A
Amsterdam university college
UC

AUC Undergraduate Journal of Liberal Arts & Sciences
Capstone Issue Vol. 14 2020

www.auc.nl

UNIVERSITY OF AMSTERDAM

VU UNIVERSITY AMSTERDAM

AUC Undergraduate Journal of Liberal Arts & Sciences

Capstone Issue Vol. 14 2020

Published by
inprint.

AUC Undergraduate Journal of Liberal Arts & Sciences
Capstone Issue Vol. 14 2020, published by InPrint

The Amsterdam University College (AUC) Undergraduate Journal of Liberal Arts and Sciences is a biannual, interdisciplinary publication showcasing outstanding undergraduate academic papers. The Journal aims to demonstrate the strength of undergraduate scholarship at AUC, to reflect the intellectual diversity of its academic programme, to encourage best research and writing practices, to facilitate collaboration between students and faculty across the curriculum, and to provide students with opportunities to gain experience in academic reviewing, editing and publishing. The Editorial of the Journal is constituted of members of the InPrint board, a registered AUCSA committee.

Editorial Board (InPrint)

Formatting

Jai Yoon Chung
Parag Dass
Sarah Martinson

Photographers (RAW)

Che Spraos Romain
Jasmin Ronach
Margherita Guida
Mirthe van Veen
Rosa Wijnen
Sanch Kuber

Capstone Team

Wade Geary, MA | Capstone co-Coordinator
Bob Kardolus | Capstone Administrator

Foreword

Welcome to the 14^{th} Capstone Issue!

Before graduation, all AUC students are required to write a Capstone thesis - an independent research paper in the disciplines of Science, Social Science or Humanities. The four-month writing process during the semester encourages students to engage with and contribute to the academic dialogue in their chosen field. In this Capstone Issue we publish six student Capstones, two from each major, written by AUC's graduating students of 2020.

The Capstones published in this issue have undergone rigorous selection and editing processes carried out by our Editorial Board. The aim of the editors is to improve the clarity and accessibility of the selected works, making them interesting to a general reader but maintaining a high standard in their academic field.

I would like to extend a word of thanks to the editors who worked tirelessly on the papers, meticulously caring for every minor detail – without them this publication would not have been possible. Thanks also goes out to the authors for their continued engagement in the process and for their patience with the long conversations regarding word choice and punctuation.

The papers in this edition cover topics ranging from Mexican cinema and the gender roles to earthworm behavioural differences, showcasing the variety of interests encouraged by a liberal arts and sciences education. I hope that this Capstone Issue can share a small slice of AUC students' academic work and interests – with our peers, our families, and more. Enjoy reading the issue!

Sarah Martinson, on behalf of InPrint

A note from the photographers

Every semester, InPrint publishes a capstone issue containing a collection of selected capstones from the previous semester. This year, they wanted to include a picture with every capstone to make the issue more visually appealing, so they approached RAW! InPrint sent the abstracts of the six selected Capstones to RAW, who distributed them over the Photographers Team. The participating photographers each interpreted the abstract in their own way, and took a picture relating to it, which can be found on the title page of each paper. If you're curious about their thought process, read the captions that accompany each photograph below!

Rosa Wijnen for Joyce den Hertog's *Artistic representation of the daily news* The photo represents the title 'Headlines as Art' quite literally, using paint instead of graphic design to create a headline. Would this catch the attention of people better than the usual bold red letters screaming 'Breaking News'?

Jasmin Ronach for Rayne Leroux's *The early bird catches the bold worm* This picture was taken during an afternoon in Flevopark, and subsequently the orange skies were created through photo editing. The focus was put on the bird and its position high up in the trees, from where it may observe the behavior of the 'bold' worm coming to the surface.

Margherita Guida for Aisha Erenstein's *In Reconciliation We Trust* The concepts of reconciliation and trust brought a lot of inspiration in me as they can be very powerful and symbolic. I juxtaposed them with a lily plant, that still has to bloom, which goes to represent Ethiopian politics.

Sanch Sen for Catherine Schulter's *Meaningful Youth Participation in Global Climate Talks* This photograph represents two ideas. One is the essential "running out of time" concept that exists for the world to solve the climate crisis. The second is the complicated framework of ideas and network that come together for the global governance of these ideas.

Mirthe van Veen for Patrīcija Keiša's *Why Are We Not Doing The Right Thing?* Most of us are akratic when it comes to diet. Whether this is about eating meat or a delicious brownie; even though eating it wouldn't be the best choice, the majority still does it.

Che Spraos Romein for Martha Echevarría González's *A Move Towards Visibility* The photo attempts to incorporate elements of a working-class aesthetic, seen through the outfit of the woman, especially in the shoes and less visible apron. It is captured in a monochromatic scheme to match visual elements of some of the films addressed in the essay, and features a woman moving behind bars as a play on the title (move towards visibility).

Contents

Sciences

Sciences

Social Sciences

Social Sciences

Humanities

Humanities

Sciences

ArttnGAN: Headlines as art

Using GANs to create an artistic representation of the daily news

Joyce den Hertog

Supervisor
Dr. Giovanni Colavizza (UvA)
Reader
Dr. Steven de Rooij (AUC)

Photographer: Rosa Wijnen

Abstract

The emergence of social media has changed the way in which we consume news. To compete, newspapers have to employ a more connecting and engaging way of representing the news, and art might be a solution. Creating an artwork for every news article by hand would be tedious and labor-intensive, thus this paper proposes the use of Generative Adversarial Networks (GANs) to generate artworks conditioned on the content of newspaper articles. Application of text-to-image GANs for artistic purposes is still limited. This paper therefore explores the effect of tweaking and implementing state-of-the-art AttnGAN for this task of news-conditioned art generation. After the input dataset is filtered and expanded, two models are proposed: *ArttnGAN2* and *ArttnGAN-F*. However, after a close analysis of the output images and a collection of online survey responses regarding the perception of the generated works, it is concluded that the proposed models are still limited in their ability to figuratively and effectively portray a newspaper description, due to the limited size of the dataset and the GPU restrictions of the current research. Nevertheless, much potential lies in the further development of the models using a larger and more coherent dataset and a longer training time for the GANs. Keywords and phrases: *text-to-image synthesis, generative adversarial network, artistic image generation, generative models, news 1*

Contents

1 Introduction

The emergence of social media has thoroughly changed the way in which we consume the news. Because social media platforms present news articles in the same environment as their entertainment options, 75% of the people that would usually ignore the news in favour of a more entertaining program now admit to reading news-related headlines when encountering them on social media [1]. However, maintaining the interest of this new audience poses difficulties. Another challenge that newspapers currently face, is that people often feel untouched by and uninterested in the many current global issues [2]. Eliasson [2] proposes art as a solution to encourage readers to feel more entertained by, connected to, and engaged with news content. Art, according to him, does not force people to act or think in a certain way, but touches the senses, body and mind, and in that way makes people contemplate [2].

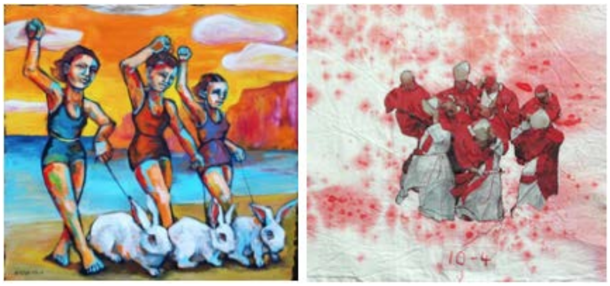


Figure 1: Left: ‘Rabbit Beach’ by Dana Ellyn, acrylic on canvas. Created on July 8th, 2013. On this day, the Pope visited the island of Lampedusa, which is a point of entry for refugees into the EU. Its Rabbit Beach won the title of the world’s best beach in 2013. Right: ‘10-04-2005’ by Seet van Hout, mixed media embroidery. Presumably inspired by the burial of Pope John Paul II.

Artists like Seet van Hout and Dana Ellyn have already created fascinating artistic representations of the daily news (Figure 1) – each of them engaged with the news daily for a certain time period and attempted to produce artworks inspired by news articles from that particular day. Nevertheless, creating unique and compelling artworks to accompany daily newspaper articles is a labor-intensive job. Automatic artistic text-to-image synthesis could therefore be a convenient and valuable alternative.

Researchers in text-to-image synthesis have been able to produce high resolution generations of realistic images like birds, flowers and faces, due to the emergence and wide exploration of Generative Adversarial Networks (GANs) in the past decade [3]. GANs have also been adopted in the artistic field, with artists like Robbie Barrat, Ahmed Elgammal and Mario Klingemann generating works ranging fr- om uncanny portraits to abstract paintings (Figure 2). However, since GANs are trained to imitate their training data, most artists merely use GANs to generate arbitrary art complementary to the artworks of their training data. The field of GAN-generated artworks from textual input is still limited. Therefore, this paper aims to implement GANs for the generation of artworks based on textual descriptions of newspaper articles.

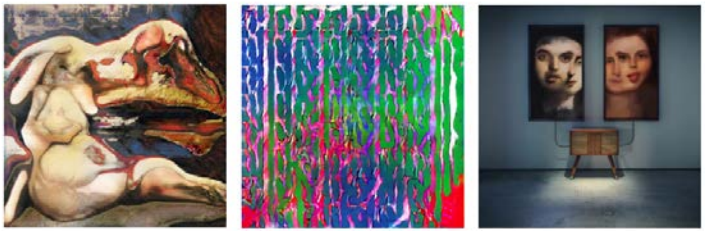


Figure 2: Examples of GAN generated paintings. From left to right: Nude painting by Robbie Barrat (2018); ‘Psychedelic’ by Ahmed Elgammal (2018); ‘Memories of Passerby I’ by Mario Klingemann (2019).

2 Research Context

2.1 Generative Adversarial Networks

GANs are the basis of most state-of-the art text-to-image synthesis models. First introduced by Goodfellow et al. [3], GANs consist of two competing neural networks: the Generator and the Discriminator. The generator tries to generate new samples that are similar to the training data from random noise. Simultaneously, the discriminator predicts whether the generated samples can be distinguished from the real input data or not [4]. By conditioning both the generator and discriminator on extra information, like class labels or text embeddings, conditional GANs are able to generate compelling images from text descriptions [5].

As an extension of traditional conditional GANs, Reed et al. [6] propose the GAN-INT-CLS, which they conditioned on the descriptions of their input images written by humans. They used a character-level text encoder to learn a function that resembles the input images. Since sentence-level text embeddings rarely contain any information about the style of an image, such as background/foregrou-nd information or a subject's pose, GAN-INT-CLS also includes a trained style encoder, which combines different information about style into new pairings to generate a larger variety of outcomes. The architecture by Reed et al. [6] is able to generate credible examples from the text descriptions, but the model is limited to flowers and birds only.

According to Xu et al. [7], as the textual embedding of GAN-INT-CLS is only encoded on the full text description, that approach lacks valuable and critical information at word level. Therefore, Xu et al. [7] propose an attentional generative net-work (AttnGAN) that first generates a low-resolution image conditioned on the global sentence vector. In the second stage, the network focuses on subregions in the image and attends to the words most relevant to those regions using the word-vector to generate an image with higher resolution. Furthermore, using the information of both vectors, Xu et al. [7] propose the Deep Attentional Multimodal Similarity Model (DAMSM) that computes the similarity between the text and the generated image and equips the generator with a more fine-grained loss for training, compared to the traditional loss, which does not have conditioning information. The results of AttnGAN were very promising for images with obvious subjects, such as a bird on a branch. However, more complex combinations of features resulted in less-realistic image generations.

OP-GAN by Hinz et al. [8] builds on AttnGAN by adding modifications focused on the specific location of the objects described in the text. Different words from the text description are given distinct weighted influence on particular parts of the generated image to generate the objects from the description at relevant positions in the image. This process is executed iteratively for every object in the image description. OP-GAN outperforms the more globally focused models, but still struggles with generating multi-domain objects. El-Nouby et al. [9] proposed another iterative approach by conditioning on semantic feedback that was given after every step of the generation process. Their training data, however, consisted of simple drawings, causing the model to be unscalable towards photorealistic images.

StackGAN, developed by Zhang et al. [10], is another multistage GAN following the architecture from Reed et al. [6]. It is a two-stage GAN; it creates a rough sketch of the text description in the first stage, and layers a stage-II GAN upon it to generate a higher resolution image, conditioned on the output of stage-I and the text description again, to correct the deficiencies of the first output. They used Conditioning Augmentation to improve their text embeddings. The quality of the resulting images outperforms GAN-INT-CLS and other state-of-the-art models, but some complex samples still lack recognisable objects or even show signs of collapse into nonsensical mode.

2.2 Artistic GANs

Since artistic images can take any shape, it is often difficult for algorithms to distinguish the background from the foreground, or even to distinguish any shapes at all. Therefore, generating compelling and meaningful artistic images is still challenging, and the research field for artistic GANs, i.e. GANs used to generate art, is limited.

Tan et al. [11] proposed ArtGAN for the generation of artistic images. They implemented back-propagation of the loss and conditioned only the Generator on labels describing the genre of the artworks to provide feedback information for better learning. Nonetheless their approach is only focused on generating art for specific genres, and is

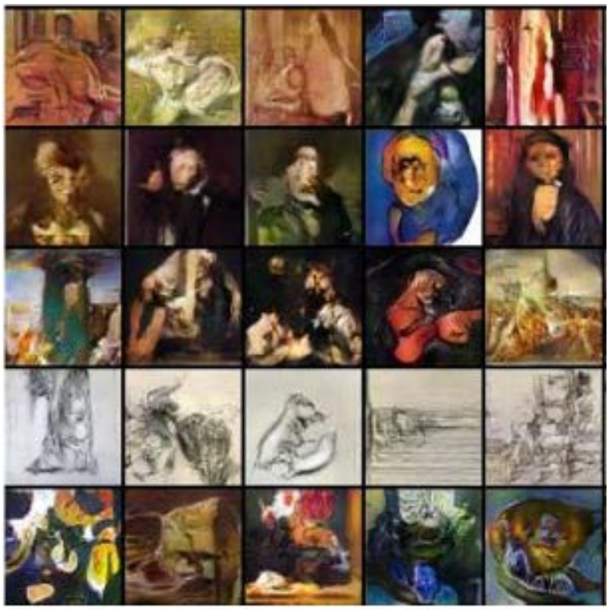


Figure 3: Examples of images generated by ArtGAN. Genres from top to bottom: (1) Nude painting, (2) Portrait, (3) Religious painting, (4) Sketch and Study, (5) Still Life.

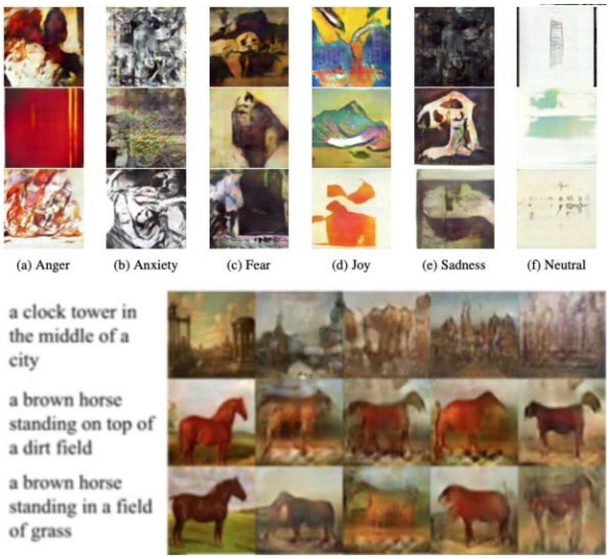

(a) Anger
(b) Anxiety
(c) Fear
(d) Joy
(e) Sadness
(f) Neutral
a clock tower in the middle of a city
a brown horse standing on top of a dirt field
a brown horse standing in a field of grass

Figure 4: Top: Examples of images generated from descriptions by Pixelbrush. Bottom: Examples of images generated by AC-GAN, conditioned on emotions.

not concerned with their textual description (Figure 3).

Pixelbrush [12], on the other hand, builds on Reed et al. [6]'s GAN-INT-CLS and is conditioned on a short text that describes the input images, which were grouped per class (e.g. horse, boat, tower). They found that the generated artistic images are largely consistent with their accompanying text descriptions (Figure 4).

Alvarez-Melis and Amores [13] trained their AC-GAN to be conditioned on the human-annotated em-otions that characterized their artistic input images. Their generated works had clear signs of features that belonged to the specific conditioned emotions, such as aggressive strokes representing negative emotions (Figure 4). Nevertheless, similar to Art-GAN and Pixelbrush, the images that AC-GAN generates are low-resolution.

2.3 GAN Art

A fascinating and thought-provoking question is, of course, whether images generated by a GAN can be acknowledged as art. One of the most important academics exploring this question is Aaron Hertzmann. He argues that art is a form of social expression; it is a medium that enables people to share, showcase and communicate their ideas [14]. Our ancestors, according to Hertzmann [14], used artistic practices like dance, theatre and music to bond socially within groups. The artistic ability of a person was also an indication of wealth, status and mating success fitness. Therefore, he argues, we can classify a work as art when it has been created by a social agent. For a computer program like GANs to be artistic, it thus has to have the social traits of being intelligent, conscious, intentional and emotional.

According to Mazzone and Elgammal [15], this is not yet the case. They argue that generative algorithms are merely a tool within the artistic process. The human artist still has a lot of influence in the creation of the artwork itself. Namely, a dataset first needs to be chosen as input for the GAN, after which the GAN algorithm is tweaked and improved to generate the desired and expected outcome images. The artist has to sort through an array of produced images to curate the resulting collection of artistic images they will eventually publish. Hence, almost the entire creative process is carried out by the artist - not the algorithm. Artists are real social agents, meaning that the artist-curated GAN creations do seem to conform with Hertzmann's theory of art being a social interaction [14].

Nevertheless, Mazzone and Elgammal [15] argue that even though GAN images may conform with certain artistic properties, they lack intent. When Picasso created his deformed and unusual cubist paintings, he did so with the intention to counter the flat, fauvist art that was in style. In contrast, when a GAN unexpectedly generates a deformed image, it is due to its failure to imitate the training data completely [15]. Although the images might seem oddly appealing to human viewers, implying an objective or emotion behind the creation of the image, the GAN had no intent for generating such curious outputs.

Despite the fact that GAN images lack intent, their aesthetics are viewed by many as profoundly novel. According to Hertzmann [16], this is because they cause 'visual indeterminacy': the promise of a visual work to signify something, juxtaposed with its impossibility to ever suggest a stable interpretation. He argues that such ambiguous works evoke a continuous interest from the viewer and never cease to stop surprising. In Art History, this is referred to as the 'elusive mask', coined by Ernst Gombrich [31]. He argues that people prefer a less detailed image, since it allows them to cooperate with the shapes and forms, and unleash their intuition. The more we look at such an image, he says, the more possibile readings emerge and thus the more interesting the image becomes [31]. GANs often portray realistic but unusual scenes, inducing these indeterministic aesthetics. Thus, according to Hertzmann [14, 16] it can be claimed that GAN-generated artistic images are appealing, aesthetically interesting and artistically grounded. Whether they can be truly classified as art is still an unresolved debate, due to the relative novelty of the field of GAN art. Nevertheless, the current research will focus mainly on the implementation of GANs for artistic purposes and will therefore leave this query up for debate.

3 Methodology

3.1 Model Architecture

3.1.1 Conditional GANs

Within the architecture of GANs there are two neural networks simultaneously at work [3]. The generator (G) tries to generate samples that are distributed ($\sim p_g$) similar to the training data ($\sim p_{data}$), so that they will 'fool' the discriminator (D). Simultaneously, the discriminator predicts the probability of whether a generated sample could come from the training data, and accurately 'imitates' the training data's distribution [5]. Thus, the generator creates a function that maps a prior noise distribution $p_z(z)$ to the new distribution of the generated images $G(z) \sim p_G$ and the discriminator returns a probability that suggests whether the output originated from the distribution of the training data rather than p_G [5]. Conditional GANs are conditioned on extra information c, which in the current paper are the captions that correspond to the images. This results in the following training objective for a conditional GAN [8]:

$$\underset{G}{\text{min}}\underset{D}{\text{max}} V(D,G) = \mathbb{E}_{(x,c)\sim p_data}[logD(x,c)]+ \\ \mathbb{E}_{(z)\sim p_{z,(c)}\sim p_{data}}[log(1-D(G(z,c),c))] \tag{1}$$

where $V(D,G)$ is the training objective of D and G (simultaneous training); $D(x,c)$ calculates the probability of datapoint x belonging to the training data conditioned on c; and $G(z,c)$ generates a new datapoint x from noise z, conditioned on c.

As established in the training objective, the generator and discriminator carry out a two-player min-max game [5]. First, the discriminator needs to adjust its parameters to maximize $logD(x,c)$, which is the log probability of training datapoint x belonging to the training data, conditioned on c. This should be maximized since x actually belongs to the training data, so D must predict that correctly. Second, the generator needs to adjust its parameters to minimize $log(1-D(G(z,c),c)$, namely, the probability that the generated image does not belong to the training data, conditioned on c. This should be minimized because G needs to fool the discriminator into thinking that its images actually belong to the training data.

For the purpose of generating artistic images based on newspaper articles, this paper implements such a conditional GAN by training it on artistic images and conditioning it on their descriptions. Firstly, the state-of-the-art OP-GAN[1] appears to be a suitable architecture for this goal, because of its iterative application of object attention layers. Nevertheless, due to expected GPU limitations, implementation of OP-GAN's preceding and less complex architecture AttnGAN[2] seems more feasible. Thus, following the architecture of the Attn-GAN, the image descriptions are first transformed into the shared semantic space of the image features, after which word-context vectors are computed and combined with the image features. The generator and discriminator are then trained on the dataset of artworks, and conditioned on their corresponding word-context feature-vectors.

3.1.2 The DAMSM encoders

To be able to accurately condition the GAN on text, the text needs to be mapped to a similar feature space as the images. To facilitate this process, Xu et al. [7] designed the DAMSM (Deep Attentional Multimodal Similarity Model) encoder. In general, the encoder consists of two neural networks that aim to map the textual input to the same semantic subspace as the image to which it relates. The encoder pays specific attention to the subregions of the image to which the text relates [7].

To encode the text, a Long Short-Term Memory (LSTM) [17] is used in two directions, which resemble the two hidden states for each word. Joined together, these two hidden states create a semantic vector representation of the word. After computing the semantic feature vector for each word, they are appended into a $D \times T$ feature matrix for each caption, where D is the dimension of the word-feature vectors, and T is the length of the sentence. Additionally, a global sentence vector is created by extracting and concatenating the last two hidden states from the LSTM.

To encode the images into similar semantic feature vectors, a Convolutional Neural Network is used. The specific model from Xu et al. [7], uses the Inception-v3 model, which is pretrained on ImageNet [18]. Firstly, a local feature matrix $f \in \mathbb{R}^{M \times N}$ is extracted from the model, where N is the number of subregions of an image, and M is the dimension of the subregions' feature vectors. Then, a global feature vector $\bar{f}$ is obtained from the last layer of Inception-v3. To convert the image feature vectors to the same semantic subspace as that of the captions, a perceptron layer W is applied:

$$e = Wf, \bar{e} = \bar{W}\bar{f},$$

which creates a matrix $e \in \mathbb{R}^{D \times N}$ of local features and global feature vector $\bar{e} \in \mathbb{R}^{D}$. Both e and $\bar{e}$ have dimension D, representing the dimension of the combined feature space of the images and the text features.

3.1.3 AttnGAN

The AttnGAN architecture from Xu et al. [7] distinguishes itself from other conditional GANs by not only focussing on the 'traditional' global sentence features, but combining this global focus with a local focus on features at word-level. Its original architecture (Figure 5) consists of 3 generators, each of which generate an image of a larger scale than the previous generator, conditioned on a hidden layer from the previously generated output, as well as the word-level feature vectors [7].

The model starts by encoding the captions of the images using the DAMSM text-encoder, which establishes a matrix e with word features and a sentence level feature vector $\bar{e}$. $F^{ca}(\bar{e})$ is the Conditioning Augmentation introduced by Zhang et al. [19] to convert $\bar{e}$ to the conditioning vector. Fur-

[1] OP-GAN: https://github.com/tohinz/semantic-object-accuracy-for-generative-text-to-image-synthesis

[2] AttnGAN: https://github.com/taoxugit/AttnGAN

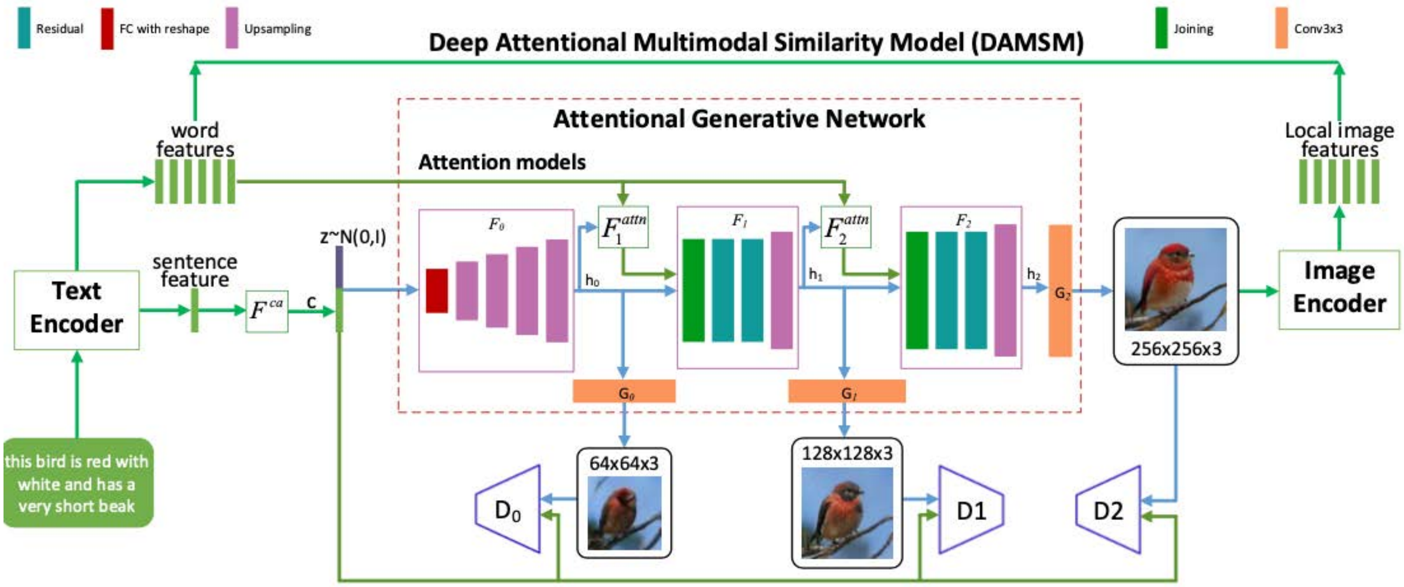


Figure 5: The architecture of AttnGAN [7].

thermore,

$$\begin{aligned} h_0 &= F_0(z, F^{ca}(\bar{e})); \\ h_i &= F_i(h_{i-1}, F_i^{attn}(e, h_{i-1})); \\ \hat{x}_i &= G_i(h_i), \end{aligned} \tag{2}$$

where h_i represent the hidden layers of the network; $\hat{x}_i$ is the final generated datapoint from h_i; and F_i, F^{ca}, F_i^{attn} and G_i are neural networks. From the first hidden layer, a low-definition image is generated that is merely conditioned on the sentence features and noise vector z, which has been randomly sampled from a standard normal distribution [7].

The hidden layer h_0 is then forwarded to the next layer and to F^{attn}, which is the attention model of the AttnGAN, established by computing a matrix of word-context features of each sub-region of the image, based on the hidden image features of the previous layer (h_i). For the j^{th} sub-region, the context vector entails:

$$c_j = \sum_{i=0}^{T-1} \beta_{j,i} e'_i, \text{ with } \beta_{j,i} = \frac{exp(s'_{j,i})}{\sum_{k=0}^{T-1} exp(s'_{j,k})}, \tag{3}$$

where $s'_{j,i} = h_j^T e'_i$, computing the relationship between the i^{th} word in e' (the word-level feature vector e reshaped to the common semantic space) to the j^{th} sub-region of hidden layer h. $\beta_{j,i}$ is a given weight for the i^{th} word of the sentence per subregion j. Thus, hidden layers h_1, h_2 are conditioned on both the previous hidden layer (h_{i-1}), as well as an extra layer that pays attention to the words of the caption within the image context of that hidden layer. Lastly, by employing the DAMSM image-encoder, the output feature vectors are combined with their corresponding image features to generate images of higher resolution and more fine-grained attention on the different sub-regions of the image; $\hat{x}_i = G_i(h_i)$ [7].

Additionally, the current generators are not lear-ned by the original conditional generator loss from Mirza et al. [5] described in equation 1, but rather by a further-developed combined loss, consisting of the conditional loss for stacked generators proposed by Zhang et al. [19] and the DAMSM-loss, defined by Xu et al. [7]. Namely,

$$\mathcal{L} = \mathcal{L}_G + \lambda \mathcal{L}_{DAMSM} \text{ where } \mathcal{L}_G = \sum_{i=0}^{m-1} \mathcal{L}_{G_i}, \tag{4}$$

where λ is a weight needed to balance the two terms of the equation. The generator loss, $\mathcal{L}_G$, is defined by

$$\mathcal{L}_{G_i} = \underbrace{-\frac{1}{2}\mathbb{E}_{\hat{x}_i \sim p_{G_i}}[log(D_i(\hat{x}_i)]}_{\text{unconditional loss}} \underbrace{-\frac{1}{2}\mathbb{E}_{\hat{x}_i \sim p_{G_i}}[log(D_i(\hat{x}_i, \bar{e})]}_{\text{conditional loss}} \tag{5}$$

[19]. Here, the unconditional loss computes the expectation of the discriminator acknowledging that the image belongs to the training data. The conditional loss checks whether the image fits correctly with the provided caption. D_i is the corresponding discriminator to generator G_i at the i^{th} stage [7].

The DAMSM loss is defined by

$$\mathcal{L}_{DAMSM} = \mathcal{L}_1^w + \mathcal{L}_2^w + \mathcal{L}_1^s + \mathcal{L}_2^s, \tag{6}$$

where w stands for 'word-level' and s for 'sentence level'; $\mathcal{L}_1^x$ is the posterior probability of the sentence matching the image on level x; and $\mathcal{L}_2^x$ is the posterior probability of the image matching the sentence on level x.

The focus on both the local word-level and global sentence-level features makes the DAMSM-loss a profound image-text matching loss [7].

Furthermore, each discriminator D_i is also trained on a combined loss [19]:

$$\mathcal{L}_{D_i} = \underbrace{-\frac{1}{2}\mathbb{E}_{x_i \sim p_{data_i}}[log(D_i(x_i)] - \frac{1}{2}\mathbb{E}_{\hat{x}_i \sim p_{G_i}}[log(1 - D_i(\hat{x}_i)] +}_{\text{unconditional loss}} \underbrace{\frac{1}{2}\mathbb{E}_{x_i \sim p_{data_i}}[log(D_i(x_i, \bar{e})] - \frac{1}{2}\mathbb{E}_{\hat{x}_i \sim p_{G_i}}[log(1 - D_i(\hat{x}_i, \bar{e})],}_{\text{conditional loss}} \tag{7}$$

where x_i is an image data point from the original data distribution p_{data_i} at the i^{th} resolution, and $\hat{x}_i$ is the generated image from distribution p_{G_i} at the same resolution. Then, the unconditional loss is based on the expectation for D to predict that actual data point x_i belongs to the data, and that generated $\hat{x}_i$ does not belong to the dataset. The conditional loss calculates these probabilities conditioned on the sentence-level feature vector $\bar{e}$. Finally, the discriminator works disjointly from the generator, meaning they can be trained simultaneously and in parallel, which advances the computational speed of the GAN [7]. As mentioned before and established by Mirza et al. [5], the generator's loss needs to be minimized and the discriminator's loss needs to be maximized to generate original, realistic images.

3.2 Data

Two datasets are needed for this research. First, a collection of artistic images that is accompanied by a textual description is used to train the GAN on the images, conditioned on the text that describes them. Providing the trained GAN with new textual input will allow it to generate novel images. Secondly, since the goal of this paper is to create art about news headlines, a dataset of newspaper headlines is needed to generate the desired output images. Namely, the trained GAN will take a headline as its input, and produce an output image by generating a new image from the first training data, while paying extra attention to certain specific words of the headlines to convey their meaning in the output image.

The BAM! (Behance Artistic Media) dataset[3] was mentioned by Zhi [12] as a promising dataset for the cause of generating images from text, since the images are paired with short descriptive captions. Nevertheless, after collecting this dataset, it emerged that the descriptions are merely focused on the way the artistic medium is applied to the associated artwork (e.g. “The strokes and vivid color looks like that of oil."). As the goal of this paper is to encapsulate the subject of a text into a generated image, the training data needs to consist of images paired with descriptive captions about their subject, rather than their medium.

Accordingly, the WikiArt Emotions dataset [20] was chosen as the training data of the GAN. This dataset contains 4,105 artworks that have been human-annotated with emotional value. From WikiAr

[3] BAM! dataset: https://bam-dataset.org

t.org, 200 works were selected per feature page of 22 selected styles, such as Realism, Expressionism, and Cubism. These styles comprise four different time periods: Renaissance, Post-Renaissance, Modern, and Contemporary. The WikiArt dataset only includes datasheets with the information (genre, art-ist, title, year, URL.) and emotional annotation of the works. Thus, the images were scraped from the provided URLs using Python's ImageScraper[4]. After locating errors and processing the file formats, the resulting dataset consists of 3,894 artworks, grouped by their genre. The title was extracted for each image to serve as the image description input for the GAN.

Furthermore, to scrape the desired newspaper article information, BeautifulSoup[5] was used. The articles were scraped from the Today's Paper page on NYTimes.com[6]. By iterating over each day of the month of February, the first article of the front page was selected, and information about its date, title, URL, and provided description were extracted. Then, Newspaper3K[7], a noted Python library for article extraction, was used to extract the main keywords of the article. This constructed a dataset with the brief descriptions and information of the 28 lead articles of February 2020.

3.3 Evaluation

A widely used evaluation metric for GAN generated images is the inception score proposed by Salimans et al. [21]. This score, however, does not measure whether an image matches the description it is conditioned on. Neither does it measure whether an image can be perceived as artistic. Since this research generates a more subjective range of images, evaluation using human annotation is considered to be more appropriate. To achieve this, an online survey (table 3 in appendix B), which shows an arbitrary generated work with its respective newspaper article title, is distributed through various media. The survey can be refreshed or reloaded to show a different artwork each time, enabling respondents to participate multiple times. A total of 256 responses were registered during the period of May 5-14, 2020.

The survey is aimed at obtaining responses that gauge the extent to which the respondents: (1) think that the artwork aesthetically and/or figuratively reflected the news article title (*To what degree do you feel the artwork is aesthetically/ figuratively related to the title?*), (2) are contemplating to a greater extent about the news article after seeing the artwork (*Does looking at the artwork make you reflect more about the title?*), and (3) appreciate the artwork in general (*What would you rate this artwork (overall) on a scale of 1-5?*). Furthermore, there is an open space for participants to elaborate on their responses if they desire to do so. The overall aim of the survey is to find out whether the generated artworks are considered to depict the overall idea and mood of their respective news articles, as well as whether the works are aesthetically and artistically appreciated.

The survey is conducted for the generated works from 3 different models described in section 4 (*ArttnGAN2* conditioned on unfiltered captions, *ArttnG-AN2* conditioned on filtered captions and *ArttnGAN-F*). Afterwards, statistical analysis is performed on the responses, where the Kruskal-Wallis H-test is used to compare the different models. Furthermore, effect sizes are reported in terms of Cohen's d.

4 Experiments

To evaluate the application of AttnGAN for artistic purposes, extensive experimentation is carried out. This section first analyses the preliminary results of the model, after which various experiments are performed by adjusting the input data. This is

[4] Using ImageScraper version 2.0.7: https://pypi.org/project /ImageScraper/

[5] Using the latest release of Beautiful Soup: version 4.8.2 (December 24, 2019). See https://www.crummy.com/softwar e/BeautifulSoup/

[6] Example of NYTimes Today's Paper for 27th February: https://www.nytimes.com/issue/todayspaper/2020/02/27/ todays-new-york-times

[7] Newspaper3k is developed and maintained by Lucas Ou-Yang: https://github.com/codelucas/newspaper

done in order to explore whether the results can be improved by focusing on different subsets and details of the model.

4.1 Preliminary results

Initially, the DAMSM encoder for the first model, *ArttnGAN1*, was trained for 400 epochs on the full dataset. Nevertheless, after 100 epochs, the loss stagnated, which is in line with the experiences of other users8 who have implemented the DAMSM encoder. Hence, the decision was made to train the future DAMSM encoders for only 125 epochs. The GAN for *ArttnGAN1* was trained for 100 epochs on the full dataset with an 80/20 random split for train/test. Furthermore, the attention network F^{attn} was trained by focusing on a limit of 5 words per image, since the average sentence length of the titles from the training data is 3.6 words. Shortening would lead to too few attention mappings (visualisations of the GAN's attention layers) for the GAN to train on, whereas expanding would cause an excess of empty data, as non-existing words are entered into the attention layer in the shape of a null vector. Regardless, the attention maps of *Arttn-GAN1* appear to be rather arbitrary and unaware of context (Figure 6). This inaccuracy could be due to the limited training time, the limited size of the dataset, the fact that the model was only trained on one caption per image, or the relatively short length of the captions.

