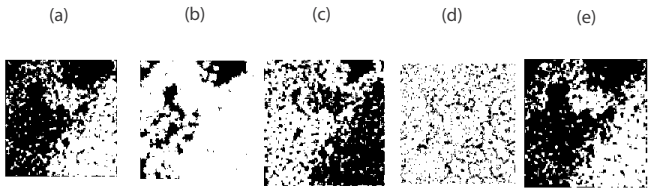

Figure 10. After a median-filtered initial condition of the experimental result of a planar wave (a) have been fed into a GH-CA model, (b) – (e) shows this planar wave's wave propagation evolution, successfully reproduced the raw data.

Discussion

Discussion of results and possible improvements

We used the FitzHugh-Nagumo model to generate data sets of simple and complex propagating waves in excitable media. We evaluated whether a cellular automaton model can reproduce these wave dynamics using two algorithmic tests, a '*fraction method*' and an '*overlap method*'. For future work, since we chose our 2D matrix cell size to be 100 for both the GH-CA and the FHN models, we could try varying the size of our simulation to see if it changes our results. For all results generated other than experimental result simulation, we relied on the condition that the GH-CA and the FHN models were initiated by the same I.C.. This was retrieved from the FHN voltage values at a certain instance and converted into discrete cell states that the GH-CA model can accept as an IC too (see Methods). We strictly used the parameter *a* from the FHN model to set the ranges for the excited and refractory cell states. In addition, we assigned a 1 for all FHN values in the "excited range" and an E+1 for all that's in the "refractory range". Both parameter assignments increase the probability that the discrete values we converted the FHN values to may not be entirely representative of the original FHN simulation, which can affect the consequent GH-CA simulation results and the comparison results between the two model.

Proposed strategies of integrating Machine Learning techniques into our GH-CA model to lessen the deviation of its simulation from experimental data

Our future strategy to resolve the challenge of spiral wave experimental data initial condition not being able to induce spirals is inspired by Gilpin's (6) recent paper "*Cellular automata as convolutional neural networks*". Gilpin (6) showed that *any* Cellular Automaton may readily be represented using a convolutional neural network (CNN) with a network-in-network architecture. They built a CNN that can learn the dynamical rules for arbitrary CA when given videos of the CA as training data. They trained ensembles of networks on randomly sampled CA and showed that CA with simpler rule tables produce trained networks with hierarchical structure and layer specialization, while more complex CA tend to produce shallower representations. This is analogous to our results. After they confirmed that arbitrary cellular automaton may be represented by convolutional perceptrons with finite layers and units, automated training of neural networks on time series of cellular automaton images was carried out. By training ensembles of convolutional neural networks on random images and random CA rulesets. More specifically, they defined a CA as an explicit mapping between an input pixel and an output pixel. For a neighborhood size of 1, there are 512 possible neighbor state groupings (i.e. a center cell has 2^9=512 neighboring state possibilities for 3-by-3-pixel groups in a binary image) for an input pixel. Therefore, their training data is an ensemble of randomly generated binary images where each image contains an equal number of black and white pixels on average. For each training input pixel, its output pixel ('label') is produced by the input pixel undergoing a randomly selected CA Conway Game of Life rule set. Subsequently, this CA map is applied to an ensemble of random binary images (the training data), in order to produce a new output binary image set (the training labels). They used a sufficient number of images and training data batches (500 images) to ensure that the training data contains at least one instance of each rule.

We may adapt this technique to train a Convolutional Neural Network (CNN) representing our CA with experimental median-filtered spiral wave data. Complex wave propagation such as a spiral wave presents dynamical rules to be learnt by the CNN like Conways' Game of Life. The input training data would be our experimental data at time with frame n, its output label data would be the experimental data at time t+1 of frame n+1. Iteratively, the next training image would be frame n+1 whose output label is frame n+2. An initial step would be to train the CNN representing our CA on synthetic wave data generated by the FHN model where its continuous values are converted into equivalent discrete states of CA values. This procedure is promising to adeptly train a CNN for leaning spiral wave propagation with an ability surpassing that of a GH-CA model.

Conclusion

We compared the dynamics of two models of cardiac propagation: a discrete Greenberg-Hastings Cellular Automaton (GH-CA) model and a continuous FitzHugh-Nagumo (FHN) model for excitable propagation. Our results were able to confirm our hypothesis that the divergence between the two models are due to initial condition (IC) complexity (i.e. the more complex the I.C. is, the faster the GH-CA simulation results deviate from the FHN model results where we treat our FHN simulation as our ground-truth surrogate experimental data). This was shown by the alignment between the CA model of a planar wave (i.e. simple) and the FHN model remaining constant, while the degree of overlap between the CA and FHN models decreasing for a spiral wave (i.e. complex) at higher time steps. Once a median-filtered frame of real-time cardiac tissue experimental data was fed to the CA model as an initial condition, the planar wave simulation propagates forward as the experimental data shows, but the spiral wave CA simulation did not successfully spiral in. This result further identifies the issue that a simpler I.C. such as a planar wave can be modelled more accurately by CA than a complex I.C. such as a spiral wave.

Acknowledgements

This project was a fulfillment for the course PHGY 461 D1/D2: Experimental Physiology at McGill University. The author would like to thank McGill M.Sc. candidate Miguel Romero Sepulvida for generously providing experimental data for simulation results of this paper and Harvard NSF-Simons Research Fellow Dr. William Gilpin for fruitful discussions and further collaboration on Convolutional Neural Network as Cellular Automata model for our cardiac experimental data. Especially, the author would like to sincerely thank her supervisor Dr. Gil Bub from McGill's Department of Physiology for continuously offering his patience, intellectual guidance, stimulating discussions and support.

References

1. Bub G, Shrier A, Glass L. Global organization of dynamics in oscillatory heterogeneous excitable media [Internet]. Physical review letters. U.S. National Library of Medicine; 2005 [cited 2020Jan15]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15698236

2. Bub G. Contents [Internet]. Optical Mapping of Pacemaker Interactions. [cited 2020Mar7]. Available from: https://www.physiol.ox.ac.uk/~gb1/cnd/bub/thesis.html#tth_sEc1.3

3. Burton RAB, Klimas A, Ambrosi CM, Tomek J, Corbett A, Entche-

va E, et al. Optical control of excitation waves in cardiac tissue [Internet]. Nature photonics. U.S. National Library of Medicine; 2015 [cited 2020Jan15]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4821438/

4. Ccoffinn: Automated Wave Tracking in Cultured Cardiac ... [Internet]. [cited 2020Jan15]. Available from: https://www.cell.com/biophysj/pdf/S0006-3495(16)30816-5.pdf

5. FitzHugh R. Impulses and Physiological States in Theoretical Models of Nerve Membrane [Internet]. Biophysical Journal. Cell Press; 2009 [cited 2020Jan15]. Available from: https://www.sciencedirect.com/science/article/pii/S0006349561869026

6. Gilpin, William. Cellular automata as convolutional neural networks [Internet]. arXiv.org. 2019 [cited 2020Jan15]. Available from: https://arxiv.org/abs/1809.02942

7. Murray JD. Mathematical Biology - I. An Introduction: James D. Murray [Internet]. Springer. Springer-Verlag New York; [cited 2020Jan15]. Available from: https://www.springer.com/gp/book/9780387952239

8. Pálsson E, Cox EC. Origin and evolution of circular waves and spirals in Dictyostelium discoideum territories [Internet]. Proceedings of the National Academy of Sciences of the United States of America. U.S. National Library of Medicine; 1996 [cited 2020Mar7]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC40047/

Submitted: 01/25/20

Review Article

[1]Department of Biology, McGill University, Montreal, QC, Canada

Keywords

Leukemia inhibitory factor, female contraceptive

Email Correspondence

hannah.dolin@mail.mcgill.ca

Hannah Dolin[1]

Leukemia Inhibitory Factor (LIF) Modulation: a Novel, Non-Hormonal Contraceptive Method

Abstract

Background: Numerous studies have demonstrated that Leukemia Inhibitory factor (LIF) plays an essential role in embryonic implantation. Vaginal application of pegylated LIF antagonists can successfully prevent implantation and pregnancy in mice. The development of non-hormonal, female-controlled contraceptives is imperative, as combined oral contraceptives are associated with depression and are not feasible for use in all women.

Methods: This paper reviews 44 studies regarding LIF, implantation, hormonal contraceptives and the use of LIF antagonists as a means to inhibit pregnancy.

Summary: Current research indicates that LIF-modulation could be effective as a non-hormonal contraceptive method, although researchers should be wary of the negative side effects associated with systemic LIF modulation. Vaginal application of LIF antagonists could decrease the risk of negative side effects.