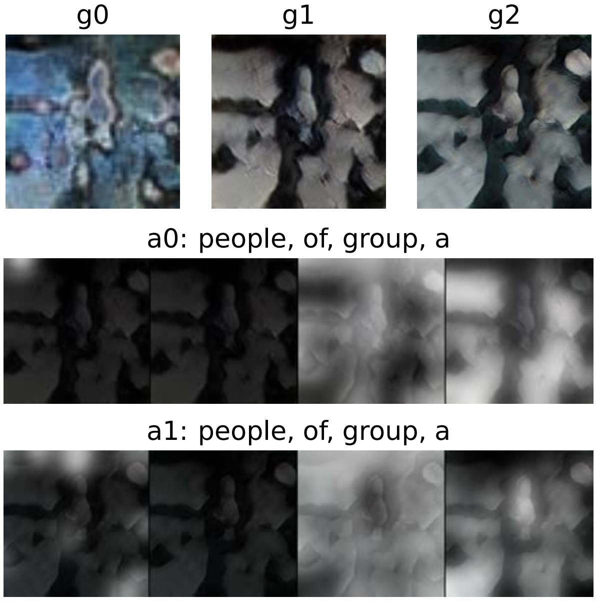


Figure 6: Images generated by *ArttnGAN1* from the sentence "a group of people" and its attention maps. The first row shows the low-to-high resolution images generated by G_0, G_1 and G_2, and the last two rows show the attention maps of F_1^{attn} and F_2^{attn} respectively.

4.2 Training on multiple captions

In an attempt to combat the problem of merely training on a single caption and thus having a low variety of descriptions to train on, the second model, $ArttnGAN2$, was trained on two captions per image. First, a new caption was generated for every image in the dataset using the image captioning code-base from Ruotian Luo [22], specifically his pre-trained top-down model, which achieved a cross-entropy loss higher than one[8]. Besides providing a larger variety of captions to train on, the generated captions also appear to be more descriptive of the artworks. This is particularly useful as artworks are usually not accompanied by a title that accurately describes the visual content of the image. Looking more closely at one of the works from the dataset, *The Marriage of Heaven and Hell* by Keith Haring: "a close up of a black and white photo of a vessel drift diamond light".

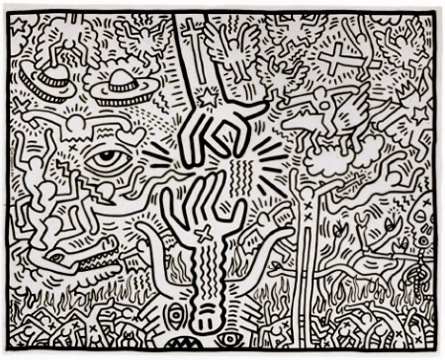


Figure 7: 'The Mariage of Heaven and Hell' by Keith Haring (1984).

[8] https://github.com/ruotianluo/ImageCaptioning.pytorch/issues/10

This caption does not describe the artwork in close detail, but figuratively, the caption makes sense. The artwork is indeed black and white; there are some apparent vessels depicted; Keith Haring has drawn motion lines around his figures, which can be interpreted as drift lines; and the 'diamond light' probably refers to the shining star positioned in the middle of the upper hand. The generated description, therefore, is probably more accurate in describing the artwork than the original title; nothing in the picture clearly refers to a 'marriage', and although the heaven above and the hell below are clearly depicted by angels and fire, more imaginative power is needed to relate these objects to their respective concepts.

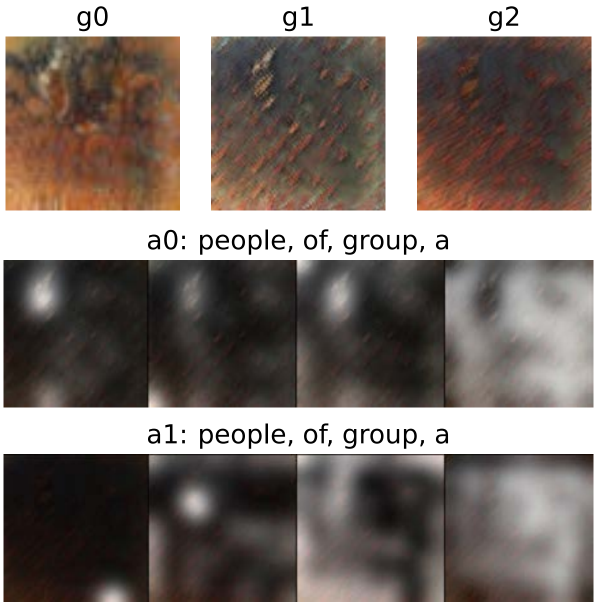


Figure 8: Images generated by *ArttnGAN2* from the sentence "a group of people" and its attention maps. The first row shows the low-to-high resolution images generated by G_0, G_1 and G_2, and the last two rows show the attention maps of F_1^{attn} and F_2^{attn} respectively.

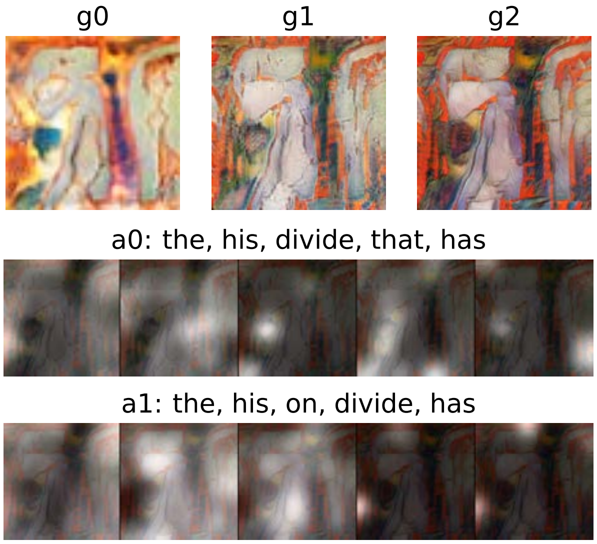

g0
g1
g2
a0: the, his, divide, that, has
a1: the, his, on, divide, has

Figure 9: Images generated by *ArttnGAN2* from the news description "The Vermont senator tightened his grip on the Democratic Party's liberal wing, benefiting from a field that has divided moderate voters." and its attention maps. The first row shows the low-to-high resolution images generated by G_0, G_1 and G_2, and the last two rows show the attention maps of F_1^{attn} and F_2^{attn} respectively.

The DAMSM encoder for $ArttnGAN2$ was trained for 75 epochs and the GAN for 250 epochs. The resulting images show more details in both their visual representation as well as their attention maps (Figure 6 vs. 8). This appears to have been achieved by training the GAN for 250 epochs rather than 100. Accordingly, the future models will be trained for 250 epochs. Results were sampled from arbitrary captions (Figure 8) as well as news article descriptions (Figure 9). Nevertheless, the attention maps of $ArttnGAN2$ still seem to be quite arbitrarily focused. Additionally, they more often address lexical words (i.e. 'the', 'his', 'that', 'has') than content words (i.e. 'divide'), thus focusing on the least meaningful words of the sentence (Figure 9). This might be due to the limited vocabulary of the captions the model is trained on. Since the model cannot recognize words that do not appear in its vocabulary, there might be no recognisable content words within the sentence to which the attention model can attend.

4.3 Filtering the training vocabulary

Because the non-descriptive artwork title and the limited vocabulary still pose trouble to the ability of the GAN to represent its input figuratively and to pay attention to more context-dense words, a more thorough filtering was applied to the titles of the input images. First, as specific names of people and places often appear only once in the vocabulary, some named entities were filtered out and replaced in the image captions. That is, using spaCy's Named Entity Recognition[9], the named entities labeled with 'PERSON', 'LOCATION', 'DATE' and 'CARDINAL' were identified and replaced respectively wi-th 'person', 'location', 'date' and 'number', to make sure that it was still apparent that such an entity appeared in the image title.

Afterwards, the captions were analysed on word-occurrences and it appeared that of the 4,095 uni-que words used in the captions, 2,961 occurred only a single time. Figure 10 shows an overview of the 30 most common words in the dataset. It was decided to filter out the single-used words and words shorter than four characters, like 'the', 'and', 'it', to make sure the captions consisted of more content words rather than lexical words. The images that were left with empty titles after this filtering, were removed from the dataset, leaving 3,178 images in the final dataset.

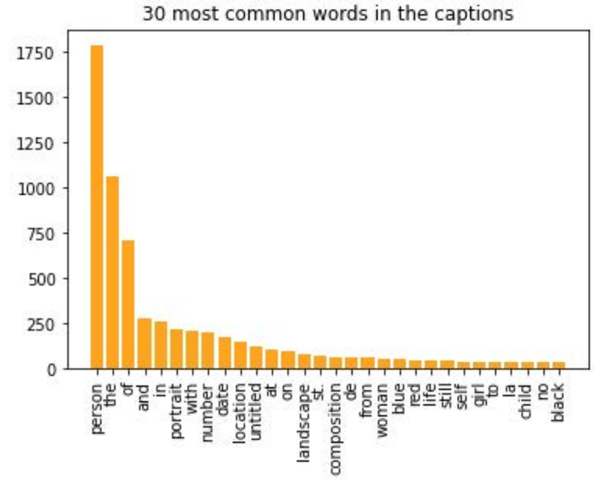


Figure 10: Bar chart showing the 30 most common words in the image captions.

For the third model, *ArttnGAN-F*, the DAMSM encoder was trained on the filtered dataset, with the single filtered title accompanying each image, for 125 epochs. The GAN was trained for 250 epochs again. Sampling was done by filtering the news articles' descriptions in the same way and generating images conditioned on those descriptions. Some results are shown in Figures 11 and 12. As expected, the attention layers are now focused more on the content words, but they still do not appear to relate concretely to any shapes within the generated image.

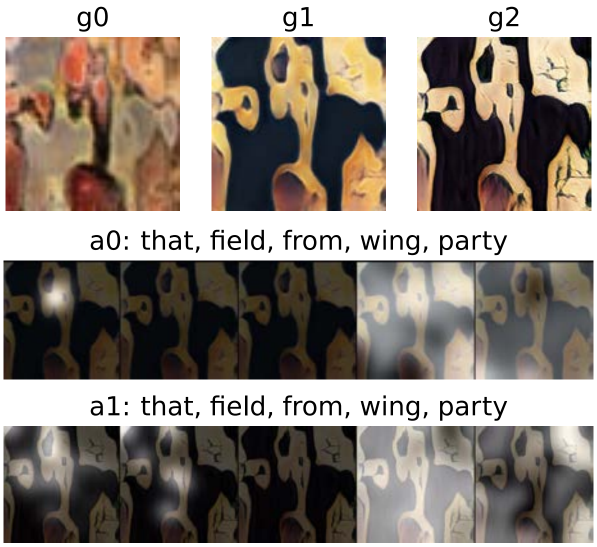

g0
g1
g2
a0: that, field, from, wing, party
a1: that, field, from, wing, party

Figure 11: Images generated by *ArttnGAN-F* from the caption "party wing from field that" , filtered from the article "Bernie Sanders Scores Narrow Victory in New Hampshire Primary", and its attention maps.

4.4 Clustering the input

Another experiment was done by clustering the images from the dataset with similar styles, to examine whether the output images would appear mo-re coherent. First, a pretrained VGG16 model was loaded from Keras Applications[10] as an image encoder. Using this model, the features from each image were extracted. Then the elbow method[11] was used to decide on the most efficient number of clusters. This was done by plotting the sum of the squared distances of the data points from their

[9] SpaCy: https://spacy.io/usage/linguistic-features

[10] Pretrained VGG16: https://keras.io/api/applications/vgg/#vgg16-function

[11] Elbow method: https://www.kaggle.com/ellecf/visualizing-multidimensional-clusters

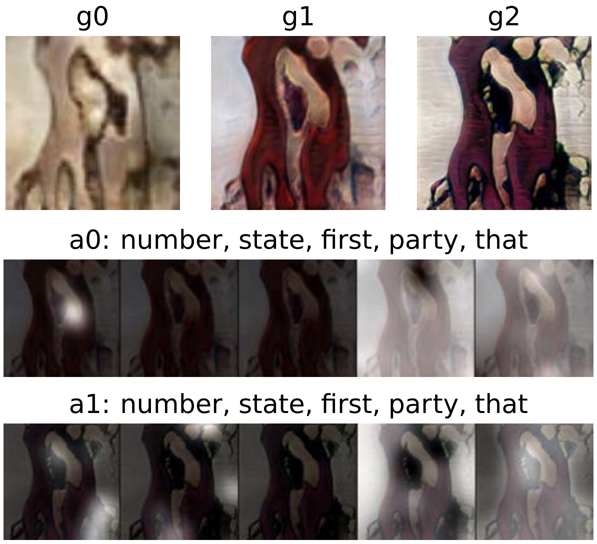

g0
g1
g2
a0: number, state, first, party, that
a1: number, state, first, party, that

Figure 12: Images generated by *ArttnGAN-F* from the caption “that party first state number", filtered from the article "How the Iowa Caucuses Became an Epic Fiasco for Democrats", and its attention maps.

cluster’s center for increasing numbers of clusters, using sklearn’s KMeans clustering. Figure 13 shows the visualisation of this method on the image features from the dataset. However, the graph shows no apparent pivot

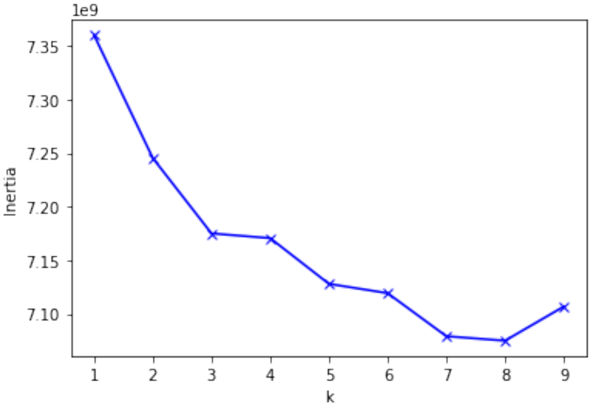


Figure 13: Visualisation of the elbow method for our dataset.

However, the graph shows no apparent pivot point, so multiple points could be observed as the ’elbow’. Nonetheless, 5 was chosen as the amount of clusters to group the dataset into, since 7 or 8 would divide the dataset in clusters that are too small for effective training, and 3 is expected to shape clusters containing images that are not as related as desired. The resulting clusters are visualised in two-dimensional space in figure 14, using PCA decomposition.

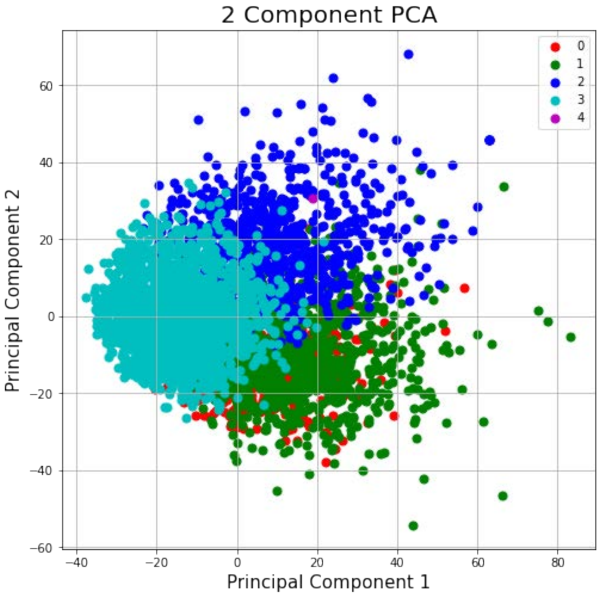


Figure 14: 2-dimensional visualisation of the 5 k-means clusters of the dataset, using PCA.

After closer examination of the clusters by labeling the data points with their art style, it appeared that the clusters were indeed representing styles of painting that were closely related to each other, e.g. Post-Impressionism and Pop-Art were clustered together, as well as Abstract art and Minimalism. Eventually, cluster 1 was chosen for training the new model – this cluster included artworks in the styles similar to Renaissance, Baroque and Classicism. This choice was made because the works from these stylistic periods are relatively more figurative and realistic, and thus might contain artworks whose contents are more related to their descriptions. Cluster 1 consists of 899 artworks.

Then, for the fourth model, *ArttnGAN-C*, the DAM-SM encoder was trained for 125 epochs on the part of the dataset consisting of cluster 1, with the normal title and the generated caption from the ImageCaptioning codebase accompanying each image. The GAN was trained for 250 epochs again. As seen in Figure 15, the results are of particularly low quality, and the attention maps seem to malfunction, as they are not paying attention to any specific region of the image at all. These failures seem to be caused by the limited size of the clustered dataset. Running this experiment with one of the other clusters will most likely not resolve this complication, as the clusters are either of smaller size (cluster 0: 457, 1: 899, 2: 764, 3: 1773, 4: 1), or contain mainly abstract works (i.e. cluster 3 contains images in the styles: Minimalism, Color field Painting, Abstract, Abstract Expressionism). Furthermore, running this experiment with fewer clusters (e.g. three), will presumably not improve the results any further, as the full dataset is already reasonably small. Splitting it up in three clusters will not increase the size of the clusters considerably (i.e. if the three clusters are uniformly distributed, they will consist of 1300 images each, whi-ch is only 400 more than the current cluster size). Therefore, the results of this approach cannot be analysed properly, because the limited size of the dataset does not allow the model to be trained to its maximum capacity. Nevertheless, as focusing on a larger cluster of more figurative and coherent styles is expected to improve the outputs of the model, there still lies potential in this approach for future research, using a more extensive and coherent dataset.

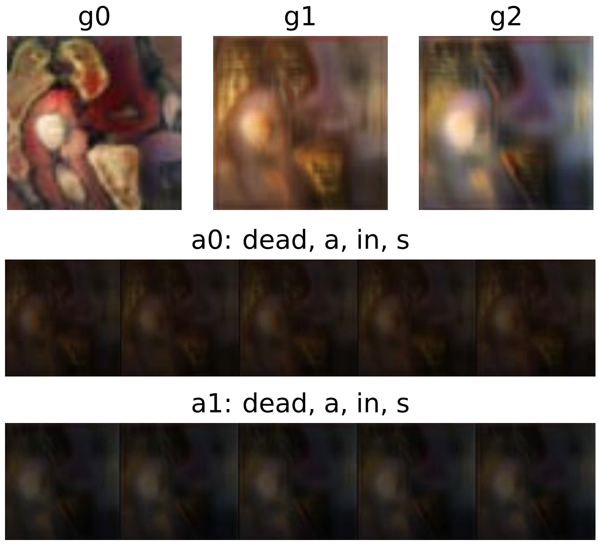


Figure 15: Images generated by *ArttnGAN-C* from the caption “a remarkable level of discord enveloped the start of the nominating process as results from Iowa’s troubled caucuses showed Pete Buttigieg and Bernie Sanders in a dead heat", extracted from the article “Muddled Democratic Race Hurtles to New Hampshire", and its attention maps.

5 Validation

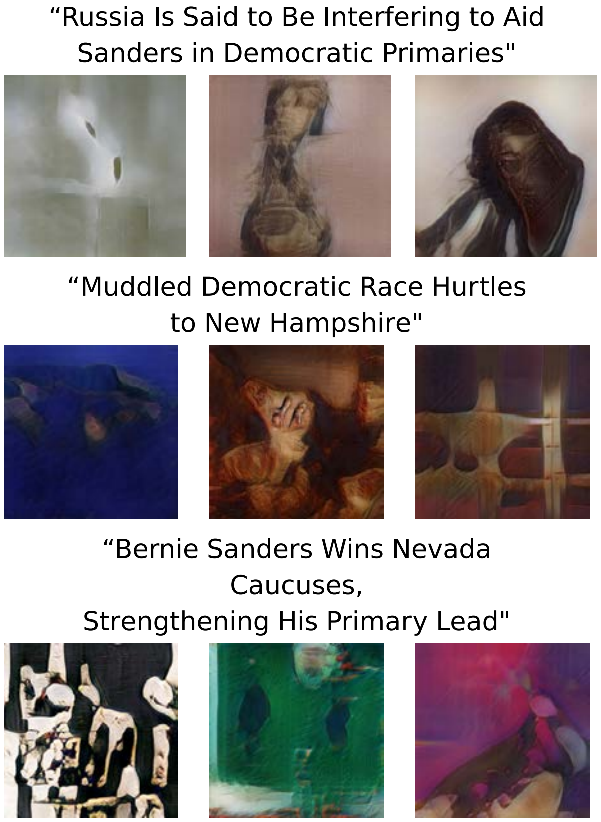
As mentioned in Section 3.3, any evaluation of the outputs of the models is essentially a subjective one, as the models produce artistic representations which can be interpreted in many different ways. Therefore, this section provides merely an intuition about the results of the different models by first taking a few individual samples and analysing them closely in terms of relatedness to the caption, content and use of color. Then, the results of the survey proposed in Section 3.3 are summarized and analysed. Table 1 shows an overview of the different DAMSM encoders and GAN models, and the number of epochs for which they were trained. In this section, the two models that generated the most detailed attention maps will be considered, namely *ArttnGAN2* and *ArttnGAN-F*.

Model	DAMSM training	GAN training
ArttnGAN	400 epochs	100 epochs
ArttnGAN2	75 epochs	250 epochs
ArttnGAN-F	125 epochs	250 epochs
ArttnGAN-C	125 epochs	250 epochs

Table 1: Overview of the training length per model

5.1 Individual sample analysis

Figure 16 shows a comparison on sampling from both models. First of all, the difference between the images generated from *ArttnGAN2* and *ArttnGAN-F* is highly noticeable, even though the images were generated from the same sentence. It can also be pointed out that the sampling of *ArttnGAN2* with filtered news descriptions (right) versus the sampling of the same model with unfiltered news descriptions (middle) gives an entirely different result, which is at least due to the fact that the attention maps are focused more on content words (i.e. ‘russia’, ‘person’) rather than lexical words (i.e. ‘is’, ‘to’, ‘in’). Furthermore, none of the generated images seem to accurately represent the content of the news articles in an identifiable way.



“Russia Is Said to Be Interfering to Aid Sanders in Democratic Primaries"

“Muddled Democratic Race Hurtles to New Hampshire"

“Bernie Sanders Wins Nevada Caucuses, Strengthening His Primary Lead"

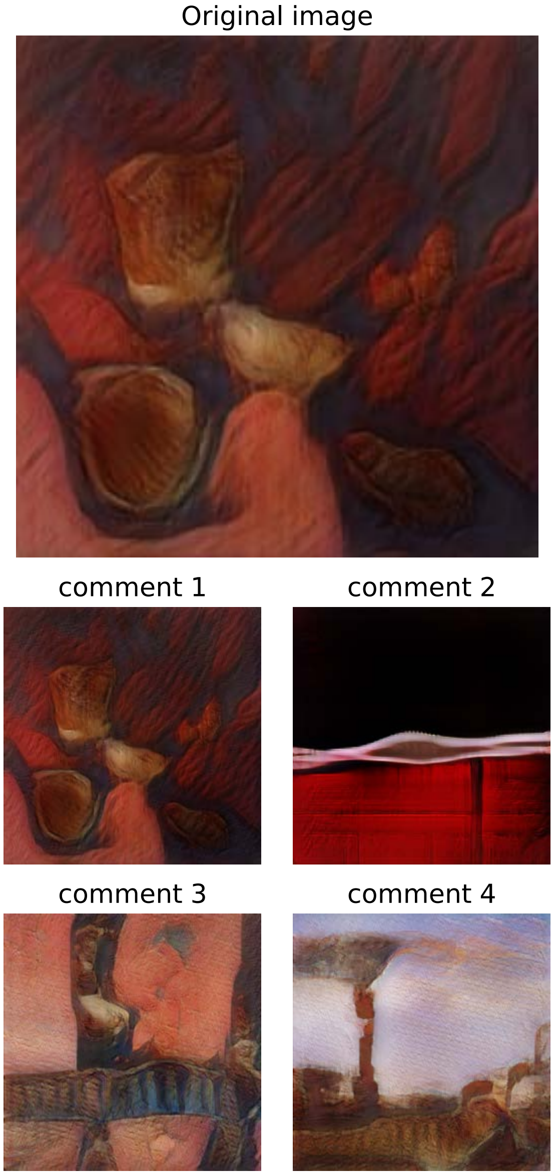
Figure 16: A comparison of *ArttnGAN-F* and *ArttnGAN2*. The images were sampled from the GANs using the description extracted from the news articles referenced above the images. From left to right: generated by *ArttnGAN-F* from the filtered news description; generated by *ArttnGAN2* from the unfiltered news description; generated by *ArttnGAN2* from the filtered news description.

Additionally, a tentative test was conducted to examine whether *ArttnGAN2* might be able to convey the mood of the article by sampling on negative and positive comments in addition to the article description of the article “Harvey Weinstein Is Found Guilty of Sex Crimes in #MeTooWater-shed". The comments used are listed in Table 2. The images were sampled from the captions after filtering them according to Section 4.3 (Figure 17). The first comment generates the exact same image as the original, which is probably due to its short length; it might not add any specific extra meaning to the original caption. Overall, the sampled artworks contain mostly organic shapes and lines, indicating a natural subject. What is remarkable is that the artworks that were generated using comments labeled ‘positive’ are darker than the ones using the ‘negative’ comments. This conflicts with the notion that dark colors are generally associated with negative emotions, whereas brighter colors are related to positive emotions [23]. It is noticeable that all images use the color red, a color which is associated with – among others – power, sex and danger

	Comment	Label
1	“Off to jail. Good."	positive
2	“Finally. Justice has been served. It is good to know this man is behind bars."	positive
3	“If the women were just normal middle class victims? Unlikely the case would even come to trial. No one is talking about this aspect of the case. It’s crucial."	negative
4	“Weinstein is a horrible human being, but this was not rape. If so, then I know many women who have been “raped” in the service of advancing their careers."	negative

Table 2: Reader’s comments on the news article “Harvey Weinstein Is Found Guilty of Sex Crimes in #MeToo Watershed", with the accompanying description: “The jury found Mr. Weinstein guilty of two felony sex crimes but acquitted him of charges that he is a sexual predator."

[24]. This seems to be in line with the subject of the article: the conviction of a sexual offender. However, no obvious figurative shapes representing an offender, conviction or #MeToo practices can be distinguished in the images. Thus, according to the current example, *ArttnGAN2* appears to be aware of the overall mood of the caption, but a focus on more detailed, defined sentiments is not achieved by model. However, this result is very anecdotal and subjective, and more examples would have to be considered to ground this statement. Future work might wish to direct a deeper analysis on the content and signification of the output.


Original image

comment 1

comment 2

comment 3

comment 4

Figure 17: Images generated by *ArttnGAN2* from the article “Harvey Weinstein Is Found Guilty of Sex Crimes in #MeToo Watershed" combined with several reader’s comments (table 2).

5.2 Survey analysis

As proposed in Section 3.3, after sampling artworks from *ArttnGAN2* and *ArttnGAN-F*, using the news descriptions from the headlines of the month of February 2020 as a condition for the GANs, a survey was conducted where an arbitrary work was shown from one of the models and respondents cou-ld answer some questions about its figurative and aesthetic relatedness to its respective news caption. The survey is described and its results are presented in Appendix B. A total of 256 responses were collected.

The survey took *ArttnGAN2* sampled on unfiltered captions as the standard model, and the responses to its generated images are compared to the responses on the images from *ArttnGAN2* sampled on filtered captions and from *ArttnGAN-F* (Table 4, Table 5 and Figure 21 in Appendix B). Overall, the images of *ArttnGAN2* (filtered) were perceived as more aesthetically pleasing than the images sampled on unfiltered captions (H=4.99, p=0.025). The same holds, less significantly, for the images of *ArttnGAN-F* (H=3.48, p=0.062). Filtering the captions thus has a moderate effect on the aesthetic perception of the artworks (Cohen’s d = 0.33 for *ArttnGAN2* and Cohen’s d = 0.28 for *ArttnGAN-F*).

A significant difference was found in the ability to portray the mood of the caption by *ArttnGAN-F* (H=5.92, p=0.015, Cohen’s d = 0.38). However, this was the only question about the relatedness to the title of the work that noted a significant difference between any of the models. The other questions about figurative relatedness and aesthetic relatedness did not report any statistical difference with *Art-tnGAN2* (unfitered). Thus, even though the attention maps of *ArttnGAN-F* might be focused on more content words than *ArttnGAN2*, there is little proof that they actually appear to relate the aesthetics or the content of the image more to the title.

For the overall likeability of the generated artworks (What would you rate this artwork (overall) on a scale of 1-5?), *ArttnGAN2* conditioned on filtered news captions scores considerably higher than the other two models (average of 3.22 vs. 2.87 for *ArttnGAN2* (unfiltered) and 2.89 for *ArttnGAN-F*), with a moderate effect between *ArttnGAN2* (unfiltered) and *ArttnGAN2* (filtered) (H=5.17, p=0.023, Cohen's d=0.34). However, there is no clear explanation as to why these images are better appreciated.

Generally, the images sampled from the article description of "Harvey Weinstein Is Found Guilty of Sex Crimes in #MeToo Watershed", generated the most interesting individual responses (Figure 22 and 23 in Appendix B). On the image generated by *AttnGAN2* conditioned on unfiltered captions, one respondent convincingly related the image's aesthetics to hopefulness on the continuity of the #MeToo movement, whereas another respondent saw no connection at all to the article's title. On the image generated by *AttnGAN2* sampled on filtered news captions, two respondents found that the title highly resonated with the image. The first claimed that the dark colors and round shapes of the artwork biased the reader in focusing more on the negative aspect of sexual assault than the positive progress of the #MeToo movement. The second respondent agreed with this visual analysis, claiming that "sexual harassment is depicted by a darker red reflecting both pain and (forced) eroticism".

On average, about 70% of the respondents indicated that the artwork they saw certainly ("yes") or possibly ("maybe") made them reflect more about the news caption. This result, as well as the interesting individual responses, demonstrate that accompanying a news article with an emotive artwork causes people to contemplate more about its content, and even gives them creative insights that may encourage a different interpretation of the content. Nevertheless, as the first respondent from Figure 23 stated, artworks might also bias the reader into interpreting the news in a possibly unrelated way.

6 Contextualisation

For this section, the models and methods used by a selection of GAN artists will be compared to that of the current paper in order to examine its possibilities and limitations with regard to the contemporary field.

One of the most influential GAN artists is Robbie Barrat, who has collaborated with, among others, Acne Studios for a fashion collection generated by one of his GANs [25]. Furthermore, his artistic implementation of *DCGAN* was used by the collective 'Obvious', who produced Portrait of Edmond de Belamy, a GAN-generated painting trained on 15,000 portraits from the 14th-20th century, which was auctioned at Christie's for $432,500 [26]. Barrat stated that they used "a ripoff of my very early, very bad work", and feels as though they have taken advantage of him [27]. Even though the portrait has become one of the most renowned portraits in the field, Barrat is among one of the many GAN artists that strongly oppose the manifests of Obvious. He argues that they are merely a marketing group, who made it seem like the computer was fully in charge of the process of creation by advertising the generated portrait as the most ṕurist' form of creativity expressed by the machine" [28], which is an intriguing but delusional statement.

Barrat has published some of the code that he used for generating his work as open-source models, such as his implementation of *DCGAN*[12]. The

[12] https://github.com/robbiebarrat/art-DCGAN

model for *DCGAN* was initially built by Radford et al.[29], but Barrat made some alterations to make it fit for artistic generation. That is, he doubled the image size, added image-scraping code to his repository and added the ability to resume training from checkpoints, to enable the possibility for the dataset to be switched-up mid-training, creating an option for style-transferring. Barrat explains that he finds that the most compelling results appear when one of his GANs misinterprets some things: “realistic results are very boring after 10 minutes" [27]. Therefore, Barrat likes to experiment with GANs by training them in detail on how small parts of the painting work, but having them fail in comprehending the overall picture. An example of this is one of the nude portraits from his early works (Figure 18), which demonstrates how the GAN interpreted that nudes consist solely of blobs of fat.

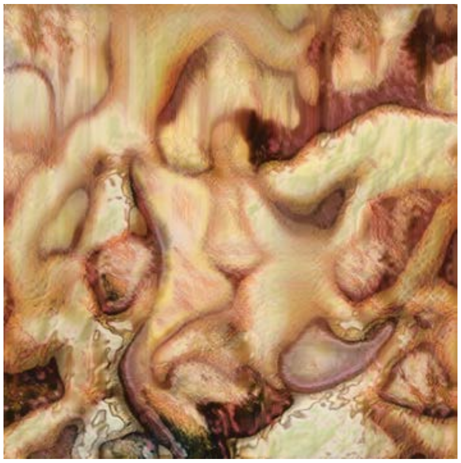


Figure 18: Nude portrait by Robbie Barrat (2018).

Another prominent AI artist is Mario Klingemann. He calls his approach with GANs “neurography", short for “neural photography", comparing it’s process to that of a photographer who wanders around the world in search of an interesting concept, after which he looks for the best way to frame and capture it [30]. In the same way, Klingemann argues, a GAN artist goes into the latent space of their algorithm to search for interesting images. However, in contrast with photography, GANs can generate unlimited worlds in which to wander, whereas the photographer is limited to that in which we live [30]. Like Barrat, Klingemann also indicates that he finds it the most interesting when accidents happen and a model produces significantly different outputs than expected [30].

In terms of models, Klingemann is a fan of *pix2pix* and *StyleGAN*, but he mostly uses his own models, which are often a hybrid of earlier developed GANs, but adapted to fit his own goals. He invests a great deal of time in curating the in- and outputs of his GAN, because in that way, he argues, it will lead to the most original outcomes: “Every creation is a search and the time spent on a search is in direct relation to the originality of that creation - the longer you search and make creative decisions the less likely it is that you will just discover what others can find, too or have come across already" (appendix A.1). According to Klingemann, only very few of the outputs are truly interesting to him, no matter how good the GANs will become. That is why the artist is so important in the process of creating GAN art. GANs, he argues, can produce limitless worlds of shapes and colors, like nature can do too. Then it is up to the artist to select and frame what these worlds have to offer.

Terence Broad, another GAN artist, agrees with this viewpoint. He values his artwork on the “amount of thought and effort that goes into the curation of the data, the choice of algorithms, the curation of the generated output and the framing and contextualization of the work" (appendix A.2). Like Barrat and Klingemann, he also works with existing GANs, which he adapts to his own liking. Sometimes, the responses to his work are those of repulsion and distaste, but he finds that those works are often the most unusual and thus the best fitted to create evoking artworks, like his series Being Foiled (Figure 19). Mario Klingemann describes this as the “Francis Bacon effect", referring to the artist’s grotesque and disturbing artworks, which often provoke mixed responses of repulsion and admiration [28].

The approach of the current paper is in line with those of contemporary GAN artists: the original *AttnGAN* model was tweaked to fit the purpose of

Figure 19: Sample of Being Foiled by Terence Broad (2020).

generating artistic rather than realistic images. Afterwards, experiments were conducted to improve the outputs. Nevertheless, more time could have been spent curating the outputs to check for more interesting results. As currently no artist is known that uses GANs to generate artworks conditioned on text, there lies much potential in the field for further exploration of this method.

7 Conclusion

This paper has implemented the conditional *AttnGAN* by Xu et al. [7] for generating artistic images from newspaper descriptions with the goal of presenting the news in a more engaging way. The *AttnGAN* is trained to pay attention to certain regions of the image described by the conditioning description. After experimenting with curating and filtering the input data from WikiArt with different approaches, this paper proposes two possible artistic models: *ArttnGAN2* and *ArttnGAN-F*. *ArttnGAN2* is trained on the images and captions from the Wiki-Art Emotions dataset, including a second caption per image generated by ImageCaptioning codebase from Ruotian Luo [22]. *ArttnGAN-F* is trained on the same dataset, but with lexical words and low-occurrence words filtered out of the captions, to pay more attention to content-dense words. Clustering of the input images was applied in an attempt to make the output more coherent, but the limited size of the dataset did not allow for the optimal training conditions, and results of *ArttnGAN-C* were thus insufficient.

The results of *ArttnGAN2* and *ArttnGAN-F* were analysed by examining individual samples, as well as through an online survey. Both these analyses concluded that even though the models seem to be aware of some textual sentiments, there is inadequate proof that the generated artworks accurately represent the news descriptions they were conditioned on. Nevertheless, a large part of the respondents answered that looking at the artwork made them reflect more about the article than if it had not been accompanied by any image, which satisfies the goal of this paper, i.e. making the news more engaging by using GAN art.

Relating the current paper to the existing artistic field of GAN art, the *AttnGAN* approach is in line with those of other artists, who are mostly engaged with modifying the architectures of GAN models as well as thoroughly curating the in- and outputs of the algorithm.

7.1 Limitations

The current paper has considerable limitations. Due to restricted GPU power, the dataset and number of epochs for training had to remain relatively small. The limited size of the WikiArt Emotions data-set, therefore, created restrictions on the extent to which the GAN could learn the figurative features of the paintings. Furthermore, the limited number of captions per image resulted in attention maps that paid insufficient attention to the content words. The regions of the image that were mapped for attention also appeared to be quite arbitrary. Therefore, the models fail to represent the news descriptions accurately, making it difficult for the audience to associate the artwork to the article.

Secondly, since the dataset consisted of artworks from a wide selection of different styles, the outputs of the models were inconsistent and incoherent. A potential solution for this problem was attempted by clustering the input data in more coherent groups, but the limited size of the dataset did not allow for clusters large enough to produce satisfying results.

Lastly, the generated works are of low resolution (256x256 pixels), and most do not portray surprisingly new or unusual visualisations in comparison to the existing works in the field of GAN art. This might clarify their mediocre scoring for appeal (Tables 4 and 5 in Appendix B).

7.2 Recommendations for future research

Most of the established limitations to the current model are due to the small size of the dataset as well as the limited size of the set of accompanying captions, which poses restrictions to representing the news captions in an accurate and identifiable way. Therefore, there is a lot of potential in expanding this model by training it on a larger and more extensive dataset, which fits the requirements of the original AttnGAN in an artistic way. Additionally, larger GPU space will most likely enable a more accurate output for the models due to the possibility for longer training. Furthermore, focusing the dataset on a more coherent set of images, as was attempted in *ArttnGAN-C* , might enable for a more figurative and comprehensible output.

As there is relatively little to no application or research to generating art from GANs conditioned on text, future research could use the insights of the current model to develop this field further. The same suggestion holds for the generation of images from news descriptions, a subject that has rare-ly been explored if at all. There lies much potential in this field not only for artistic purposes mentioned in the current paper, but also to encourage society to engage with the news in a more reflective and less disconnected way.

References

[1] Nicolas M Anspach. The new personal influence: How our Facebook friends influence the news we read. *Political Communication*, 34(4):590 –606, oct 2017.

[2] Olafur Eliasson. Why art has the power to change the world, 2016.

[3] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mir-za, Bing Xu, David Warde-Farley, Sherjil Oz-air, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 3, pages 2672–2680, jun 2014.

[4] Ian J. Goodfellow. NIPS 2016 Tutorial: Generative Adversarial Networks. 2016.

[5] Mehdi Mirza and Simon Osindero. Conditional Generative Adversarial Nets. nov 2014.

[6] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Ho-nglak Lee. Generative adversarial text to image synthesis. *33rd International Conference on Machine Learning, ICML 2016*, 3: 1681–1690, 2016.

[7] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1316–1324, 2018.

[8] Tobias Hinz, Stefan Heinrich, and Stefan Wermter. Semantic Object Accuracy for Generative Text-to-Image Synthesis. pages 1–21, 2019.

[9] Alaaeldin El-Nouby, Shikhar Sharma, Hannes Schulz, Devon Hjelm, Layla El Asri, Samira Ebrahimi Kahou, Yoshua Bengio, and Graham W. Taylor. Tell, Draw, and Repeat: Generating and Modifying Images Based on Continual Linguistic Instruction. 2018.

[10] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris N. Metaxas. StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks. *IEEE Transactions on Pattern Analysis and Machine In-*

telligence, 41(8):1947–1962, 2019.

[11] Wei Ren Tan, Chee Seng Chan, Hernan E. Aguirre, and Kiyoshi Tanaka. ArtGAN: Artwork synthesis with conditional categorical GANs. *Proceedings - International Conference on Image Processing*, ICIP, 2017-Septe: 3760–3764, 2018.

[12] Jiale Zhi. PixelBrush : Art Generation from text with GANs. *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Sprint 2017*, page 256, 2017.

[13] David Alvarez-Melis and Judith Amores. The Emotional GAN: Priming Adversarial Generation of Art with Emotion. (Nips), 2017.

[14] Aaron Hertzmann. Can computers create art? *Arts*, 7(2):18, 2018.

[15] Marian Mazzone and Ahmed Elgammal. Art, creativity, and the potential of Artificial Intelligence. *Arts*, 8(1):26, 2019.

[16] Aaron Hertzmann. Visual indeterminacy in generative neural art. 2019.

[17] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997.

[18] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for co-mputer vision. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[19] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris N Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 5907– 5915, 2017.

[20] Saif M. Mohammad and Svetlana Kiritchenko. An annotated dataset of emotions evoked by art. In *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, Miy-azaki, Japan, 2018.

[21] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training GANs. *Advances in Neural Information Processing Systems*, pages 2234–2242, 2016.

[22] Ruotian Luo. An image captioning codebase in pytorch. https://github.com/ruotian luo/ImageCaptioning.pytorch, 2017.

[23] Michael Hemphill. A note on adults' color–emotion associations. *The Journal of Genetic Psychology*, 157(3):275–280, 1996. PM ID:8756892.

[24] Herman Cerrato. The meaning of colors. *The Graphic Designer*, 2012.

[25] Helen Papagiannis. Acne studios x robbie barrat, Jan 2020.

[26] Is artificial intelligence set to become art's next medium?, Dec 2018.

[27] Ruby Boddington. "these are important visual moments": artist robbie barrat pushes, tests and breaks ai in his works, Feb 2020.

[28] Tim Schneider and Naomi Rea. Has artificial intelligence given us the next great art movement? experts say slow down, the 'field is in its infancy', Sep 2018.

[29] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks, 2015.

[30] Interview with mario klingemann, Mar 2019.

[31] P. Magli. How things look. The 'Physiognomic Illusion'. *Journal of Art Historiography*, (5), 1, 2011.

All code used in this paper can be found on https://github. com/joycedh/ArttnGAN.

8 Appendices

A Interviews

This appendix contains an overview of the interview survey conducted with GAN artists Mario Klingemann and Terence Broad. The questions were aimed at finding out what the rationale of the artists was behind using GANs for artistic purposes and finding out which methods they use in the process of creating their artworks. Furthermore, their opinions were collected on whether GAN output can be seen as art as well as who the artist of the work is. The last goal of the interview was to find out what the public perception is on GAN art and whether it is valued the way it should.

A.1 Interview with Mario Klingemann

M = Mario Klingemann, I = Interviewer (Joyce den Hertog).

I: Why did you decide to use GANs for the purpose of creating art?

M: I have always been searching for new ways to have machines surprise me and GANs happened to be quite successful at that for a while. Since GANs have become very common now I have reached a point where the surprise factor has reached a plateau and I am on the search again.

I: Which GAN(s) are you using, and did you code it yourself?

M: I am using anything that allows me to explore latent spaces in an intuitive way. I am not exclusively married to GANs, though at least at the moment they are the most convenient tools available. Some of my favorite architectures are pix2pix and StyleGAN, for most of my work in the past years I have used my own architecture and training process which is my own hybrid of different GAN architectures that evolved over time.

I: Do you spend a lot of time curating the in- and output of the GAN?

M: Yes, you have to invest your time in the process, otherwise it’s not art but just a commodity. Every creation is a search and the time spent on a search is in direct relation to the originality of that creation - the longer you search and make creative decisions the less likely it is that you will just discover what others can find, too or have come across already.

I: Are you always content with the output that your algorithm generates?

M: No, I am almost never content. That is the whole point. No matter how good and expressive GANs or other neural architectures get, the ratio of interesting output will always be a tiny percentage of what the models generate. That is simply how interestingness works. Of course the average quality will always go up so that even results that I consider mediocre will work for most people, but most of them do not satisfy me.

I: Do you consider your work as art, why so?

M: Of course I do. It is art because I say it is art. Fortunately enough people including gallerists, curators and collectors already believe me which is how art works - it’s a belief system that is constantly being renegotiated. I: Who do you consider the ‘artist’/creator of your work (i.e. you or the algorithm), why so?

M: I am the artist and creator. The models are my tools, like a painter uses brushes or a pianist plays a piano. GANs and other models are able to produce a whole universe of forms and colors, but so is nature and it is the artist who selects and cuts out a small part of what they have to offer and puts it into a frame. I: How are people generally responding to your work?

M: The average reaction is "Nightmare fuel, reminds me of Francis Bacon." But it looks like most people find it at least interesting in some way, at least the ones that tell me.

I: What is the most interesting response that you have gotten?

M: I am still waiting for it. People seem to be even less surprising than machines in that aspect.

I: Do you think GAN art is valued the way it should (i.e. in terms of money but also artistic/social validation)?

M: I really do not like the term "GAN art" and it is not what I am practicing. Every art form gets exactly the amount of validation that it deserves at a given point in time - it is responsible for its own fate. When it comes to art made with the help of computers it is good that it is getting more and more accepted compared to the typical prejudices that it faced maybe 20 or 30 years ago, but it still has a long way to go to be on equal terms with other art forms in the art market.

A.2 Interview with Terence Broad

T = Terence Broad; I = Interviewer (Joyce den Hertog).

I: Why did you decide to use GANs for the purpose of creating art?

T: I have been experiment with GAN’s for generating art since 2015, their ability to generate new ’unseen’ things fascinates me, and with the amazing developments in their improvement in quality over the past couple of years its a very exciting time to be experimenting with them.

I: Which GAN(s) are you using, and did you code it yourself?

T: So in my work I usually take implementations of existing GANs and then adapt them for my own purposes, either adapting the code for training them to train them in new ways or to alter the way that they perform their forward pass after training to get new and different kinds of effects.

I: Do you spend a lot of time curating the in- and output of the GAN?

T: So unlike a lot of other artists working with GAN’s I don’t make or curate my own datasets. I am currently doing PhD and the focus of my research is how to manipulate GANs (and other generative models) that have already been trained. I do however spend a lot of time curating the output of a GAN, when I’m working I tend to generate thousands of samples, and then handpick a handful of my favorites. I’ve recently done this on a commission for a series of EP artworks for the band 0171 (which will be revealed on the 20th of May I believe, I can send you info to that when the press release is out if it would be useful).

I: Are you always content with the output that your algorithm generates?

T: No, not at all. It usually takes a lot of iterating and tweaking of the algorithms before I get the desired output with my work.

I: Do you consider your work as art, why so?

T: Yes, the way I view my working practice is that there is a lot of skill and tacit knowledge that I have developed over time by working with and experiment with these models, and I think the results reflect that. So I don’t see working with GANs as being much different from the way artists would work with other tools or software packages.

I: Who do you consider the ‘artist’/creator of your work (i.e. you or the algorithm), why so?

T: I would consider myself to be the artist in the creation of my works. Some people are of the opinion that GANs or other AI systems are capable of autonomous generation of art (or at least artifacts), but I think this opinion (especially when I consider my own work) belies to amount of thought and effort that goes into the curation of the data, the choice of algorithms, the curation of the generated output, and the framing and contextualizing of the work that the (human) artist performs.

I: How are people generally responding to your work?

T: I have been getting good response to my work (which always surprises me a bit) especially from people with no technical background, as a lot of my work is often the result of technical research and experimentation I worry that it would only be of interest to people who have a quite a deep understanding of deep learning and generative models, but to my surprise some of my recent work (like (un)stable equilibrium—where I trained GANs from scratch with no training data) I have gotten the most interest from people with non-technical backgrounds.

I: What is the most interesting response that you have gotten?

T: So a funny anecdote is that when I first generated images that would later be used to generate the series of artworks Being Foiled, I showed them to my girlfriend and another woman from my PhD office, and they both absolutely hated them, found the images completely repulsive. So I didn't show it to anyone else for about 3 months because I thought I had made something so horrible. Eventually, I revisited the results and realised that they were very unusual and could be a good series of artworks. I eventually wrote a paper about it and it has been accepted to this years xCoAx conference which you can read here: https://arxiv.org/abs/2002.06890

I: Do you think GAN art is valued the way it should (i.e. in terms of money but also artistic/social validation)?

T: I personally think in some ways there has been too much hype about GAN generated art, and people have been wilfully overselling the power and autonomy of these systems. I do worry that because of this in a few year's time it will be seen as a naff fad and it won't be valued in the future (not that I have ever made much money selling my GAN generated art anyway). But I still believe these systems are amazingly powerful and are here to stay, it's just that he broader understanding of them will change as they become easier to use and more commonplace in all sorts of different ways (i.e being integrated into software like photoshop). I believe there is still a huge amount of potential for people to make interesting work with GANs, it just requires new ways of thinking about these systems to try things that other people aren't doing.

B Online survey

Table 3 describes the questions and structure of the survey, figure 2 shows some examples of the artworks shown in the survey. Tables 4, 5 and figure 21 present the results of the survey.

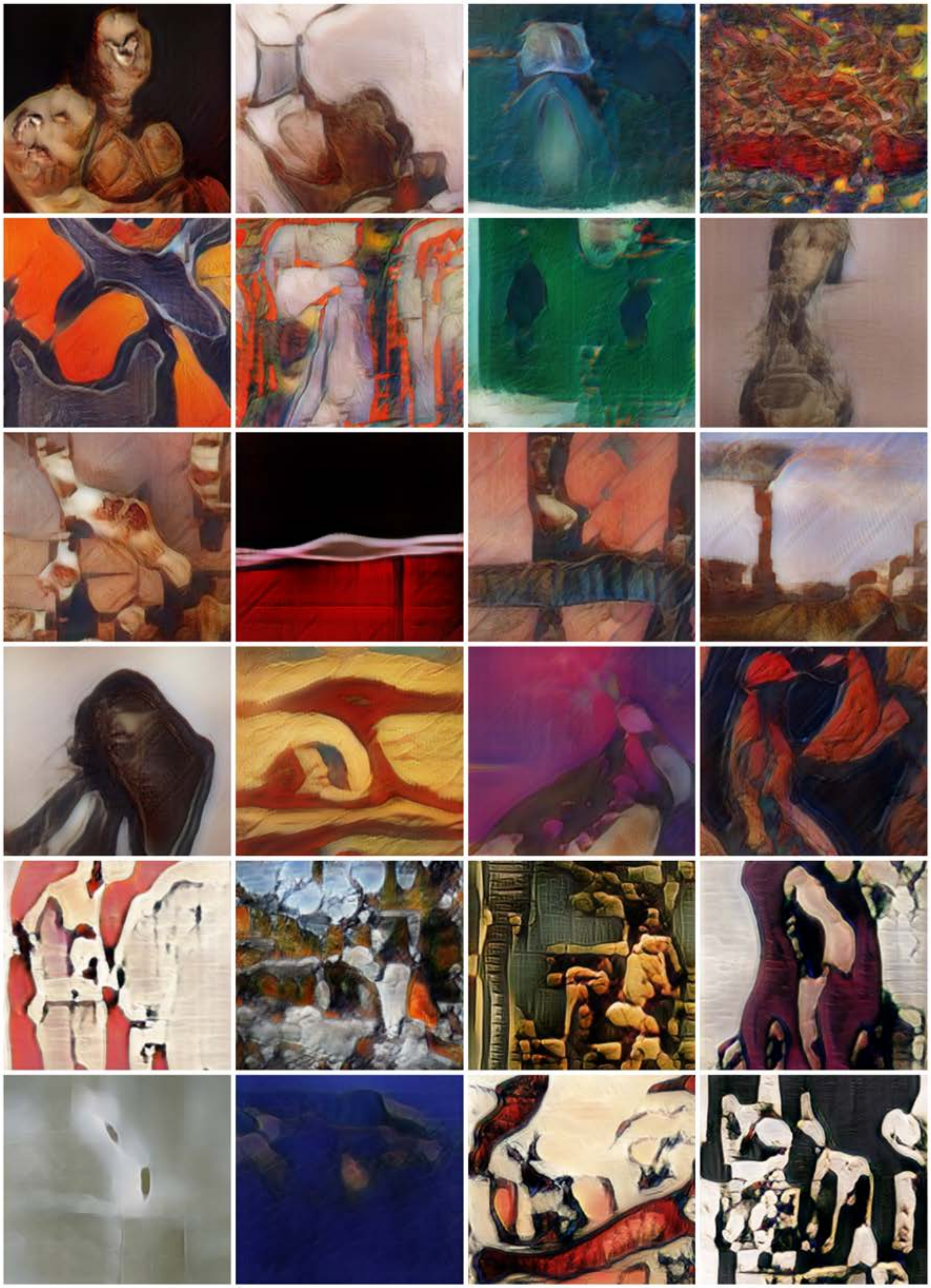


Figure 20: Examples of images shown in the survey. Row 1 & 2: *ArttnGAN2 (unfiltered)*, row 3 & 4: *ArttnGAN2 (filtered)* and row 5 & 6: *ArttnGAN-F*.

This artwork has been generated using *AttnGAN*, conditioned on a news headline. <Read more>about my thesis project. **<a random generated image and the title it was conditioned on>** Please answer the following questions.		
Q1	Do you think the artwork is aesthetically pleasing?	Yes / No / Maybe
Q2	Do you think the artwork portrays the mood of the title?	Yes / No / Maybe
Q3	Does looking at the artwork make you reflect more about the title?	Yes / No / Maybe
Q4	To what degree do you feel the artwork is *aesthetically* related to the title?	(1) Not at all (5) A lot
Q5	To what degree do you feel the artwork is *figuratively* related to the title?	(1) Not at all (5) A lot
Q6	What would you rate artwork (overall) on a scale of 1-5?	(1) I don't like it (5) I really like it
Q7	More comments (elaboration on the questions, comments on the artwork, impressions, questions, etc..	Text input
Refresh this page if you want to view a different work of art!		

Table 3: The questions of the survey.

Measure	**ArttnGAN2, unfiltered** (N = 125)		**ArttnGAN2, filtered** (N = 68)		**Test**		**Effect size**
	Mean	SD	Mean	SD	KW H	p-value	Cohen's d
Aesthetically pleasing	0.50	0.43	0.64	0.45	4.993	0.025	0.330
Mood portrayal	0.55	0.47	0.50	0.45	0.543	0.461	0.107
Enhance reflection	0.65	0.45	0.64	0.45	0.032	0.859	0.017
Aesthetically related to title	2.80	1.20	2.85	1.25	0.084	0.772	0.041
Figuratively related to title	2.67	1.17	2.74	1.30	0.080	0.777	0.052
Overall rate (1-5)	2.85	1.02	3.22	1.18	5.171	0.023	0.341

* KW = Kruskal-Wallis H test; SD = Standard Deviation

Table 4: Difference between the survey answers on *ArttnGAN2* conditioned on unfiltered captions vs. *ArttnGAN2* conditioned on filtered captions.

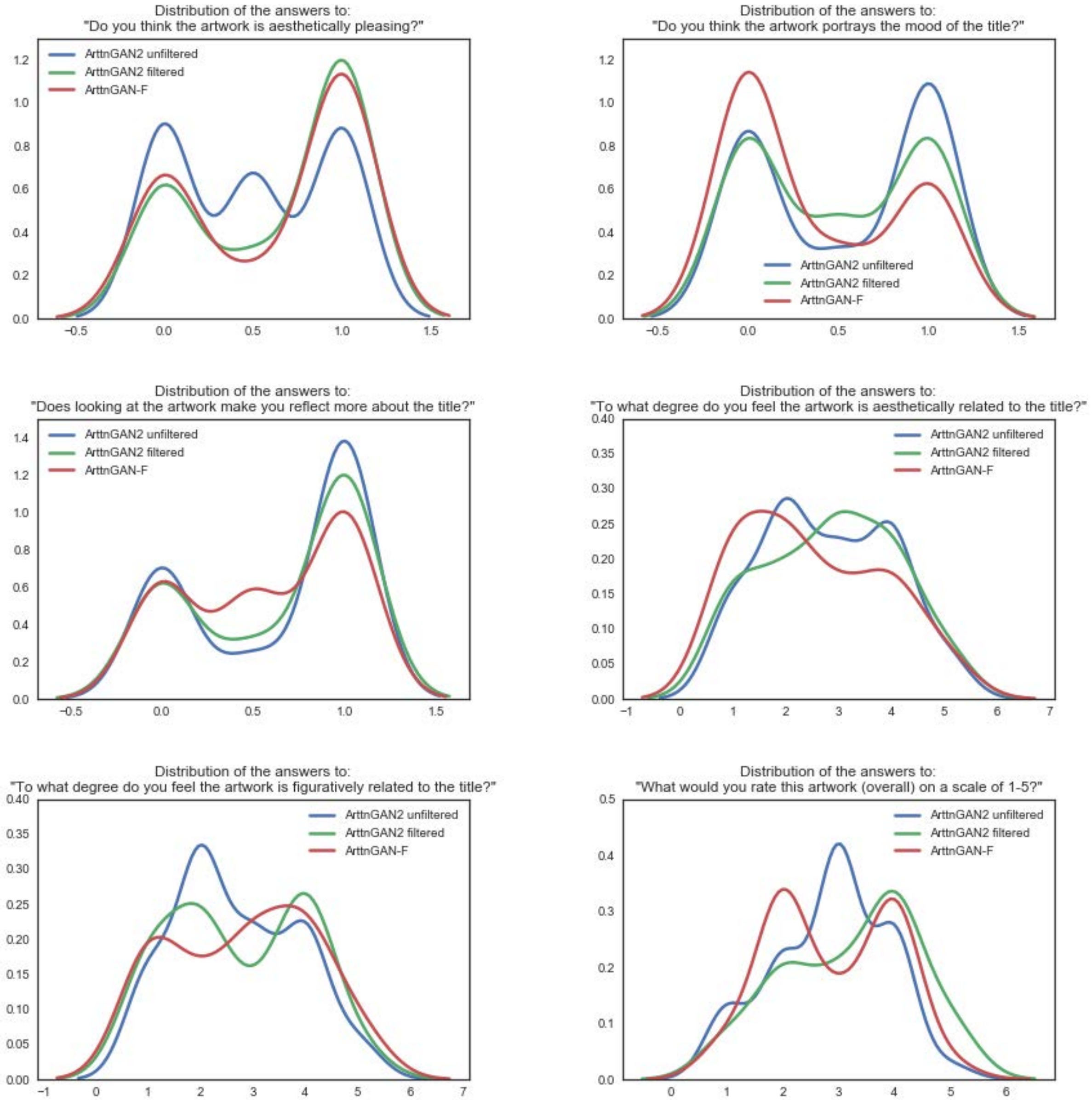


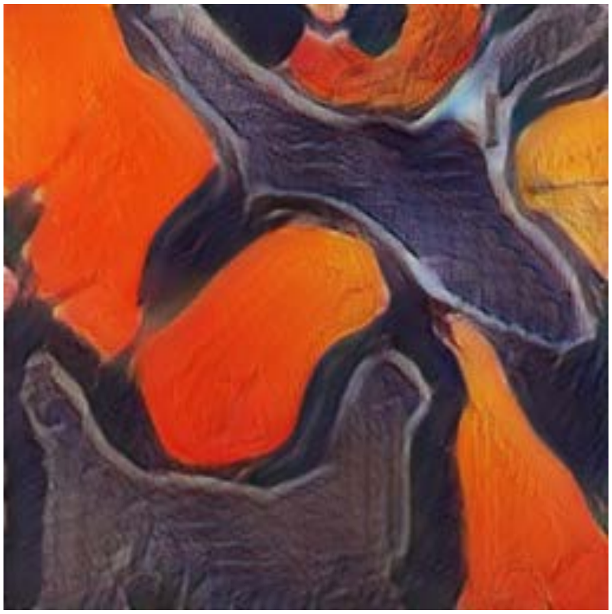
Figure 21: Distribution plots of the answers to the survey questions per model.

Measure	**ArttnGAN2, unfiltered** (N = 125)		**ArttnGAN-F** (N = 63)		**Test**		**Effect size**
	Mean	SD	Mean	SD	KW H*	p-value	Cohen's d
Aesthetically pleasing	0.50	0.43	0.62	0.47	3.478	0.062	0.279
Mood portrayal	0.55	0.47	0.37	0.45	5.926	0.015	0.383
Enhance reflection	0.65	0.45	0.59	0.43	1.227	0.268	0.136
Aesthetically related to title	2.80	1.20	2.52	1.32	2.348	0.125	0.225
Figuratively related to title	2.67	1.17	2.81	1.34	0.475	0.491	0.112
Overall rate (1-5)	2.85	1.02	2.89	1.09	0.025	0.876	0.035

* KW = Kruskal-Wallis H test; SD = Standard Deviation

Table 5: Difference between the survey answers on *ArttnGAN2* conditioned on unfiltered captions vs. *ArttnGAN-F*.

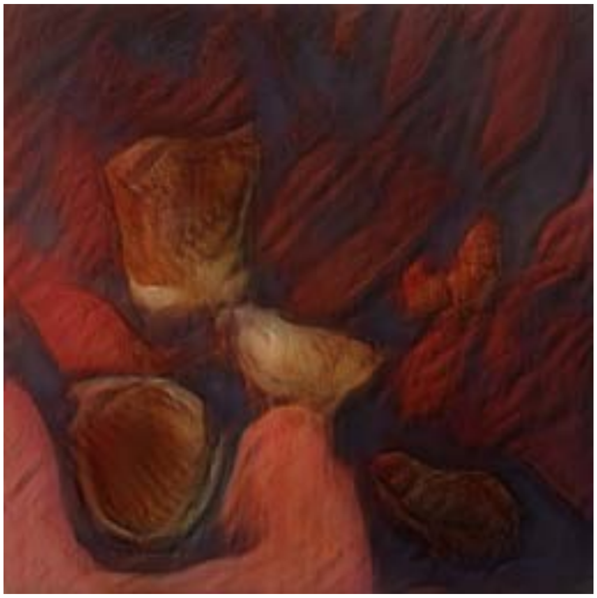
AttnGAN2, unfiltered captions



comment 1	“the artwork makes me think that the weinstein conviction is not the end of the me too movement, there still is more to be done :)"
comment 2	“I don’t really see the connection, and I’m not the biggest fan of the color scheme. Otherwise it’s a cool piece"

Figure 22: Individual comments on the artwork generated from “Harvey Weinstein Is Found Guilty of Sex Crimes in #MeToo Watershed" by *ArttnGAN2* conditioned on the unfiltered news caption.

AttnGAN2, filtered captions



comment 1	“In my opinion, I feel like the artwork is immediately biasing the reader, as the dark red and round shapes are often associated with representations of crime and sex respectively. This bias could on one hand reinforce the will of the writer of the article, or on one hand unconsciously shape the opinion of the reader before reading the whole article. In this case for instance, by reading the title and seeing the artwork, I see more the "sex crime" aspect, related to sentencing Harvey Weinstein, than the "#metoo watershed" aspect, related to shake some aspect of our society (which would be a more optimistic way of seeing things). But that’s my point of view, which is maybe already biased :)"
comment 2	“I think the colours and tone really match the title. Sexual harassment is depicted by a darker red reflecting both pain and (forced) eroticism. (The colour reminds me of the red lights district too)"

Figure 23: Individual comments on the artwork generated from “Harvey Weinstein Is Found Guilty of Sex Crimes in #MeToo Watershed" by *ArttnGAN2* conditioned on the filtered news caption.

Sciences

The early bird catches the bold worm

Individual behavioural differences in the common earthworm (*Lumbricus terrestris*)

Rayne Leroux

Supervisor
Dr. Wouter Halfwerk (VU)
Reader
Dr. Cor Zonneveld (AUC)

Photographer: Jasmin Ronach

Abstract

In recent years, behavioural ecologists have become increasingly focused on animal personalities; individual differences in behaviour that are consistent across time and contexts. The majority of these studies focus on vertebrate species and little attention has been given to invertebrates. In this study, individual behaviour differences will be evaluated in the common earthworm (*Lumbricus terrestris*) by observing their activity levels and boldness as well as evaluating whether body weight is a factor related to these traits. The results indicate interindividual differences in activity and boldness, however, no significant relationship was found between the body weight of the earthworms, their activity levels, or boldness.

Keywords and phrases: *personality, individual variation, boldness, activity, earthworms*

Contents

1 Introduction

Animal personalities have become of increasing interest in behavioural ecology (Sih et al. 2004; Réale et al. 2007; Sih and Bell 2008; Réale et al. 2010); however, research on individual variation in invertebrate species remains limited (Kralj-Fiser and Schuett 2014). In this study, individual variation in the common earthworm (*Lumbricus terrestris*) will be analysed by measuring earthworm responses to a predator risk context and observing their activity levels. Earthworms have a significant impact on soil profiles by modifying their biological, chemical, and physical properties (Edwards and Bohlen 1996), and are as such frequently used in ecological research (Singh et al. 2019). Finding individual variation in the behaviour of *L. terrestris* will be applicable for future experiments and understanding how personality impacts the ecology and evolution of the species.

2 Research Context

The study of animal personality, also known as temperament, can be traced back to the early 20th century; however, it is only in the last few decades that this phenomenon was incorporated into the field of ecology (Sih et al. 2004; Réale et al. 2007; Réale et al. 2010; Kralj-Fiser and Schuett 2014; Ahlgren et al. 2015). It is found that in many species individuals vary in their behaviour from each other and this variation influences how they interact with their environment. Studies on animal personality are mainly based on the measurement of traits, su-ch as sociality or exploration. For instance, a species may exhibit an explorative trait where some individuals are more inclined to explore new areas than others. As a new concept to behavioural ecology, the definition of personality has been extensively debated (Réale et al. 2010; Koski 2011). The current consensus describes animal personalities as individual behavioural differences that are consistent over time and across contexts (Dall et al. 2004; Réale et al. 2007; Réale et al. 2010; Kralj-Fiser and Schuett 2014). In this definition, consistency implies that the differences between individuals will remain similar, but the trait values can change in individuals over time or across situations (Dall et al. 2004). Furthermore, personality does not only involve differences at an individual level, but can also describe differences between families or populations (Hayes and Jenkins 1997; Sih et al. 2004; Réale et al. 2007). The maintenance of a behavioural trait across different environments is called a behavioural carryover (Sih et al. 2004). For instance, an individual that is highly active in an environment with no predators will also show high activity in an area with predators. Behavioral carryover is also used to describe the consistency of a trait over different developmental stages, which is comparable to a trait that is displayed by an individual in two different environments (Réale and Dingmans 2010).

Individual variation in behaviour has often been assumed to correspond to an absence of behavioural plasticity, where plasticity refers to a change in behaviour in response to exposure to stimuli, such as a change in environmental conditions (Sih et al. 2004; Dingemanse et al. 2010; Koski 2011). However, research demonstrates a link between personality and individual plasticity (Koolhaas et al.
1999; Sih and Bell 2008), where highly consistent individuals express a limited part of the phenotypic variation of the population, and less consistent individuals exhibit most of the variation within the population. Therefore, variation in an individual's consistency corresponds to individual variation in plasticity (Bergmuller 2010; Réale and Dingemanse 2010). It should be noted that a highly consistent individual can also exhibit phenotypic plasticity, as the individual's behaviour can still change to adjust to its environment (Réale and Dingemanse 2010). Several theories have been proposed to explain why repetitive behaviour exists. The costs involved in maintaining flexibility are large in terms of energetics and acquiring information about the environment (DeWitt et al. 1998; Dall et al. 2004). Furthermore, environments are rarely predictable, which can lead to an unreliable assessment of cues and can give rise to individuals expressing phenotypes that poorly match their environment (DeWitt et al. 1998).

Réale et al. (2007) proposed five categories for animal traits to measure personality: (1) boldness, which encompasses an individual's reaction to a situation involving risk, such as encountering a predator; (2) exploration, which involves how an individual reacts to a new situation such as a novel object; (3) activity, which concerns the activity level of the individual and can influence the measurements of the two previous categories; (4) aggressiveness, which is where the agnostic response of an individual to their conspecifics is analysed; (5) sociability, which is where individuals may seek or avoid the presence of conspecifics. A phenotypic correlation between two behavioural traits is termed a behavioural syndrome (Gosling 2001; Sih et al. 2004; Krams et al. 2014). Syndromes that occur due to a common mechanism are important to identify because selection on one trait may shape behaviour in other contexts (Sih et al. 2004).

In several species, the aggressive-bold syndrome has been documented whereby bold individuals are more inclined to be aggressive to their conspecifics (Huntingford 1976; Bell 2005; Reaney and Backwell 2007). Boldness has also been associated with exploration and activity where bolder individuals tend to be explorative in novel situations and more active than shy conspecifics (Huntingford 1976; Koolhaas et al. 1999; Bell 2005). The correlation between boldness, activity, and aggression has been formalized as the reactive-proactive axis: a behavi-oural syndrome of the coping strategies of individuals to stressful situations. In this syndrome, proactive individuals are more bold, aggressive, active, explorative, and insensitive to environmental chan-ges in comparison to reactive individuals (Sih et al. 2004; Dal et al. 2004; Dingemanse et al. 2010). The aforementioned correlations coincide with research that has found several evolutionary and ecological consequences of the bold-shy continuum in different species. Studies have shown bold individuals to have more mating opportunities (Reaney and Backwell 2007), higher dispersal ranges (Dingemanse et al. 2003) and higher foraging rates (Ioannou et al. 2008), but also a higher mortality rate due to predation, in comparison to their shy conspecifics (Biro et al. 2004). There has been little research on the compromise of risk from predators for beneficial foraging, dispersal, and reproduction in bold individuals. A study by Ahlgren and others (2015) found that bold individuals can compensate for this by expressing phenotypic traits that reduce the risk of predation. Their results found a strong correlation between the shell shape of the aquatic wandering snail (*Radix balthica*) and their tendency towards risk-taking.

The study of personality differences has been valuable to society in a variety of ways, from improving animal welfare to predicting disease risk in humans (Réale et al. 2007). Furthermore, studying the personality of non-human animals can provide us with a better understanding of the effects of genetics, development, and the environment on human personality and its evolutionary origins (Gosling 2001; Bergmuller 2010). Individual variation in behaviour is commonly distributed in a non-random manner across specific axes (Gosling 2001), suggesting its likeness to have significant consequences to the ecology and evolution of species (Dall et al. 2004; Réale et al. 2007; Killen et al. 2017). One of these consequences is the tendency of a species to be invasive; dispersal plays an important role in the invasiveness of a species and it has been associated with boldness (Dingemanse et al. 2003), aggression, and high activity levels (Rehage and Sih 2004). Therefore, measuring the behavioural traits of an invasive species could be helpful in understanding their dispersion patterns and potential to invade new areas. Personality can also influence how well an individual may respond to a change in the environment. For example, reactive individuals respond better to environmental changes than proactive individuals as they are more sensitive to changes in their environment and approach novel situations with more caution (Sih et al. 2004; Dingemanse et al. 2009). Furthermore, personality can affect the distribution of individuals. For example, individuals with high activity and limited plasticity may be restricted to environments with low predation risk, whilst low activity types could make use of high predation risk areas (Sih et al. 2004). With regards to research on animal behaviour, accounting for personality when conducting research can avoid the issue of generating sampling bias (Biro and Dingemanse 2009). For instance, a study using only individuals that exhibit a particular trait, such as high aggression, would lead to bias results.

There has been a significant increase in the number of publications on animal personalities over the last few decades (Réale et al. 2010), however, the majority of personality studies have been conducted on vertebrates whilst invertebrate species have received little attention (Kralj-Fiser and Schuett 2014). Over 98% of all animal species are invertebrates and they have a wide range of characteristics and behaviours that are rare in vertebrate species, such as asexual reproduction and parasitism (Mather and Logue 2013; Kralj-Fiser and Schuett 2014). Investigating invertebrate personalities is essential to bro-adening our understanding of patterns of individual behavioural differences

(Mather and Logue 2013) and could provide explanations to the ultimate and proximate underpinnings of individual variation in personality where vertebrate studies have not been able to deliver (Kralj-Fiser and Schuett 2014). A literature search by Mather and Logue (2013) found only 32 papers that observed individual differences in invertebrates, including species within the phyla Arthropoda, Nematoda, and Mollusca. Currently, there are no studies on the personality of a species within the phylum Annelida.

Lumbricus terrestris, the common earthworm, is an anecic organism that builds deep vertical burrows in the soil and moves to the soil's surface to feed (Edwards and Bohlen 1996). It has been theorized that the function of their negatively phototactic behaviour is to guide them away from areas that experience strong light to avoid encounters with predators as well as desiccation risks (Sandhu et al. 2018). Earthworms are predated by a wide range of animals including the red fox (*Vulpes vulpes*), the European badger (*Meles meles*), and herring gulls (*Larus argentatus Pontoppidan*) (Catania 2008). Th-ese predators use their vision and olfaction to capture earthworms. Therefore, negative phototaxis may reduce the risk of predation from diurnal spec-ies, such as birds (Sandhu et al. 2018). Another predator of the common earthworm is the mole (*Tal-pidae*) that digs underground to forage and create vibrations that the earthworms detect and respond to, allowing them to escape to the soil's surface (Catania 2008).

Another antipredator behaviour exhibited by *L. terrestris* is the defence mechanism tonic immobility (TI), which is a state of reversible paralysis where the organism appears to be dead and is unresponsive to its surroundings (Ruxton et al. 2004). The behaviour is also known as death-feigning or thanatosis; however, these terms are misleading as animals exhibiting TI often display a position different from dead animals (Honma et al. 2006). Furthermore, TI is a secondary anti-predatory strategy as it occurs after the prey has been detected and physical contact has taken place, whereas thanatosis attempts to avoid initial detection from a predator (Humphreys and Ruxton 2018). Several hypotheses have been proposed for the functionality of this defence strategy, of which three concern the behaviour of the common earthworm. By exhibiting paralysis, the predator may struggle to detect the prey after dropping it or the predator may lose interest (Miyatake et al. 2004); this is particularly successful for evading predators such as birds that are attracted to prey movement (Jones et al. 2007). A second potential function for TI is that a paralysed individual is less likely to be predated than nearby non-paralysed conspecifics. In other words, the attention of the predator is diverted to prey that is not exhibiting TI (Miyataka et al. 2009). Lastly, TI can make the individual appear dead; some predators have an aversion to dead prey, as the assumed death may be related to disease, leading them to avoid consuming the prey (Humphreys and Ruxton 2018).

The consistency of TI within individuals has been associated with metabolism and activity. A study by Krams et al. (2014) observed a population of mealworm beetle larvae (*Tenebrio molitor*) and found that individuals with a higher metabolic rate exhibited a shorter duration of immobility when encountering a predator. A study on two avian species (*Euplectes afer* and *Passer montanus*) reported a negative correlation between the duration of TI and activity levels, where individuals exhibiting shorter durations of TI were more active (Edelaar et al. 20-12). Similarly, a study on the flour beetle (*Tribolium confusum*) found a negative correlation between activity levels and the duration of TI (Nakayama et al. 2010). TI is a suitable behavioural trait for studying boldness (Edelaar et al. 2012), where individuals that do not enter a state of immobility or enter TI for a short duration are considered bold individuals, and individuals that exhibit TI and remain in the state for a longer duration are considered shy individuals.