Introduction

Reciprocal molecular communication between an implantation competent blastocyst and a receptive uterus is imperative for proper embryo implantation. Problems during implantation can result in miscarriage or pregnancy associated disorders. (1) Proper implantation depends on both maternally and embryonically produced factors. The best known maternal factors are ovarian estrogen and progesterone, but factors secreted from the uterine glands such as Leukemia Inhibitory Factor (LIF) are also vital for successful implantation. (2,3)

This review aims to provide a brief overview of current knowledge regarding the structure of uterine glands and their role in establishing pregnancy, specifically through their secretion of leukemia inhibitory factor (LIF) during implantation and decidualization[1]. LIF may also act as a trophic factor for the embryo. (6,7) Next, this review aims to discuss whether LIF could be a viable alternative to current forms of hormonal contraception, specifically, combined oral contraceptives. (COC) Various adverse side effects of COC have been identified including depression, which is the most commonly cited reason for discontinuation. (36) Other side effects like thromboembolism and hypertension have been described but these are less common. (8) Moreover, COC usage is contraindicated in women with known familial thrombophilia and other health conditions such as migraines with aura. Therefore, women with these conditions could benefit from the development of alternative contraceptives. (44)

Uterine Structure

Before discussing the uterine glands and LIF, the general structure of the uterus must be outlined. The uterine wall can be divided into two distinct compartments: the inner mucosal lining (the endometrium) and the smooth muscle outer wall (the myometrium). The endometrium comprises two epithelial cell types: the luminal epithelium (LE) and glandular epithelium (GE). (10) The adult human endometrium can also be stratified into two distinct structural zones: the stratum functionalis and the subjacent stratum basalis. The stratum functionalis contains LE, stroma and branched uterine glands and, unlike the stratum basalis, is shed during menstruation. The stratum basalis thus serves as the generative base for adenogenesis, or uterine gland development, which forms tubular glands. During the proliferative phase of the menstrual cycle, these tubular glands undergo extensive branching to form coiled, branched structures termed "glands" within the loosely packed region of the stratum functionalis called the stratum spongiosum. (10,11)

Due to the ethical constraints related to human research, most experimental studies on the role of uterine glands and their secretions are conducted in mice. Unlike humans, mice are born with a simple epithelial uterus that lacks endometrial glands. Invagination of the luminal epithelium at postnatal day 7 results in the formation of tubular uterine glands. (10,11) As mice do not menstruate, they do not rebuild these glands cyclically.

Leukemia Inhibitory Factor

Vital to implantation

The vital role of uterine glands in establishing pregnancy is largely attributed to their expression and secretion of leukemia inhibitory factor (LIF). (2) LIF is a cytokine of the interleukin-6 family and one of the few molecules that are obligatory for fertility in mice. (9) In a seminal study by Stewart *et al.*, homozygous LIF-null mutant female mice were found to be infertile, even when mated with wild type males.(12) Infertility was found to be caused by failure in implantation, rather than in ovulation or fertilization as fertilized, unimplanted blastocysts could be recovered from LIF-null uteri on GE day 4-7. These recovered mutant blastocysts could implant successfully and develop to term when transferred to wild type females, demonstrating that maternal LIF, rather than embryonic LIF, is required for proper implantation.

LIF is also important for implantation in humans. Examination of fluid within the uterine lumen demonstrated that infertile women secrete significantly less LIF than fertile women. (13) Similarly, women undergoing in-vitro fertilization (IVF) with relatively high levels of endometrial LIF during the window of implantation are more likely to become pregnant than their counterparts. (14,15)

Expression and function

Uterine LIF levels peak biphasically. First, LIF is expressed in the murine uterus at ovulation. However, LIF deficiency has not been shown to affect ovulation. Next and seemingly more important, LIF peaks in the GE during the window of receptivity, which immediately precedes implantation and occurs on day 4 of gestation in mice. This second peak occurs in response to the nidatory estrogen pulse[2] and is essential to rendering the luminal epithelium receptive to embryo attachment. In fact, LIF repletion in ovariectomized mice can compensate for missing nidatory estrogen, demonstrating that vital uterine changes triggered by the nidatory estrogen surge in mice are largely mediated through LIF. (16) LIF is also expressed in the decidualizing stroma during the attachment reactions. (17) In humans, LIF is most abundantly expressed in the GE of fertile women during the secretory or postovulatory phase between days 18 and 28 of the menstrual cycle, analogous to the murine "window of implantation" (18,19)

Secreted LIF functions by binding to its cognate transmembrane receptor (LIFR), which is localized on the apical membrane of the LE. The LIFR-LIF complex then binds to glycoprotein 130 (gp130) to form the activated LIF-LIFR-gp130 complex. (5,20) This complex modulates myriad downstream pathways in the endometrium, such as JAK-stat, wnt/B-catenin, notch and mTOR signalling. (5) For instance, upon LIF-LIFR-gp130 binding, STAT proteins are tyrosine-phosphorylated, homodimerized, and translocated to the nucleus, where they modulate the expression of target genes. Deletions of the gp130 region responsible for STAT phosphorylation also results in failures during implantation, demonstrating that this LIF-activated pathway is vital. (21,22)

LIF-effected pathways are implicated in diverse processes allowing for successful implantation. For example, LIF is postulated to modulate adhesion between endometrial epithelial cells and trophoblast cells. Further, LIF plays a vital role in stromal cell decidualization as LIF-/- mice do not undergo decidualization. (5) While it is still not clear which of LIF's many functions are truly essential to implantation, LIF has repeatedly and conclusively been shown to be vital for successful adhesion, decidualization and thus, implantation. (2,3,23)

Leukemia Inhibitory Factor Antagonists as an Alternative to Hormonal Contraceptives

Because of LIF's vital role in implantation, LIF-modulation to prevent implantation is being explored as a novel, non-hormonal contraceptive method. (5,12)

Previously, Fairlie *et al.* used alanine scanning mutagenesis[3] to target residues within the C and D helices of LIF's four helical structure which are responsible for the species specific binding affinity of human LIF (hLIF) to human LIFR.(20) Using this method, key protein regions and amino acids essential for LIF-LIFR and LIF-gp130 binding were identified. Importantly, this study found that abolition of the identified gp130 binding site in hLIF required for productive LIF-signalling, combined with mutations to increase the protein's affinity to LIFR by >1000-fold generated a potent antagonist of LIF action. This protein will henceforth be referred to as LIF antagonist (LA).

LA has been shown to have a half-life of 10 to 30 minutes. Consequently, effective inhibition of LIF activity and, thus, blastocyst implantation in mice using LA requires continuous LA administration. This can be accomplished by using miniosmotic pumps implanted within the peritoneal cavity, in combination with 4 hourly intraperitoneal LA injections. (21) Evidently, continuous LA administration would not be practical. Therefore, the instability of LA could negate its utility as an alternative, non-hormonal contraceptive.

Fortunately, as shown by Reddy *et al.*, covalent attachment of a branched 40-kd polyethylene glycol moiety (PEG) through a process referred to as "PEGylation" can increase the stability of circulating cytokines. (24) Reddy *et al.* showed that PEGylation of an interferon (IFN) used to treat Chronic Hepatitis C (CHC) resulted in sustained IFN absorption, reduced IFN clearance from the plasma, and higher IFN concentrations in blood plasma when compared to non-PEGylated IFN. (24) By increasing molecular size, PEGylation reduces renal ultrafiltration of attached compounds, decreases uptake of the compound by the liver and protects the compound from proteolytic cleavage. (21) In the study described above, the increased stability of PEGylated IFN meant that the cytokine could be administered less frequently and still be effective in treating CHC; PEGylated IFN was effective when administered once weekly, whereas IFN alone was required three times a week. (24)

In order to increase the stability, and thus, potential utility of LA as a contraceptive, White *et al.* conjugated LA with two PEG molecules (henceforth referred to as PEGLA). (21) Strikingly, mated mice who received only 3 intraperitoneal injections of PEGLA and did not receive a transplanted miniosmotic pump had no implantation sites. Thus, while LA administration for contraceptive purposes may not be feasible in humans, PEGLA administration could be both effective and feasible.

Rather than injecting PEGLA, some researchers have attempted to develop a "birth control vaccine". This would allow one's own immune system to target and neutralize endogenous LIF. Notably, Lemon and Naz induced a LIF- or LIFR-directed immune response by conjugating LIF or LIFR peptides to T cell carrier proteins. (6) While this approach significantly decreased fertility, it did not totally abolish implantation. However, this may not be the case in humans because they are less fertile than mice.

While most studies to date have been conducted in mice, LIF antagonism using antibodies has been shown to inhibit implantation in rhesus monkeys. (25) An in-vitro study by Lalitkumar *et al.* demonstrated that PEGLA can inhibit embryo attachment in humans. In this study, three-dimensional cell culture models on endometrial tissue and excess IVF embryos were combined either with or without PEGLA. (7) PEGLA-treated samples saw failure in embryo attachment to culture while control samples did not. Further, PEGLA treated blastocysts failed to hatch, or emerge from the zona pellucida, and expand. Unlike control blastocysts, PEGLA-treated blastocysts lost phospho-AKT-1 which is a vital factor for cell survival. Further, blastocysts exposed to PEGLA underwent apoptosis at significantly higher rates as indicated by increased caspase-3 activity. Taken together, these points allude to the trophic role of LIF on human embryos as well as affirm its effects on human uterine receptivity. (7) This finding supports that inhibition of LIF activity could prevent implantation in humans. However, LIF-modulation for contraceptive purposes could have certain risks. LIF is implicated in a plethora of non-reproductive pathways. Modulation of these pathways via systemic LIF-modulation, either through a vaccine or intraperitoneal PEGLA administration, could result in adverse side effects. (9,12,18,26,27) These risks would likely increase when LIF is modulated over an extended period as would be necessary for contraceptive purposes.

For example, LIF is produced by astrocytes in response to autoimmune insults within the central nervous system and has been shown to increase oligodendroglial survival in-vitro. Similarly, LIF has been shown to prevent oligodendrocyte death in animal models of multiple sclerosis (MS). (28) Systemic administration of LIF antagonists over four days has been shown to worsen experimental autoimmune encephalomyelitis (EAE), which is an experimental model of autoimmune disease induced by immunization against myelin epitopes in mammals. EAE is a standard animal model for multiple sclerosis (28), doubling observed oligodendrocyte loss. (29) Thus, any LIF-targeting contraceptive which would systematically inhibit LIF action could have negative effects on oligodendrocyte survival and thus, myelin integrity.

Moreover, LIFR-null mutant mice have been shown to have a reduced number of spinal and brainstem astrocytes. (9)

Further, LIF may play a key role in modulating hematopoiesis. LIF has been shown to be constitutively expressed in bone marrow stroma. Exogenous LIF administration has been shown to increase platelet counts in a dose-dependent manner. (26) Thus, systemic LIF modulation for contraceptive purposes could potentially alter blood composition. To my knowl-

edge, this effect has yet to be studied in LIF-null mice.