Studies have shown a link between personality differences and energy metabolism of individuals whereby individuals with a fast-paced life, such as a high metabolism, show high risk-taking behaviour (Réale et al. 2010b; Krams et al. 2014). For example, a study on the adzuki bean beetle (*Callosobruchus chinensis*) reported the duration of TI observed was influenced by body size (Hozumi and Miyataka 2005). Additionally, studies have found bold individuals to express phenotypic traits that reduce predation risk (Ahlgren et al. 2015); in the case of the earthworm, this trait could potentially be its body weight.

In this study, individual behavioural differences will be tested for in *L. terrestris* by observing one of their antipredator responses and activity levels. The earthworms will be tested under two different contexts: inducing the earthworm's TI response to predation using a shake method and observing the activity level of the earthworm. By analysing the results of these experiments, this paper aims to answer the following questions: Does *L. terrestris* exhibit individual variation in TI, TI duration, and activity levels? If so, do the findings support the existence of a bold-activity syndrome, and is there a relationship between the body weight of *L. terrestris* and the behaviours studied?

3 Materials and Methods

Twenty-five *L. terrestris* were obtained from an online bait shop and housed individually in opaque plastic containers filled with compost. To avoid 'motivational states', where the differing hunger level of the earthworms may have a confounding effect on their behaviour, the earthworms were fed ad libitum in a standardized manner to avoid hunger affecting their risk-taking behaviour and activity levels (Koolhaas et al. 1999). The body weight of each earthworm was recorded once before conducting trials with a digital scale ($\pm$0.001g) to analyse the effect of body weight on the behavioural traits boldness and activity.

Trials were conducted in an open field arena with dimensions 24 x 19 x 8 cm filled with 4 cm of soil. To distinguish between activity and exploratory behaviour, this soil was used to imitate the natural environment of the earthworm, meaning that they would exhibit activity rather than exploratory behaviour (which is instead investigated by testing animals in novel settings) (Réale et al. 2007). To measure boldness, a shake stimulus was applied to mimic the attack of a predatory bird. Tonic immobility was induced by seizing the earthworm at its midbody with forceps, shaking the individual side-to-side five times, and dropping it into the arena from 10cm above the soil. In each trial, TI was provoked and the duration of this behaviour was timed using a stopwatch. To measure activity, the locomotor activity of the earthworms was observed individually for five minutes after the boldness test. Activity levels were rated on a scale of 1 to 4 which are defined as: 1 is the lowest activity level and describes that the individual's head and/or tail moved but the body remained in the same starting area; 2 denotes that the individual's body moved but remained partially in the same starting area; 3 describes that the individual moved to a different area of the container; 4 is the highest activity level where the individual moved across $\geq$ 50% of the arena. To measure the consistency of the earthworms' behaviour, four trials were performed on each individual. To avoid habituation occurring, trials were conducted every three days. Repeating the same measures multiple times can lead to individuals habituating to the shake stimuli and their behaviour could become more or less responsive, which may bias the results (Martin and Réale 2008).

Data was collected from all twenty-five earthworms for each trial totalling to 100 observations for each behavioural trait measured. Statistical tests were conducted in R (R Core Team 2017). The intraclass correlation coefficient was used to estimate the consistency of the worm's behaviour; it is the most commonly used statistic to estimate repeatability in animal behaviour (Hayes and Jenkins 1997) and is a good indicator of individual consistency within a population (Réale and Dingemanse 2010). Repeatability is expressed as $r = \frac{S_A^2}{S_A^2+S^2}$ where the variables S_A^2 and S^2 stand for the variance among individuals and the variance within individuals respectively. The estimate ranges between one to zero; with an estimate of one, it is possible to predict an individual's exact behavioural value in future trials, whereas an estimate of zero indicates that it is not possible to make a prediction.

The aims of the experiment were to determine individual variation in exhibiting TI, TI duration, and activity levels in *L. terrestris* and whether there is a relationship between the body weight of individuals and these behaviours. A secondary aim was to determine if a bold-activity syndrome exists in *L. terrestris*.

To test for consistent individual differences in the behaviour of *L. terrestris*, the repeatability estimate of the behavioural measurements was determined. The r package $rptR$ (Stoffel et al. 2017) was used to implement the intraclass correlation coefficient with generalized linear mixed-effects models fitted. To test for consistency of TI duration within individuals a Poisson generalized linear mixed-effects model was fitted. In the model, TI duration was used as the response, body weight as a fixed effect, and the earthworm ID and trial number as random effects. The model was fitted within the $rptR$ function and bootstrapped 100 iterations. A similar approach was conducted to test for the repeatability of exhibiting TI using a binomial generalized linear mixed-effects model. To test for the repeatability of activity in the earthworms, the ordinal scale was converted to binomial data, as ordinal data was not suitable for this analysis. The scale of 1 to 4 was reduced to two categories: low activity (originally activity levels 1 and 2) and high activity (originally activity levels 3 and 4).

To test for a relationship between the earthworm's body weight and TI duration, a Poisson generalized linear mixed-effects model was fitted using the r package $lme4$ (Bates et al. 2015). In the model, TI duration was the response, and the earthworm ID was nested in the trial number and included as a random effect. Body weight was included as a fixed effect and a second model was fitted without body weight as an effect. To test for the significance of body weight on TI duration, the models were compared using ANOVA. A similar approach was conducted to test for the significance of the body weight of the earthworms on exhibiting TI and their activity levels. Binomial generalized linear mixed-effects models were fitted with TI or activity level as the response and the same random and fixed effects previously used.

To determine the existence of a bold-activity syndrome in *L. terrestris*, a correlation between TI duration and activity levels is required. Binomial generalized linear mixed-effects models were fitted with activity levels as the response, the earthworm ID as a random effect with trial number nested, and TI duration as a fixed effect in one of the models. A point-biserial correlation was conducted to find the strength of the correlation between the two traits.

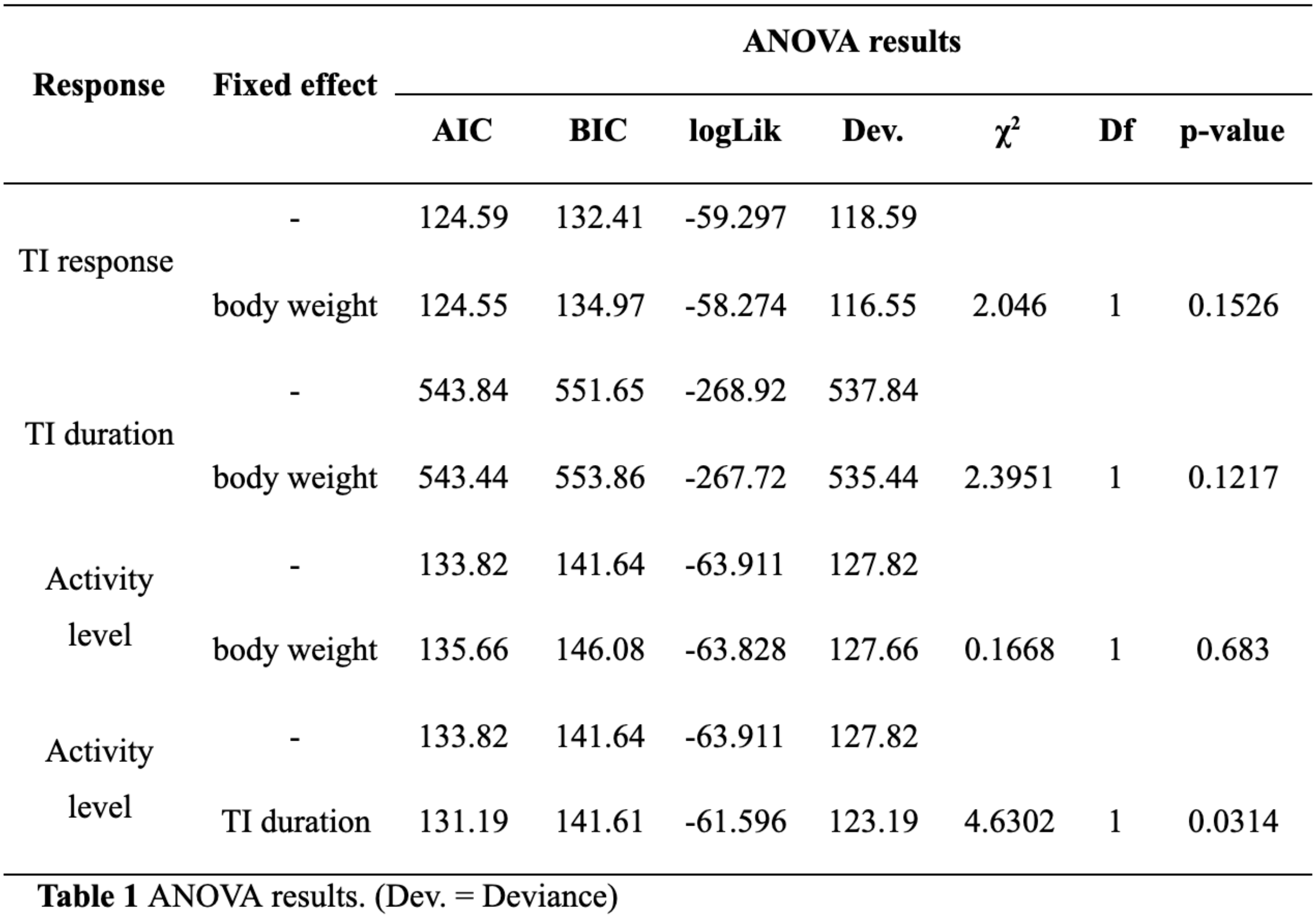
4 Results

Repeatability of individuals exhibiting TI showed 50.8% of total variation attributed to difference among individuals (r = 0.508, standard error = 0.141, confidence interval = [0.22, 0.73], p-value < 0.001). Repeatability for the duration individuals would remain immobile was also high with 52.3% variation among individuals (r = 0.523, standard error = 0.112, confidence interval = [0.228, 0.68], p-value < 0.001). The repeatability for activity level was lowest with 31.8% variation among individuals (r = 0.318, standard error = 0.136, confidence interval = [0.038, 0.529], p-value = 0.00197). The results of the ANO-VA tests (Table 1) indicate that body weight is not a good predictor of exhibiting TI, TI duration, or activity level (p-value 0.121, 0.1526, 0.6832 > significance level 0.05). However, TI duration is a predictor for activity levels (p-value 0.0314 < 0.05). A point-biserial correlation between these two variables indicates a weak correlation (r_{pb}=-0.20601, p-value = 0.0398) where individuals with high activity show shorter durations of TI.

5 Discussion

Personality traits have been reported in numerous species across a wide range of taxa. In this report, personality in *L. terrestris* has been demonstrated by analysing the repeatability of boldness and activity traits. A meta-analysis on the repeatability of behaviour used 759 estimates from 114 studies on vertebrate and invertebrate species to determine a repeatability range of 0.35 $\leq$ r $\leq$ 0.52 with an average of 0.37 (Bell et al. 2009). In comparison to the meta-analysis, the results indicate the presence of a shy-bold axis in *L. terrestris* (repeatability estimate for exhibiting TI: r = 0.508, repeatability estimate for the duration of TI: r = 0.523). These high values for the repeatability of TI could be explained by three factors. Firstly, the repeatability of behaviour in invertebrates has been repor-ted to be higher than vertebrates for some behav-iours, with the meta-analysis showing repeatability values for invertebrates to be closer to the end of the range specified above (Bell et al. 2009). Secondly, research has shown that consistency is higher in individuals when trials are conducted over short intervals in comparison to long intervals (Bell et al. 2009). This study was conducted within three weeks, which is considered short for studies of animal behaviour. Lastly, TI is only one of the antipredator behaviours of *L. terrestris* and consists of two outcomes – the animal will either enter a state of TI or it will move. An experiment examining a different antipredator behaviour such as their response to vibrations caused by foraging moles would have more possible outcomes which could lead to a repeatability estimate for boldness different from the one found by this study.

In this study, shy individuals are individuals that



Response	Fixed effect	ANOVA results						
		AIC	BIC	logLik	Dev.	χ^2	Df	p-value
TI response	-	124.59	132.41	-59.297	118.59			
	body weight	124.55	134.97	-58.274	116.55	2.046	1	0.1526
TI duration	-	543.84	551.65	-268.92	537.84			
	body weight	543.44	553.86	-267.72	535.44	2.3951	1	0.1217
Activity level	-	133.82	141.64	-63.911	127.82			
	body weight	135.66	146.08	-63.828	127.66	0.1668	1	0.683
Activity level	-	133.82	141.64	-63.911	127.82			
	TI duration	131.19	141.61	-61.596	123.19	4.6302	1	0.0314

Table 1 ANOVA results. (Dev. = Deviance)

exhibited TI most frequently and bold individuals are individuals that immediately moved after encountering the stimulus. Shy individuals are more likely to exhibit TI during a predation attack; this behaviour deters the predator to attack further, wh-ich increases the probability of the individual's survival (Miyatake et al. 2004). A review of TI in beetles and moth larvae reported individuals who entered an immobile state were discovered and consumed less frequently by predatory birds than active conspecifics (Steiniger 1936). In contrast, bold individuals that do not exhibit TI may benefit when encountering a slow-paced predator such as some carnivorous insects (Ritter et al. 2016). However, it is not possible to conclude how this behaviour affects the fitness of *L. terrestris* as there are a multitude of behaviours that affect the likelihood of a predator capturing and consuming an individual (Lind and Cresswell 2005). For example, in this study, earthworms remained immobile for 5 seconds on average; it is plausible that this amount of time is sufficient for individuals to survive an attack. However, the actual outcome of an attack depends on the type of predator and their attention span, which differs between species.

The repeatability estimate for activity levels in this study was relatively low (r = 0.318) compared to the meta-analysis (Bell et al. 2009), but is accounted for in the repeatability value range of 0.30 - 0.50 from a study by Réale and colleagues (2007). To avoid the issues arising in the analysis of ordinal data, a future experiment to measure the activity of worms could collect continuous data, such as measuring the overall distance or average speed of the earthworms using motion tracking equipment. In this study, active individuals expressed high activity levels with a fast pace and frequent movement, whereas less active individuals expressed low activity levels with a slow pace and minimal movement. These activity traits of earthworms relate to behaviours in their natural environment; active individuals may have a higher chance of encountering predators in the wild (Killen et al. 2017), however, it is also possible that active individuals dig more burrows which would enable them to escape fossorial predators, such as moles, quicker than less active individuals.

Although all three repeatability estimates are high, there is the possibility that some individuals were more consistent in their behaviour than others (Dall et al. 2004; Bell et al. 2009). Differences in the consistency of an individual's behaviour can affect the repeatability estimate; individuals that are less consistent will reduce the repeatability estimate whilst individuals that are highly consistent will increase the estimate (Réale and Dingemanse 2010). A future study could involve measuring the individual plasticity and the repeatability of boldness and activity of *L. terrestris* using a framework based on the theory of behavioural reaction norms proposed by Dingemanse and others (Dingemanse et al. 2010). This would involve examining the relationship between the anti-predatory response of individuals across an environmental gradient (i.e. different risks of predation).

Studies have established an association between morphological and behavioural traits in species, such as the size of shell relating to boldness in wandering snails (Ahlgren et al. 2015). In this study, no relationship was found between the body weight and boldness of *L. terrestris* or the body weight and activity levels of *L. terrestris*, which suggests body weight is not related to boldness or activity in earthworms. A plausible reason that no relationship was found is due to the small range in body weight of the twenty-five individuals sampled, where 84% of individuals fell within a 3g range (4g - 7g). All of the earthworms sampled were obtained in a single order from a bait shop. Assuming the earthworms are raised in the same environment, it would suggest environmental conditions have less influence on variation in TI and activity than other factors, such as genetics, because there would be no variation in environmental effects between the worms.

As previously mentioned, correlated relationships between behavioural traits such as boldness and aggression have been observed in a wide variety of species. A weak negative correlation (r_{pb} =-0.20601) was found between the activity levels and the TI duration of the earthworms, indicating that some individuals who exhibited high activity levels presented shorter durations of TI. The result corresponds qualitatively with the studies previously presented; however, it cannot be stated that a bold-activity syndrome exists in *L. terrestris* due to an experimental error. During the trials, the observations of activity levels were conducted directly after the stimulus test. Therefore, the behaviour induced by the stimulus may have continued into the observations of activity, meaning that the two observations may not have been entirely independent.

When the results of this study are placed in a broader context, the consequence of individual variation in the boldness and activity of *L. terrestris* on their ecology can be inferred. Earthworms are important ecosystem engineers; they influence their environment through burrowing, producing casti-ngs, and litter fragmentation, all of which affect functions of an ecosystem such as nutrient cycling, soil carbon sequestration, and water infiltration (Si-ngh et al. 2019). The presence of an activity trait in *L. terrestris* has the potential to affect the function rates of an ecosystem: an ecosystem with a substantial portion of low activity types may have lower rates of functions, such as infiltration, than an area with high activity types. This reduced functioning would affect the growth of plants in the soil as well as the distribution and abundance of other soil fauna, having large-scale effects on the entire system.

Earthworms use vibrational cues to detect predatory moles (Catania 2008); it can be assumed that bold individuals would be slower or less likely to attempt to flee to the surface during an encounter with a mole than those less bold. If bold individuals are less responsive to vibrations, then it is plausible they are more tolerant of certain vibration levels than shy individuals. A difference in responses to vibrations could create a species distribution of bolder individuals inhabiting areas with higher levels of vibrations such as locations near wind turbines. Additionally, environments with different predation pressure on earthworms may select for more bold individuals if predators are generally slow pa-ced. As previously stated, boldness and high activity have been associated with proactive coping strategies where individuals are less responsive to changes in the environment. Bold or high activity earthworms may not efficiently adapt to new conditions in comparison to reactive individuals. For example, earthworms are poikilothermic, meaning their activity and metabolism are affected by temperature (Edwards and Bohlen 1996). Therefore, if the temperature of their environment increases, highly active individuals may struggle to adapt their behaviour, increasing their probability of mortality due to starvation or physical exhaustion. The limited plasticity that is often associated with proactive individuals is important to consider in our current climate crisis as global temperatures are predicted to increase and droughts will become more frequent (Singh et al. 2019).

At present, no research has been published on the personality traits of species within the phylum Annelida. The present findings provide supporting evidence for the existence of individual differences in the behaviour of the common earthworm. Individual variation with regards to boldness and activity traits of *L. terrestris* has been identified in this study. No relationship was detected between the body weight and boldness, or the body weight and activity levels of *L. terrestris*. Furthermore, a bold-activity syndrome was not determined in *L. terre-stris*. Whilst these relationships were not detected in this study, it is conceivable that they exist but require different experimentation methods to be identified. The findings from this study are a useful addition to the growing body of research on animal personalities. More specifically, these results are an important step towards addressing the ecological and evolutionary consequences of personality in the common earthworm, a species that has a significant impact on the biological, chemical, and physical properties of the soil on a global scale. This study has given rise to many questions in need of further investigation. Future studies should investigate the interindividual variation of antipredator responses and activity levels in a range of contexts to test the plasticity of these traits in *L. terrestris* and examine the mechanisms behind them.

References

Ahlgren J, Chapman BB, Nilsson PA, Bronmark C. 2015. Individual boldness is linked to protective shell shape in aquatic snails. Biol. Lett. 11: 20150029. doi: 10.1098/rsbl.2015 .0029

Bates D, Machler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using {lm e4}. J. Stat. 67(1): 1-48. doi: 10.18637/jss .v067.i01

Bell AM. 2005. Differences between individuals and populations of three-spined stickleback. J. Evol. Biol. 18: 464–473. doi: 10.1111/j.142 0-9101.2004.00817.x

Bell AM, Hankison SJ, Laskowski KL. 2009. The repeatability of behaviour: a meta-analysis. Anim. Behav. 77:771-783.

Bergmuller R. 2010. Animal personality and behavioural syndromes. In: Kappeler, P, editor. *Animal Behaviour: Evolution and Mechanisms*. Springer-Verlag Berlin Heidelberg. p. 587 - 621.

Biro PA, Abrahams MV, Post JR, Parkinson EA. 2004 Predators select against high growth rates and risk taking behaviour in domestic trout populations. Proc. R. Soc. Lond. B 271: 2233–2237. doi:10.1098 /rspb.2004.2861

Biro PA, Dingemanse NJ. 2009. Sampling bias resulting from animal personality. Trends Ecol. Evol. 24:66-67. doi: 10.1016/j.tree.2008.11 .001

Catania KC. 2008. Worm Grunting, Fiddling, and Charming—Humans Unknowingly Mimic a Predator to Harvest Bait. PloS ONE. 3(10): e3472. doi: 10.1371 /journal.pone.0003472

Dall SRX, Houston AI, McNamara JM. 2004. The behavioral ecology of personality: consistent individual differences from an adaptive perspective. Ecol. Letters. 7:734-739. doi: 10.1111/j.1461-0248.2004 .00618.x

DeWitt TJ, Sih A, Wilson DS. 1998. Costs and limits of phenotypic plasticity. Trends Ecol. Evol. 13:77-81. doi: 10.1016/s0169-5347(97)012 74-3

Dingemanse NJ, Both C, Van Noordwijk AJ, Rutten AL, Drent PJ. 2003. Natal dispersal and personalities in great tits. Proc. R. Soc. Lond. Ser. B. 270: 741–747. doi: 10.1098/rspb. 2002.2300

Dingemanse NJ, Kazem A, Réale D, Wright J. 2010. Behavioural reaction norms: animal personality meets individual plasticity. Trends Ecol. Evol. 25(2): 81–89. doi: 10.1016/j.tree.2009 .07.013

Edelaar P, Serrano D, Carrete M, Blas J, Potti J, Tella JL. 2012. Tonic immobility is a measure of boldness toward predators: an application of Bayesian structural equation modelling. Behav. Ecol. 23:619–626. doi: 10.1093/beheco/ars006

Edwards CA, Bohlen PJ. 1996. Biology and ecology of earthworms, Volume 3. Springer Science & Business Media

Gosling SD. 2001. From mice to men: What can we learn about personality from animal research? Psychol. Bull. 127(1): 45-86. doi: 10.1037/0033-2909.127.1.45

Hayes JP, Jenkins SH. 1997. Individual variation in mammals. J. Mammal. 78(2): 374-293. doi: 10.2307/1382882

Honma A, Oku S, Nishida T. 2006. Adaptive significance of death feigning posture as a specialized inducible defence against gapelim-ited predators. Proc. R. Soc. B. 273(1594): 1631–1636. doi: 10.1098/rspb.2006.3501

Hozumi N, Miyatake T. 2005. Body-size dependent difference in death-feigning behavior of adult Callosobruchus chinensis. J. Insect. Behav. 18:557–566. doi: 10.1007/s10905-005-5612-z

Humphreys R, Ruxton G. 2018 A review of thanatosis (death feigning) as an anti-predator behaviour. Behav. Ecol. Sociobiol. 72(2): 1–16. doi: 10.1007/s00265-017-2436-8

Huntingford FA. 1976. The relationship between anti-predator behaviour and aggression am-ong conspecifics in the three-spined stickleback. Anim. Behav. 24: 245–260. doi: 10.1016/S0003-3472(76)80034-6

Ioannou CC, Payne M, Krause J. 2008 Ecological consequences of the bold–shy continuum: the effect of predator boldness on prey risk. Oecologia. 157: 177–182. doi: 10.1007/s0044 2-008-1058-2

Jones MP, Pierce KE, Ward D. 2007. Avian vision: a review of form and function with special consideration to birds of prey. J. Exot. Pet. Med. 16:69–87. doi: 10.1053/j.jepm.2007.03.012

Killen SS, Calsbeek R, Williams TD. 2017. The Ecology of Exercise: Mechanisms Underlying Individual Variation in Behavior, Activity, and Performance: An Introduction to Symposium. Integr. Comp. Biol. 57(2): 185-194. doi: 10.1093/icb/icx083

Kralj-Fiser S, Schuett W. 2014. Studying personality variation in invertebrates: why bother? Anim. Behav. 91:41-52. doi: 10.1016/j.anbehav.2014.02.016

Krams I, Kivleniece I, Kuusik A, Krama T, Freeberg T, Mänd R, Ljubova S, Rantala MJ, Mänd M. 2014. High Repeatability of Anti-Predator Responses and Resting Metabolic Rate in a Beetle. J. Insect Behav. 27(1): 57–66. doi: 10.1007/s10905-013-9408-2

Koolhaas JM, Korte SM, De Boer SF, Van Der Vegt BJ, Van Reenen CG, Hopster H, De Jong IC, Ruis MAW, Blokhuis HJ. 1999. Coping styles in animals: current status in behavior and stress-physiology. Neurosci. Biobehav. Rev. 23: 925-935. doi: 10.1016/S0149-7634(99) 00026-3

Koski SE. 2011. How to Measure Animal Personality and Why Does It Matter? Integrating the Psychological and Biological Approaches to Animal Personality. In: Inoue-Murayama, M, Kawamura, S, Weiss, A, editors. *From Genes to Animal Behavior Social Structures, Personalities, Communication by Color*. Spr-inger. p. 115-137.

Lind J, Cresswell W. 2005. Determining the fitness consequences of antipredator behaviour. Behav. Ecol. 16(5): 945-956. doi: 10.1093/beheco/ari075

Martin JGA and Réale D. 2008. Temperament, risk assessment and habituation to novelty in eastern chipmunks. *Tamias striatus*. Anim. Behav. 75: 309–318. doi: 10.1016/j.anbehav.2007.05.026

Mather JA, Logue DM. 2013.The bold and the spineless: invertebrate personalities. In: Carere, C, Maestripieri, D, editors. *Animal personalities: Behavior, physiology, and evolution*. Chicago, IL: University of Chicago Press. p. 13-35.

Miyatake T, Katayama K, Takeda Y, Nakashima A, Sugita A, Mizumoto M. 2004. Is deathfeign-ing adaptive? Heritable variation in fitness difference of death-feigning behaviour. Proc. R. Soc. B. 271:2293–2296. doi: 10.1098/rspb.2004.2858

Miyatake T, Nakayama S, Nishi Y, Nakajima S. 2009. Tonically immobilized selfish prey can survive by sacrificing others. Proc. R. Soc. B. Biol. Sci. 276:2763–2767. doi: 10.1098/rspb.2009.0558

Nakayama A, Yusuke N, Takahisa M. 2010. Genetic Correlation Between Behavioural Traits in Relation to Death-Feigning Behaviour. *Population Ecol*. 52(2): 329–335. doi: 10.1007/s10144-009-0188-7

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.o rg

Réale D, Dingemanse NJ. 2010. Personality and individual social specialisation. In: Szekely, T, Moore, AJ, Komdeur, J, editors. *Social Behaviour: Genes, Ecology and Evolution*. Cambridge University Press. p 417 - 441

Réale D, Dingemanse NJ, Kazem AJN, Wright J. 2010. Evolutionary and ecological approaches to the study of personality. Phil. Trans. R. Soc. B. 365: 3937 - 3946. doi: 10.1098/rstb.201 0.0222

Réale D, Garant D, Humphries MM, Bergeron P, Car-eau V, Montiglio PO. 2010b. Personality and the emergence of the pace-of-life syndrome concept at the population level. Phil. Trans. R. Soc. B. 365(1580): 4051-4063. doi: 10. 1098/rstb.2010.0208

Réale D, Simon M. Reader SM, Sol D, McDougall PT, Dingemanse NJ. 2007. Integrating animal temperament within ecology and evolution. Biol. Rev. 82:291-318. doi: 10.1111 /j.1469-185X.2007.00010.x

Reaney LT, Backwell PRY. 2007. Risk-taking behavior predicts aggression and mating success in a fiddler crab. Behav. Ecol. 18: 521–525. doi: 10.1093/beheco/arm014

Rehage JS, Sih A. 2004. Dispersal behavior, boldness, and the link to invasiveness: a comparison of four Gambusia species. Biol. Invasions. 6: 379-391. doi: 10.1023/B:BINV. 0000034618.93140.a5

Ritter C, De Mol F, Richter E, Struck C, Katroschan KU. 2016. Antipredator Behavioral Traits of some Agriotes Wireworms (Coleoptera: Elateridae) and their Potential Implications for Species Identification. J. Insect Behav. 29:214–232. doi: 10.1007/s10905-016-955 5-3

Ruxton GD, Sherratt TN, Speed MP. 2004. Avoiding attack: the evolutionary ecology of crypsis, warning signals and mimicry, Oxford Univ Press. doi: 10.1093/acprof:oso/978019852 8609.001.0001

Sandhu P, Shura O, Murray RL, Guy C. 2018. Worms make risky choices too: the effect of starvation on foraging in the common earthworm (*Lumbricus terrestris*). Can. J. Zool. 96: 1278-1283. doi: 10.1139/cjz-2018-0006

Sih A, Bell AM. 2008. Insights for Behavioral Ecology from Behavioral Syndromes. Adv. Study Behav. 38 227–281. doi: 10.1016/S0065-3454(08)00005-3

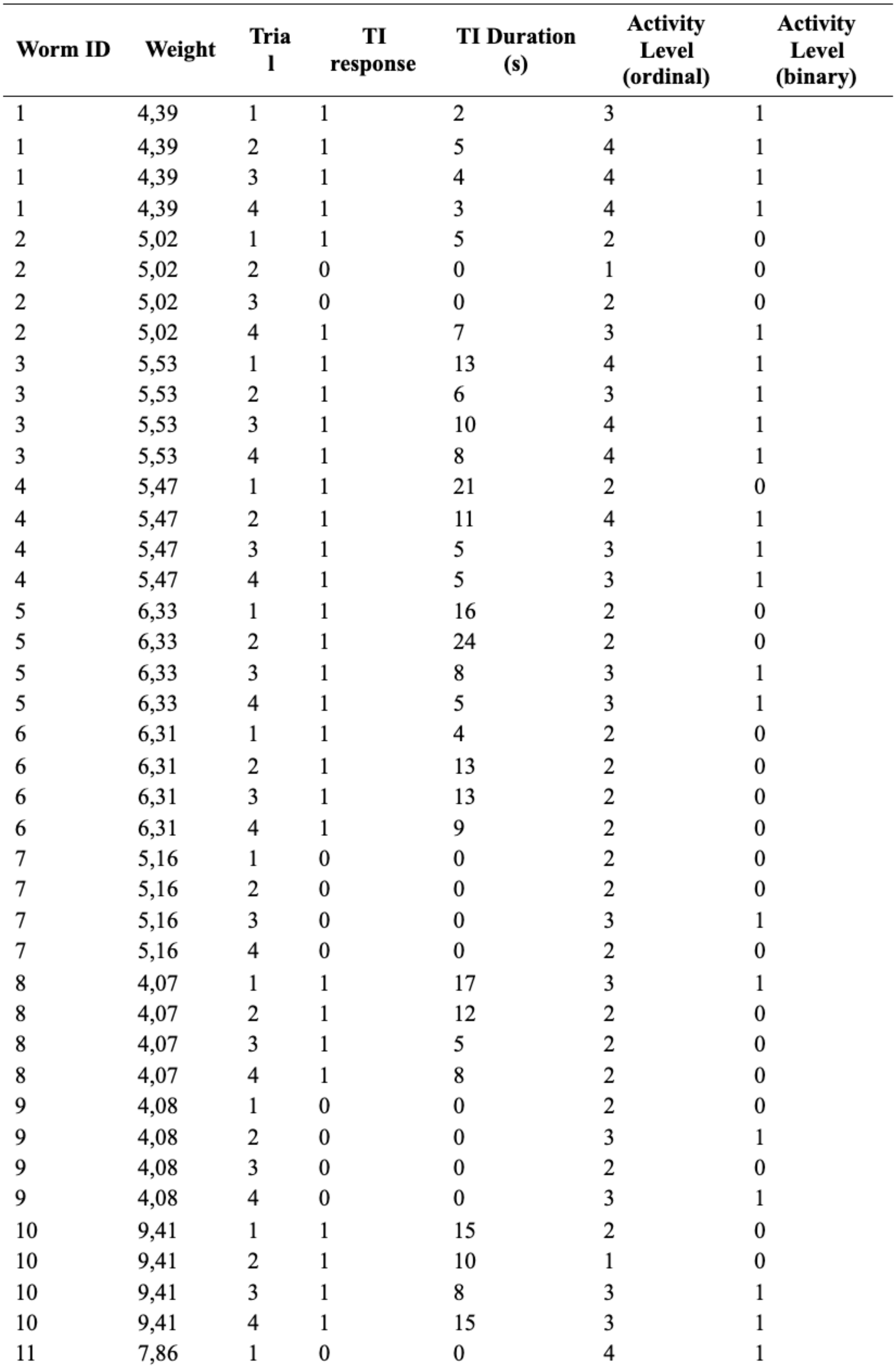
Sih A, Bell AM, Johnson JC. 2004. Behavioral syndromes: an ecological and evolutionary ove-rview. Trends Ecol. Evol. 19(7): 372-378. doi: 10.1016/j.tree.2004.04.009

Singh J, Schädler M, Demetrio W, Brown GG, Eisenhauer N. 2019. Climate change effects on earthworms - a review. *Soil Org*. 91(3):114 – 138. doi:10.25674/so91iss3pp114

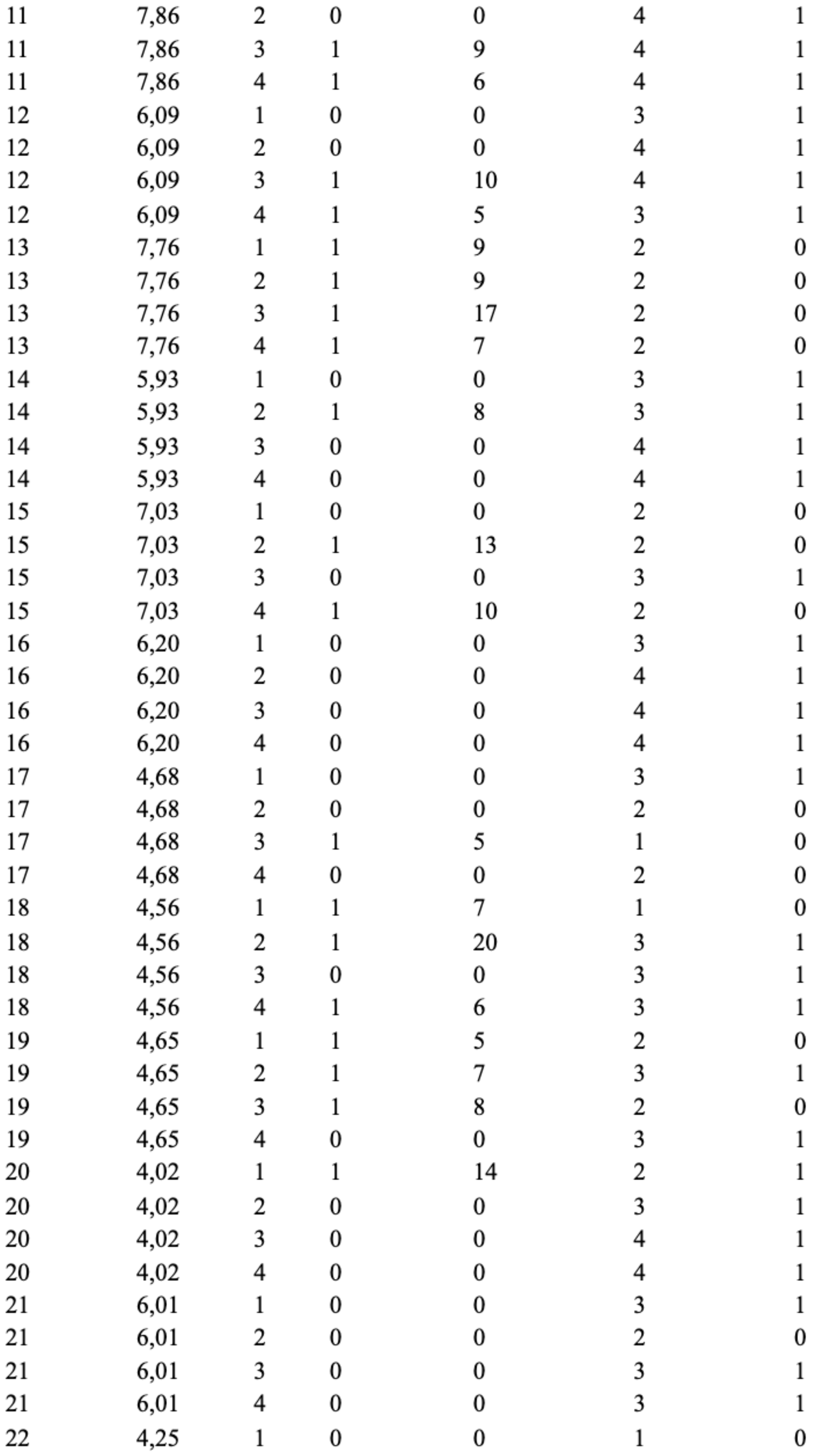
Steiniger F. 1936. Die Biologie der sog. „tierischen Hypnose“. In: Frisch, K, Goldschmidt, R, Ruhland, W, Vogt, W, editors. *Ergebnisse der Biologie*. Springer, Berlin, Heidelberg. p. 348-451

Stoffel MA, Nakagawa S, Schielzeth H. 2017. rptR: Repeatability estimation and variance decomposition by generalized linear mixed-effects models. Methods Ecol. Evol. 8(110): 1639-1644. doi: 10.1111/2041-210X.12797