However, LIF exhibits significant homology with Oncostatin M and Ciliary Neurotrophic Factor, both of which are secreted factors that can bind the LIFR. Therefore, the effect of LIF modulation on the non-reproductive systems described above could be dampened by other homologous factors binding to LIFR.(12) In fact, despite the many potential issues of systemic LIF modulation, LIF-null mice in Colin Stewart's seminal LIF-knockout study appeared normal, although slightly smaller than their wild type counterparts. (12) However, in Stewart's study, the non-reproductive health of the mice was not examined closely. (12)

Nevertheless, Menkhorst *et al.* demonstrated that intraperitoneal injections of PEGLA (like those performed by White *et al.*) affect bone remodelling. (9,21) Specifically, intraperitoneal injections of PEGLA resulted in increased cancellous bone volume and thickness. Further, Menkhorst *et al.* observed decreased osteoclast levels in females injected with PEGLA. (9) Non-mated females treated with PEGLA also had less osteoid[4] and osteoblasts than controls. This effect was not observed in mated females, as control-mated females already had low osteoblast and osteoid levels. Taken together, these results indicate that intraperitoneal injection of PEGLA decreased bone remodelling. This is a notable finding with regards to PEGLA's potential use in women as a contraceptive as low levels of bone remodelling increase the risk of fracture in adult humans. (29) More generally, this finding demonstrates that PEGLA and LIF-modulating vaccinations can affect systems outside of the uterus.

Previous research has shown that vaginally administered drugs preferentially localize to the uterus in what is termed as the "uterine first pass effect." (9,30) The preferred explanation for this effect is that drugs administered via the vagina are absorbed through veins within the upper third of the vagina and are transported by countercurrent exchange (exchange by diffusion between two fluids flowing in opposite direction) to the uterine arteries. However, the exact mechanism of the "uterine first pass effect" is still unclear. (9) Possible explanations for this phenomenon include direct diffusion of the compound through tissues, passage of the compound to the uterine lumen through the cervical lumen, and transport of the compound via venous or lymphatic circulatory systems (the lymph system has been recognized as an important carrier for steroid hormones). (30) Regardless of the exact mechanism of the "uterine first pass effect", this phenomenon can be exploited to mitigate the potential systemic effects of PEGLA application. In fact, vaginal application of PEGLA results in its localization to the uterine luminal epithelium and the basal surface of the glandular epithelium as opposed to its localization to the liver, ovary, oviduct, spleen and thyroid, which was the case in IP injected non-mated mice. (9) Strikingly, in Menkhorst *et al.*'s study, mice that received PEGLA vaginally did not have any change in cancellous bone volume or thickness or in osteoclast/blast number and size when compared to controls, while mice receiving PEGLA intraperitoneally did. (9) This demonstrates that systematic effects of LIF modulation by PEGLA can be mitigated by vaginal administration of the compound. Furthermore, vaginal administration of PEGLA still effectively inhibited implantation, demonstrating that PEGLA can be absorbed through the vagina to inhibit LIF signalling and therefore, implantation in the uterus. (9)

If LIF-modulating contraceptives are to be brought to market, future studies on LIF modulation or using LIF-knockout mice, ought to more thoroughly investigate the overall health of the organism rather than only focusing on implantation. For instance, while Menkhorst *et al.* showed that both intraperitoneal and vaginal administration of PEGLA did not exacerbate murine EAE, the effects of long-term systemic PEGLA exposure, or LIF modulation in general, on the CNS are yet to be explored. (9) Given that the effects of LIF modulation could be mitigated by other homologous factors, analysis of LIF-null or LIF-modulated organisms would be the best way to determine the true extent to which LIF modulation affects non-reproductive systems in-vivo. Additionally, given that Menkhorst *et al.* showed systemic PEGLA administration altered bone density, researchers who aim to develop novel non-hormonal contraceptives should move away from systemic LIF modulation and towards uterus-targeted LIF modulation via vaginal administration (9) as a more cautious approach.

Hormonal Contraceptives

A discussion of the need for and merits of LIF modulation as a non-hormonal alternative to current contraception necessitates a brief overview of hormonal contraceptives: their formulation, basic mechanism of action, and adverse effects. This will help elucidate whether or not the development of non-hormonal contraceptives in general, and LIF modulators specifically, can bring about more social benefit than research regarding the amelioration of hormonal contraceptives.

While various different compositions and forms of hormonal contraceptives are presently available, the most common are combined oral contraceptives (COC). As such, they are a reasonable standard against which novel contraceptives such as LIF antagonists can be compared. Given that COC is used by over 100 million women worldwide as a form of birth control, its associated adverse side effects should be viewed as an important public health issue. (31)

Combined oral contraceptives pills contain both synthetic estrogen (typically containing synthetic ethynyl derivatives, ethinyl-estradiol and mestranol) and synthetic progesterone (typically containing synthetic desogestrel, ethynodiol diacetate, gestodene, levonorgestrel, lynestrenol, norethisterone, norethisterone acetate, norgestimate or norgestrel). (8) The progesterone and estrogen analogs inhibit LH and FSH peaks, which normally occur prior to ovulation, via negative feedback. Inhibition of the LH surge stops ovulation, or the release of an egg from the ovaries into the peritoneal cavity. Further, the progestin component thickens cervical mucus, which decreases sperm penetration, and reduces endometrial proliferation, which decreases uterine receptivity to implantation. Most combined oral contraceptives contain placebo pills without estrogen or progestin to stimulate menstrual bleeding while keeping women in the habit of taking the pill every day.

Beyond combined oral contraceptives containing both synthetic estrogen and progestins, progestin-only forms of contraception are also widely available. (32) Progestin-only contraceptives are often prescribed when estrogen administration is problematic. For example, exogenous estrogen is thought to negatively affect milk production in breastfeeding mothers (8,33).

While hormonal contraceptives are effective in preventing pregnancy, their use has been associated with an increased risk of thromboembolism, hypertension, and gallstones. (8) Both combined oral hormonal contraceptives (COC) and progestin-only contraceptives (POC) are associated with altered bone metabolism. Decreased bone turnover is observed in both formulations of OC, while decreased bone resorption is observed only in POC. (43) Further, COC use is associated with a small, yet significant increased risk of breast cancer. However, COC use has been associated with a decreased risk of endometrial, ovarian and colorectal cancer. (34)

Aside from physical ailments, a postulated side effect of hormonal oral contraceptive use is depression or dysthymia. Yet, studies regarding the extent of this effect have been inconsistent; some studies find that birth control is not associated with worsened effect (36,37), while others show the opposite. (8,38) Skovlund *et al.* found that COC users were 1.8X more likely that non-users to start using antidepressants for the first time. (38) This trend was even more pronounced for POC users, who were 2.2X more likely to start antidepressant use. Inconsistency in this research may be due to varied contraceptive formulations and approaches to measuring depressive symptoms. Further, the multifactorial nature of depression complicates studies. (37)

Controversy notwithstanding, estrogen and progesterone have been shown to have psychoactive properties. (39,40) The precise mechanism of and extent to which estrogen and progesterone lead to depressive symptoms is still unclear. (8,41) Current theories generally attribute depressive symptoms to progestins. (37) For example, progesterone is postulated to decrease serotonin levels by increasing the activity of monoamine oxidase, which degrades serotonin[5]). (8,38) Progesterone metabolites have also been shown to act on the γ-aminobutyric acid A (GABA) receptor complex, a major inhibitory system in the human CNS. (38) As follows, various studies have also shown that oral contraceptives with high progestin

doses are more frequently associated with depressive symptoms. (41,38) For instance, Lawrie et al found that women taking progestin-containing injectable contraceptives postpartum scored significantly higher than the placebo group in terms of depressive symptoms as measured by the Montgomery-Asberg Depression Rating Scale. (41)

Vaginal administration of synthetic estrogens and progestins via a vaginal ring (somewhat analogous to vaginal PEGLA administration (9)), may alter the severity or incidence of negative side effects associated with combined oral contraceptive use in women. Select studies have shown that ring usage is associated with a lower incidence of negative COC-associated side effects, such as depression. Unfortunately, ring-usage is also associated with increased incidence and severity of local side effects such as leukorrhea and vaginitis. (42)

Conclusion/Future Directions

The development of novel non-hormonal birth-control methods is imperative as hormonal contraceptives have been associated with an increased risk of depression (8,38) and are not suitable for women with certain pre-existing conditions. (44) LIF modulation could potentially be effective for this purpose.

However, one should be wary of the effects of systemic LIF modulation, as LIF is involved in many non-reproductive systems. No studies to date have fully characterized the effects of systemic LIF modulation in mice, primates, or humans. Systemic inhibition of LIF activity, via LIF-antagonists or immunization, could therefore have adverse side effects. Future studies should aim to better characterize non-reproductive health (bone density, blood composition and myelin integrity) in mice or primates receiving systemic or local inhibition of LIF, and in LIF-null mice. This would allow researchers to better understand the potential side effects of LIF modulation, and likely further support the stipulation made by Menkhorst *et al.* (9) that systemic LIF-modulation may not be appropriate for contraceptive purposes.

Menkhorst *et al.* demonstrate that adverse side effects of systemic LIF-modulation, such as altered bone density, could be mitigated by administering PEGLA vaginally and thus, focusing its LIF-inhibitory activity to the uterus. (9) They also propose that a vaginal gel delivering PEGLA could also be used to simultaneously deliver microbicides to inhibit sexually-transmitted infections (STIs). (9) However, much remains to be done before such a gel can be developed. In addition to performing further animal studies, research should be performed to determine if women would actually want to use a gel as their primary contraceptive method, or if it seems too messy or difficult to administer. If the latter were the case, PEGLA could potentially be delivered via a vaginal ring.

Further, more research ought to be done regarding the adverse side effects associated with hormonal birth control usage. Specifically, given that previous research regarding depression and oral contraceptive usage has yielded inconsistent results and sparked disagreement amongst scientists (37,38), further meta-analysis of human data as was done by Skovlund *et al.* ought to be done to conclusively demonstrate that COC usage increases the risk of depression. Further, more direct methods should be employed in mouse models to try and understand the molecular mechanism of this observed effect.

A deeper understanding of the mechanism by which hormonal contraceptives bring about adverse side effects will serve as a good starting point for remodulating these drugs. For instance, some studies have shown that the depressive side effects of COC are mostly due to the progestin component (37), so future formulations might be improved by simply decreasing the amount of progestin. Remodulated oral contraceptives could potentially decrease the severity of negative side effects and would likely be publicly available much sooner than LIF-modulating contraceptives which are still in the preliminary stages of development.

Alternatively, a better characterization of the negative effects of hormonal contraceptives could further support the argument that society ought to invest heavily in the development of novel, non-hormonal contraceptives, namely PEGLA. Present studies suggest that vaginal PEGLA administration could be associated with fewer negative side effects than COC.(9) Additionally, several non-LIF genes have been identified in the uterus which are also vital for maintaining pregnancy. (2) Like LIF, these genes could potentially be modulated for contraceptive purposes.

To conclude, present research suggests that vaginally administered PEGLA could be as effective as hormonal contraceptives in terms of preventing pregnancy in women. (7,9,21) While systemic LIF-modulation has been shown to be associated with altered bone density, vaginal administration of PEGLA seems to mitigate this risk without enhancing local side effects, which was the case when hormonal contraceptives were administered vaginally (9,42) Therefore, in terms of mitigating negative side effects and thereby increasing clinical desirability, vaginally administered LIF-modulating contraceptives could be preferable to hormonal contraceptives.

[1] Decidualization is the process by which stromal cells differentiate, resulting in the formation of a "decidua", which provides the blastocyst with nutrients until the functional placenta is formed, while simultaneously restraining trophoblast-uterine invasion. (4,5)

[2] A secondary, small ovarian estrogen pulse, occurring on GE day 4 in mice.

[3] A biochemical technique that identifies vital residues of a studied protein by replacing single residues with alanine, via targeted mutagenesis, and subsequently assaying protein function.

[4] Organic component of bone.

[5] Monoamine oxidase inhibitors are a form of antidepressant. (39)

References

1. Cross JC, Werb Z, Fisher SJ. Implantation and the placenta: key pieces of the development puzzle. Science. 1994 Dec 2;266(5190):1508–18.

2. Kelleher AM, Milano-Foster J, Behura SK, Spencer TE. Uterine glands coordinate on-time embryo implantation and impact endometrial decidualization for pregnancy success. Nature Communications. 2018 Jun 22;9(1):1–12.

3. Stewart CL, Kaspar P, Brunet LJ, Bhatt H, Gadi I, Köntgen F, et al. Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. Nature. 1992 Sep 3;359(6390):76–9.

4. Okada H, Tsuzuki T, Murata H. Decidualization of the human endometrium. Reprod Med Biol. 2018 Feb 1;17(3):220–7.

5. Rosario GX, Stewart CL. The Multifaceted Actions of Leukaemia Inhibitory Factor in Mediating Uterine Receptivity and Embryo Implantation. Am J Reprod Immunol. 2016 Mar;75(3):246–55.

6. Lemons AR, Naz RK. Birth control vaccine targeting leukemia inhibitory factor. Molecular Reproduction and Development. 2012;79(2):97–106.

7. Lalitkumar S, Boggavarapu NR, Menezes J, Dimitriadis E, Zhang J-G, Nicola NA, et al. Polyethylene glycated leukemia inhibitory factor antagonist inhibits human blastocyst implantation and triggers apoptosis by down-regulating embryonic AKT. Fertility and Sterility. 2013 Oct 1;100(4):1160-1169.e2.

8. Kulkarni J. Depression as a side effect of the contraceptive pill. Expert Opin Drug Saf. 2007 Jul;6(4):371–4.

9. Menkhorst E, Zhang J-G, Sims NA, Morgan PO, Soo P, Poulton IJ, et

al. Vaginally Administered PEGylated LIF Antagonist Blocked Embryo Implantation and Eliminated Non-Target Effects on Bone in Mice. PLOS ONE. 2011 May 18;6(5):e19665.

10. Gray CA, Bartol FF, Tarleton BJ, Wiley AA, Johnson GA, Bazer FW, et al. Developmental biology of uterine glands. Biol Reprod. 2001 Nov;65(5):1311–23.

11. Spencer TE. Biological Roles of Uterine Glands in Pregnancy. Semin Reprod Med. 2014 Sep;32(5):346–57.

12. Stewart CL, Kaspar P, Brunet LJ, Bhatt H, Gadi I, Köntgen F, et al. Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. Nature. 1992 Sep 3;359(6390):76–9.

13. Wu M, Yin Y, Zhao M, Hu L, Chen Q. The low expression of leukemia inhibitory factor in endometrium: possible relevant to unexplained infertility with multiple implantation failures. Cytokine. 2013 May;62(2):334–9.

14. Zhang W, Kim SM, Wang W, Cai C, Feng Y, Kong W, et al. Cochlear Gene Therapy for Sensorineural Hearing Loss: Current Status and Major Remaining Hurdles for Translational Success. Front Mol Neurosci [Internet]. 2018 Jun 26 [cited 2019 Nov 18];11. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6028713/

15. Serafini PC, Silva ID, Smith GD, Motta EL, Rocha AM, Baracat EC. Endometrial claudin-4 and leukemia inhibitory factor are associated with assisted reproduction outcome. Reprod Biol Endocrinol. 2009 Apr 19;7:30.

16. Chen JR, Cheng JG, Shatzer T, Sewell L, Hernandez L, Stewart CL. Leukemia inhibitory factor can substitute for nidatory estrogen and is essential to inducing a receptive uterus for implantation but is not essential for subsequent embryogenesis. Endocrinology. 2000 Dec;141(12):4365–72.

17. Salleh N, Giribabu N. Leukemia Inhibitory Factor: Roles in Embryo Implantation and in Nonhormonal Contraception [Internet]. The Scientific World Journal. 2014 [cited 2020 Jan 17]. Available from: https://new.hindawi.com/journals/tswj/2014/201514/

18. Stewart CL. Leukaemia inhibitory factor and the regulation of pre-implantation development of the mammalian embryo. Mol Reprod Dev. 1994 Oct;39(2):233–8.

19. Dimitriadis E, Menkhorst E. New generation contraceptives: interleukin 11 family cytokines as non-steroidal contraceptive targets. J Reprod Immunol. 2011 Mar;88(2):233–9.

20. Fairlie WD, Uboldi AD, McCoubrie JE, Wang CC, Lee EF, Yao S, et al. Affinity maturation of leukemia inhibitory factor and conversion to potent antagonists of signaling. J Biol Chem. 2004 Jan 16;279(3):2125–34.

21. White CA, Zhang J-G, Salamonsen LA, Baca M, Fairlie WD, Metcalf D, et al. Blocking LIF action in the uterus by using a PEGylated antagonist prevents implantation: a nonhormonal contraceptive strategy. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(49):19357–62.

22.Ernst M, Inglese M, Waring P, Campbell IK, Bao S, Clay FJ, et al. Defective gp130-mediated signal transducer and activator of transcription (STAT) signaling results in degenerative joint disease, gastrointestinal ulceration, and failure of uterine implantation. J Exp Med. 2001 Jul 16;194(2):189–203.

23. Cullinan EB, Abbondanzo SJ, Anderson PS, Pollard JW, Lessey BA, Stewart CL. Leukemia inhibitory factor (LIF) and LIF receptor expression in human endometrium suggests a potential autocrine/paracrine function in regulating embryo implantation. Proc Natl Acad Sci U S A. 1996 Apr 2;93(7):3115–20.

24. Reddy KR, Wright TL, Pockros PJ, Shiffman M, Everson G, Reindollar R, et al. Efficacy and safety of pegylated (40-kd) interferon alpha-2a compared with interferon alpha-2a in noncirrhotic patients with chronic hepatitis C. Hepatology. 2001 Feb;33(2):433–8.

25. Sengupta J, Lalitkumar PGL, Najwa AR, Ghosh D. Monoclonal anti-leukemia inhibitory factor antibody inhibits blastocyst implantation in the rhesus monkey. Contraception. 2006 Nov;74(5):419–25.

26. Estrov Z, Talpaz M, Wetzler M, Kurzrock R. The modulatory hematopoietic activities of leukemia inhibitory factor. Leukemia & Lymphoma. 1992 Jan 1;8(1–2):1–7.

27. Hudson KR, Vernallis AB, Heath JK. Characterization of the Receptor Binding Sites of Human Leukemia Inhibitory Factor and Creation of Antagonists. J Biol Chem. 1996 May 17;271(20):11971–8.

28. Butzkueven H, Zhang J-G, Soilu-Hanninen M, Hochrein H, Chionh F, Shipham KA, et al. LIF receptor signaling limits immune-mediated demyelination by enhancing oligodendrocyte survival. Nat Med. 2002 Jun;8(6):613–9.

29. Martin TJ, Seeman E. New mechanisms and targets in the treatment of bone fragility. Clin Sci. 2007 Jan;112(2):77–91.

30. Cicinelli E, de Ziegler D. Transvaginal progesterone: evidence for a new functional "portal system" flowing from the vagina to the uterus. Hum Reprod Update. 1999 Aug;5(4):365–72.

31. Christin-Maitre S. History of oral contraceptive drugs and their use worldwide. Best Pract Res Clin Endocrinol Metab. 2013 Feb;27(1):3–12.