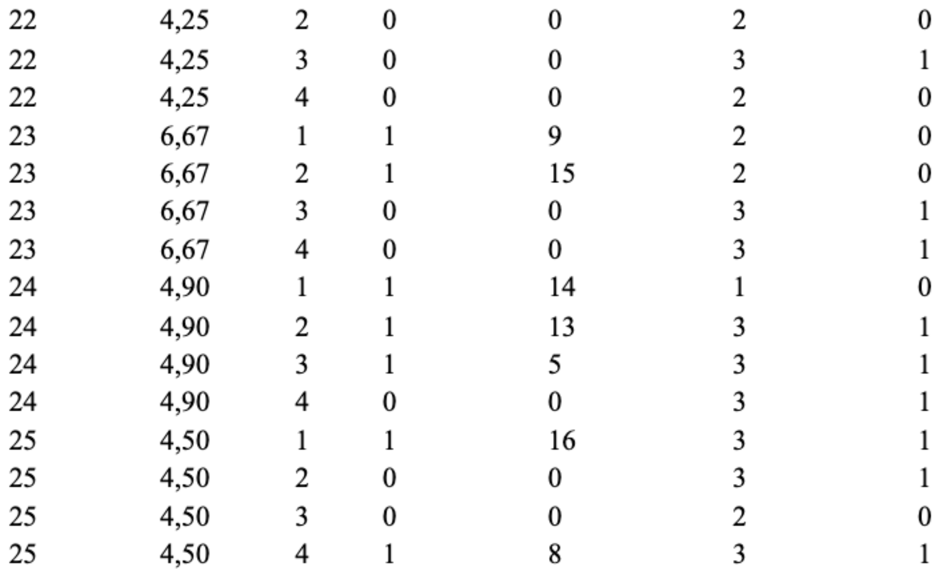
6 Appendix



Worm ID	Weight	Tria l	TI response	TI Duration (s)	Activity Level (ordinal)	Activity Level (binary)
1	4,39	1	1	2	3	1
1	4,39	2	1	5	4	1
1	4,39	3	1	4	4	1
1	4,39	4	1	3	4	1
2	5,02	1	1	5	2	0
2	5,02	2	0	0	1	0
2	5,02	3	0	0	2	0
2	5,02	4	1	7	3	1
3	5,53	1	1	13	4	1
3	5,53	2	1	6	3	1
3	5,53	3	1	10	4	1
3	5,53	4	1	8	4	1
4	5,47	1	1	21	2	0
4	5,47	2	1	11	4	1
4	5,47	3	1	5	3	1
4	5,47	4	1	5	3	1
5	6,33	1	1	16	2	0
5	6,33	2	1	24	2	0
5	6,33	3	1	8	3	1
5	6,33	4	1	5	3	1
6	6,31	1	1	4	2	0
6	6,31	2	1	13	2	0
6	6,31	3	1	13	2	0
6	6,31	4	1	9	2	0
7	5,16	1	0	0	2	0
7	5,16	2	0	0	2	0
7	5,16	3	0	0	3	1
7	5,16	4	0	0	2	0
8	4,07	1	1	17	3	1
8	4,07	2	1	12	2	0
8	4,07	3	1	5	2	0
8	4,07	4	1	8	2	0
9	4,08	1	0	0	2	0
9	4,08	2	0	0	3	1
9	4,08	3	0	0	2	0
9	4,08	4	0	0	3	1
10	9,41	1	1	15	2	0
10	9,41	2	1	10	1	0
10	9,41	3	1	8	3	1
10	9,41	4	1	15	3	1
11	7,86	1	0	0	4	1



11	7,86	2	0	0	4	1
11	7,86	3	1	9	4	1
11	7,86	4	1	6	4	1
12	6,09	1	0	0	3	1
12	6,09	2	0	0	4	1
12	6,09	3	1	10	4	1
12	6,09	4	1	5	3	1
13	7,76	1	1	9	2	0
13	7,76	2	1	9	2	0
13	7,76	3	1	17	2	0
13	7,76	4	1	7	2	0
14	5,93	1	0	0	3	1
14	5,93	2	1	8	3	1
14	5,93	3	0	0	4	1
14	5,93	4	0	0	4	1
15	7,03	1	0	0	2	0
15	7,03	2	1	13	2	0
15	7,03	3	0	0	3	1
15	7,03	4	1	10	2	0
16	6,20	1	0	0	3	1
16	6,20	2	0	0	4	1
16	6,20	3	0	0	4	1
16	6,20	4	0	0	4	1
17	4,68	1	0	0	3	1
17	4,68	2	0	0	2	0
17	4,68	3	1	5	1	0
17	4,68	4	0	0	2	0
18	4,56	1	1	7	1	0
18	4,56	2	1	20	3	1
18	4,56	3	0	0	3	1
18	4,56	4	1	6	3	1
19	4,65	1	1	5	2	0
19	4,65	2	1	7	3	1
19	4,65	3	1	8	2	0
19	4,65	4	0	0	3	1
20	4,02	1	1	14	2	1
20	4,02	2	0	0	3	1
20	4,02	3	0	0	4	1
20	4,02	4	0	0	4	1
21	6,01	1	0	0	3	1
21	6,01	2	0	0	2	0
21	6,01	3	0	0	3	1
21	6,01	4	0	0	3	1
22	4,25	1	0	0	1	0



22	4,25	2	0	0	2	0
22	4,25	3	0	0	3	1
22	4,25	4	0	0	2	0
23	6,67	1	1	9	2	0
23	6,67	2	1	15	2	0
23	6,67	3	0	0	3	1
23	6,67	4	0	0	3	1
24	4,90	1	1	14	1	0
24	4,90	2	1	13	3	1
24	4,90	3	1	5	3	1
24	4,90	4	0	0	3	1
25	4,50	1	1	16	3	1
25	4,50	2	0	0	3	1
25	4,50	3	0	0	2	0
25	4,50	4	1	8	3	1

Social Sciences

In Reconciliation We Trust

Examining the Central Role of ’Trust’ in Reconciliation Efforts

Aisha Erenstein

Supervisor
Dr. Siniša Vuković (AUC)
Reader
Dr. Michael Eze (AUC)

Photographer: Margherita Guida

Abstract

This thesis analyses and examines the importance of trust in reconciliation processes and the misunderstanding of its role in past reconciliation efforts. Drawing from the body of literature on reconciliation, it synthesizes an understanding of the vitality of trust in the reconciliation process. In so doing, it identifies, analyses, and specifically outlines the role of trust as a catalyst for the starting of any reconciliation efforts. It differs from past interpretations of trust, which posited trust merely as one of three pillars in the theories of reconciliation (Rosoux, 2008), by proposing that establishing trust is the necessary precondition for any reconciliation process to begin. Having established this framework, the research then seeks to apply it to the creation and work of reconciliation commissions, examining the role these commissions have in fostering trust, as well as their inherent reliance upon it. The Ethiopian National Reconciliation Commission serves as a demonstrative case study. In analysing the founding proclamation of the Commission, its initial workings, as well as its current and intended role in Ethiopian politics, we see that the lack of trust in this new institution, and the democratic transition heralded by the new government, are fundamentally obstructing the course of reconciliation in Ethiopian society. Although the policy contribution of this analysis is limited to the particular case of the Ethiopian National Reconciliation Commission, the understanding of the role of trust in reconciliation is one that can contribute to the successful establishment and work of future commissions.

Keywords and phrases: *trust, reconciliation, reconciliation commissions, transitional justice, Ethiopia*

On recent developments in Ethiopia

The following paper covers an issue that is current and consistently evolving, and uses a case study which has proven to be very volatile. Whilst I was writing and researching this Capstone in the Spring of 2020, events were already in motion that were changing the very fabric of Ethiopian politics. As we were editing the Capstone for publication, news of tensions boiling over between the local Tigrayan authorities and the Ethiopian government hit headlines. The paper cannot take into account these latest developments, thus some of the situations described may seem outdated given the current local socio-political climate. That being said, the analysis of the background situation and its translation into policy remains relevant. The events of November 2020 are worrying and reflect many of the fears and conclusions presented in the paper. However, they also reiterate the importance of trust when trying to build from a painful past towards an idealistic, radically different future.

Acknowledgements

For his unconditional support, perceptive guidance, kindness and patience, I'd like to thank my supervisor Dr. Siniša Vuković. For his time and enthusiasm, I'd like to thank my reader Dr. Michael Eze. The feedback received from both was heartwarming and reassuring. A thank-you also goes out to Genet Ashebir who was kind enough to follow up on my questions about the development of the ENRC with her own network.

Additional gratitude goes out to my friends, family, committee-colleagues and teammates, all of whom I'm incredibly grateful to have in my life. And last but not least, I'd like to acknowledge the wonderful editorial team behind this Capstone Issue, and the hours of meticulous work that they put into improving these papers – I'm honoured to have written one of the Capstones you chose for this publication.

Contents

Acronyms

ANC	African National Congress.
ENRC	Ethiopian National Reconciliation Commission.
EPRDF	Ethiopian People's Revolutionary Democratic Front.
NP	National Party.
PP	Prosperity Party.
RPF	Rwandan Patriotic Front.
RTT	Red Terror Trials.
SATRC	South African Truth and Reconciliation Commission.
TPLF	Tigrayan People's Liberation Front.

1 Introduction

In 2018, a “bloodless revolution” occurred in the Horn of Africa (Dersso, 2018). In Ethiopia, the first peaceful transition of power in almost a century took place with the election of Prime Minister Abiy Ahmed (Dersso, 2018). PM Abiy was one of the leaders of the ‘reformist camp’ within the Ethiopian People’s Revolutionary Democratic Front (EPRDF), and since his election his regime has led a series of sweeping reforms centred around the rapid political liberalization of the country (Yusuf, 2019b, pg. v). This included the release of over 13,000 political prisoners, the decriminalization of rebellious and exiled groups, the expansion of media freedoms, and the proclamation of peace with Eritrea—the latter winning the PM the 2019 Nobel Peace Prize (Mekonnen, 2019). However, the rapid liberalization process has inadvertently contributed to destabilizing the country, with reports of violent protests erupting along ethnic and territorial lines and estimates of 3 million internally displaced people (Gebreluel, 2019). The transition, although welcomed by many, is uprooting a history of authoritarian rulers and state policing of everyday life. The legacy of the EPRDF’s human rights abuses, the strains on the social fabric caused by ethnic federalism, and the potential “threat” from the predecessor regime (Mekonnen, 2019) weigh heavily upon the new regime led by PM Abiy and the newly-founded Prosperity Party (PP).

A significant portion of the new regime’s reforms target the culture of impunity against opposition upheld by its predecessors, and aim to help the nation recover from the collective traumas suffered as a result of political oppression and state violence (Allo, 2018). This research looks at one specific instrument that is intended to shape and facilitate recovery: the Ethiopian National Reconciliation Commission (henceforth referred to as ‘the Commission’[1], or the Ethiopian National Reconciliation Commission (ENRC)). Established in February 2019 with the adoption of Proclamation No.1102/2018 (henceforth referred to as ‘the Proclamation’), the Commission was intended to guide and harbour a national discussion about the past, bringing “peace justice (sic), national unity and consensus and also Reconciliation among Ethiopian Peoples (sic)” (Article 5 of Proclamation No.1102/2018[2]). It constitutes the first national attempt at transitional justice since the Red Terror Trials (RTT) of the early 1990’s, which had tried the leaders of the Marxist-Leninist militia regime known as the Derg. The RTT’s manner of retributive justice emphasized limited and punitive justice, targeting only the leaders of the Derg for their crimes – it did not target the larger context of suffering the country found itself in following the civil war. The legal framework of the Commission emphasizes the tools of ‘soft governance’ (Yusuf, 2019a, pg. 5) and restorative justice practices in the reconciliation it pursues, allowing the scope of justice it pursues to be broader and more inclusive. However, crucial gaps in the founding proclamation of the Commission continue to hinder its development and thereby the progress towards its goals.

The key issue identified in the Ethiopian case is the lack of trust between the government and its citizens. This absence of trust – and the lack of recognition of this as a major issue – means that the Commission is running the risk of dying whilst still in infancy. In most reconciliation efforts, the facilitating party is the party in government at the time of the process; they, too, are usually those establishing the commissions. Structural measures such as the establishment of a reconciliation commission can help start (re)building trust between parties. However, this is not enough to resolve the conflicts themselves, particularly not in the case of protracted, complex conflicts (Wilmer, 1998, pg. 93; as cited in Rosoux, 2008, pg. 545). This paper hypothesizes that trust is the aspect fundamentally necessary for a reconciliation process to begin, and that without it, such a process is doomed from the start. The definition of ‘trust’ used in the paper is based on a synthesis of understandings (notably those of Rosoux, 2008; Kelman, 2004; Zartman & Kremenyuk, 2000; Gurr, 1996). ‘Trust’ is then linked with the ideas of ‘truth’ and ‘identity’, presented in Rosoux (2008) as the three pillars of reconciliation. This conception will be visually reformulated into a triangular, spectral model.

There is no “one-size-fits-all” approach to matters like reconciliation (African Union, 2019, Section E, pg. 6); thus, no one research project could realistically aim to find the perfect model. The case study for this capstone is also very specific, and finds it-

[1] As was also done in the Commission’s founding document, under Article 3(1).

[2] In citations, should the source be the Proclamation itself, I will be referring directly to the article numbers only.

self in a particularly unique set of circumstances. The ENRC is the result of a hybridized transitional model, as opposed to a negotiated or forced transition (Dersso, 2018). However, there are general models through which the process can be understood, both within academia and within policymaking. The aim of this research is to contribute additional nuance to those models, shed light on what does not work, and provide insight into possible improvements.

2 Research Context

The research at hand finds itself at the nexus between several interlinked fields of study: transitional justice, restorative justice, and reconciliation processes. Reconciliation is an important aspect of restorative justice that serves both as a goal and as a process for the participating community. The reconciliation process includes the establishment of institutions to facilitate the shift away from a culture of conflict (African Union, 2019, pg. 12, 19-20), towards a state of peaceful coexistence and mutual acceptance (Ignatieff, 2003, pg. 326; as cited in Rosoux, 2008, pg. 549). Reconciliation also carries varied notions of scale: it may be relatively small, focused on the agreements made between formerly opposed parties and the relationship between them (Long & Brecke, 2004, pg. 1); or it may carry broader connotations which target the society as a whole (Lederach, 1997). Finally, there are three pillars of reconciliation upon which our understanding of reconciliation is based: the (re)building of trust between parties, the creation of a shared truth by them, and the changes in their conceptions of identity (Rosoux, 2008, pg. 544).

Each pillar has influence over the others, whilst still having its own distinctive features. The identity of a group refers to the social parameters by which a group separates themselves from an 'other'; during (protracted) conflicts this is often done by asserting one's identity as fundamentally unlike the opposition, commonly also acquiring a self-appraising moral tone (Kelman, 1978, pgs. 170-171; Kelman, 1999; both as cited in Rosoux, 2008, pg. 550). Prior research has asserted that the integrity of an identity is often seen as being at risk following a conflict, thus steering parties away from engaging in reconciliation efforts for the sake of protecting their identity (Kelman, 1999; as in Rosoux, 2008, pg. 550). This can be addressed by building a shared 'truth narrative' that is mutually acceptable and does not threaten either party (Asmal et al., 1997, pg. 46). It is crucial for both parties to accept this truth, as it provides a common ground for a future peaceful coexistence. Trust enables these processes to occur, as parties grow to feel safe and secure enough to engage in reconciliation (Govier & Verwoerd, 2002). It can take many forms, but trust requires the mutual acknowledgement of suffering and responsibility, although the exact nature of the actions that foster trust will likely be unique to the reconciliation context (Magarditsch, 2005, pg. 172; as cited in Rosoux, 2008, pg. 552).

Reconciliation processes can focus on different aspects of these pillars, as well as different aspects in the approaches to the process. Relationship-based approaches place the private realm at the center, and attempt to form a bridge between the personal experiences of those involved and peaceful coexistence (Wilson, 2001). Issue-based approaches emphasise systematic changes, thus operating on a macroscopic level to create preventative measures on an institutional level against a return to the state of conflict or oppression (Rosoux, 2008, pg. 544). The reconciliation commission mechanism is a way in which societies can effectively engage in aspects of both approaches (Wilson, 2001). The role of a commission is to create an official, recognized platform through which people can address past trauma in a structured, 'forward-looking' form (Bolocan, 2004, pg. 396; Zartman & Kremenyuk, 2005, pg. 294). These commissions typically focus on one of the three aforementioned pillars of reconciliation: truth (such as the South African Truth & Reconciliation Process), identity (such as the Rwandan National Unity and Reconciliation Commission), or trust. In looking at past research and commissions, we see that the understanding of trust in the process is shallow in comparison to that of truth and identity. Deepening this understanding of trust is thus the aim of the present research. This will be done by examining the literature, and then applying these to the case study.

Regarding the ENRC, it is important to note that all the political reforms in Ethiopia have been introduced fairly recently, starting in 2018. As such, there is relatively little academic literature that bases itself upon the current socio-political changes ongoing in the country, and their corresponding applications in government. This is especially true for the Commission, as this particular effort is a little over a year old, and has been notably quiet in its development. The latter is an issue in and of itself that will be discussed later in this paper. Because of the contemporary nature of the subject, I will be using a mixture of academic literature concerning the relevant theories, as well as news sources and articles to develop an understanding of the Commission. Through this I hope to be able to create an accurate and current picture of the case being studied, whilst still being able to scrutinize and analyse it with the rigour befitting of academia.

The context of the ENRC is vital to understanding its potential in reconciliation and in scope. Not constituting the first transitional justice mechanism the country has engaged with, it follows the Red Terror Trials which punished the leaders of the Derg regime of the 1970's and 1980's for the atrocities committed during the Qey Shibir (Tadesse, 2007). However, these trials were a notable 'victor's justice', meaning that the crimes and atrocities perpetrated by the rebel forces of what became the EPRDF were left unaddressed. The EPRDF's regime also became known for ruling with an iron fist after the Derg's downfall, as reports of political, civil and human rights violations were common – yet the accusations themselves remained unaddressed.

The election of reformist PM Abiy Ahmed in early 2018 saw violence erupt along ethnic lines as sweeping liberalising reforms were implemented (Yusuf, 2019a, pg. 2). Following their victory in the civil war, the EPRDF institutionalised their delocalised guerilla administration into a system of ethnic federalism, in which the country was split up into semi self-governing regions based on ethnicity (Yusuf, 2019b, pg. v; Gebreluel, 2019). Amongst these reforms was the desire to reconcile the separate groups under a unified national identity, which the ENRC will theoretically contribute to by opening up a national discussion of the past. The new regime hopes that through the ENRC, Ethiopian society can finally address social traumas, foster new values and reconcile across the ethnic divides through national discussion (Yusuf, 2019b, pg. 36). The Commission was established with the passing of Proclamation No. 1102/2018 by Parliament, which outlines a commission whose chief objective is "to maintain peace justice (sic), national unity and consensus and also Reconciliation among Ethiopian Peoples (sic)" (Article 5). However, the Commission's structure is uneven, particularly given its socio-political context, a position echoed by the analyses of specialists (e.g. Yusuf, 2019a; Yusuf, 2019b; Mekonnen, 2019; Allo, 2018; Gebreluel, 2019) and members of the Commission (e.g. Dersso, 2019a; Dersso, 2019b; Dersso, 2018). Building upon this previous body of work and examining the case through a trust-focused lens will highlight key issues, and give insight into previously unexplored potential methods to address them.

3 Methodology

The present research used a combination of methods to develop upon pre-existing theories and apply them to real life scenarios. This was done through a series of literature reviews which built upon each other, and an analysis of the sum of their findings which were then applied to the specific case study of the ENRC.

3.1 Chapter 1 – Reconciliation: Theory

This first chapter dives into the theory of reconciliation through a literature review. In doing so, it examines key concepts and understandings related to reconciliation, as various definitions of reconciliation are consulted.[3] This literature review synthesizes the key 'pillars' of reconciliation as identified by Rosoux (2008) – the creation of a shared truth, changes in identity, and the rebuilding of trust – which are codified into a triangular model that will be used to help visually and theoretically contextualise various reconciliation attempts. Sources delving deeper into the three pillars were found on the basis of snowball sampling, starting at the initial Rosoux (2008) paper. Each of the three pillars will also be critically examined and evaluated in relation to the concept of 'trust'; thereby arriving at the conclusion that trust plays the linchpin role for the three in terms of the starting of reconciliation processes. This conclusion is the central thesis of this research, and will be used as the primary analytical lens throughout the rest of the paper.

[3] Of particular note here are Long & Brecke's 2004 definition of reconciliation on the basis of peace agreements and Lederach's 1997 definition of wide-scale, 'transcendent' reconciliation.

3.2 Chapter 2 – Reconciliation: Practice

The second chapter builds on the first chapter's literature review, but rather focuses on the practical approaches to the reconciliation process, placing a particular emphasis on the commission as a reconciliation mechanism. Once again using snowballing in compiling a review, this chapter highlights the important distinctions between relationship-based reconciliation efforts and issue-based reconciliation efforts, as well as top-down (Bar-On, 1996) vs. bottom-up (Bargal & Sivan, 2004) approaches. Then it follows to introduce the commission as a reconciliation mechanism, and its role in shaping the nature of a reconciliation process as a bridge between the personal and the public as both attempt recovery (Wilkinson, 2001). Two examples of notable African commissions – the Rwandan Gacaca Courts and the South African Truth and Reconciliation Commission – are also analysed to establish points of comparison for the case study.

3.3 Chapter 3 – Case study: The Ethiopian National Reconciliation Commission

The third and final chapter presents the demonstrative case study through which the theory developed in Chapter 1 can be applied, thus following a logic of confirmation. It begins by establishing the background and socio-political context of the ENRC through the consultation of local scholarly work and news sources. Then a thorough examination of local (e.g. Dersso, 2018; Borkena & Ezega articles[4]; Mekonnen, 2019), governmental (e.g. Addisu, 2020) and international (e.g. Allo, 2018; Dersso, 2019a; Dersso, 2019b; Gebreluel, 2019; Yusuf, 2019a; Yusuf, 2019b) media coverage of the commission is used to create as accurate as possible a picture of the current status and workings of the Commission. The works of Dersso (2019a, 2019b) are given particular weight in this analysis given that they were written after his appointment to the Commission.

However, it must be noted that the resources available were limited, and although the consultation of available sources was exhaustive, it was restricted to what was available online. This research was further constrained by my inability to understand Amharic and the other Ethiopian languages, which narrowed down the accessible literature to texts written or translated in English[5]. The examination of secondary sources regarding the Commission was supplemented by an in-depth analysis of the founding Proclamation itself, as the Articles are used to create an understanding of the potential future workings of the Commission – which is then analysed through the lens of trust as formulated in Chapter 1, in a form of tentative, putative testing.

4 Discussion/Analysis

4.1 CHAPTER 1 – Reconciliation: Theory

4.1.1 Reconciliation processes

Reconciliation processes are an instrument in the tool belt of restorative justice, resulting from the cooperation and collaboration of individuals and their communities in finding a way to live side by side following a socially traumatic moment in history (African Union, 2019, pg. 12). Both a goal and process, the reconciliation process – if and when successful – leads to a state of reconciliation. On a national level, this goal can be seen in national unity, and the process is reflected in the actions of, for example, reconciliation commissions. In the process institutions are built to help reinstate the rule of law and foster a culture of respect for human rights (African Union, 2019, pgs. 12, 19-20). The political goal of reconciliation is to reach a state of 'acceptance' regarding the state of affairs (Ignatieff, 2003, pg. 326; as cited in Rosoux, 2008, pg. 549). Traditionally not understood to be a part of state business or politics, the reconciliation process is now recognized as vital to transitioning states; for some it is even recognized as "probably the most important condition for maintaining a stable peace" (Bar-Siman-Tov, 2000, pg. 237; as cited in Rosoux, 2008, pg. 543).

[4] Shortened for ease of reading. Full citations are as follows: Ethiopia gets National Reconciliation Commission legislation, 2018; Ethiopia named members of National Reconciliation Commission, 2019; Ethiopian Reconciliation Commission Announces Three-Year Plan, 2019.

[5] I did informally follow this up with a family friend who worked in the Ethiopian media industry for decades, to see if there was a significant difference between the work available between languages. She told me she had not heard anything beyond the news of the Commission's establishment, but had asked her network of journalists, media-professionals and acquaintances-—who knew similarly little.

Reconciliation assumes a prior state of conflict and/or trauma, and the existence of at least two opposing parties. However, as the Ethiopian case demonstrates, it can also be understood in the context of more complicated situations such as protracted conflicts, decolonisation processes, or the presence of international third parties. Subsequently, definitions range from establishing reconciliation as a "mutually conciliatory accommodation between former protagonists" (Long & Brecke, 2003, pg. 1) to defining reconciliation as a more "transcendent", arduous process targeting the society as a whole, entrenched in restorative practices (Lederach, 1997). In light of this variation, I have taken to understanding it as a spectrum between, and including, these two definitions, with various reconciliation efforts taking different points of emphasis (Figure 1 provides a rough visualization of this).

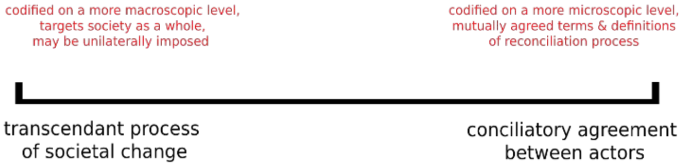

codified on a more macroscopic level,
targets society as a whole,
may be unilaterally imposed
codified on a more microscopic level,
mutually agreed terms & definitions
of reconciliation process
transcendant process
of societal change
conciliatory agreement
between actors

Figure 1: A spectrum of reconciliation definitions

These points of emphasis have also been more traditionally construed as the 'pillars' of reconciliation: the (re)building of trust between actors, the creation of a shared truth, and the changes in conceptions of identity (Rosoux, 2008, pg. 544). That is to say that although they are all important to reconciliation processes, some are given a more central role in the process of a given case or framework. The model below (Figure 2) illustrates this spectral model of understanding reconciliation. However, here it is developed from a mere linear spectrum to a triangular one, where each pillar forms a leg of the triangle. This allows the focus to lie on the aforementioned three key pillars of the process, and their relations to each other. Because the model forms a connected, cohesive shape, it embodies the way in which these aspects encapsulate any (successful) reconciliation process. Specific cases can be placed within the triangle to help visualize their priorities; an ideal commission would thus find itself fairly central within the triangle, perhaps leaning towards the necessary emphasis for its case.

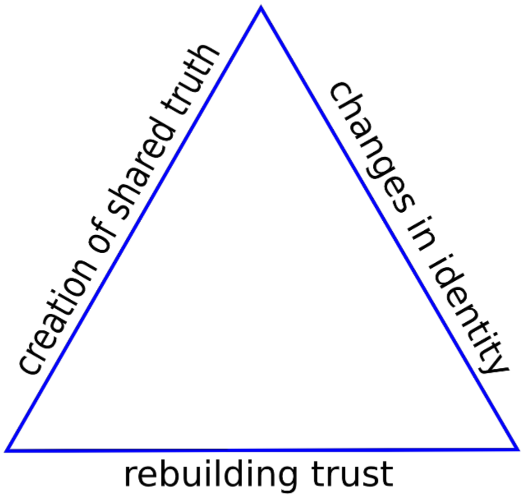


Figure 2: A triangular model of reconciliation

First we have the pillar of 'identity'. In cases of social trauma, parties will often hold onto their perception of their identity as integral to their being; engaging in reconciliation or even negotiations with the 'other' would require sacrificing the integrity of this identity (Kelman, 1999). Thus, reconciliation efforts engage in shifting the psychology around that identity so that following its reformulation the basis of the identity changes and it is no longer threatened by coexisting with the other (Bar-Tal & Bennink, 2004, as cited in Rosoux, 2004, pg. 544; Kelman, 2004).

On the left-hand side of 'identity' we find the pillar of 'truth'. Truth-focused reconciliation efforts work towards building a unitary understanding of the history within the transitioning society, attempting to "harmonize incommensurable world views" (Asmal et al., 1997, pg. 46). Through discussion, the various contradictory understandings of the past can eventually, "progressively" become understood in a cohesive fashion – but this does take time and work (Basalou & Baxter, 2007; as cited in Rosoux, 2008, pg. 551). For some, it simply creates "a single universe of comprehensibility" (Asmal et al., 1997, pg. 46) for future conflicts; for others, it is in this shared truth that one can find the basis for a lasting peace.

Last but not least, in the 'trust' pillar of the triangle, we find understandings and reconciliation efforts that emphasize the rebuilding of trust between parties following the historical trauma (e.g. Govier & Verwoerd, 2002; Amstutz, 2005; Nadler

& Liviatan, 2006; Marrow, 1999 on establishing friendships to bridge the gap; as cited in Rosoux, 2008, pg. 544). This could be between parties that are enemies in a protracted conflict, for example, or between a people and their new government following a change of regime. For reconciliation to occur in such scenarios, some form of trust must be established to overcome the "traditional split" between the parties; the reconciliation process would then build on – and if successful – strengthen that trust (Marrow, 1999, pg. 132; as cited in Rosoux, 2004, pg. 544).

Key to understanding these concepts is realizing that they are interconnected—all three are ultimately needed for a successful reconciliation process to occur. Accepting an altered truth from your own, requires a party of a particular identity to be able to trust their surroundings. It also requires the basis of that identity to be built on more than just a single truth. Accepting an altered truth will also inherently shift the boundaries of an identity, as parameters adapt to fit these new contexts. Similarly, engaging with other parties whose identities at some point were opposed to yours requires an implicit acknowledgement of the possibility of there being multiple truths. The strong connection between trust and the other pillars secures its role as a catalyst for the overall reconciliation process.

4.1.2 The power of trust

Trust comes from a place of perceived security; such security can be found if one's identity is no longer under constant threat by the existence of an 'other'. It is vital that the crimes committed are separated from those who committed them to allow for a basis upon which trust can be built (Montville, 2001, pg. 132; Bar-Tal, 2000; as cited in Rosoux, 2008, pg. 550) so as to overcome the "traditional split" between the parties (Marrow, 1999, pg. 132; as cited in Rosoux, 2004, pg. 544). That being said, trust is also subject to great influence from outside the reconciliatory triumvirate. The issue of timing is crucial to matters of 'trust' in a manner that cascades onto 'truth' and 'identity'. In other words, reconciliation efforts need to be well-timed, and the society needs to desire a state of reconciliation before engaging in the process[6] or participating in 'reconciliatory events'. The contested nature of a perceived truth following a conflict is why some argue that leaders should demonstrate a "partial amnesia" right after a transition (Rosoux, 2008, pg. 550; also reflected in Bargal & Sivan, 2004, pg. 144). 'Partial amnesia' means that a blind eye is turned to certain minor offences, whilst the largest crimes are prosecuted - thus demonstrating that justice is being served, but also setting priorities for a new normal (Krondorfer, 1995; as cited in Rosoux, 2008, pg. 550). If the timing is wrong, engaging in reconciliation could be perceived as making a mockery of the victims' suffering, and falling under 'too little, too late' (Rosoux, 2008, pg. 552). The verdict would be largely dependent on the public's trust in those leading the society after the transition, the role of leadership in the transition, and the (perceived) handling of the power associated with it. A distrusted leader and/or regime is seen as manipulative; a trusted leader and/or regime is visionary, leading their people in learning "how to remember *and* forget, in order to move forward" (Garton Ash, 2003, pg. 415; as cited in Rosoux, 2008, pg. 551-552).

The process inherently requires the shifting of 'identity markers' of the parties involved, as the understanding of the 'self' and that of the 'other' are reassessed and redefined. In the state of conflict, this definition is dependent on the relationship of 'the self' to the 'other', and reinforced by a state of perpetual threat (Kelman, 1978, pgs. 170-171; as cited in Rosoux, 2008, pg. 550). A conflicted state is important because it informs a tendency towards self-sympathy, where the party viewed as 'the self' is inherently considered universally 'good', and their testimony 'trustworthy' (Kelman, 2004, pg. 121). This polarized perception becomes intrinsic to the dignity of the party, such that the conception of the identity of 'the self' within the context of the prior conflict is the only information that can be trusted (Gurr, 1996, pg. 63). Reality, however, is of course much more dappled, and coming to terms with it requires accepting a diverse, more holistic view of the situation. It requires altering the perception of 'the self' and 'the other' so that they no longer depend on something as volatile as another group of people. With a more stable basis of identity, the parties will then be able to engage in 'forward-looking' processes (Zartman & Kremenyuk, 2005), where collaboration and empathy can be used as mutual aides in building trust between them. However, to be able to initiate such

[6] Similar to the concept of 'ripeness' in conflict-resolution efforts (Zartman, 2000).

shifts in identity, there must be some form of inherent trust in the stability of the party's position, societal context, and relative security in the present and the future.

This trust in the preservation of the integrity of group identity is significantly informed by the respect afforded to the 'truths' considered integral to the identity. Arguably the most well-known of the 'pillars of reconciliation', the basis of unitary understanding developed by truth-telling practices is based on trust between the party giving testimony and the party entrusted with it (Asmal et al., 1997, pg. 46). This created truth will inherently be a compromise, somewhat incomplete to most parties involved – however, it needs to be acceptable to the greater majority. In understanding 'truth' in a reconciliation context, we also need to acknowledge the different types of truth (viz., factual, personal and official) and how they relate to one another. Factual truth, in particular, can often be conflated with the personal and official; rather, it's helpful to interpret it as the 'skeleton' of the 'body' of the 'truth'. It's made of concrete facts, such as events, whose occurrence cannot be changed (Ricoeur, 2000, pg. 496; as cited in Rosoux, 2008, pg. 556). However, the meaning of these events is ever-changing, and it is in the interpretation of events that one finds subjectivity. Subjectivity is not antagonistic to the 'truth'; if given the space, varying intersubjective truths can coexist—such as the personal truths entrusted to the truth-telling process, and the official truths into which they are molded. Thus, in pursuing 'the truth' in reconciliation, we need to understand that the aim should be to find "an agreed description of the basic factual landscape of the past [which is the] factual framework within which the vital healthy and unending battle of interpretations must go on" (Ash, 2003, pgs. 416-417; as cited in Rosoux, 2008, pg. 556). However, the aim is also to ensure that these "unending battles" do not compromise the trust placed in the authority figures who control the official narrative (Ash, 2003, pgs. 416-417; as cited in Rosoux, 2008, pg. 556).

4.2 CHAPTER 2 – Reconciliation: Practice

4.2.1 Reconciliation approaches

The aim of any reconciliation process is to engage with the past so as to make the experience more "manageable" in the present for those involved. In so doing, the future is liberated for prospects of peaceful coexistence (Rosoux, 2008, pg. 543-544). The recovery process usually uses one of these: recovery from the past by addressing the suffering of the victims, or recovery from the past by addressing and changing the previous political system (Dersso, 2018). Alternatively, these can be described as relationship-based approaches and issue-based approaches, respectively.

Relationship-based approaches create change through engaging the personal and private realm of the involved parties (Wilson, 2001). This can be done through the social-psychological method, in which reconciliation works with the cognitive and emotional state(s) of the involved parties to try to create 'deep change' within a society (Bart-Siman-Tov, 2004). Such 'deep changes' would alter the psychological presence of the past in the collective consciousness of the majority as beliefs and emotions are invited to adjust and to shift gradually throughout the reconciliation process (Bar-Tal & Bennink, 2004, pg. 17; as cited in Rosoux, 2008, pg. 545). The alternative is what Rosoux (2008) has termed the 'spiritual' method, which concentrates on the collective healing process of both victim and perpetrator, with the end goal being seeking and granting forgiveness (e.g. Shriver, 1995; Lederach, 1998; Staub, 2000; Philpott, 2006; as in Rosoux, 2008, pg. 545).

Issue-based approaches zoom out of the personal and private in favour of examining the public realm, changing the structural and institutional mechanisms through which the society operates (Rosoux, 2008, pg. 544). In doing so, different aspects of the state may be specifically targeted, such as: security, economic interdependence, (re)distribution of wealth, political/democratic inclusion, the rule of law, and the reinstatement of the protection of peoples' rights (Rosoux, 2008, pg. 545; Kacowicz, 2000). The main aim is thus to enact justice by removing the threat of a return to the previous state of violence and/or trauma from within the system, and in this process resolving major areas of disagreement (Rosoux, 2008, pg. 544). However, it should be noted that although structural measures can help begin (re)building trust between a people and their government, these alone will not succeed (Wilmer, 1998, pg. 93; as cited in Rosoux, 2008, pg. 545).

The tertiary approach to the concept is that which emphasizes the relationship of the process to the people. This is of particular importance with regards to the implications it may have on trust between the parties. In a bottom-up approach (e.g. Bar-On, 1996; as cited in Rosoux, 2008, pg. 552), groups are considered to be networks of individuals, each having an influence over the outcome of reconciliatory efforts; action taken locally results in consequences on a national level. Top-down (e.g. Bargal & Sivan, 2004), on the other hand, views groups as abstract entities divided based on identity politics; action is taken on a national level, which then alters local realities. The latter model lends itself most naturally to the reconciliation process when pushed for by state leadership, as authorities try to facilitate the reconciliation process through the creation of national reconciliation bodies and require the public to trust in the official mechanisms. However, both top-down and bottom-up approaches are necessary for a successful reconciliation effort as they strengthen each other; creating the supporting institutional infrastructure(s) only serves as a catalyst to furthering the process. In the words of Rosoux: “the outcome of the process depends above all on popular support. For, even if a rapprochement seems necessary to the representatives of each party, it cannot be imposed by decree” (2008, pg. 552).

4.2.2 Reconciliation commissions

Following a transition from a socially traumatic state, victims can be expected to demand ‘justice’ for their suffering. A commission can be used in the transitional and reconciliation process to provide the public with a platform and a mechanism to be able to look into, address, and bring forward the trauma(s) of the past and to discuss why they occurred (Bolocan, 2004, pg. 396). Justice, in this case, is found through a ‘forward-looking’ attempt to improve relations and thus build trust between parties in the long term (Zartman & Kremenyuk, 2005, pg. 294; Cobban, 2007). It demonstrates an acquiescence to the demands for the exercising of ‘rights of recountability’, where previously suppressed memories and experiences are now given space to be memorialized (Werbner, 1998b as discussed in Colvin, 2018). This is done by creating a formalized platform from which the substance of these rights can be enjoyed (Werbner, 1998; in Ross, 2003, pg. 326), such that the memories of victims become “socially validated” (Ross, 2003, pg. 337). It requires a “concerted act of will” on behalf of the participants to trust their communities and authority figures with their histories, and to trust that such histories are treated with the necessary respect (Ross, 2003, pg. 337). In exchange, they receive public and official acknowledgement of their experiences as well as potential reparations.