32. Young EA, Kornstein SG, Harvey AT, Wisniewski SR, Barkin J, Fava M, et al. Influences of hormone-based contraception on depressive symptoms in premenopausal women with major depression. Psychoneuroendocrinology. 2007 Aug;32(7):843–53.

33. Kelsey JJ. Hormonal contraception and lactation. J Hum Lact. 1996 Dec;12(4):315–8.

34. Schneyer R, Lerma K. Health outcomes associated with use of hormonal contraception: breast cancer. Curr Opin Obstet Gynecol. 2018;30(6):414–8.

35. Robinson SA, Dowell M, Pedulla D, McCauley L. Do the emotional side-effects of hormonal contraceptives come from pharmacologic or psychological mechanisms? Med Hypotheses. 2004;63(2):268–73.

36. Oinonen KA, Mazmanian D. To what extent do oral contraceptives influence mood and affect? J Affect Disord. 2002 Aug;70(3):229–40.

37. Bitzer J. Hormonal contraception and depression: another Pill scandal? Eur J Contracept Reprod Health Care. 2017;22(1):1–2.

38. Skovlund CW, Mørch LS, Kessing LV, Lidegaard Ø. Association of Hormonal Contraception With Depression. JAMA Psychiatry. 2016 Nov 1;73(11):1154–62.

39.Biegon A, Reches A, Snyder L, McEwen BS. Serotonergic and noradrenergic receptors in the rat brain: modulation by chronic exposure to ovarian hormones. Life Sci. 1983 Apr 25;32(17):2015–21.

40. Toffoletto S, Lanzenberger R, Gingnell M, Sundström-Poromaa I, Comasco E. Emotional and cognitive functional imaging of estrogen and progesterone effects in the female human brain: a systematic review. Psychoneuroendocrinology. 2014 Dec;50:28–52.

41. Lawrie TA, Hofmeyr GJ, Jager MD, Berk M, Paiker J, Viljoen E. A double-blind randomised placebo controlled trial of postnatal norethisterone enanthate: the effect on postnatal depression and serum hormones. BJOG: An International Journal of Obstetrics & Gynaecology. 1998;105(10):1082–90.

42. Lete I, Arrilucea MP de, Rodríguez M, Bello E. Efficacy, safety, and patient acceptability of the etonogestrel and ethinyl estradiol vaginal ring

[Internet]. Open Access Journal of Contraception. 2014 [cited 2020 Jan 17].

43. Herrmann M, Seibel The effects of hormonal contraceptives on bone turnover markers and bone health Clinical Endocrinology (2010) 72, 571–583 doi 10.1111/j.1365-2265.2009.03688.x

44. ACOG Practice Bulletin No. 206. Obstetrics & Gynecology [Internet]. 2019 Feb [cited 2020 Mar 4];133(2):e128–50. Available from: https://journals.lww.com/greenjournal/Fulltext/2019/02000/ACOG_Practice_Bulletin_No__206__Use_of_Hormonal.41.aspx

Submitted: 01/20/20

Review Article

[1]Department of Microbiology and Immunology, McGill University, Montreal, QC, Canada

Keywords

T cell immunotherapy, CAR-T cells

Email Correspondence

laura.rendon@mail.mcgill.ca

Laura Rendon[1]

Delivering the "Living Drug": T Cell Immunotherapy

Abstract

Background: Cancer is one of the most lethal diseases worldwide. Traditional approaches such as chemotherapy have toxic side effects. New therapies on the rise are more target specific. One such therapy, immunotherapy, has become increasingly attractive in the field. However, to ensure the modulated and controlled manipulation of the immune system, delivery methods for drugs cells and biomaterials must be developed.

Methods: In this review, we analyze the literature to discuss the recent advances in T cell immunotherapy as well as four delivery technologies that address the issues of safety and efficacy associated with this treatment.

Summary: We conclude that the CAR-T approach could be a step towards overcoming the inaccessibility of poorly vascularized tumors and the evasion mechanisms of tumor cells. Delivery methods such as surface conjugated nanoparticles, DNA nanocarriers, scaffolds and artificial antigen-presenting cells aim for a more tumor-targeted approach rather than a systemic one, making this therapy applicable in the clinic.

Introduction

Despite a century of scientific advancement, cancer remains one of the most lethal and challenging diseases worldwide. Tumorigenic cells arise from the accumulation of mutations that collectively result in the acquisition of two cellular properties: the ability to grow and divide in defiance of normal cellular restraints and the capacity to colonize territories normally inhabited by different cell types. To this day, surgery, chemotherapy and radiation predominate as the main treatments for cancer. While surgery is aimed at eliminating local tumor masses, both chemotherapy and radiation therapy operate via non tumor-specific mechanisms that often result in off-target toxicity. Furthermore, chemotherapy seems to have reached a developmental plateau. (1) These concerns have led scientists to seek alternative therapies that could potentially replace or be used in combination with chemotherapy to optimize outcomes and minimize toxic side effects. (2)

Immunotherapy has become an increasingly attractive clinical strategy over the past decade, shifting the paradigm of cancer therapy from a chemical to a biological approach. The fundaments of immunotherapy come from our better understanding of the regulatory mechanisms by which the immune system keeps malignant cells in check. According to the *stranger-danger* model, the immune system not only recognizes foreign entities but also altered-self. However, malignant cells often evolve strategies to evade immunosurveillance and become resilient. Immunotherapy aims to counteract this by strengthening and restoring the ability of the immune system to fight cancer.

Despite being a promising strategy for the cure of cancer, immunotherapies face challenges related to their safety and efficacy. Safety concerns come from the serious adverse effects engendered by these therapeutics: off-target effects and autoimmunity. Efficacy concerns originate from the lack of response to treatment in some patients, likely due to the phenotypic and morphological heterogeneity of tumor cells. (3) Reduction of adverse off-target effects can be achieved by targeting the tumor with higher specificity; efficacy for all patients can be increased by developing a more patient-specific approach that addresses the issue of cancer heterogeneity. To achieve such a controlled and patient-specific immunomodulation, scientists have developed novel delivery technologies for immunotherapy. (2)

The most popular immunotherapies currently used are monoclonal antibodies, non-specific immunotherapies with interferons/interleukins, virus therapy, and T cell therapy. Monoclonal antibodies are used to detect and tag a cancer protein to make cancer cells visible to the immune system or as checkpoint inhibitors of cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and programmed cell death protein 1/programmed death-ligand 1 (PD-1/PD-L1) to unleash the immune system. Interferon/interleukin immunotherapies help with the activation and proliferation of immune cells that fight cancer. Virus therapy consists of a genetically modified virus injected into the tumor to kill the cells. The antigens released from the dead tumor cells trigger the immune system which can then target all tumor cells displaying these antigens. (4)

In this review, we will discuss in detail one type of immunotherapy: T cell immunotherapy, as well as four delivery technologies that address the issues of safety and efficacy to improve the implementation of this treatment.

T Cell Immunotherapy

Immunotherapy brief historical review

Even though immunotherapy only gained popularity in the last decade, it has its roots back in the 19th century. The success of vaccination and the idea that weakened pathogens could provide protective immunity fuelled scientific discovery at the time. Medical scientists became optimistic that the acquired ability to manipulate the immune system could be used to treat cancer. Dr. William B. Coley was the first to attempt such an endeavor. Coley dug deep into the literature and came across 47 cases that portrayed a strange phenomenon: the unexpected remission of incurable neoplastic malignancies following a streptococcal dermal infection, erysipelas. Reluctant, he injected patients with heat-killed erysipelas causing agent and achieved long-term cure for many of them. Yet, despite Coley's success, the lack of understanding of his results at the time doomed his strategy. (5) His findings lingered in the dark until modern immunotherapy resurfaced a century later with the theory of cancer immunosurveillance by Brunet and Thomas. According to this theory, lymphocytes act as sentinels to identify transformed cells. (6) Yet, it wasn't until the discovery of interleukin-2 (IL-2) in 1976 that immunotherapy truly spread it wings. Other therapeutics began to rise: in 2010, cancer vaccine sipuleucel-T (cell based vaccine to treat prostate cancer) was approved by the U.S. Food and Drug Administration (FDA), followed by Ipilimumab, a monoclonal antibody against CTLA-4, in 2011. Along with CTLA-4, PD-1 and PD-1L inhibitors are a group of checkpoint inhibitors that act on the programmed cell death ligand; they entered clinical trials in 2006. (7) Finally, the first T cell therapy with chimeric antigen receptor (CAR)-T cells was the breakthrough of 2013 and a turning point in cancer immunotherapy. (2)

T cell immunity and tumor evasion

To fight cancer through immunotherapy, it is crucial to first understand how the immune system operates. The body's immune system has developed its own biological mechanisms to detect and eliminate potentially tumorigenic cells. However, tumors often outsmart the immunoregulatory processes that destroy them and become resilient. In this section, we will primarily focus on T cell mechanisms against altered-self cells and how they get evaded.

The tumor micro environment (TME) enables a three-phase cancer progression in the context of immunity: elimination, equilibrium, and escape. (8) The first phase is characterized by host immunosurveillance: the ability of the immune system to distinguish self from altered self. Professional antigen presenting cells (pAPCs) such as dendritic cells (DCs) engulf tumor cells and present tumor-specific antigens to naïve T cells in an major histocompatibility complex (MHC)-restricted manner. T cells with specificity to the tumor antigens get activated, expand, differentiate and become the effectors of cell-mediated adaptive immunity. (9) The two main T cell lineages, CD4+ and CD8+, operate through different mechanisms. CD8+ T cells recognize specific antigens presented on MHC class I at the surface of tumor cells and interact through Fas-FasL to kill the cells via perforin and granzyme. CD4+ T cells are less well understood in the context of cancer immunity, but some lineages are known to act through inhibition of angiogenesis and eosinophil recruitment. (10)

In the second phase, the tumor cells that have escaped the initial immune attack can neither expand nor be eradicated and therefore cohabitate in a state of equilibrium with the effector T cells. During this long-lasting phase, the tumor does what it knows best: it rapidly mutates until it acquires resistance abilities. These include the downregulation of their MHC class I, without which CD8+ T cells cannot target tumor cells, (9) and the cytolytic ability to kill CD4+ T cells. (10) In a Darwinian-selection fashion, new tumor variants arise with acquired abilities to fight the immune system and progress into the escape phase. During this final and critical phase, tumor cells escape the state of equilibrium and expand uncontrollably under the impotence of the immune system.

Engineered T Cells immunotherapy: CAR-T Cells

In an attempt to strengthen cell-mediated immunity in the fight against the newly weaponized tumorigenic cells, several immunotherapeutic approaches have been developed, most of which involve small molecules such as checkpoint inhibitors and bigger protein complexes such as monoclonal antibodies. Yet perhaps the most revolutionary immunotherapeutic invention was the CAR-T cell. The uniqueness of this therapy compared to those involving molecules comes from the fact that cells are capable of intelligent sensing-response behaviors, making their manipulation far more complex. (11)

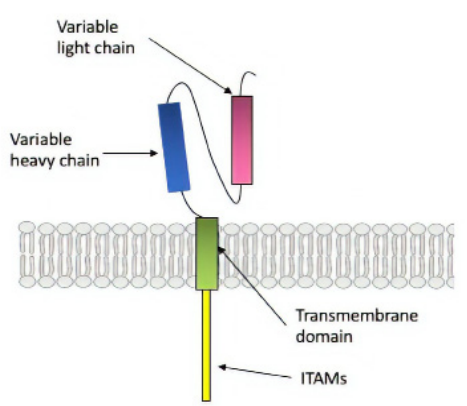

Figure 1. The structure of the Chimeric Antigen Receptor (CAR). On the extracellular side the CAR has the variable domain which consists of a light chain and a heavy chain, which is connected to the transmembrane domain which in turn is connected to the intracellular which has signaling capability such as the immunoreceptor tyrosine-based motif (ITAM).

To make CAR-T cells, T cells are initially isolated from the blood of either the patient (autologous) or a healthy donor (allogenic) through the process of leukocyte apheresis. T cells proliferation is then induced via IL-2 and anti-CD3 antibody. The expanded T cell population is genetically engineered to express a CAR of interest (Figure 1) using the CRISPR/Cas9 mechanism to knock out inhibitory genes such as the T cell receptor (TCR). (12) Viral vector transfection is then used for the expression of the CAR. After the CAR-T cell population has been expanded to the desired numbers, patient undergoes lymphodepletion chemotherapy to destroy and prevent the original T cells to compete with the incoming CAR-T cells for resources. Finally, the CAR-T cells are delivered to the patient. (13)

CAR-T cells offer many advantages over native T cells. For instance, CAR targeting is human leukocyte antigen (HLA) independent, which means antigen recognition in the context of MHC is no longer required (Figure 2). This prevents the evasion of tumors that have downregulated MHCI. HLA-independent recognition also overcomes the issue of MHC allorecognition: CAR-T cells don't need to be specific to the patient's HLA expression profile. Furthermore, unlike native T cells which can only respond to one antigen-MHC complex, CAR-T cells can be engineered to respond to a broad array of targets and to differentiate into various effector subtypes such as CD4+, CD8+ or memory cells. (11)

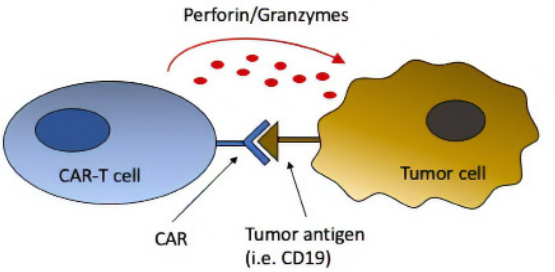

Figure 2. The mechanism of action of CD8+ CAR-T cells. CAR-T cells can target cells independently of HLA, upon which they release perforin and granzyme which induces target cell death.

CAR-T cell therapy was first approved by the FDA in 2017 for the treatment of acute lymphoblastic leukemia. CAR-T cells were engineered to target CD19, a molecule highly expressed in both normal and malignant B cells. The therapy was a success as more than 80% of patients entered remission with few treatable on-target/off-tumor side effects such as B cell aplasia. (14) Despite the optimism for T cell therapies, their rise remains tempered by safety and efficacy concerns. First, with respect to safety, T cell immunotherapy can cause serious side effects: cytokine release syndrome (CSR) and neurological toxicity. Second, with respect to efficacy, the infused cells are unable to conquer solid tumors and do not persist in the TME for extended time periods, resulting in relapse. (15) These hurdles may be overcome by the emerging delivery technologies that will be discussed in the following section.

Delivery technologies for T Cell Immunotherapy

Surface conjugated surface nanoparticles

In their paper, *Therapeutic cell engineering with surface-conjugated synthetic nanoparticles*, (16) Matthias T. Stephan and his colleagues describe a strategy to enhance cell therapy via adjuvant-loaded nanoparticles conjugated to the T cell surface. It has been observed that CAR-T cells often fail to persist and require the systemic administration of adjuvant drugs. Drug administration via the systemic route is not targeted to their site of action, the TME, and therefore higher doses must be used which can lead to global toxic side effects. The authors argue that CAR-T cell conjugated-nanoparticles would provide a more targeted, effective and safe pseudo-autocrine drug stimulation.

They initially designed liposome-like synthetic nanoparticles coated with thiol-reactive maleimide headgroups (T cells contain high amounts of free thiols at their surface). They incubated the nanoparticles with the T cells to allow the two to covalently bind via maleimide-thiol coupling. According to the results, coupling was non-toxic and only blocked 17.2% of the cell surface thiol groups. Furthermore, unlike lipid-coated polymers which get removed in washing steps, maleimide-linked particles remained bound to cells.

They then determined the maximum number of particles that could be linked to the T cell surface without jeopardizing cellular functions. They assessed CD8+ key functions: cell killing via cell surface interactions, proliferation, cytokine release, transmission across endothelial tissue, and tissue homing. According to their results, up to ~100 nanoparticles could be conjugated without affecting CAR-T cell proliferation, killing of target cells or cytokine release. Furthermore, they showed that the CAR-T cells carrying 100 nanoparticles had the same transmission efficiency through endothelial layers as native cells. However, following migration the cells only retained ~83% of the nanoparticles initially conjugated. Finally, they assessed tumor homing capacities of nanoparticle-conjugated T cells compared to unmodified T cells with the specificity to the same EL4 tumor antigen. No difference was observed.

In their next experiment, Stephan and his colleagues tested if the adjuvant-drug containing nanoparticles could enhance their carrier T cell action. They encapsulated cytokines known to promote T cell proliferation and effector functions: IL-15 and IL-21. The cytokines had a greater proliferative effect when loaded onto the T cells compared to systemically infused cytokines. Nanoparticle-conjugated T cells also displayed a more proliferative behavior than unmodified cells and persisted for longer after the contraction period. Treatment with nanoparticle–conjugated T cells completely eradicated the tumor, whereas treatment with unmodified T cells and/or systemic cytokine infusion lead to lower survival rates. These experiments were done in mouse models.

In conclusion, this study reveals the advantage cell carriers have over inanimate molecules in delivering a drug to tissues with difficult access. The challenge of pleiotropic activity and toxicity of drugs required for the proper function of CAR-T cells can be overcome by encapsulating the drug in cell-conjugated nanoparticles. This technique limits toxic side-effects, requires lower drug doses and improves T cell functions such as proliferation and homing due to enhance paracrine secretion of IL-15, without interfering with normal cellular activities.

DNA nanocarriers for *in situ* T cell engineering

In their paper, *In situ programming of leukemia-specific T cells using synthetic DNA nanocarriers*, (17) Tyrel T. Smith *et al.* describe a strategy to improve the application of T cell immunotherapy by engineering T cells in vivo via gene therapy. An obstacle to the implementation of personalized CAR-T cells is the high costs and complicated manufacturing procedures associated with their production. To solve this, Smith and his colleagues propose a DNA-carrier nanotechnology that can efficiently introduce the CAR gene into native T cells and reprogram them *in vivo*.

The first step was to design nanocarriers that could induce CAR expression in T cells. The carriers must effectively be taken up by the T cell and import the CAR DNA into the nucleus. Endocytosis was achieved by coupling anti-CD3e f(ab')2 to the surfaces of nanoparticles. CAR DNA nuclear import was facilitated by integrating microtubule-associated sequences (MTAS) and nuclear localization signal (NLS) into the nanoparticle. To achieve CAR expression by the host's gene expression machinery, the authors designed two plasmids, one containing the CAR cassette flanked by piggyBac transposable elements and the other containing the sequence encoding piggyBac transposase. The transposon machinery enables the integration of vectors into chromosomes.

To test the functionality of the designed nanoparticle in the production of leukemia-specific CAR-T cells, the authors conducted a series of *in vitro* experiments on mouse splenocytes. They first confirmed by flow cytometry that CD3-targeted nanoparticles effectively bound T cells with few off-targets. They then observed by confocal imaging that particles were rapidly internalized into the cytoplasm. Shortly after administration, leukemia-specific receptors were detected on the surface of T cells. They demonstrated the advantage of NLS and MTAS containing nanoparticles compared to negative controls by measuring nuclear transfection rates. Finally, piggyBac transposons sustained high levels of CAR gene expression over a longer period of time compared to nanoparticles that lacked the transposon machinery.