Reconciliation commissions are often combined with, or quite similar to, truth commissions. Truth commissions seek to establish an “undeniable, irreversible truth” of the tragic history as their main priority (Bolocan, 2004, pgs. 396-397). Whereas a truth commission prioritizes the creation of a national truth, the reconciliation commission prioritizes the (re)creation of national unity – which can involve reconciling a national truth. This development of a more or less unified lens through which a community can view, interpret and understand its history can be vital to its recovery (Dwyer & Aukerman, as cited in Bolocan, 2004, pg. 397). However, it also requires the teller to relinquish some control over their narrative as they share it. The strength of the relationship between reconciliation and truth-telling is not necessarily symbiotic, and experiences will vary by case (Allen, 1999, pg. 316-317). There is always the risk of authority figures sounding relativistic in their treatment of personal information such as a person’s ‘truth’; thus, the conversion of the personal truth-telling into the public, narrative truth is one to be handled with care.

In essence, the commission forms a bridge between the personal, private experiences of victims, and the shared public knowledge used by the country in forming a national narrative about the history (Wilson, 2001). Forging this official narrative requires a body that can verify and integrate it into the “full historical record” of the period, which is inclusive of the experiences and roles of all actors involved (African Union, 2019, pg. 10). As it pieces together this national narrative, a commission should provide instruments through which the reconciliation process can be continued on both the individual and institutional level. On the individual level, this would involve mechanisms for and support in social reintegration or reparations. For such mechanisms to be effective, a victim must trust in the constancy and accountability of, for example, the commission. On the institutional level it includes the issuing of recommendations, the introduction of legislations, the creation of policies, and the upholding of new best practices (African Union, 2019, pg. 10). This requires the public to trust the government to seriously deliberate such recommendations and take a positive, active role in realizing the process beyond the commission.

4.2.3 Past reconciliation commissions

4.2.3.1 The South African Truth and Reconciliation Commission The South African Truth and Reconciliation Commission (SATRC) was the product of a political compromise between the rising African National Congress (ANC) and the descending National Party (NP) after the end of the apartheid regime (Rosoux, 2008, pg. 548). Intended to address the 'grey zones' of the apartheid regime, the SATRC invited parties to participate in a form of "large scale storytelling" (Mbembe, 2000; as in Ross, 2003). In so doing it engaged with the discourses of powerlessness and the horrors of the past, allowing the victims to restore their dignity by sharing stories from their perspectives and providing official mechanisms for reparation and rehabilitation.

That being said, the SATRC was successful in galvanizing a national process towards the constitutional goals of "understanding" and "reparation" (Rosoux, 2008, pg. 548). Although it was conceived without (much) input from the general public, which is not uncommon in the formation of such commissions (Wilson, 2001, pgs. 198-200), the result was a fairly robust system. Furthermore, what we see in the SATRC is an emphasis on those responsible for crimes acknowledging their trespasses, apologizing for their actions, and seeking some form of forgiveness (Graybill & Lanegran, 2004, pg. 6). Although the personal acceptance of truth in this context would naturally vary depending on the nature of individual cases, the mechanism worked well enough on a national level for trust to be successfully placed in the government, which could then lead the country towards a shared future. The mechanism of amnesty, although controversial, successfully allowed for a peaceful political transition between regimes. The NP was assured that members would be free from prosecution by the ANC, and the ANC could begin reconstructing the nation without fear of obstruction or repercussions from the NP.

Subsequently, recalling the triangular model of reconciliation processes introduced in Chapter 1, we can place the SATRC in the near-centre, although shifted more towards the 'truth' leg. This is more fitting because of the commission's emphasis on truth-telling, yet still significant engagement in the changing of identity parameters as non-white citizens moved from a time of severe inequality to a position of equal citizenship.

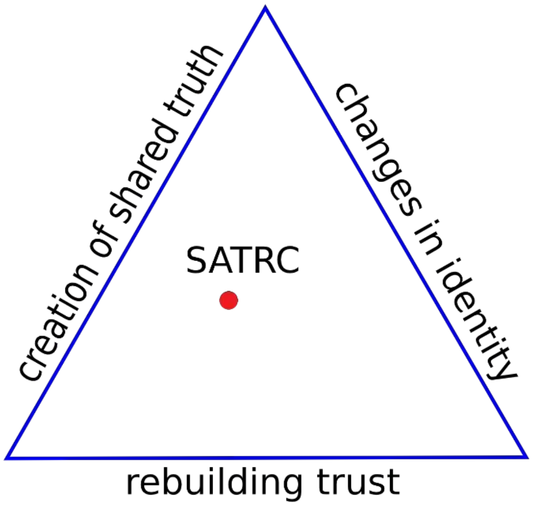


Figure 3: The SATRC's placement within the triangle model

In the context of trust, we find an odd dynamic at play here. The disproportionality of the power imbalance during apartheid was so drastic that the transition required a form of inherent trust in the 'other'. That is to say, once the actions were set in motion for the ANC to accede to power, the majority supporting it was so massive and carried such scale that any minority would have to trust in the ANC's actions—and thereby the method of the SATRC.

4.2.3.2 The Rwandan National Unity and Reconciliation Commission The Rwandan National Unity and Reconciliation Commission (RNURC) was established to provide systematic, nationalised support following the horrific mass slaughter and rape of the Tutsi, moderate Hutu and Twa populations in Rwanda during the Rwandan genocide. Within Rwandan society the RNURC has several functions and initiatives, but for the purposes of this research, we will be focussing on only one of its mechanisms: the Gacaca Courts (GC). Due to the massive scale of the conflict, the new Rwandan government of the Rwandan Patriotic Front (RPF) needed a way to process and prosecute over 100,000 people accused of having participated in the killings without flooding the legal system and ruining the economy[7]. Gacaca was a Rwandan tradition revived to help provide justice on a national scale, while keeping the process localized to the communities in which the atrocities had occurred. The GC were a "decentralized, community-based system of courts inspired by local traditions" (Bolocan, 2004, pg. 355), but also a "legal-social" experiment in the field of transitional justice (Uvin in Bolocan, 2004, pg. 356) intended to punish the génocidaires whilst also providing case-by-case mechanisms for reconciliation. An "ambitious" network of local courts, the GC was intended to enact both restorative and retributive justice (Rettig, 2008, pg. 25) on an unparalleled scale—providing "mass justice for mass atrocity" (Waldorf, 2006, pg. 1).

The exact influence of the GC on the reconciliation process has been unclear and remains controversial within Rwanda (Rettig, 2008, pg. 25; Rosoux, 2008, pg. 558). The GC were able to bring more people to trial than other more standard legal approaches. Due to their local and traditional nature, the transitional justice schemes were also attuned to the practices and theoretical needs of the population, allowing for "reintegrative shaming" (Drumbl, 2000, pg. 1263; also reflected in Cobban, 2002)[8]. On the other hand, the fairness and best practices of the GC have often been called into question (e.g. African Rights, 2003; Penal Reform International, 2006; in Rettig, 2008, pg. 26), particularly in relation to the treatment and protection of the rights of defendants (e.g. Avocats Sans Frontières, 2007; Penal Reforms International, 2006 in Rettig, 2008, pg. 26; Bolocan, 2004, pg. 356). By focusing on individual accountability for the atrocities committed, the Courts "failed to promote justice and societal reconciliation", even backfiring in some communities as the hearings deepened the divides as accusations came to light and wounds were reopened in an already torn society (Bolocan, 2004, pg. 395). Furthermore, the GC were erected out of necessity due to the sheer scale of the genocide, and thus pushed for "confession, apology, and forgiveness" to allow life to return back to 'normal' as soon as possible (Rettig, 2008, pg. 44). Further prioritizing a return to normal affairs, the GC's would not hear cases against the RPF or its members, so as to protect the legitimacy of its claim over the government.

The subsequent lack of accountability for the human rights violations of the Rwandan Patriotic Front during the genocide and the civil war (Bolocan, 2004, pg. 396; Amnesty International, 2002, as in Rettig, 2008, pg. 26) has been a significant issue for the legitimacy of the commission. Not only is the commission a demonstration of 'victor's justice,' but it also undermines the reconciliation process and the 'quality' of the reconciliation within the new Rwandan society (Bolocan, 2004, pg. 396). Crucially, it impaired the relationship of trust in the government as it showed that they are not to be held accountable to their own people. Furthermore, the uneven quality of the justice enacted through the courts casts doubts upon the skill and ability of the government. In the local communities where these courts were active, distrust in the courts and the governmental system remains common, and the divides within the community are still prevalent. As a result, the GC can be placed approximately between the 'truth' and 'identity' legs, but quite far from 'trust' on the theoretical triangle:

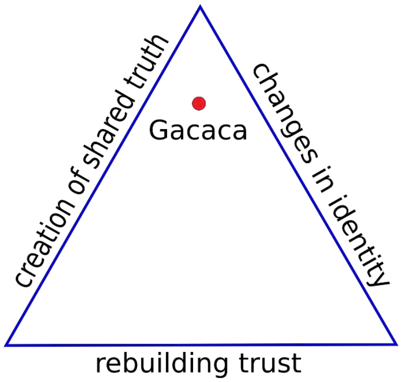


Figure 4: The Gacaca Courts' placement within the triangle model

[7] It is estimated that 'properly' prosecuting all crimes committed would have taken 200 years (Powers, 2011).

[8] Reintegrative shaming being the restorative justice practice of providing convicted perpetrators of crimes alternative acts through which they can serve their sentence and their community.

4.3 CHAPTER 3 – Case Study: The Ethiopian National Reconciliation Commission

4.3.1 Understanding the Ethiopian Socio-Political Context

4.3.1.1 The Red Terror Trials The Reconciliation Commission is not Ethiopia's first engagement with the practices of transitional justice. From 1974 to 1987, the country was under the rule of the Communist-Marxist Leninist military junta known as the Derg. The regime was notorious for its mismanagement of aid resources, for human rights abuses, and for a bloody period known as the Qey Shibir, which is more commonly referred to as the Red Terror. The Red Terror was a systemic of mass elimination of political opponents to the Derg and its newly elected leader, Mengistu Hailemariam[9]. As a result of the horrific process, 30,000 (Harff & Gurr, 1988) and 500,000 (Andrew & Mitrokhin, 2006; Trial Watch, 2016) people were estimated to have been killed, alongside uncounted cases of systematic rape (Courtois, 1999). During the Red Terror Trials (RTT), which ran from 1994 to 2007, Mengistu and other high-level members of the regime were convicted of genocide, war crimes, and crimes against humanity (Tadesse, 2007). Important to note is that the violence of the Red Terror was ruled to have been a political genocide[10] [11].

4.3.1.2 The Ethiopian People's Revolutionary Democratic Front (EPRDF) That being said, the period of time following the Derg regime wasn't rosy either. The EPRDF regime, which ousted the Derg, became notorious for political and civil repression, as well as human rights violations. The single-party democracy, led by the Tigrayan People's Liberation Front (TPLF) sub-group of the EPRDF, was known for running the Ethiopian state with an iron fist. Using the words of Theodore Vestal,

> "...not a single important political or organizational question is decided by government officials or mass organizations without guiding direction from the party. The Front [TPLF] stands above all, and the leaders do not test their policies in a forum of free speech and fair elections. Instead they mobilize and enforce consent." (as found in Yusuf, 2019b, pg. 29)

As previously mentioned, the first peaceful major transition of power in Ethiopia in decades occurred with the election of Prime Minister Abiy Ahmed (Dersso, 2018)[12] . The transition has been a hybrid between two more traditional models: transitions based on a negotiated change of regime,[13] and transitions based on the overthrow of one regime by another (Dersso, 2018)[14]. In this particular case, the transition came out of an "ad hoc alliance" between the 'old guard' in the EPRDF and the reformists now in power. In practical terms, we see the use of the old regime's institutions and structures whilst they are simultaneously being reformed (Dersso, 2018). PM Abiy was a popular leader in the 'reformist camp' of the EPRDF, who had been championing radical reform to the party and the country for years. Following his election, his new regime has pushed through several political liberalization measures (Yusuf, 2019b, pg. v). However, transitions into more liberal states tend to have violent conflicts erupt as restraints against freedom of expression are lifted without the neces-

[9] Spelling may vary in sources; the current spelling was chosen due to its usage in the Red Terror Trials.

[10] The conviction of genocide was based on the domestic penal code (Article 281 of the 1957 Penal Code, as found in Prevent Genocide International, sd). The definition found in the Convention on the Prevention and Punishment of the Crime of Genocide (CCPCG) does not include 'political affiliation' as the identifying grounds for genocidal persecution (Cassese, 2009; Staub, 1992, pg. 8; Gellately & Kiernan, 2003, pg. 267). However, the CCPCG does allow for states to tailor judicial measures to their own cases and integrate the CCPCG into their own legal frameworks (Articles V and VI, CCPCG, 1948). A more detailed list of convictions can be found on (Trial Watch, 2016).

[11] The conviction of genocide was based on the domestic penal code (Article 281 of the 1957 Penal Code, as found in Prevent Genocide International, sd). The definition found in the Convention on the Prevention and Punishment of the Crime of Genocide (CCPCG) does not include 'political affiliation' as the identifying grounds for genocidal persecution (Cassese, 2009; Staub, 1992, pg. 8; Gellately & Kiernan, 2003, pg. 267). However, the CCPCG does allow for states to tailor judicial measures to their own cases and integrate the CCPCG into their own legal frameworks (Articles V and VI, CCPCG, 1948).
A more detailed list of convictions can be found on (Trial Watch, 2016).

[12] Technically, this would have been the second, with the first being following the death of Meles Zenawi in 2012. However, the transition between Zenawi and Desalegn was more akin to a continuation of policy, whereas Abiy has demonstrated a marked turn away from the previous status quo—thus my counting it as the 'first' transition.

[13] As seen in post-Apartheid South Africa in the 1990s.

[14] As seen in Ethiopia in 1974, following the end of imperialism, and in 1991, following the end of the Derg regime.

sary institutionalised means to act (Gurr, 1996, pg. 69). We see this in the violent protests and clashes which have been occurring across the country as residual ethnic tensions are given the freedom to be acted on (Gebreluel, 2019).

In order to understand Ethiopian politics, one must understand the concept of ethnic federalism. Used by the TPLF when creating the EPRDF during the revolution, and subsequently by the EPRDF in governance, ethnic federalism was devised to grant the various peoples of Ethiopia self-governance rights (Gebreluel, 2019). It functions as a system of federated political institutions, feeding into the central, national government (Yusuf, 2019b, pg. v). Under the EPRDF ethno-federalist system, certain larger ethnic groups were given the room for their own administrative (executive, judiciary, legislative) and political bodies whilst still falling under the umbrella of the EPRDF (Yusuf, 2019b, pg. 6). Not all parties were granted these rights: the distinction was kept for only some of the larger groups, which exacerbated inequalities between populations (Gebreluel, 2019). Claims of "ethnic obstruction" increased internal resentment against the TPLF-led EPRDF as some policies were viewed as intentionally disadvantaging others (Yusuf, 2019b, pg. 7). Following the death of PM Meles Zenawi in 2012, the previously appointed elite began acting more independently from the central authority, creating "a condition of decentralized autocracy, and elite competition for power [which was] mobilized along ethno-national lines" (Gebreluel, 2019). In more concrete terms, this meant the weakening of federal institutions, undermining of the central command/power, the loss of control over key sectors of society, and a political division within the EPRDF on the basis of ethnicities (Yusuf, 2019a, pg. 2). The result of the increasing ethnic mobilisations since 2018 and the apparent "incoherence of the state and ruling party" in responding to them, have contributed to a sense of "perceived party and state fragility" (Yusuf, 2019a, pg. 2, 3). The cycle can be summarized in the diagram in Figure 5.

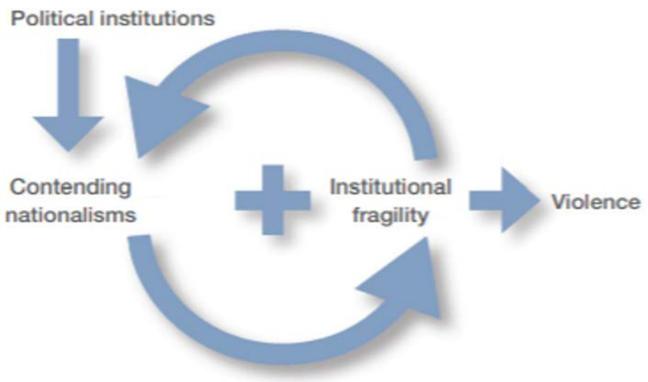


Figure 5: Interaction between nationalism and institutions producing ethnic violence in Ethiopia (Yusuf, 2019a, pg. 6; also found in Yusuf, 2019b, pg. 34)

This cycle has been exacerbated by the insistence upon the use of soft governance methods of peace enforcement by PM Abiy's new regime, which are intended to bring about a "climate of freedom" and encourage dialogue and reconciliation (Yusuf, 2019a, pg. 5). Whilst using these methods has succeeded in calming the onslaught of "anti-regime struggles", some argue that it has made the new government appear weak compared to its more hard-handed predecessor (Yusuf, 2019b, pg. 2, 30-33). Nevertheless, PM Abiy's regime has continued efforts to create a unified Ethiopian identity—one of the ways being through the establishment of the Ethiopian National Reconciliation Commission.

4.3.2 The Ethiopian National Reconciliation Commission

Dersso writes: "the choice of the transitional justice measure that a society in transition adopts constitutes an outcome of and a vehicle for the implementation of a (new) political settlement" (2019b, pg. 1). The Ethiopian National Reconciliation Commission is a 'forward-looking' product of a hybrid transition (Zartman & Kremenyuk, 2005; Dersso, 2018). It provides a mechanism that translates the rhetoric of unity and reconciliation employed by the new regime into action (Yusuf, 2019b, pg. v). Through the Commission, the government aims to address social traumas, foster new values and open a reconciliatory discussion within the society (Yusuf, 2019b, pg. 36). If the Commission succeeds, it has the potential to demonstrate the strength of the new regime regardless of its preference for 'soft measures' (Yusuf, 2019a, pg. 5). Arguably more importantly, its potential success could help build trust between the government and the peoples of Ethiopia (Amani Africa, 2019, pg. 1) in a relationship that is currently on shaky ground due to the nature of past regimes and the as-of-yet untested reforms of the new regime.

Regarding ethnic federalism in the reconciliation context, the Commission will need to be capable of addressing the ethnic diversity as it is relevant to their case work, but also make itself available to and inclusive of these populations. This is reflective of the ideal shape of the national dialogue to which it is intended to contribute: not ethnic in basis, but inclusive of the diversity of experience and the various ethnicities in Ethiopia (Yusuf, 2019a, pg. 2). Transitional justice processes have no "one-size-fits-all approach", but need to be tailored to the society undergoing a transition—and include the target societies in their entirety (African Union, 2019, Section E, pg. 6). If set up well, and executed successfully, the Commission could be instrumental in rewriting the national narrative and the dialogue with the past that is finally opening up. Doing this will require that citizens have some trust in the Commission, as citizens are being asked to share stories and accusations that would have likely previously resulted in retaliatory action from the government. As such, the government is requesting the trust of its citizens, and subsequently the responsibility to demonstrate that this trust is not ill-placed.

However, looking at Proclamation No. 1102/2018, which contains the legal framework that established the Commission, we see vague language that primarily sets "broad outlines regarding the transitional justice objectives, role and mechanisms of the Commission" (Dersso, 2019b, pg. 1). The success of the Commission will rely upon the dedication of the state to the transitional justice process, and their executing of the associated mechanisms in good faith (African Union, 2019, pg. 25). It will also require the continued support of the people (Rosoux, 2008, pg. 552) – it is in between these two aspects of the Commission that the importance of trust comes into play.

4.3.2.1 Mandate Dersso (2019a; 2019b) identifies two primary conceptual 'pillars' around which the Commission has been shaped. The first places a "particular premium [on the] peace and reconciliation dimension" (Dersso, 2019b, pg. 2). Within the Ethiopian context, the meaning of this is outlined in Articles 2(3) and 5 of the Proclamation:

> Article 2(3): [reconciliation includes] "establishing values of forgiveness for the past, lasting love, solidarity and mutual understanding by identifying reasons of conflict, animosity that are (sic) occurred due to conflicts, misapprehension, developed disagreement and revenge"
> Article 5: the "objective of the Commission is to maintain peace, justice, national unity and consensus and also reconciliation among Ethiopian peoples."

The second concerns itself with the gross violations of human rights. These are not to be treated in isolation from the first pillar, yet rather form a subsidiary but still substantive aspect within the greater sphere of peace and reconciliation. Highlighted in the second preambulatory clause, this is conceived as:

> ". . . necessary to identify and ascertain the nature, Cause (sic) and dimension of the repeated gross violation of human rights so as to fully respect and Implement (sic) basic human rights recognized under the Constitution of the Federal Democratic Republic of Ethiopia and international and continental agreements which Ethiopian (sic) ratified and since it is important for the reconciliation;"

Crucially, the role of addressing previous human rights abuses is recognized as having a key place in the Ethiopian reconciliation. However, it is important to note that the Proclamation does not include an actual definition of what the Commission will be considering as 'gross violations of human rights', rather, they seem to assume it as an apparent given. This is risky, as a failure to define such key concepts undermines the integration of such a concept into the social consciousness, and makes it harder to concretely target.

Given that the system of ethnic federalism historically has a particular propensity for the violation of group rights and the targeting of the human rights of specific groups, the trust of the people will likely be determined by the Commission's ability to address this second pillar of addressing human rights violations. Success in this will feed into any potential success in meeting their primary goal of peace and reconciliation, which arguably is the primary goal of the government due to the continuous emphasis on reconciliation throughout the

Proclamation. This is also reflected in the addressing of remedies and reparations for human rights violations: the remedies as of yet are primarily the creation of narrative justice by providing a platform for victims (and offenders) to be heard, and to have their suffering publicly acknowledged (Article 6(2)). In doing so, the Proclamation argues that it will contribute to institutional respect for human rights and the implementation thereof (Preambulatory clause 2). However, there are no guarantees included within the clause, nor the rest of the Proclamation, to ensure this or to give the public the means to hold the Commission accountable. Thus, it requires a blind trust from the public, without prevalent history to justify such trust (Brooks, 2020).

4.3.2.2 Temporal Scope The Commission's temporal scope is unclear from the Proclamation and any updates about its development since. The Proclamation seems to suggest the "formation of the unitary state structure in Ethiopia" as a starting point (Dersso, 2019a). Thus, potential starting points could reasonably include 1936 and the Italian invasion, Ethiopia joining the United Nations and the international human rights regime, after the end of the imperial reign of Emperor Haile Selassie, or after the start of the EPRDF regime. On the basis of what he calls 'hints', Dersso (2019b) offers several suggestions of the temporal scope. According to him, the use of the phrase "for years" in the first preambulatory clause indicates conflicts before the establishment of the Commission (Dersso, 2019b, pg. 7). As for 'endpoints', there is some more clarity. An additional hint is found in the third preambulatory clause, which discusses "gross human rights abuses in different time and historical event." Dersso (2019b) argues that this effectively implies that the Commission will not be hearing ad hoc cases as they occur (Dersso, 2019b, pg. 7). Doing so would compromise its role in addressing the past for the sake of the future, and could risk the stability of the new organization; it also exempts a significant portion of the new regime's time in office from being discussed.

However, in my opinion, this reading still seems largely speculative at best. No other sources have shared or alluded to this reading of the Proclamation. The Commission runs the risk of being required to address an impossibly long period of time, and subsequently being saddled with a backlog of past cases[15]. For the Commission to be trusted by the public and to work both practically and socially, it will need to engage with its cases in a timely fashion and with sensitivity regarding the time during which they occurred. Moreover, should the Commission in its working choose to issue acts of contrition for perpetrators to make amends with their victims, there should be a time limit for doing so (Kritz, as in Bolocan, 2004, pg. 398)[16]. Such a solution does not require amnesty, but rather provides a process through which individuals can take the burden of blame for their actions, whilst also providing a way in which the communities can also benefit from their return (Stahn, as in Bolocan, 2004, pg. 397).

4.3.2.3 Inclusivity In his research on the Rwandan Gacaca Courts, Rettig (2008) emphasized the importance of "community trust" in the process (pg. 46). Yet the manner in which the Commission was created is cause for concern. The Commission was pushed through parliament, where opposition groups held no seats, and was approved without consultation of other parties (Yusuf, 2019a, pg. 3; Dersso, 2019a).[17] It is concerning that the Commission was formed blind to the will of the people it aims to unify. That is not to say it is undesirable, or undesired within Ethiopia – we simply do not know what the citizens would have wanted because the question wasn't asked to them. To that effect, the Commission seems imposed upon the society, which rightly prompts the question of how its work will be regarded within Ethiopia. Articles 6(2), 6(10), and 6(3) of the founding Proclamation contain measures for how an inclusive reconciliation process can take form. 6(2) requires the Commission to make its work accessible and interactive, through the use of technology or other participatory opportunities; 6(3) asks the Commission to use this interaction to "identify principles and values which will be base (sic) for national Reconciliation (sic)".[18] 6(10) emphasizes the need for reconciliation to be between conflicting parties, "to

[15] Although, it should be noted that the South African and Kenya commissions each also covered extensive periods of time.

[16] As was done with the Commission for Reception, Truth and Reconciliation in East Timor.

[17] The vote which passed the establishing proclamation stood at 545 for, 1 against, and 1 abstention.

[18] Both also discussed at length in (Dersso, 2019b, pg. 3).

narrow the difference created and to create consensus." Although the Commission didn't need the "confidence and support of Ethiopia's diverse social and political groups" to be approved, it will need it to succeed in its stated objectives (Dersso, 2019a) – whereas currently it is faced with a "foundational gap" (Dersso, 2019b, pg. 2).

Past experience on the continent shows the value of "public consultations" through "the creation of adequate platforms that solicit the input of various sectors of society including victim groups on the draft law establishing the transitional justice mechanism" (Dersso, 2019b, pg. 2). Encorporating such input allows for an early setting of "normative expectations" and "best practice standards", and sets the political tone of a commission (Dersso, 2019b, pg. 2). For example, such public consultations could have been applied during the selection of the commission members. Rather, the members were selected by PM Abiy, after which they were submitted to parliament for approval. Also notable is that the members themselves, and the work of the Commission, are ultimately accountable to the PM (Article 3(4); Ethiopia gets National Reconciliation Commission legislation, 2018). The members were "drawn from different faith groups, former politicians, thought leaders, intellectuals, artists and actors, authors, legal experts, philanthropists, politicians and elders, among others"; their introductory list seems to support this claim (Ethiopia named members of National Reconciliation Commission, 2019).[19] The clear effort at including representatives from key demographics in Ethiopia is commendable. However, because it is unknown how exactly the members were selected, and whether or not there were opportunities for public input, the selection process may be rightly questioned (Ethiopia named members of National Reconciliation Commission, 2019; Addisu, 2020).

An additional key aspect of inclusivity – and thus public engagement – would need to be the timely and frequent sharing of information by the Commission. This is outlined in 6(9) as a duty to "notify to the public (sic) and concerned government organs the conclusions reached through the examination as appropriate". The most recent update stemming from the Commission is dated April of 2019, when the Commission announced a three-year plan of action (Ethiopian Reconciliation Commission Announces Three-Year Plan, 2019). However, neither I nor those I asked for help could find a digital copy of this plan or a record of its publication. In the Proclamation, three years was also the original timespan given to a single term of the Commission (Article 14(1)), although this could "be prolonged as may be necessary" (Article 14(2)). Some worry that the Commission isn't "as active as they should be to handle the work they were assigned to do", and that the three-year plan will be "too little, too late" to achieve its task (Ethiopian Reconciliation Commission Announces Three-Year Plan, 2019). The worrisome lack of available information is further highlighted by the fact that the Commission is supposed to meet at least twice a month (Article 7(1)). If the Commission has indeed been meeting as frequently as they should, it would be reasonable to assume that more accessible work should have since been made available, or at least that more information about its progress is public. Importantly, this removes an important source of external accountability found in the public.

The Commission needs to demonstrate the supposed values of the new regime: "professional[ism] and politically impartial [work]" (Yusuf, 2019b, pg. 39). So the Commission will need to start working with visible transparency, independence and compliance with due process standards (Article 13; Dersso, 2018). However, the Commission is still in its starting phase, and has yet to begin hearing actual cases, which means that there likely isn't much work readily available for publication. Nevertheless, we should then see these standards of transparency being reflected in the work done so far, or at least in the process of the work (Yusuf, 2019b, pg. 39). However, what has been produced to this point has been relatively anonymous and unknown to the public. Furthermore, there has been no known "public participation in the development of the commission's enabling law, nor in the nomination and appointment of the commission's members" (Dersso, 2019a). This is worrisome as it inherently already undermines several of the principles of the Commission, and in the future it could call into question the legitimacy of the Commission (Articles 13, 6(2)) – all this before its public work has even begun.

4.3.2.4 Truth The Commission embraces the idea that "there may not be a single truth about conflicts in Ethiopia" (Dersso, 2019b, pg. 3). That

[19] Available via: https://borkena.com/2019/02/05/ethiopia-named-members-of-national-reconciliation-commission/

being said, the Commission has the authority to access all information, data and documentation, [20] as well as institutions. It also has the power of subpoena which should enable it to access most, if not all, necessary factual information to determine a 'truth' (Article 6(5-8)). The Commission then needs to be able to turn the so-called factual/forensic truth into a narrative/restorative truth, which can contribute to the history of the nation. This is no small feat, and no small power for such a Commission to have – giving it this power is a considerable token of trust from the government.

It is also important to note that the 'truth' is not completely ignored in the Proclamation; rather it has a less prominent position than in most comparable cases. Truth is included thematically in the preambulatory clauses, where it is construed as "necessary" for reconciliation. The first preambulatory clause states that reconciliation should be attempted on the basis of truth and justice (Preambulatory clause 1), which is followed with the assurance that the Commission will "free[ly] and independen[ly]. . . inquire and disclose" the truth of the cases they are brought (Preambulatory clause 4). This shows an important duality in the understanding of the 'truth' as it applies to our context. There is an implicit understanding that on a macroscopic level, there is no single story befitting of Ethiopia in the past decades; however, in individual cases, there is a truth to be found. It is the duty of the Commission, then, to bridge the gap between the former and the latter. They can do so by using the stories that come out of their work to create a complex narrative of history – one which is acceptable for the various social groups within Ethiopia.

4.3.2.5 Addressing gross human rights violations As mentioned previously, the Commission is meant to target specifically the gross human rights violations of the past decades within the greater reconciliation scheme (Preambulatory clause 2). Article 6 details the powers of the Commission and its commissioners, whilst Article 6(4) specifies the lenses to be taken into account when addressing gross human rights violations accusations. These are: the power to "make examination (sic) to identify the basic reasons of disputes and violations of human rights by taking into consideration of (sic) political, social and economic circumstances and the views of victims and offenders" (Article 6(4)). However, overall the powers and scope thereof remain broad-brushed and vague. In the comparable examples of the South African and Rwandan commissions, the establishing laws were detailed; not having this detail potentially leaves gaps in the Ethiopian case. This is important because although there are commonalities in the definition and general interpretation of terms such as 'gross human rights violations', there may also be some crucial differences.[21] However, it should be noted that under Article 19 of the Proclamation, the Council of Ministers or the Commission itself can change the regulations as needed "for the effective implementation of the proclamation". This does not mean that the Commission has the authority to create new powers for itself, or to create new laws, but it does have the authority to create new regulations, such as the understanding or definition of a term within the scope of the Commission (Dersso, 2019b, pg. 5).

The lack of definition aside, the Commission will be tasked with figuring out fact-finding with regard to the stories they are presented, which parties were involved in violations, the implications for the victims, and whether violations were the result of deliberate planning by the perpetrator(s) (Preambulatory clause 2, Article 6(4)). Subsequent actions should be performed in the pursuit of justice as it applies to reconciliation, peace, national unity and consensus. It is important to note that justice in this case does not necessarily refer to punishment for a crime, but rather to a concerted effort to bring the country to a 'just' state of being, which is the ideal "end state of the work of the Commission" (Dersso, 2019b, pg. 5). Accountability in this model is achieved through finding the truth about the "nature, Cause (sic) and dimension" of the violations (Preambulatory clause 2); providing a forum for perpetrators to confess and victims to be heard (Preambulatory clause 3); the political, social and economic context of the violations, and the multiple perspectives relating to the case (Article 6(4)); and the creation and sharing of an official public record of the suffering, and public recognition through that (Article 6(9)). Importantly, the

[20] Below the level 'secret' for state security (Article 6(5)).

[21] I remind you of the inclusion of political groups in the definition of genocide under Ethiopian law, which allowed for the conviction of Mengistu for attempted genocide in the Red Terror Trials. The international definition does not include this distinction.

Commission will only be accepting cases brought forth on the basis of “attempts to achieve reconciliation, rather than criminal accountability” (Dersso, 2019a). The Commission will need to demonstrate how it will walk the tightrope of balancing reconciliation-focussed action, allowing for open discussion, and building a shared truth.

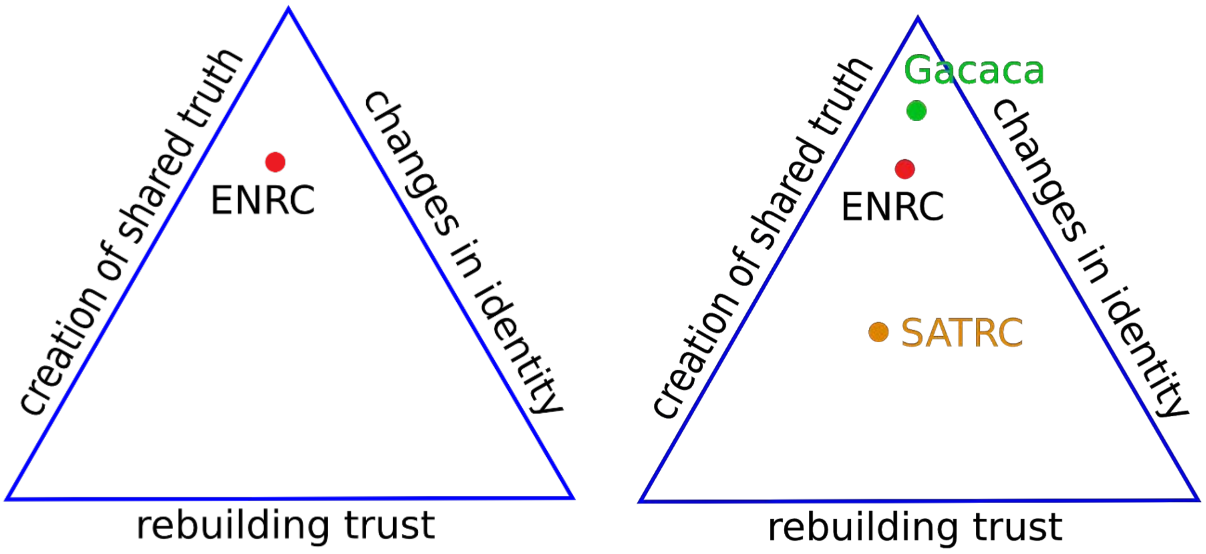
4.3.2.6 Practicalities The Commission is headquartered in Addis Ababa, as per Article 10(3). However, there may be merit to opening more regional or local options should the Commission aim to hear all voices and, therefore, all sides to the truth. The Commission has the power (Article 10(3)) and presumably the capacity to do so (Articles 10(1) & 10(2)). It may also be prudent to follow a decentralized approach similar to that of the Gacaca courts. That will, of course, depend on the scale of the Commission, which is as of yet unclear. However, if Ethiopia does truly want to experience a reconciliation process on a *national* level, the Commission will need to hear stories from all corners of the federation. Given the system of ethnic federalism, it may thus be useful to employ a decentralized process of justice, allowing the different experiences to speak within a context of understanding. This may seem counterintuitive to the reformists’ aim of creating a unified Ethiopian history and identity. It is, however, more inclusive of the fact that ethnic federalism remains popular amongst the general populace (Brooks, 2020). To facilitate a truly *national* discussion and recovery, the government will need to appreciate that a significant portion of the country does not live within the capital, and that there may be a need for region-specific approaches to national reconciliation. The ethnic divides that have been upheld for decades will not immediately disappear with a single unifying sweep – or with the creation of the Prosperity Party. Thus, the Commission will need to appreciate the divides in order to bridge them properly. In other words, these identities, along with the truths that set their parameters, need to be respected before they are challenged. Respecting the will of the people might just foster the necessary trust to unify them.

Of particular note is the fact that the Commission cannot refer any of their cases to the Office of the Attorney General or the state’s judicial bodies. Nor can any testimony given to the Commission be used as evidence against a party (Article 18(1)), in addition to the whistle-blower and witness protection laws already included in the domestic legal system (Article 18(2)). However, given the Ethiopian government’s history of repressing freedom of speech and the jailing of thousands of journalists in the past, keeping true to this in particular will be a linchpin to public trust in the government, and to the success of the Commission overall. If this remains true, the Commission will likely engage with many more participants as they will be able to trust that the risk and ramifications of sharing their experiences are relatively low. This scenario would allow for an actual discussion to exist, rather than one puppeteered by the government.