The next step was to further examine CD3-mediated targeting of T cells as well as the efficacy of cell-reprogramming mechanisms, this time *in vivo* with mice. They obtained the same results for CD3+ T cell targeting and the few off-target effects decreased over the course of treatment. Furthermore, they demonstrated that targeted nanoparticles effectively localized to lymphatic organs (spleen lymph nodes and bone marrow) as opposed to the non-targeted ones which localized to the liver. Subsequently, they tested for *in vivo* toxicities. This time, they used a prostate-targeting CAR to ensure that the observed effects originated from the nanoparticles themselves and not the DNA editing activity. Cell counts and blood profiles were normal; inflammatory cytokine levels increased minimally. *In vivo* data also evidenced the need for co-delivery of the transposon machinery to achieve efficient expansion of CAR-T cells as well as tumor eradication. Finally, to assess the efficacy of cancer treatment with DNA-carrying nanoparticles compared to conventional therapy, the authors treated a group of mice following conventional CAR-T cell therapy protocol, described in section 2.3. The results showed that survival rates were slightly higher in the conventional therapy group, but overall very similar in both approaches.

In conclusion, the use of nanoparticles to reprogram gene expression *in vivo* is an attractive alternative to the costly and labor-expensive T cell therapy, with comparable results as the latter. Nonetheless, some challenges have yet to be overcome. First, the recurring issue of solid tumors remains a concern as their accessibility is limited. Second, with respect to clinical translation, the safety issues of off-target gene transfer must be further addressed. For instance, the inclusion of a T cell-specific promoter upstream of the CAR sequence could ensure that despite there being off-target integration, CAR is only expressed in T cells.

Biomaterial based implants for engineered T cell delivery

In a study by Tyrel T. Smith *et al.*, a possible solution to the recurrent hurdle of solid tumors is put forward. (18) Solid tumors have been the lingering nightmare of T cell immunotherapies for two reasons. First, solid tumors create an immunosuppressive environment that impedes normal T cell function. Second, solid tumors are highly heterogeneous and include cells that lack CAR-targeted antigens. As a solution, the authors proposed a biopolymer scaffold comprised of CAR-T cells and co-stimulator stimulator of interferon genes (STING) agonist. They hypothesized two functions: the physical delivery of ex-vivo engineered CAR-T cells and the STING induction of native T cells with different tumor antigen specificities as CAR-T cells. The latter would help overcome the obstacle of tumor heterogeneity.

In their experiments, the authors used pancreatic cancer mouse models. They first designed a CAR specific to RAE1, a pancreatic tumor antigen. They then manufactured a bioactive scaffold using polymerized alginate and macromolecules important for migration and stimulation of T cells. The scaffolds were porous matrices made from ultrapure sodium alginate powder and were delivered to the animals through surgery. Initially, they tested the T cell expansion and capacity to clear pancreatic tumor in two conditions: direct injection of CAR-T cells into the tumor, and direct implantation of CAR-T cell scaffold onto the tumor. In the first group, cells did not persist and only temporarily delayed cancer progression. In contrast, robust CAR-T cell proliferation and reduced tumor growth was observed for the second group. Nevertheless, neither group achieved complete clearance and RAE1 negative tumor cell variants emerged, resulting in tumor evasion.

Since the CAR-T cells that targeted single antigens did not halt tumor progression, the authors refined their strategy. The natural recognition of

various tumor antigens relies on the proper priming of T cells by DCs, a mechanism that is impaired in cancer. To recruit and stimulate DCs, the authors added STING agonist cyclic di–GMP (cdGMP) to the scaffold. They assessed DC activity by measuring their CD86 and MHC class II expression profiles, which are indicative of DCs' activated phenotype. The cdGMP-CAR-T cell scaffold treatment group presented higher numbers of activated DCs than the group stimulated by only CAR-T cell scaffolds. The next step was to visualize the localization and magnitude of T cell activation in the same two treatment conditions: CAR-T cell scaffold treatment and cdGMP-CAR-T cell scaffold treatment. They created mice carrying a transgenic gene composed of nuclear factor of activated T cell (NFAT) linked to a luciferase reporter gene. In the CAR-T cell scaffold condition luminescence was detected in the tumor area but not as strongly as in the cdGMP/CAR-T cell scaffold condition. In the latter, other organs such as the spleen and lymph nodes showed induction of native T cells.

Finally, the authors evaluated the host anti-tumor immunity in metastases and the side effects on pancreatic functions. The most effective results were again seen in the cdGMP-CAR-T cell scaffold condition. When they re-challenged the cured mice with pancreatic tumor cells to further test for global anti-tumor immunity in lung metastases, no tumor formation was seen. On the downside, decreased serum amylase and lower lipase activity were indicative of defective pancreatic exocrine function.

All in all, this delivery technique provides a solution to solid tumors that do not respond to conventional T cell therapies. Moreover, targeted delivery of adjuvant STING reduces off target exposure and requires lower drug dosage. However, in the case of pancreatic cancer, some adverse effects should be carefully watched for in clinical trials. Furthermore, the implant-to-tumor size ratio in mice could account for some of the findings in this study that might not be reproducible in human tumors.

Artificial antigen-presenting cells

Artificial antigen-presenting cells (aAPCs) constitute a promising platform for the stimulation and amplification of T cell responses. T cells are typically activated by signals 1 (survival) and 2 (proliferation) through cell surface interactions with APCs. The T cell-APC interaction induces the formation of an immune synapse characterized by the clustering of TCRs. Previous data have shown that conventional spherical micron-sized aAPCs function efficiently in vitro but not in vivo. Based on previously observed advantages of nanoparticles and benefits of ellipsoid over spherical particles, Randall A. Meyer and his colleagues developed a nanoellipsoidal aAPC model that could overcome the limitations of conventional aAPCs immunotherapy. This approach is described in their study *Biodegradable Nanoellipsoidal Artificial Antigen Presenting Cells for Antigen Specific T-Cell Activation*. (19)

They first synthesized PLGA nanoparticles by emulsion and PVA film stretch to obtain spherical and ellipsoid shapes of different curvatures. Then, to turn the ellipsoid nanoparticles into aAPCs, they conjugated peptide-MHC-IgG dimers and anti-CD28 antibody to the PLGA polymer. They verified particle stability by incubating the artificial cells in physiological conditions. The optimal particle/antigen dosages and curvature of aAPCs were determined by titration and carboxyfluorescein succinimidyl ester (CFSE) dilution analysis. 2-fold ellipsoid particles with the highest particle/antigen ratios were the most efficient at stimulating CD8+ T cells and surpassed their spherical counterparts.

The first feature they studied was nanoparticle non-specific cell uptake in ellipsoid nanoscale aAPCs (naAPCs) compared to spherical naAPCs. They used two different models of uptake: by macrophages and by human umbilical vein endothelial cells (HUVEC). *In vitro*, both macrophages and HUVECs preferentially phagocytosed spherical naAPCs. The *in vivo* circulation and half-life of naAPCs was assessed via a biodistribution experiment. Ellipsoid naAPCs moved faster, remained in the periphery for longer, and had a longer half-life than spherical naAPCs.

Next, they sought to evaluate the stimulatory capacity of naAPCs *in vivo* in the context of immunotherapy. To this end, mice were administered either ellipsoid naAPC with CAR-T cells, spherical naAPC with CAR-T cells, or CAR-T cells alone. According to blood analysis carried throughout the experiment, ellipsoid naAPCs elicited a notably higher CAR-T cell proliferation rate than the two other treatments. They also observed a higher number of ellipsoid naAPC in the dissected spleen and lymph nodes.

In conclusion, aAPCs delivery systems offer an attractive alternative to conventional T cell therapies. The conventional aAPCs technology was successful *in vitro* but not *in vivo*. In this experiment, the authors improved the technology by using nanoellipsoid design which not only provides higher efficacy but also reduces unwanted cellular uptake and displays an enhanced biodistribution compared to conventional aAPCs. Nonetheless, other aAPC features such as membrane fluidity and particle rigidity should be further assessed to optimize this delivery system for clinical immunotherapy.

Discussion

Over the past ten years, immunotherapy has taken the lead over traditional cancer treatments such as radiotherapy and chemotherapy. The developing field of immunology has allowed scientists to better understand and hence exploit our immune system for the treatment of cancer which to this day, remains a major challenge for health practitioners. Initially, scientists employed small molecules such as IFNα and IL-2 as well as checkpoint inhibitors for the application of immunotherapy. However, this methodology took a revolutionary turn with the invention of CAR engineered T cells. Often referred to as the "living drug", CAR-T cells adopt and enhance the anti-tumor functions of native T cells, the sentinels of cell-mediated adaptive immunity. This cellular approach tackles some of the hassles associated with poorly vascularized tumors via trafficking mechanisms superior to those of small molecules. As well, and perhaps most importantly, artificial T cells overcome the evasion and immunosuppressive powers tumor cells hold over native T cells.

Nonetheless, to exploit the high potential of CAR-T cells in a clinical setting, scientists must first jump the hoops of safety and efficacy this therapy may have. The immune system is a powerful weapon against cancer, yet it can just as easily turn into a bomb against our own body. Uncontrolled stimulation of the immune system in the hopes of eradicating cancer can result in lethal auto-immune responses and cytokine release syndrome. It is crucial to tailor the immune stimulation to the tumor, reduce off-target effects and minimize potential toxicities. To address these issues, scientists have come up with different delivery techniques for the treatment with CAR-T cells. These include surface conjugated surface nanoparticles, DNA nanocarriers for *in situ* T cell engineering, biomaterial-based implants for engineered T cell delivery and artificial antigen-presenting cells. All of these techniques aim for a more tumor-targeted approach rather than a systemic one, but all therapies target different aspects of tumor progression to enhance treatment efficacy: surface conjugated nanoparticles aim to improve adjuvant drug delivery to the CAR-T cells, DNA nanocarriers render the treatment more accessible and available in the market, CAR-T cell bio-scaffold address the hurdle of solid tumors, and aAPCs fight the immunosuppressive effects of tumors in the evasion phase.