4.3.3 Synthesis & Analysis

The task set out for the Commission is by no means easy, and the founding Proclamation, which should outline the mechanism, is filled with vague language and gaps. Importantly, many of these gaps may be retroactively ‘filled’ through the use of Article 19; however, an overemphasis on the potential of Article 19 places the analysis in a speculative and ultimately unhelpful position. As such, this section will focus on synthesizing the findings of Chapter 3.2, and analysing them through the theoretical framework established in Chapters 1-3, but particularly Chapter 2. Returning to the triangular model of reconciliation, the ENRC is placed somewhat in between the SATRC and the Gacaca Courts (see Figure 6).

This placement is due to the Commission’s focus on truth – although the conception of the truth as it pertains to the ENRC is different from what is usually seen as it embraces multiplicity. This differentiated it from the SATRC, which emphasized truth-telling but in the context of a single shared truth, and the GC which was more selective with the truths it accepted. Furthermore, although there are some changes in identity occurring and intended by the Commission, these are not fully addressed by its framework. ‘Identity’ seems to fall back in favour of frequent referral to the creation of “national unity” (Article 5) – thus leading to the placement of the ENRC slightly off-center between the ‘truth’ and ‘identity’ legs. This is something the ENRC has in common with the SATRC and GC; thus they are placed relatively in line with each other. Now, crucially, the ENRC is placed fairly far from the ‘trust’ branch, although less far than the GC as



(a) ENRC's placement within the triangle model

(b) ENRC's placement within the triangle model, relative to the Gacaca and SATRC

Figure 6: Relative placement of the ENRC in the triangular model

it less blatantly serves a victor's justice. This is because although the Commission will need the trust of all parties involved, its framework does not create an environment that would foster it, and those measures that do have already been undermined with the Commission's relative silence since its inception (Articles 6(2), 6(9)). It should be noted that the circumstances brought on by the global COVID-19 pandemic may have played a role in this; nevertheless, the Commission has been in development since far before the pandemic, and so substantial progress is a reasonable expectation at this stage.

Given the socio-political context of Ethiopian society, history, and current politics, the lack of visible progress is particularly worrisome. A decades-long history of political repression and single-party rule, and where civil and human rights violations were commonplace, provides no foundation for a culture of trust in the establishment the new regime can make use of. The work of the new regime is too recent and too uncertain to be considered reliable by the Ethiopian public (Brooks, 2020), and trust in the current government and its leaders is already shaky. Although the Commission promises to open up the discussion on past human rights abuses and to remain inclusive to all, the method and reality of these vital options are obfuscated by the lack of concrete address in the Commission itself. The situation is exacerbated by the apparent inaccessibility of the Commission, with its lack of publicly available information and visible progress, which is needed for its work to be truly the product of a national reconciliation process, and to result in the national reconciliation and unity that the government desires.

The ENRC also demonstrates how trust is needed to effectively start a reconciliation effort. In this case study, it is clear that trust is the key missing ingredient: trust in the Commission, and in the government in particular. The lack of clarity and practicality in the Proclamation and the relative silence on the progress of the Commission arguably also allude to a lack of trust in the public from the government. On the other hand, we see a government demonstrating – to varying degrees of efficacy – what seems to be a lasting change. The second party, the Ethiopian peoples, are also changing, although it is currently still uncertain as to how that will affect the potential workings of the Commission. Furthermore, we see a groundbreaking approach being used towards the understanding of truth. The multifaceted, flexible version of the truth of Ethiopian history as constructed by the Commission is truly commendable and noteworthy. Unfortunately, we see that the Commission is struggling to get off the ground, held back by a lack of trust in it from citizens that would enable the process to begin in earnest.

5 Conclusion

This research aimed to examine the significance of trust within reconciliation processes, and ultimately argues for the role of trust as a necessary catalyst for the effective beginning of reconciliation processes. In doing so, it built upon Rosoux's (2008) three-pillar model of reconciliation, where the concepts of 'trust', 'truth' and 'identity' are considered the key concepts vital to any reconciliation process. This three-pillar model was translated into a triangular model, with each pillar represented by a leg, and the interconnected nature of the three pillars to the process as a whole demonstrated through the use of a unified shape. Reconciliation efforts, such as the key examples of the South African Truth and Reconciliation Commission and the Rwandan National Unity and Reconciliation Commission's Gacaca Courts, can be placed at different positions within the triangle to help understand how different attempts at reconciliation processes demonstrate different priorities. A reconciliation process must be adapted to its context, and thus there is no perfect location within the triangle - although the more central it finds itself, the more balanced it's likely to be. Similarly, it must be recognized that although the concepts the pillars embody are referred to in their abstract form throughout the theory, their practical translations will also inherently vary from case to case, and from conflict to conflict.

In our key case study with the ENRC, we saw that the practical version of 'truth' could be multifaceted - incorporating intersubjective truths into the official 'Ethiopian' narrative could be a part of the solution to unifying a very divided country. Nevertheless, trust also plays a vital role in this more practical application of reconciliation. By applying the findings from the theory-based literature reviews to our case study, we were able to look into the formation of trust at the start of such a commission's work. Through an exhaustive review of English-language local and international literature, media and news on the current state of Ethiopia, we were able to gain an understanding of a transitioning and, as of yet, unstable society, where there is little basis for trusting the government. The ENRC is meant to facilitate this trust by providing a mechanism for reconciliation – a truth-telling platform that allows for ethnic divides to be bridged in favour of national unity. However, the various structural gaps of the ENRC and the vague wording of its founding Proclamation No. 1102/2018 provide little substance in which the public can place their trust. Its mandate to address national reconciliation and peace, and the gross violations of human rights through the years do address the core issues of Ethiopian reconciliation – yet lack any reassuring follow-through. Similarly the temporal scope the Commission received in its mandate is generous, but too vague to be of practical use, forcing us to rely on 'hints' (Dersso, 2019b). The potential of the Commission is further undermined by the apparent lack of public involvement and the ominous absence of published updates in local, governmental and international publications.

There is of course no "one-size-fits-all" model for reconciliation (African Union, 2019, Section E, pg. 6), and each process needs to be tailored to its context of application. That being said, there are recurring themes and concepts around which most approaches are built that facilitate a holistic process. The aim of the current research was to help develop a more nuanced understanding of trust, the pillar of reconciliation that is most often overlooked. The case study provided a demonstrative example of the importance of trust, showing the potential for negative consequences when trust is seemingly neglected. By extension it also illustrates the risks commissions can run when insufficient attention is paid to establishing, fostering and nourishing trust. Future research can and should test the hypothesis of trust as a catalyst against more case studies, in particular more positive examples, by which I mean cases that actively attempt(ed) to foster trust. Examining a broader range of case studies and policies – not only those limited to the African continent – would help refine the nuance this research has tried to bring to the conception of trust in the reconciliation context.

The short title of this paper is 'In Reconciliation We Trust', alluding to the significance of trust in the reconciliation process. Yet, perhaps, it would have been more accurate to have the title reflect the order of necessity: 'In Trust We Reconcile'.

6 References

Addisu, A. (2020, February 26). *"Stand together, we can overcome our adversities".* Retrieved from FDRE, House of Peoples' Representatives: https://www.hopr.gov.et/newsreader/-/asset_ publisher/HDTI2PX7G3dd/content/-stand-together-we-can-overcome-our-adversities-?_101_INSTANCE_HDTI2PX7G3dd_viewMode=view

African Union. (2019, February). Transitional Justice Policy. *African Union's Transitional Justice Policy.* Addis Ababa, Ethiopia: African Union. Retrieved from: https://au.int/sites/default/files/documents/36541-doc-au_tj_policy_eng_web.pdf

Albin, C. (2008). Peace vs. Justice – and Beyond. In J. Bercovitch, V. Kremenyuk, & I. W. Zartman (Eds.), *The SAGE Handbook of Conflict Resolution* (pp. 580-594). SAGE Publications Ltd.

Allen, J. (1999). Balancing Justice and Social Unity: Political Theory and the Idea of a Truth and Reconciliation Commission. *University of Toronto Law Journal, 49*(3), 315-353.

Allo, A. K. (2018, November 20). *Navigating Ethiopia's journey towards reconciliation and justice.* Retrieved from Al Jazeera: https://www.aljazeera.com/indepth/opinion/navigating-ethiopia-journey-reconciliation-justice-181119135649004.html

Amani Africa. (2019). *Insights on the Peace & Security Council: Ministerial session on 'National Reconciliation, Restoration of Peace, Security and Rebuilding of Cohesion in Africa'.* Media and Research Services. Addis Ababa: Amani Africa. Retrieved from: http://www.amaniafrica-et.org/images/Reports/MinisterialsessiononNationalReconciliation.pdf

Andrew, C., & Mitrokhin, V. (2006). *The World Was Going Our Way: The KGB and the Battle for the Third World.* Basic Books.

Asmal, K., Asmal, L., & Roberts, R. S. (1997). *Reconciliation Through Truth: Reckoning of Apartheids Criminal Governance.* Cape Town: David Philips.

Babbitt, E. F. (2008). Conflict Resolution and Human Rights: The State of the Art. In J. Bercovitch, V. Kremenyuk, & I. W. Zartman (Eds.), *The SAGE Handbook of Conflict Resolution* (pp. 613-629). SAGE Publications Ltd.

Bargal, D., & Sivan, E. (2004). Leadership and Reconciliation. In Y. Bar-Siman-Tov (Ed.), *From Conflict Resolution to Reconciliation* (pp. 125-147). Oxford: Oxford University Press.

Bar-Siman-Tov, Y. (2000). Israel-Egypt Peace: Stable Peace? In A. M. Kacowicz, & O. E. Yaacov Bar-Siman-Tov (Eds.), *Stable Peace Among Nations* (pp. 220-238). Boulder: Rowman and Littlefield Publishers.

Bar-Siman-Tov, Y. (2004). *From Conflict Resolution to Reconciliation.* Oxford: Oxford University Press.

Biggar, N. (2003). Making Peace or Doing Justice: Must We Choose? In N. Biggar (Ed.), *Burying the Past, Making Peace and Doing Justice after Civil Conflict* (pp. 3-24). Washington DC: Georgetown University Press.

Bolocan, M. G. (2004). Rwandan Gacaca: An Experiment in Transitional Justice. *Journal of Dispute Resolution, 2*(2), 355-400. Retrieved from https://scholarship.law.missouri.edu/jdr/vol2004/iss2/2

Brooks, A. (2020, February 27). *How popular is Abiy Ahmed in Ethiopia as election looms?* Retrieved from The East Africa Monitor: https://eastafricamonitor.com/how-popular-is-abiy-ahmed-in-ethiopia-as-election-looms/

Call, C. (2004). Is Transitional Justice Really Just? *Brown Journal of World Affairs, 11*(1), 101-113.

Cassese, A. (Ed.). (2009). *The Oxford Companion to International Criminal Justice.* Oxford: Oxford University Press.

CCPCG. (1948, December 9). Convention on the Prevention and Punishment of the Crime of Genocide. United Nations. Retrieved from https://treaties.un.org/doc/publication/unts/volume%2078/volume-78-i-1021-english.pdf

Central Intelligence Agency. (2019, April 06). *Ethiopia.* Retrieved from CIA World Factbook: https://www.cia.gov/library/publications/the-world-factbook/geos/et.html

Cobban, H. (2002). The Legacies of Collective Violence: The Rwandan Genocide and the Limits of the Law.*Boston Review, 27*(2), 4-15.

Cobban, H. (2007). *Amnesty after Atrocity? Healing Nations After Genocide and War Crimes.* Boulder: Paradigm Publishers.

Colvin, C. J. (2018). *Traumatic Storytelling and Memory in Post-Apartheid South Africa: Performing Signs of Injury.* Routledge.

Constitution of the Federal Democratic Republic of Ethiopia. (1995, August 21). Retrieved from https://www.wipo.int/edocs/lexdocs/laws/en/et/et007en.pdf

Courtois, S. (Ed.). (1999). *The Black Book of Communism: Crimes, Terror, Repression.* (&. M. J. Murphy, Trans.) Cambridge, Massachusetts: Harvard University Press.

Dersso, S. A. (2018, December 14). *Pursuing Transitional Justice and Reconciliation in Ethiopia's Hybrid Transition.* Retrieved from Addis Standard:http://addisstandard.com/oped-pursuing-transitional-justice-and-reconciliation-in-ethiopias-hybrid-transition/

Dersso, S. A. (2019a, September 23). *Ethiopia's Experiment in Reconciliation.* Retrieved from United States Institute of Peace: https://www.usip.org/publications/2019/09/ethiopias-experiment-reconciliation

Dersso, S. A. (2019b, August 2). Ethiopia's transitional justice framework: Defining the boundaries of the mandate of the Ethiopian Reconciliation Commission. *Dialogue Forum of the Justice Sector Joint Forum.*

(A. Verjee, Ed.) Addis Ababa, Ethiopia: United States Institute of Peace. Retrieved from https://www.usip.org/sites/default/files/20190923-Dersso_Presentation-AC.pdf

Drumbl, M. (2000). Punishment, Postgenocide: From Guilt to Shame to Civis in Rwanda. *New York University Law Review, 75*(5), 1221-1326.

Ethiopia gets National Reconciliation Commission legislation. (2018, December 25). Retrieved from Borkena.com: https://borkena.com/2018/12/25/national-reconciliation-commission-legislation/

Ethiopia named members of National Reconciliation Commission. (2019, February 5). Retrieved from Borkena.com: https://borkena.com/2019/02/05/ethiopia-named-members-of-national-reconciliation-commission/

Ethiopian Reconciliation Commission Announces Three-Year Plan. (2019, April 30). Retrieved from Ezega News: https://www.ezega.com/News/NewsDetails/7075/Ethiopian-Reconciliation-Commission-Announces-Three-Year-Plan

Gebreluel, G. (2019, April 05). *Should Ethiopia stick with ethnic federalism?* Retrieved from Al Jazeera: https://www.aljazeera.com/indepth/opinion/ethiopia-stick-ethnic-federalism-190401092837981.html

Gellately, R., & Kiernan, B. (2003). *The Specter of Genocide: Mass Murder in Historical Perspective.* Cambridge: Cambridge University Press.

Govier, T., & Verwoerd, W. (2002). Trust and the Problem of National Reconciliation. *Philosophy of the Social Sciences, 32*(6), 178-205.

Graybill, L., & Lanegran, K. (2004). Truth, Justice, and Reconciliation in Africa: Issues and Cases. *African Studies Quarterly, 8*(1), 1-18.

Gurr, T. (1996). Minorities, Nationalists and Ethnopolitical Conflict. In C. Crocker, F. Hampson, & P. Aall (Eds.), *Managing Global Chaos: Sources of and Responses to International Conflict.* Washington DC: United States Institute of Peace Press.

Harff, B., & Gurr, T. R. (1998). Systematic Early Warning of Humanitarian Emergencies. *Journal of Peace Research, 35*(5), 551-579.

Kacowicz, A. M., & Bar-Siman-Tov, Y. (2000). Stable Peace: A Conceptual Framework. In A. M. Kacowicz, Y. Bar-Siman-Tov, O. Elgaström, & M. Jerneck (Eds.), *Stable Peace Among Nations* (pp. 11-35). Lanham: Rowman & Littlefield.

Kelman, H. C. (1999). Transforming the Relationship between Former Enemies: A Social-Psychological Analysis. In R. L. Rothstein (Ed.), *After the Peace: Resistance and Reconciliation* (pp. 193-205). London: Boulder.

Kelman, H. C. (2004). Reconciliation as identity change: A social psychological perspective. In Y. Bar-Siman-Tov (Ed.), *From conflict resolution to reconciliation* (pp. 111-124). Oxford: Oxford University Press.

Lederach, J. P. (1997). *Building Peace: Sustainable Reconciliation in Divided Societies.* Washington DC: United States Institute of Peace Press.

Long, W. J., & Brecke, P. B. (2003). War and Reconciliation. *In Reason and Emotion in Conflict Resolution.* Cambridge: The MIT Press.

Louw-Vaudran, L. (2018, March 06). *Why the African Union needs a stable Ethiopia.* Retrieved from Institute for Security Studies: https://issafrica.org/iss-today/why-the-african-union-needs-a-stable-ethiopia

Mekonnen, D. R. (2019, February 01). *Ethiopia's transitional justice process needs restoration work.* Retrieved from Ethiopia Insight: https://www.ethiopia-insight.com/2019/02/01/ethiopias-transitional-justice-process-needs-restoration-work/

Ness, D. W., & Strong, K. H. (2010). *Restoring Justice – An Introduction to Restorative* (4th ed.). New Province, New Jersey: Matthew Bender & Co., Inc.

Powers, S. E. (2011). Rwanda's Gacaca Courts: Implications for International Criminal Law and Transitional Justice. *Insights,* 1-6.

Prevent Genocide International. (n.d.). *Article 281 of the Ethiopian Penal Code.* Retrieved from Prevent Genocide International: http://preventgenocide.org/law/domestic/ethiopia.htm

Proclamation No. 1102/2018. (2019, February 05). *(27),* 10982-10989. Addis Ababa, Ethiopia: Federal Negarit Gazette of the Federal Democratic Republic of Ethiopia. Retrieved from: https://www.hopr.gov.et/documents/20181/94381/A+PROCLAMATION+TO+ESTABLISH+RECONCILATION+COMMISSION/c46c936e-cf61-4390-9c7c-66ffe7578bd0?version=1.0

Reardon, B. A., & Hans, A. (Eds.). (2010). *The Gender Imperative: Human Security VS State Security.* New Delhi, Abingdon: Routledge.

Rettig, M. (2008, December). Gacaca: Truth, Justice, and Reconciliation in Postconflict Rwanda? *African Studies Review, 51*(3), 25-50. doi:10.1353/arw.0.0091

Rosoux, V. (2008). Reconciliation as a Peace-Building Process: Scope and Limits. In J. Bercovitch, V. Kremenyuk, & I. W. Zartman (Eds.), *The SAGE Handbook of Conflict Resolution* (pp. 543-563). SAGE Publications Ltd.

Ross, F. C. (2003, September 1). On having Voice and Being Heard: Some after-Effects of Testifying Before the South African Truth and Reconciliation Commission. *Anthropological Theory, 3*(3), 325-341. https://doi.org/10.1177/14634996030033005

Rothe, D. L., & Friedrichs, D. O. (2006). The State of the Criminology of Crimes of the State. *So-*

cial Justice, 33(1), 147-161. Retrieved from https://www.jstor.org/stable/29768358

Staub, E. (1992). *The Roots of Evil: The Origins of Genocide and Other Group Violence.* Cambridge: Cambridge University Press.

Tadesse, T. (2007, January 21). *Ethiopia's ex-ruler Mengistu sentenced to life.* Retrieved from Reuters: https://www.reuters.com/article/us-ethiopia-mengistu-sentence/ethiopias-ex-ruler-mengistu-sentenced-to-life-idUSL115356820070111

Tavuchis, N. (1991). *Mea Culpa: a Sociology of Apology and Reconciliation.* Stanford: Stanford University Press.

Trial Watch. (2016, May 4). *Mengistu Haile Mariam.* Retrieved from Trial International: https://trialinternational.org/latest-post/mengistu-haile-mariam/

UNODC. (2006). Handbook on Restorative Justice Programmes. Vienna, New York: United Nations Office on Drugs and Crime. Retrieved from: http://www.unodc.org/pdf/criminal_justice/06-56290_Ebook.pdf

Waldorf, L. (2006). Mass Justice for Mass Atrocity: Rethinking Local Justice as Transitional Justice. *Temple Law Review, 79*(1).

White, M. (2011). *Atrocitology.* Edinburgh: Canongate.

Wilson, R. (2001). *The Politics of Truth and Reconciliation.* Cambridge: Cambridge University Press.

Yusuf, S. (2019a). *What is driving Ethiopia's ethnic conflict?* ISS East Africa. Institute for Security Studies. Retrieved from https://issafrica.s3.amazonaws.com/site/uploads/ear-28.pdf

Yusuf, S. (2019b). *Drivers of ethnic conflict in contemporary Ethiopia.* Institute for Security Studies. Retrieved from https://issafrica.s3.amazonaws.com/site/uploads/mono-202-2.pdf

Zartman, I. W. (2000). Ripeness: The Hurting Stalemate and Beyond. In P. Stern, & D. Druckman (Eds.), *International Conflict Resolution after the Cold War* (pp. 225-250). Washington DC: National Academy Press.

Zartman, I. W., & Kremenyuk, V. (2005). *Peace versus Justice: Negotiating Forward- and Backward-Looking Outcomes.* Lanham, Maryland: Littlefield Publishers.

Zehr, H. (1990). *Changing Lenses – A New Focus for Crime and Justice* (3rd ed.). Scottdale, Pennsylvania, USA: Herald Press.

Zehr, H., & Gohar, A. (2002). *The Little Book of Restorative Justice.* Intercourse, Pennsylvania, USA: Good Books. Retrieved from https://www.unicef.org/tdad/littlebookrjpakaf.pdf

Social Sciences

Exploring Meaningful Youth Participation in Global Climate Talks

Recognition, Influence and Empowerment

Catherine Schulter

Supervisor
Dr. Siniša Vuković (AUC)
Reader
Thijs Etty (AUC)

Photographer: Sanch Kuber

Abstract

The immense media attention and public support that the recent youth climate strike movement has received, alongside ample critique on multilateral negotiation settings, lead to a questioning of the role and influence of young, non-state actors in Global Climate Governance (GCG). This thesis seeks to establish a theoretical foundation to study youth participation in GCG by considering both inside and outside participatory strategies. Through an extensive literature review, it explores definitions and concepts of participation, power relations, power sources, and stakeholders' interests. Based on these theoretical insights, this research defines meaningful involvement as a codependent triangle of recognition, influence, and empowerment. A multidisciplinary approach that combines global governance and policymaking with youth and community studies incentivizes a thought-provoking assessment of meaningful youth involvement in GCG. Existing frameworks and typologies of participation focus on participatory forms and do not consider the quality of involvement, which is a highly determining factor for meaningful involvement. According to an assessment of transformative participation, such as youth participation in Article 6 negotiations, can be meaningful. Lower forms, in contrast, gradually decrease in meaningfulness and nominal forms, such as side-events organized by youth and plenary interventions, are not meaningful. However, as this research shows, the reality is not that clear-cut, and in-depth, on-site observation and research are necessary to study the extent, quality, and influence of youth participation in GCG in a more comprehensive manner. This thesis is of high relevance to discussions in GCG, in theory and practice, as well as youth and social movement studies, and can support policymakers' and negotiators' understanding of youth participation and its implication in other governance areas.

Keywords and phrases: *youth participation, global climate governance, multilateral negotiations, meaningful involvement, non-state actors*

Contents

Acronyms

COP Conference of the Parties.

GCG Global Climate Governance.

IYCM International Youth Climate Movement.

NGO Non-Governmental Organization.

NSA Non-State Actor.

SA State Actor.

SB40 UNFCCC Subsidiary Bodies.

SIDS Small Island Developing States.

UN United Nations.

UNCED United Nations Conference on Environment and Development.

UNDP United Nations Development Programme.

UNFCCC United Nations Framework Convention on Climate Change.

UNICEF United Nations Children’s Fund.

YOUNGO Youth Non-Governmental Organizations.

1 Introduction

Devastating fires, record floods, extreme heat waves, severe storms, and other unprecedented weather conditions at the end of the past decade have led the world to debate and negotiate climate change in an attempt to deal with this global challenge. While an increasing number of states and world leaders recognizes the urgency of the climate crisis, collaboration, consensus-seeking, and, above all, efficiency at multilateral negotiations remain difficult (Falkner, 2016; Thew, 2018). Besides ambitious negotiation talks, the amount of greenhouse gases in the atmosphere keeps rising (United Nations Environment Programme, 2019). We find ourselves amid a crisis, somewhat unsure how to tackle it efficiently on a global level, somewhat hesitant to step out of the comfort of the current economic and political systems. In response to states' incapacity or reluctance to pursue effective, collaborative action in preventing irreversible environmental (and thus, human) damage, many young people, in fear of a viable future, have raised their voices to demand action (O'Brien, Selboe & Hayward, 2018).

The Youth Non-Governmental Organizations (YOUNGO) are one of nine civil society constituencies in the United Nations Framework Convention on Climate Change (UNFCCC) (Thew, Middlemiss & Paavola, 2020). YOUNGO gives young people a unique, yet rather limited, opportunity to participate in global climate change negotiations - limited because of their minor status in society and therefore, structural barriers for participation. Thus, this thesis explores how meaningfully involved young people truly are in such negotiations, particularly in the Conference of the Parties (COP) of the UNFCCC. The COP is the central international environmental institution, and is thus at the heart of GCG. Two key contemporary developments make studying meaningful youth participation in GCG highly relevant: the critique on multilateralism (e.g. Falkner, 2016; Hampson & Heinbecker, 2011) and the tremendous rise and popularity of the youth climate strikes as part of the International Youth Climate Movement (IYCM) (e.g. Dirth, 2019; O'Brien, Selboe & Hayward, 2018; Orr, 2016). Together, these two developments open a window of opportunity to rethink common procedures in multilateral negotiations. They can give an incentive to think of meaningful youth involvement and challenge its narrative in GCG. It is crucial to comprehend the IYCM's role and potential to contribute to negotiation talks, on the one hand, and to challenge the UNFCCC's legitimacy on the other. Besides the relevance for GCG, both in theory and practice, the discussion in this thesis is pertinent for social movement studies and youth studies. It may also bring insights for political scientists, policymakers and negotiators in other areas of governance.

Including the voice of youth in decision-making processes has arguably become a norm in global governance as can be witnessed in growing youth involvement in one form or another in governance processes (Bersaglio, Enns & Kepe, 2015; Sukarieh & Tannock, 2008; Yunita, Soraya & Maryudi, 2018). The value of including youth, particularly in decisions concerning the environment and (sustainable) development, was first brought forward in the 1992 United Nations Conference on Environment and Development (UNCED). However, youth participation can also be interpreted as a necessity in the earlier 1987 Brundtland definition of sustainable development, which highlights the urgency for future generations to meet their needs (Yunita, Soraya & Maryudi, 2018). Youth have been participating in United Nations (UN) climate change negotiations specifically since COP5 in 1999. Ten years later, in 2009, YOUNGO gained constituency status from the UNFCCC secretariat. These developments together indicate that the conception of youth as a "meaningful contributor and agent of change" (Yunita, Soraya & Maryudi, 2018, p. 53) had already gained widespread recognition by the 1990s (e.g. Ward & Parker, 2013). However, such formalities do not guarantee that youth will meaningfully participate in decision-making processes. Specifically, at the global level, multiple barriers (such as lack of opportunities and limited individual power within a large group of actors) restrict or inhibit youth from contributing (UNESCO, 2020; as cited in Yunita, Soraya & Maryudi, 2018, p. 53).

This research adds to a growing literature on youth participation in governing activities (e.g. Määttä & Aaltonen, 2016; Yunita, Soraya & Maryudi, 2018) as well as to the literature on civil society in global environmental governance (e.g. Newell, 2000; Orr, 2006; Rietig, 2016; Vormedal, 2008), where the focus on young individuals lacks attention. The discussion expands on pioneering works (Thew, 2018; Thew, Middlemiss & Paavola, 2020) in the novel field of youth involvement in

GCG. The rise of the IYCM should not be ignored by academics or policymakers. Thus, this thesis seeks to highlight a need for empirical research in youth participation in GCG, both from an inside and an outside approach. The youth climate strike movement has received plenty of attention from the media and the public (Thew, Middlemiss & Paavola, 2020). Its presence and influence should not be separated from a discussion on formal youth participation inside negotiations. Moreover, this research seeks to demonstrate a need for a thorough understanding of the concepts of *participation* and meaningful involvement. Having an assumptive perception thereof runs the risk of assessing the reality of public involvement in political affairs in a misguided manner.

After defining the central terms *youth* and *global climate* governance, the research context provides background knowledge and situates this research within existing scholarly discussions on *Non-State Actor (NSA) participation in global governance* (Albin, 1999; Bulkeley & Newell, 2015; Jordan et al., 2015; Lisowski, 2005; Nasiritousi, Hjerpe & Linnér, 2016; Rietig, 2016; Thew, 2018) as well as within debates on the conceptualization of *'participation'* and *'youth participation'* more specifically (Checkoway, 2011; Ekman & Amnå, 2012; Hart, 1992; O'Donoghue, Kirshner & McLaughlin, 2003; Percy-Smith & Thomas, 2010; Villa-Torres & Svanemyr, 2015; White, 1996; Wong, Zimmerman & Parker, 2010; Yunita, Soraya & Maryudi, 2018). Chapter 4 presents different models of civic participation which allow for a general understanding of different forms and degrees of participation. The subsequent chapters try to dismantle various dynamics and concepts that work within the presented models of participation: power sources, influence and meaningful youth involvement. Chapter 5 acknowledges the main power difference between State Actor (SA) and NSA, i.e. formal legislative and executive power, and explores alternative sources of power used by NSAs. Chapter 6 presents White's (1996) typology of interests that examines nominal, instrumental, representative and transformative forms of participation both from a top-down and a bottom-up perspective. Chapter 7 highlights the importance of meaningfulness in the participation discourse and develops a model that allows for operationalizing the term. Chapter 8 then analyses the case of youth participation in GCG. Data from Thew's (2018) research and insights from discussed literature allow for an analysis and categorization of different forms of participation based on White's (1996) typology of interests. Concluding remarks are given in the final chapter. Finally, the last chapter elucidates concluding thoughts on meaningful youth participation in GCG.

1.1 Definitions

The UN defines youth as people aged 15 to 24 years old (UNESCO, 2020). However, youth is a vastly fluid category that depends greatly on culture, country, and context. Hence, the age limitation is rather flexible; some definitions expand to 35 years (Yunita, Soraya & Maryudi, 2018). In 2018, 16 percent of the world population, i.e. 1.2 billion people, fell within the UN definition of youth (Yunita, Soraya & Maryudi, 2018). Bersaglio, Enns and Kepe (2015) argue that two dominant perspectives on the meaning of youth have developed: youth as a stage of life and youth as a separate social and cultural category (also see Ansell, 2005).

The first, youth as a stage of life, represents the transition from dependency, i.e. childhood, to independency, i.e. adulthood, and from *becoming* to *being* a contributing member of society (see e.g. Adams, 2007). The second, youth as a distinct social and cultural category, focuses on how young people themselves perceive the world, rather than comparing it with adults' experience. This latter perception builds the base for youth agency as it constitutes an intrinsic motivation (Bersaglio, Enns & Kepe, 2015). Essentially, youth, like most group identities, are not objectively real but socially constructed, and are hence best understood through social constructivism (e.g. Ansell, 2005; as cited in Bersaglio, Enns & Kepe, 2015, p. 60). The youth identity can be (re)constructed depending on socio-political, economic, spatial and temporal factors. Thus, the conception of youth is subject to how it is framed much more than to the characteristics of a youth population itself (Bersaglio, Enns & Kepe, 2015). When trying to understand who 'youth' refers to, it is essential to ask "who is talking about youth, in what contexts, and towards what larger economic and political ends" (Sukarieh, 2012, p. 427; as cited in Bersaglio, Enns & Kepe, 2015, p. 60).

Global governance as defined in the first issue of the journal *Global Governance* by Rosenau (1995) is "conceived to include systems of rule at all levels of human activity – from the family to the international organization – in which the pursuit of goals through the exercise of control has transnational repercussions" (p. 13). Global governance evolves from international law in that it considers levels besides the national and international as relevant (Rosenau, 1995). Hence, the capacity to exercise authority, i.e. formal recognition of power, expands from the traditional perception of nation-states to private authority. Governments remain undeniably central actors; however, a diverse range of NSAs can exercise agency and exert influence as well (Jagers & Stripple, 2003). *Global climate governance*, therefore, refers to "all the purposeful mechanisms and measures aimed at steering social systems toward preventing, mitigating, or adapting to the risks posed by climate change" (Jagers & Stripple, 2003, p. 388). Global climate talks are an integral manifestation of global climate governance, where state and non-state entities come together to set goals, pursue policy-making (Jagers & Stripple, 2003), and discuss the challenges of the climate crisis in the twenty-first century.

2 Research Context

The UN landmark agreement reached in Paris in 2015 epitomizes the difficulties faced in reaching agreements, and the inefficiency of these processes in a multilateral, nation-state centric negotiation setting (Falkner, 2016; Thew, 2018). The ambitious Paris Agreement replaced the Kyoto Protocol of 1997 as the main instrument for global emission control. While this treaty is a significant milestone, considering 195 countries reached a compromise, it is vital to bear in mind its non-binding and hybrid nature that may restrain its effectiveness (Falkner, 2016). It is hybrid in a way that it blurs the lines between top-down/bottom-up activity and state/non-state action and consequently complicates issues of legality and voluntarism among other global governance principles (Kuyper, Linnér and Schroeder, 2018). The Paris Agreement, whether considered a success story or a failure to commit to real actions, has led to intensified academic discussion on the role of a growing number of diverse NSAs in international negotiations and global governance (Bulkeley & Newell, 2015; Jordan et al., 2015; Thew, 2018; building on Albin, 1999). Much academic work explores the authority and agency of NSA participation in international and intergovernmental negotiations (e.g. Lisowski, 2005; Nasiritousi, Hjerpe & Linnér, 2016; Rietig, 2016). However, the works which recognize the heterogeneous nature of Non-Governmental Organization (NGO) involvement in GCG are limited (e.g. Betsill, 2008); very few scholars have started exploring smaller, less influential groups of NGOs such as those representing youth (e.g. Nasiritousi, Hjerpe & Linnér, 2016; Thew, 2018).

Considerable research has been conducted on youth participation on local and national levels (e.g. Finlay, 2010; Mitra, Serriere & Kirshner, 2014; Phiri, 2019; Theis, 2007). Some studies have focused on specific governance areas, such as forestry (Yunita, Soraya & Maryudi, 2018), health (Villa-Torres & Svanemyr, 2015; Wong, Zimmerman & Parker, 2010), or disaster risk reduction (Cumiskey et al., 2015). Others have explored the meaning of youth participation in conferences organized explicitly for youth, such as the Global Forum on Youth, Peace and Security (Kwon, 2019). However, very little work has explored youth participation in global environmental and climate governance (Kwiatkowski, 2017; Thew, 2018), and even the concept of participation itself has barely received any attention. What does 'participating' mean? What does participation look like for children and the youth, what are the dynamics between youth and adults, and what are the crucial differences between various forms of participation? These are questions that have been asked by scholars in the fields of community development and children, and youth studies (Checkoway, 2011; Ekman & Amnå, 2012; Hart, 1992; O'Donoghue, Kirshner & McLaughlin, 2003; Percy-Smith & Thomas, 2010; White, 1996; Wong, Zimmerman & Parker, 2010), but have not been transferred to discussions of youth participation in GCG, and have been left out in other global governance areas.

2.1 The complexity of multilateral negotiations

Two decades of multilateral climate change negotiations have instigated much debate on the limitations and challenges of "UNFCCC-style multilateralism" (Falkner, 2016, p. 87), and have brought forward ideas of new and innovative approaches on climate diplomacy (e.g. Brenton, 2013; Giddens, 2009; Victor, 2006; as cited in Falkner, 2016). Scholars have suggested various alternative approaches from “bottom-up policy processes and experimentalist governance to transnational regime complex and multi-actor governance networks” (Abbott, 2012; De Burca, Keohane & Sabel, 2014; Hoffmann, 2011; Vasconcelos, Santos & Pacheco, 2013; as cited in Falkner, 2016, p. 300). Minilateralism, involving many different concepts and ideas, has received considerable attention. It could be a “smarter, more targeted” approach (Naím, 2009; as cited in Falkner, 2016, p. 87) than multilateralism because it includes fewer countries that share similar characteristics, such as economic strength and interests, and hence facilitate consensus seeking. Such an approach is, however, highly controversial in terms of its inclusiveness, representativeness, and universality because it entirely excludes the voices of small countries and those with fewer resources (Falkner, 2016; Hampson & Heinbecker, 2011). Because of the limited nature and exclusivity of minilateralism, NSAs particularly are excluded to an even greater extent compared to multilateral settings; youth have limited agency, influence, and power. Hjerpe and Nasiritousi (2015) find that minilateral approaches that have been attempted in the past have never gained enough momentum, recognition, and legitimacy to replace UNFCCC multilateralism.

Moreover, Hampson and Heinbecker (2011) highlight that present international governance institutions emerged after World War II and that there has been little debate for change ever since. Today’s world, however, is very different from the post-war situation: there are more nation-states, more people, but above all, a greater diversity of NSAs who seek to voice their opinions, desires, and needs (Hampson & Heinbecker, 2011). The Paris Agreement indeed takes a step towards modernizing global governance by assigning significant roles to non-state parties and therefore intensifying, enabling, and constraining formal and informal NSA participation (Kuyper, Linnér & Schroeder, 2018). Restricted public participation fuels “issues of legitimacy, accountability, social justice, and effectiveness” (Hampson & Heinbecker, 2011, p. 299), posing significant challenges to global governance (Okereke & Coventry, 2016). There are two strands of multilateralism regarding the issue of a public participation deficit: on the one hand, traditional multilateralists argue that democracy is not and has never been at the heart of intergovernmental institutions; on the other hand, bottom-up multilateralists believe that the inclusion of civil society is crucial to global governance, especially in an ever more globalized world (Hampson & Heinbecker, 2011). , Whether the inclusion of NSAs in the Paris Agreement facilitates or constrains participation, the mere fact that they are included shows that participation in global governance is changing.