Even though the field of immunotherapy is rapidly progressing, these delivery methods are still at nascent stages and need to be further investigated and optimized to be used in clinical trials. *Ex-vivo* manufacture and expansion remains an arduous process that is both time-consuming and costly. Improving manufacturing methods can greatly increase the applicability of these delivery methods in a clinical setting. With respect to surface conjugated nanoparticles, the release of adjuvant drugs could be extended so that less treatment administrations are required. The DNA carriers, despite being more cost effective, pose a higher risk of off-target gene insertion, an effect that should be addressed perhaps through the inclusion of a T cell specific promoter, (16) or a more targeted vector delivery system. In the treatment with biosynthetic scaffolds, the authors do not specify whether or not the subjects undergo lymphocyte eradication by chemotherapy like in conventional CAR-T cell therapy before administrating the scaffold, which is meant to prevent CAR-T and native T cell competition. In the scaffold approach, the authors aim to both deliver the engineered cells and stimulate the native ones and the interplay between

the two should be further investigated for the implementation of this delivery method. Lastly, design specificities for aAPCs currently under study slowly set the path towards the perfectly shaped, rigid and sized particles to achieve optimal induction of T cell immunity.

Acknowledgements

I thank Dr. T.M.S Chang and Dr. Dr. S. Prakash for teaching the class Artificial Cells at McGill and assisting in the writing of this paper.

References

1. Nurgali K, Jagoe RT, Abalo R. Editorial: Adverse Effects of Cancer Chemotherapy: Anything New to Improve Tolerance and Reduce Sequelae? Frontiers in Pharmacology. 2018;9.

2. Riley RS, June CH, Langer R, Mitchell MJ. Delivery technologies for cancer immunotherapy. Nature Reviews Drug Discovery. 2019Aug;18(3):175–96.

3. Schuster SJ, Svoboda J, Chong EA, Nasta SD, Mato AR, Anak Ö, et al. Chimeric Antigen Receptor T Cells in Refractory B-Cell Lymphomas. New England Journal of Medicine. 2017;377(26):2545–54.

4. "Understanding Immunotherapy." Cancer.Net, January 29, 2019. https://www.cancer.net/navigating-cancer-care/how-cancer-treated/immunotherapy-and-vaccines/understanding-immunotherapy.

5. McCarthy EF. The toxins of William B. Coley and the treatment of bone and soft-tissue sarcomas. The Iowa orthopaedic journal. 2006;26:154–8.
6. Ribatti D. The concept of immune surveillance against tumors: The first theories. Oncotarget. 2016;8(4).

7. Iwai Y, Hamanishi J, Chamoto K, Honjo T. Cancer immunotherapies targeting the PD-1 signaling pathway. Journal of Biomedical Science. 2017.

8. Dunn GP, Old LJ, Schreiber RD. The Three Es of Cancer Immunoediting. Annual Review of Immunology. 2004;22(1):329–60.

9. Naing A, Hajjar J. Immunotherapy. Advances in Experimental Medicine and Biology. 2018.

10. Ostroumov, D., Fekete-Drimusz, N., Saborowski, M., Kühnel, F., & Woller, N. (2018). CD4 and CD8 T lymphocyte interplay in controlling tumor growth. Cellular and molecular life sciences: CMLS, 75(4), 689–713. doi:10.1007/s00018-017-2686-7

11. Lim WA, June CH. The Principles of Engineering Immune Cells to Treat Cancer. Cell. 2017;168(4):724–40.

12. Chenggong Li, Heng Mei, Yu Hu. Applications and explorations of CRISPR/Cas9 in CAR T-cell therapy. Briefings in Functional Genomics. 2020.

13. General cancer information [Internet]. CAR T-cell therapy | Immunotherapy | Cancer Research UK. 2019 [cited 2020Jan20]. Available from: https://www.cancerresearchuk.org/about-cancer/cancer-in-general/treatment/immunotherapy/types/CAR-T-cell-therapy

14. CAR T Cells: Engineering Immune Cells to Treat Cancer [Internet]. National Cancer Institute. [cited 2020Jan20]. Available from: https://www.cancer.gov/about-cancer/treatment/research/car-t-cells

15. Srivastava S, Riddell SR. Chimeric Antigen Receptor T Cell Therapy: Challenges to Bench-to-Bedside Efficacy. The Journal of Immunology. 2018Aug;200(2):459–68.

16. Stephan MT, Moon JJ, Um SH, Bershteyn A, Irvine DJ. Therapeutic cell engineering with surface-conjugated synthetic nanoparticles. Nature Medicine. 2010;16(9):1035–41.

17. Smith TT, Stephan SB, Moffett HF, Mcknight LE, Ji W, Reiman D, et al. In situ programming of leukaemia-specific T cells using synthetic DNA nanocarriers. Nature Nanotechnology. 2017;12(8):813–20.

18. Smith TT, Moffett HF, Stephan SB, Opel CF, Dumigan AG, Jiang X, et al. Biopolymers codelivering engineered T cells and STING agonists can eliminate heterogeneous tumors. Journal of Clinical Investigation. 2017;127(6):2176–91.

19. Meyer RA, Sunshine JC, Perica K, Kosmides AK, Aje K, Schneck JP, et al. Biodegradable Nanoellipsoidal Artificial Antigen Presenting Cells for Antigen Specific T-Cell Activation. Small. 2015Dec;11(13):1519–25.

The Rare Disease Interdepartmental Science Case Competition (RISCC) is a McGill-based and student-run interdisciplinary case competition. In its first year of operation, RISCC has provided undergraduate students with the opportunity to understand the pathophysiology and socioeconomic challenges associated with a rare disease. In teams of 5 to 7 members, students generate potential research and treatment solutions before delivering an oral presentation for evaluation by faculty members.

MSURJ is proud to collaborate with RISCC by publishing the following abstract from the 2019-2020 RISCC winning team.

Submitted: 04/11/20

Jude Balit[1], Aanya Bhagrath[1], Marina Nysten[2], Farida Rahman[3], Joyce Wu[4], Kejin Zhu[2]

Clinical Manifestations of Hereditary Leiomyomatosis and Renal Cell Carcinoma: a Case Report

[1]Department of Microbiology and Immunology, McGill University, Montreal, QC, Canada
[2]Department of Physiology, McGill University, Montreal, QC, Canada
[3]Department of Biology, McGill University, Montreal, QC, Canada
[4]Department of Psychology, McGill University, Montreal, QC, Canada

Keywords

HLRCC, uterine leiomyomas, fumarate hydratase, renal cell carcinoma

Email Correspondence

marina.nysten@mail.mcgill.ca

Abstract

Introduction: Hereditary leiomyomatosis and renal cell carcinoma (HLRCC) is a rare genetic disorder etiologically stemming from mutations in the FH gene. This disorder has been reported in 300 families worldwide. We were presented with the following case as part of the Rare Disease Interdepartmental Science Case Competition (RISCC) in order to successfully diagnose the patient and examine the pathophysiology, socioeconomics, and treatment strategy of this disorder.

Case Presentation: A 27-year-old woman presented to the emergency room with a two-week history of hematuria. Her medical history included a single benign skin lesion. She did not report a family history of cancer but stated that her mother underwent a hysterectomy for unknown reasons.

Clinical Investigations: An abdominal ultrasound revealed an isoechoic nodular formation on the right kidney, and CT scans confirmed the presence of a 7-cm renal tumour and extensive leiomyomas in the patient's uterus. Genetic testing revealed a mutation in the fumarate hydratase (FH) gene, leading to the diagnosis of hereditary leiomyomatosis and renal cell carcinoma.

Pathophysiology: HLRCC is a condition caused by an autosomal dominant germline mutation in the FH gene, characterized by skin lesions, uterine leiomyomas, and aggressive renal cell carcinoma. FH acts as a tumour suppressor gene and encodes the fumarase protein. HLRCC arises from a variety of missense mutations in the FH gene that prevent the formation of a full fumarase tetramer. The disease is likely related to the dosage of the gene, rather than the site of mutation. Consequently, there is no known genotype/phenotype correlation. Fumarate inhibits HIF-α hydroxylases, so loss of the FH gene causes fumarate accumulation within cells and prevents the degradation of HIF-α. This induces a state of pseudohypoxia and leads to upregulation of genes involved in cell growth and tumorigenesis. Fumarate accumulation also modulates the Keap1-Nrf2 pathway by inducing activation of Nrf2 and allows for the protection of cancer cells from oxidative and electrophilic stressors. Furthermore, deficiency of fumarase prevents oxidative phosphorylation, thus promoting aerobic glycolysis, the most common method of metabolism used by cancer cells.

Socioeconomics: The risk of infertility and developing an aggressive cancer makes it essential to have frequent and thorough screening procedures early in life for children with a family history of HLRCC. Medical costs, infertility risk, and frequent hospital visits breed anxiety and stress for patients and family members. Fortunately, organizations like HLRCC Family Alliance exist to provide accessible information and represent a welcoming support community.

Treatment Strategy: To treat the patient's renal tumour, we suggest a radical nephrectomy with wide margins due to the aggressive nature of renal cell carcinoma. Additional imaging should be performed to identify and excise any metastatic tumours. The patient should also consult a gynecologist regarding her extensive uterine leiomyomas and will likely undergo a hysterectomy or myomectomy. A full dermatological exam should be performed to locate additional cutaneous leiomyomas. We also strongly suggest that family members undergo genetic testing. Lastly, we recommend the patient connect with support groups such as HLRCC Family Alliance to receive continued social support following her diagnosis and throughout her life.

Acknowledgements

We would like to thank the Rare Interdepartmental Science Case Competition (RISCC) team for this opportunity. Additionally, thank you to David Ji, Dr. Terry Hébert, Dr. Andrew Churchill, and all the judges for your valuable feedback and support.