2.2 NSA participation in global climate governance

Despite high recognition of the diverse population in the climate regime complex (Abbott, 2012, 2014; Bulkeley et al., 2012; Kuyper, Linnér & Schroeder, 2018; Nasiritousi & Linnér, 2016; Schroeder & Lovell, 2012), only a few studies have investigated the differences between NSA groups in terms of authority, agency, and influence. Many have explored the influence of dominant environmental NGOs (e.g. Betsill, 2008) and business NGOs (e.g. Falkner, 2010; Vormedal, 2008), but very little attention has been given to marginal groups such as indigenous peoples (Schroeder, 2010) and youth (Thew, 2018). This limitation is detrimental to a nuanced understanding of NSA authority as it runs the risk of exaggerating their positions and influence by assuming them to be identical to other, stronger NSAs (Betsill, 2008; Nasiritousi, Hjerpe & Linnér, 2016; Thew, 2018). Building on findings from Betsill (2008), Nasiritousi, Hjerpe and Linnér (2016) and Thew (2018) propose that the availability of various power sources is significant for NSA influence, although not all constituencies have the same access, particularly to material and social power sources. For example, Betsill (2008) points out that “pre-established relationships with negotiators” (Thew, 2018, p. 373) elevate social power and can thus allow NSAs to overcome barriers. However, insider connections are a valuable asset that are not naturally and equally available, certainly not to a marginal, rather inexperienced group of young people.

Schroeder and Lovell (2012) identify distinct formal and informal modes of participation of NSAs in the UNFCCC and find that side-events are most valued as NSA participation (also see Hjerpe & Linnér, 2010). Side-events are “a platform for admitted observer organizations, which have limited speaking opportunities in the formal negotiations, to engage with Parties and other participants for knowledge sharing, capacity building, networking and exploring actionable options for meeting the climate challenge" (UNFCCC, 2019, para. 11). On the other hand, Thew (2018) emphasizes a lack of understanding regarding what category of NSAs chooses which modes of participation and why. Nasiritousi, Hjerpe and Linnér (2016) link the concept of power sources with the concept of recognition and argue that "agency can be understood by studying the different roles that actors are perceived to perform" (p. 112). However, the relation of power sources to modes of participation and the ability to gain and exercise agency, especially for smaller groups of NSAs, remain rather unexplored. Nasiritousi, Hjerpe and Linnér (2016) give a valuable contribution by conducting quantitative research at COP17 and COP18. Their research clarifies the relation of participation modes, power sources, and agency. They also address the gap between ego and alter perceptions of recognition, and suggest how such perceptions affect agency.

Generally, NSAs do not have a formal voice within international climate negotiations; however, there are different strategies they use to make their voices heard (Betzold, 2013). Their actions are typically classified as part of inside and outside strategies (e.g. Betzold, 2013; Binderkrantz, 2005; Kollman, 1998). Inside advocacy aims to influence policy and decision-making directly through delegations and representatives. In contrast, outside advocacy seeks to put pressure on decision-makers through media and public opinion (Betzold, 2013). Depending on the access to power sources, constituents use different strategies to exert their influence (Nasiritousi, Hjerpe & Linnér, 2016). Inside advocacy is generally preferred as it is a more direct and straightforward approach; however little or no access to policymakers, i.e. low social and material power, does not allow for inside advocacy (Betzold, 2013).

Thew (2018) pioneered a discussion on youth participation and agency in the UNFCCC, which was followed by another in-depth study on youth "climate justice claims" showing the evolution from "articulated injustices based on perceived future risk for their generation" towards "solidarity claims about injustices experienced by other groups in the present" (Thew et al., 2020, p. 1). Besides these two key studies on youth participation in GCG, scholarly discussion has focused on the ways youth has expressed their dissent, challenged power structures, and expressed agency through activism, (e.g. O'Brien, Selboe & Hayward, 2016) in addition to how emotions such as fear, hope and anger drive actions in the youth climate movement (e.g. Kleres & Wettergren, 2017; Ojala, 2012, 2013). However, little work has explored the impact of youth climate activism on the UN and negotiation talks in the UNFCCC, neither has there been much discussion around the pressure created through the grassroots movements outside the formal negotiation processes. Orr (2016) considers civic engagement, particularly youth activism, at the COP21 negotiations in Paris, and calls for future research to address restrictions on participation by the UN in relation to outcomes of negotiations in the UNFCCC.

This research aims to build a theoretical understanding of the notion of youth participation in GCG. By reviewing diverse literature on participation, power, and interests, it provides an understanding of what meaningful involvement constitutes and explores the current level of meaningful involvement of the youth in global climate talks. Scholarly attention on the nature of multilateral negotiations (e.g. Falkner, 2016; Hampson & Heinbecker, 2011), the tremendous rise of the IYCM (e.g. Dirth, 2019; O'Brien, Selboe & Hayward, 2018; Orr, 2016), and youth participation in recent climate negotiations (e.g. Thew, 2018), particularly after Paris in 2015, are contemporary factors that require a more sophisticated understanding of the quality of youth involvement in UNFCCC multilateralism and GCG in general.

This thesis argues that youth participation in GCG is a struggle over recognition, which poses limits to meaningful involvement. There is a gap in the understanding of how the IYCM operates within formal negotiations as part of a constituency group, and outside as part of global climate strikes. The connection between the two forms of youth participation must be explored to comprehend the space in which youth operates to assert their claims and exercise agency and to assess how meaningful their participation is. Moreover, this research argues that both inside and outside approaches can challenge the traditional, state-centric, multilateral negotiation structures, and both experience struggles over recognition that have implications on youth agency. However, the restricted availability of empirical research on youth participation in

GCG poses a limitation to this research. Nonetheless, the theoretical understanding constructed in this thesis is fundamental to exploring the phenomenon of young NSAs participating in global climate talks and their role in GCG. To achieve a nuanced understanding of this phenomenon, detailed on-site observation and investigation in global climate talks in the future are necessary.

3 Methodology

This thesis employs an in-depth literature review to synthesize academic work on youth participation, power relations, power sources, and stakeholders' interests. The analysis of these concepts facilitates the understanding of meaningful involvement, which subsequently allows for a discussion on meaningful youth participation in GCG. Youth participation in multilateral climate talks has only recently become a topic of academic debate and is thus highly unexplored (Thew, 2018; Thew et al., 2020). Therefore, this research takes a conceptual and definitional approach and attempts to create a basis for further investigation on youth participation in global climate talks. The reviewed literature is from highly diverse academic fields; it blends studies from youth and children studies, youth and community development, and community psychology with research on NSAs in global governance, global environmental and climate governance, as well as political science and policy-making. Through this multidisciplinary approach, insightful notions on meaningful youth participation in GCG can be elucidated.

The underlying assumptions of the research question *How meaningfully involved are young people in global climate talks?* are that (1) generally NSAs play an increasingly vital role as contributors, partners and influencers in multilateral climate negotiations; (2) marginal groups of NSAs have developed strategies to exercise agency and influence decision-making processes; and (3) youth wants to participate actively and meaningfully in decision-making processes in GCG. Thus, this research suggests that certain preconditions empower or disempower youth to participate and contribute meaningfully, and hence, exercise agency in GCG.

This thesis contributes to a growing literature on youth participation in governing activities as well as to the literature on NSAs in GCG, where the focus on young individuals lacks attention. Literature is accumulated and reviewed by searching for secondary sources in the field of governance and policy-making, specifically in global climate and environmental governance, as well as youth and children studies, community development and community psychology. Google Scholar served as an initial database source and led to the focus the following journals: *International Environmental Agreements, Global Environmental Politics, Climate Policy*, and *International Negotiation*. Searched keywords include but are not exclusive to: youth participation, global climate governance, multilateral negotiations, UNFCCC, non-state actors, agency, power sources, participatory strategies.

4 What is participation?

Participation is a broad term that has become a buzzword, and can mean anything and everything that involves people (Cornwall, 2008; O'Donoghue, Kirshner & McLaughlin, 2003; White, 1996). The vagueness and undefined nature of the term can lead to misguided expectations and hence, diminish possibilities for participation, particularly of minor, less established groups of actors (Cornwall, 2008). In order to distinguish various kinds of participation and the quality of involvement, several scholars have deliberated different frameworks of political participation and civic engagement (Arnstein, 1969; Connelly, Smith, Benson & Saunders, 2012; Ekman & Amnå, 2012; Newell, 2005; Price, 1990). More specifically, Thew (2018) identified conference access, side-events, exhibits, plenary interventions, high-level meetings, and actions as the main modes of NSA participation. Others have developed theories and frameworks on young people's participation specifically (Hart, 1992; Kwiatkowski, 2017; Shier, 2001; Treseder, 1997; Wong, Zimmerman & Parker, 2010). Public participation has been associated with higher legitimacy, greater quality of decisions made, and enhanced democratic governance (Checkoway, 2011; Yunita, Soraya & Maryudi, 2018). However, participation not only reflects representation – who is participating? – but also deals with actors' empowerment (O'Donoghue, Kirshner, McLaughlin, 2003; Yunita, Soraya & Maryudi, 2018) – are the voices of the less privileged and more marginal groups being heard?

4.1 Youth participation

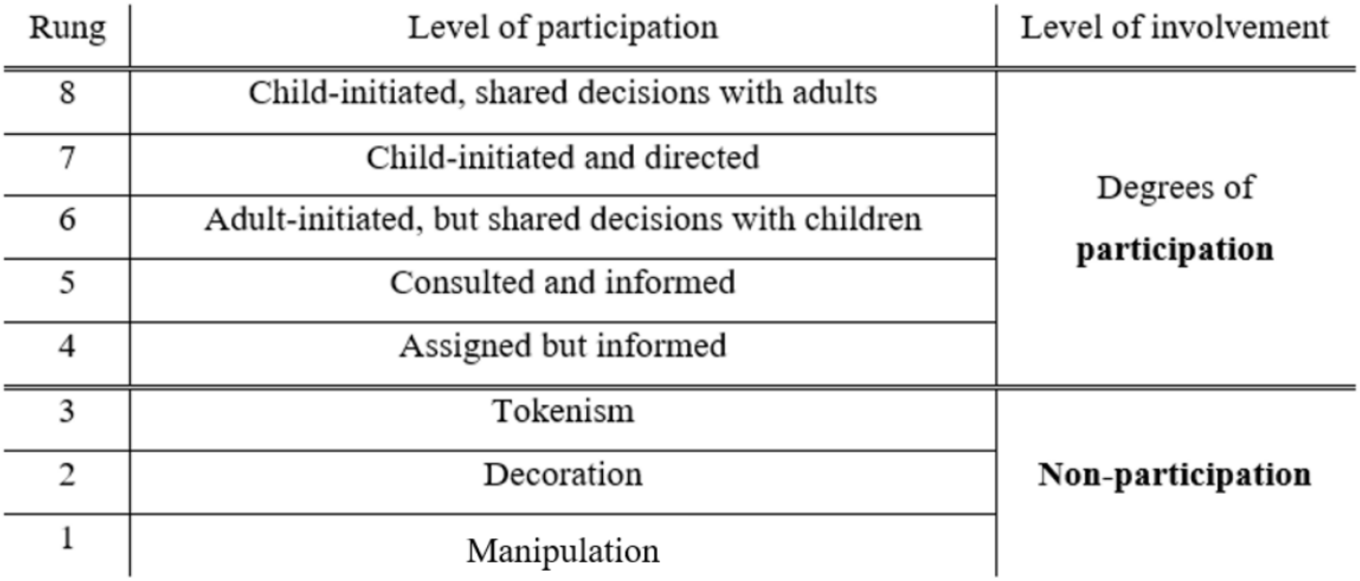
Youth participation is concerned with the involvement and influence of young people on political and decision-making processes (O'Donoghue, Kirshner, McLaughlin, 2003; Yunita, Soraya & Maryudi, 2018) as well as the personal and social development and empowerment of youth (Checkoway, 2011). However, while youth participation implies homogeneity of young people, it is crucial to acknowledge the heterogeneous nature of youth - just as we would do with other actors. For example, we do not perceive nation-states as a homogenous group; there are small, developing as well as big, developed ones and everything on a spectrum in between - the same recognition should be given to the youth. "Young people engage with the public world as individuals, not as representatives of all youth, African American youth, or gay youth, for example" (O'Donoghue, Kirshner & McLaughlin, 2003, p. 21). Percy-Smith and Thomas (2010) suggest that participation should never be considered in isolation or apart from the context in which it occurs. They argue that an interpretation of participation broader than speaking in public decision-making fora may allow for a better contextualization of meaningful involvement (Percy-Smith & Thomas, 2010). The following section examines different frameworks of participation concerning the quality of youth participation, power relations between youth and adults, and recognition as equal constituents, closely related to participants' critical awareness and to youth empowerment.

4.2 Ladder frameworks

Arnstein (1969) defines a *ladder of citizen participation*, which, although coined half a century ago, is still discussed and remodeled by several scholars today (e.g. Cornwall, 2008; Hart, 1992; Kwiatkowski, 2017; Pretty, 1995; White, 1996). Arnstein's (1969) ladder progresses from forms of non-participation and tokenism to higher degrees of citizen power in eight steps. He argues that "there is a critical difference between going through the empty ritual of participation and having the real power needed to affect the outcome of the process" (Arnstein, 1969, p. 216), and that power can only be gained and exercised when those who already have it, share it. Arnstein's (1969) model clearly indicates that participation is inherently linked to power and control.

Hart (1992) adopts Arnstein's (1969) ladder framework to children participation (Figure 1). His work is part of research commissioned by the United Nations Children's Fund (UNICEF) International Child Development Centre. Hart's (1992) ladder ranges from non-participation to degrees of participation, also in eight steps. The lowest rungs are non-participatory: (1) *manipulation*, (2) *decoration*, and (3) *tokenism*. Manipulation and decoration are stages in which children do not understand the meaning and consequences of their actions (as cited in Kwiatkowski, 2017, p. 12). The difference between the two is that in decorative forms, "adults do not pretend that the cause is inspired by children" whereas in manipulative forms they do (Hart, 1992, p. 9). Tokenism emerges when children are seemingly given a voice but that voice is, in fact, highly constrained, unrepresentative, and uninformed. The voices of a few children are used as a token to prove youth involvement in decision-making, which often is the case at conference panels (Hart, 1992).

Towards the upper part of the ladder, on degrees of participation, Hart (1992) defines (4) *assigned but informed*, which is the first stage towards children being aware of their actions and the intention behind them. Children move from a decorative to a meaningful role (Hart, 1992). For example, young people could be involved in a conference by being assigned to specific negotiators, diplomats or politicians and follow them around - in that way they would learn about the processes in decision-making; they would, however, not be able to actively contribute with their ideas. In the next stage, (5) *consulted and informed*, children's opinions are given serious consideration. They can be involved in a project and used as a resource that is taken seriously and informed of the outcome resulting from their input. Hart's (1992) framework differs from Arnstein's (1969) in so far that Arnstein (1969) defines a category between participation and non-participation: (3) *informing*, (4) *consultation* and (5) *placation* are degrees of tokenism. For Arnstein, in the degrees of tokenism, citizens have a voice; however, they lack power and are thus constrained from full participation. According to Hart (1992), the sixth rung, (6) *adult-initiated*,



Rung	Level of participation	Level of involvement
8	Child-initiated, shared decisions with adults	Degrees of **participation**
7	Child-initiated and directed	
6	Adult-initiated, but shared decisions with children	
5	Consulted and informed	
4	Assigned but informed	
3	Tokenism	**Non-participation**
2	Decoration	
1	Manipulation	

Figure 1: Ladder of participation, as defined by Hart (1992, p. 8)

shared decisions with children, is true participation because the decision-making is shared between adults and children. Arnstein's (1969) rung (6) *partnership* communicates similar processes in which power is redistributed to some degree. The last two stages of Hart's (1992) ladder are both projects by children: (7) *child-initiated and directed* and (8) *child-initiated, shared decisions with adults*. Children become the main power holders, adults serve as support figures, and decision-making powers are shared equally between children and adults. However, projects on the highest rung of the ladder are rare because of "the absence of caring adults attuned to particular interests of young people" (Hart, 1992, p. 14).

Hart's (1992) ladder framework suggests a progression from adult to youth control and implies that youth-driven participation is ideal. Treseder (1997) challenges Hart's (1992) model by placing the five forms of participation in equal nodes, thus indicating nonlinearity. Treseder (1997) argues that youth-driven participation may be inappropriate in some cases and that the ideal form of participation varies based on the context and circumstances of the discussion. However, he does not address forms of non-participation, as discussed by Hart (1992). Because of this, Hart's (1992) linear model provides more structure and hence, is used as a reference model in this paper.

4.3 TYPE pyramid

Wong, Zimmerman and Parker (2010) present a model that uses a theoretical framework of empowerment to analyze five types of youth-adult participation: the TYPE pyramid. Critical awareness or critical consciousness, i.e. the ability to perceive critically one's position and surroundings, including broader social and historical implications, is created through empowerment and, thus, is crucial to the empowerment process (Wong, Zimmermann & Parker, 2010; cf. Freire, 1970/2003). The TYPE pyramid framework suggests that an egalitarian approach to critical consciousness from both youth and adults may empower both sides. While adults' involvement may often be necessary for supervision or support, the responsibility of creating space to develop critical consciousness and a sense of empowerment lies with both youth and adults based on a shared co-learning relationship (Wong, Zimmerman & Parker, 2010). This framework distinguishes between the variables of engagement and control. It suggests that engagement can be initiated in three different forms: by adults (adult-driven), by youth (youth-driven) or together (shared control). However, the degree of control can vary within those forms. Therefore, the TYPE pyramid encompasses five types of participation (Figure 2): *vessel* and *symbolic* (adult-driven), *pluralistic* (shared control), and *independent* and *autonomous* (youth-driven).

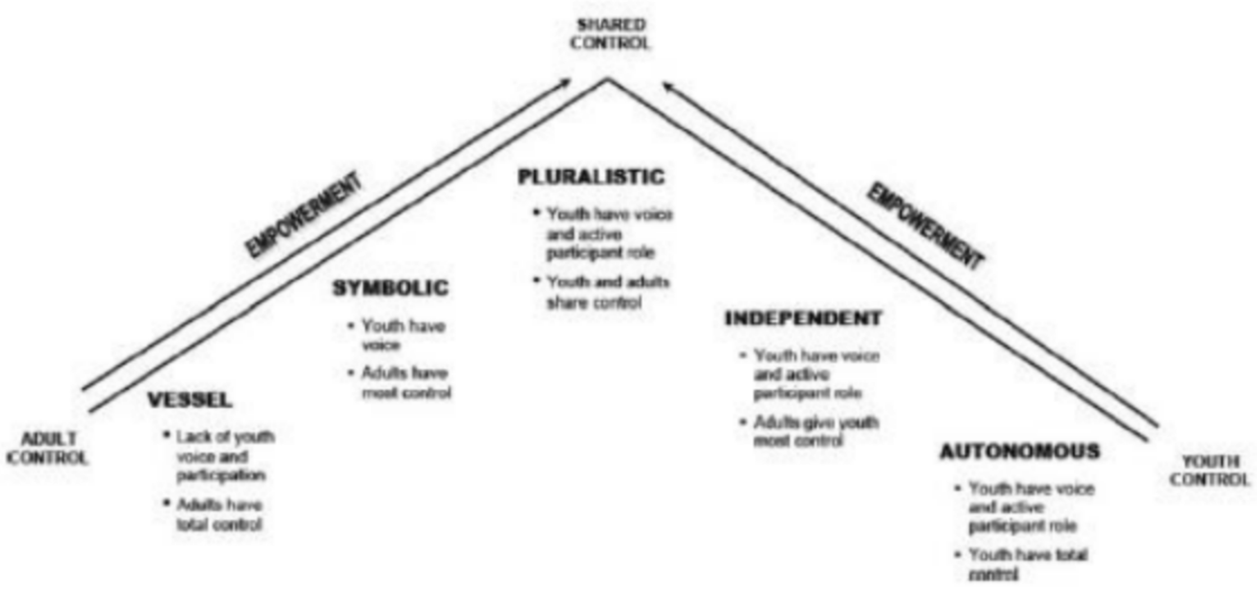

SHARED CONTROL
EMPOWERMENT
EMPOWERMENT
PLURALISTIC
• Youth have voice and active participant role
• Youth and adults share control
SYMBOLIC
• Youth have voice
• Adults have most control
INDEPENDENT
• Youth have voice and active participant role
• Adults give youth most control
VESSEL
• Lack of youth voice and participation
• Adults have total control
ADULT CONTROL
AUTONOMOUS
• Youth have voice and active participant role
• Youth have total control
YOUTH CONTROL

Figure 2: TYPE pyramid (Wong, Zimmerman & Parker, 2010, p. 105)

Similar to the ladder framework, Wong, Zimmerman and Parker (2010) argue that adult-driven participation can lead to “decoration, manipulation, or tokenism” of young people. In some cases, adult-driven participation modes can be beneficial as they are accompanied by support and expertise. Often, however, when young people are not involved or able to contribute meaningfully, they express frustration and anger (Wong, Zimmerman & Parker, 2010); this potential reaction makes adult-driven participation detrimental to youth empowerment. The participation type *vessel* is driven and controlled by adults and requires little to no participation from young people. *Symbolic* participation is also adult-driven; young people’s opinions are voiced but not truly considered. On the other end of the pyramid are youth-driven participation types: Adults who provide space for youth-driven and youth-led projects consider young people to be valuable assets, able to contribute meaningfully with creative and innovative ideas (Camino, 2000; Larson, Walker & Pearce, 2005). Independent participation allows for plentiful opportunities for youth to participate and contribute with little guidance and control from adults. In the *autonomous* participation type, youth “create their own spaces for voice, participation, and expression of power regardless of adult involvement” (Wong et al., 2010, p. 110). An example of autonomous participation is youth gangs (Wong et al. 2010, p. 110): they organize independently and make their own decisions, however they might feel somewhat out of their depth at some point as they lack experience, expertise, and connections.

At the top of the TYPE pyramid is *pluralistic* participation controlled by both youth and adults. However, Wong, Zimmerman and Parker (2010) emphasize that shared control does not mean shared decision-making powers on each aspect. Instead, it may be more appropriate for either youth or adults to contribute ideas and have decision-making powers depending on the specific topic and their strengths and interests.

> It may, for example, be advantageous for youth to brainstorm new ideas and adults to recommend a timeline and procedure for carrying out the ideas. In this situation, youth might come up with ideas that adults may not have considered whereas adults can draw upon experience to suggest how long the idea will take to implement, strategies for implementation and where to find resources. (Wong, Zimmerman & Parker, 2010, p. 108)

The *pluralistic* approach considers both adults’ and youth’s strengths and weaknesses and seeks cooperation and partnership to achieve optimal results. This participation type is similar to Hart’s rung (6) *adult-initiated, shared decisions with children* or Arnstein's rung (6) *partnership*. The TYPE model indicates that cooperative, partnership or pluralistic approaches may be ideal for the development and empowerment of young people, and best for the advancement of the community. This stands in opposition to the implication of Hart's (1992) ladder framework, which suggests youth-led participation as ideal.

When talking about participation, and youth participation in particular, it is crucial to understand the following: First, participation is a broad-encompassing and highly political term. 'To be participating' only has meaning when it is investigated and we know who is participating, as well as how, where and why (e.g. Cornwall, 2008; White, 1996). Second, forms of participation do not have to be linear with absolute power of either adults or youth on either end of the spectrum, as implied in Arnstein's (1969) and Hart's (1992) frameworks, but can be conceptualized in different ways, such as the pyramid model suggested by Wong, Zimmerman and Parker's (2010), or even less hierarchical, in equal forms of participation (Treseder, 1997). Third, assuming that youth-led participation is optimal can be detrimental to the expectations of young people if they find themselves confined in the social structures of society. The expectation that they can reach their goals and pursue their interests without the support of adults can be highly discouraging (Kwiatkowski, 2017; Treseder, 1997; Wong, Zimmerman & Parker, 2010).

The debate to consider youth as an equal and valuable constituent is an important one because there is clear potential for young people to be involved and to contribute meaningfully. However, the assumption that young people are entirely equal to adults with regards to their expertise, their experience and their social networks is untrue. Rarely can a child or a teenager know how to plan, coordinate, and manage the way a professional with many years of experience in the field could. This is not to say that age is an indicator of productivity and effectiveness; a 50-year old professional who has been working in the field for decades may in fact be limited by a narrow perception on the issue and blindsighted by the opportunities that lie underneath a mountain of challenges. A young individual with much less experience and expertise can look at it with fresh eyes and contribute new ideas and perspectives. However, only with the help of adults with greater experience, resources, and connections can young people execute their ideas in reality. If they lack that support, their ideas will often be left underdeveloped and hence, frustration will emerge.

5 Alternative power sources

Generally, children and youth do not have equal access to the same forms and strengths of power as adults because of their limited experience, minor status, and restricted resources (cf. Wong, Zimmerman & Parker, 2010). What distinguishes all NSAs from SAs is a lack of formal legislative and executive power, which is reserved for states only. Hence, NSAs make use of alternative power sources to exercise agency in global governance, as can for example be seen in the organizing of side-events (e.g. Falkner, 2010; Gulbrandsen & Andresen, 2004; Keck & Sikkink, 1999; Nasiritousi, Hjerpe & Linnér, 2016; Newell, 2000). Such forms of power can stem from "intellectual, [sic] membership, political, and financial bases" (Gulbrandsen & Andresen, 2004; as cited in Nasiritousi, Hjerpe & Linnér, 2016, p. 113). Nasiritousi, Hjerpe and Linnér (2016) build a typology of distinct power sources used by NSAs to gain recognition and authority in global governance. They define *symbolic* (making moral and representative claims), *cognitive* (holding relevant information and expertise), *social* (having access to networks), *leverage* (having access to decision-making processes and influence on key agents within them), and *material* (possessing sufficient financial resources) powers (Figure 3). However, different NSAs hold unique combinations of power sources that shape their position in global governance and thus, their agency (Nasiritousi, Hjerpe & Linnér, 2016). Youth have different access to power sources, and thus, a different potential to exercise agency than business or environmental groups. For example, youth often lack access to networks and hence, lack the opportunity to make use of social power sources.

A noteworthy notion that Nasiritousi, Hjerpe and Linnér (2016) add is *recognition* as an indicator for agency, which explains the distinction of *actors versus agents*. Agents are "actors with authority" (Nasiritousi, Hjerpe & Linnér, 2016, p. 112) who are recognized, and hence have the ability to influence agenda-setting and decision-making pro-

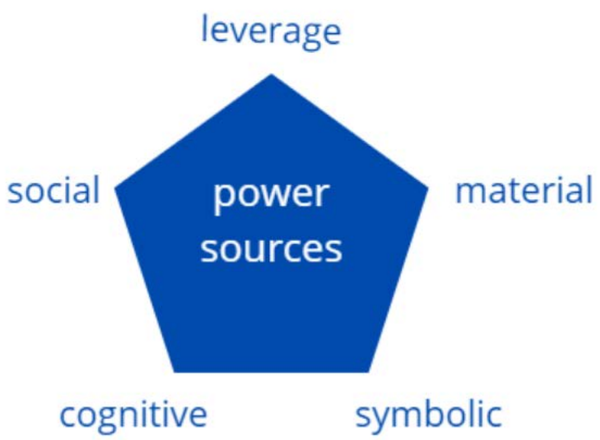


Figure 3: Concept of power sources, as operationalized by Nasiritousi, Hjerpe & Linnér (2016)

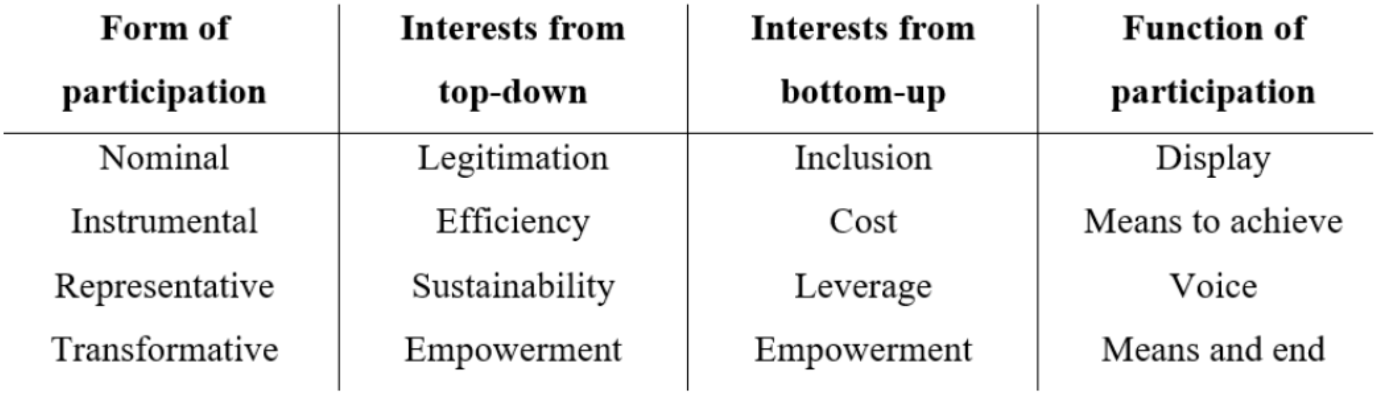
cesses (Dellas, Pattberg & Betsill, 2011). Depending on the governance area, many actors can participate but only a few agents have the ability to influence events (Nasiritousi, Hjerpe & Linnér, 2016). In principle, all actors can gain agency and become agents (Braun, Schindler & Wille, 2019). “What matters for their empowerment is that they receive explicit or implicit recognition for the roles that they claim to fulfil” (Nasiritousi, Hjerpe & Linnér, 2016, p. 114; see Andonova, Betsill & Bulkeley, 2009; Dellas, Pattberg & Betsill, 2011). Thus, agency is empowered by a strong set of power sources and the recognition of authority.

6 Interests

An alternative perspective to examine NSA agency is by analyzing the intentions and interests of the participation. The objectives of people to participate (bottom-up) and those of institutions to encourage or discourage their participation (top-down) vary depending on the form or degree of participation. White (1996) suggests a framework that maps different forms, interests, and functions of participation (Figure 4). Cornwall (2008) acknowledges this framework as a “useful tool to identify conflicting ideas about why and how participation is being used at any particular stage in the process” (p. 271). White (1996) distinguishes between four forms of participation: *nominal, instrumental, representative*, and *transformative*. She argues that participation is a dynamic process; thus, the distinctions are fluid yet seek to highlight specific differences (White, 1996).

Nominal participation functions as a display, showing that an institution is merely doing something to support its claims and policies but this is not particularly meaningful. Thus, their objective is to appear more legitimate without true contribution. Participants barely fulfill any tasks but are included “to keep their names on the books” (White, 1996, p. 8). For *instrumental* forms, people’s participation is necessary to achieve a goal that is both at interest for the institution and the people. The people’s participation comes at a cost because they have to make time out of their daily lives; however, it is the only way they can achieve what they want. For the institution, on the other hand, it is cost-effective as it does not have to invest less resources; hence, the people’s participation serves for instrumental objectives. This imbalance can be explained by an unequal power distribution between institution and participants. In *representative* participation forms, participants take on more active roles: they voice their own interests, opinions, and ideas, and hence have leverage over the decision-making and policies. By giving the people a more active and influential role, the institution benefits from greater sustainability.

Lastly, *transformative* participation embodies empowerment. “The idea of participation as empowerment is that the practical experience of being involved in considering options, making decisions, and taking collective action to fight injustice is itself transformative” (White, 1996, p. 8). White (1996) highlights that empowerment is often considered something for and from the people. It involves bottom-up action; however, “outsiders can only facilitate it, they cannot bring it about” (White, 1996, p. 9). She argues that initially empowerment is not a bottom-up but rather a top-down interest because participants have more immediate and concrete goals as interests. Only through experience do participants start to consider empowerment as a factor also in their interest. Therefore, in transformative forms, participation is a two-fold concept: it is used to achieve empowerment and at the same time its goal is to be empowered, hence it breaks down “the division between means and ends” of empowerment which distinguishes it from the other forms (White, 1996, p. 9). This framework implies a linear progression of people’s voice, active participation, empowerment, and recognition;



Form of participation	Interests from top-down	Interests from bottom-up	Function of participation
Nominal	Legitimation	Inclusion	Display
Instrumental	Efficiency	Cost	Means to achieve
Representative	Sustainability	Leverage	Voice
Transformative	Empowerment	Empowerment	Means and end

Figure 4: Typologies of interests, defined by White (1996, p. 7)

however, it is different from the linearity suggested by Hart (1992) in the ladder framework because it differentiates between the stakeholders' interests rather than their power. White's (1996) typology is more nuanced compared to the ladder framework because it considers both parties' interests simultaneously rather than focusing on who has more or less power at a given place on the spectrum.

In the same article, White (1996) emphasizes that this one-dimensional typology of interests does not allow for a complete and inclusive analysis of real-life situations because it fails to present overlaps and dynamic intersections. Therefore, she presents the following model (Figure 5) which contains the same information as the table but arranged differently. Groups of people are commonly portrayed as homogeneous, for example European youth, however, in reality they are diverse with different expectations and interests. Such differences can create internal tensions around decisions like how to proceed and whose interests to prioritize. Moreover, there can be conflicts between the form and function of participation. Besides the actors and dynamics in the context, other influential actors and power holders may lay claims in accordance with their interests from the outside and hence, diffuse the model entirely. In practice, rarely anything can be placed in black and white boxes; pure forms are theoretical manifestations. The figure seeks to portray all the possible dynamics. Nonetheless, the original framework serves as an analytical tool. By setting out clear divisions, the distinctions between different forms of interests should become more comprehensive (White, 1996). Hence, the analytical strength of the framework should not be undermined.

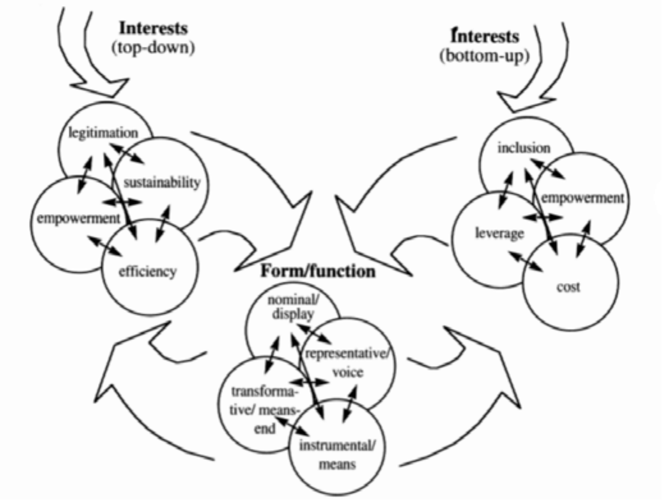


Figure 5: The politics of participation (White, 1996, p. 10)

7 Meaningful youth involvement

Initially, the obstacle to youth participation was a question of quantity - why should youth (out of all actors) be involved (Kara, 2013). Now, young people are more present; they are participating in decision-making processes in different forms and degrees on local, regional, and global levels, as shown by the founding and accreditation of YOUNGO and other organisations. Thus, the question has become an issue of quality – how can youth be involved meaningfully, and what characterizes meaningful involvement (e.g. Finlay, 2010; Hart, 1992; Kara, 2013; Kwiatkowski, 2017). The United Nations Development Programme (UNDP, 2014) states that meaningful youth political participation can occur in the following forms:

1. *Consultative* participation: young people

have a voice in an adult-assigned project. They are informed and aware of the intentions of their actions

2. *Youth collaborative* participation: young people are involved in decision-making processes together with adults

3. *Youth-led* participation: young people initiate projects and have direct impact on the decision-making processes

This definition by the United Nations Development Programme (UNDP) aligns with Wong, Zimmerman and Parker's (2010) argument that meaningful youth participation can not only be found on top of the ladder but can take on different degrees of participation. Kwiatkowski (2017) argues that young people are not the only actor group seeking meaningful participation and thus, youth-initiated and youth-led decision-making is highly unrealistic, impractical and perhaps inappropriate (cf. Villa-Torres & Svanemyr, 2015); particularly, in the context of GCG where many different actors such as Small Island Developing States (SIDS), NGOs, indigenous peoples, and least developed countries struggle to exercise agency alongside big, influential power holders. Wong, Zimmerman and Parker (2010) suggest that a collaborative, pluralistic approach between youth and adults "provides optimal levels of youth empowerment and positive youth development" (as cited in Kwiatkowski, 2017, p. 12). This resonates with the above-mentioned quality (2) *youth collaborative* participation defined by the UNDP, rung (6) *adult-initiated, shared decisions with children* of Hart's (1992) ladder of participation, and *transformative* participation of White's (1996) typology.

By comparing these different approaches that analyze young people's participation, it becomes clear that there is a consensus regarding what type of engagement is most productive and meaningful for all parties involved. Young people need to be taken seriously and recognized in order to feel empowered to contribute; they need to be given the freedom to make decisions on 'trial and error' to gain experience, and they need access to adults' expertise and resources to do this. However, when youth is merely consulted and not taken seriously, they will not be able to thrive personally nor will they be able to contribute meaningfully to the wider discussion among adults. When adults acknowledge the potential resource that lies in young people's participation, and power is shared with them, meaningful participation can unfold that benefits all parties. This is the understanding that all of the reviewed models share; they use different typologies and frameworks, but the underlying meaning is very similar and ultimately leads to more meaningful involvement of the youth.

7.1 What does meaningful participation mean to young people at COPs?

In her master's thesis, Kwiatkowski (2017) asks young people at COP22 in Marrakesh what meaningful youth involvement means to them. She consequently identifies the following criteria:

Recognition as equal political actors
Being recognized as an equal constituent is a precondition for being taken seriously and hence, being able to contribute meaningfully (Kwiatkowski, 2017). Such recognition also comes with the understanding that young people want to contribute with their perspectives and ideas on various topics rather than only being recognized concerning youth issues, i.e. issues directly concerning and affecting young people specifically (Kwiatkowski, 2017).

Influence on policy and agenda
Young people who have a voice and are listened to is a good starting point; however, when it stops there, their voices, opinions, perspectives, and ideas do not act as meaningful involvement but rather as decorative or manipulative sideshows. The youth want to contribute to policy and decision-making processes (Kwiatkowski, 2017). Many young people themselves emphasize that the quality of involvement is much more important than having a great number of young people present – they prioritise meaningful participation over recognition and forms of tokenism. Young people "can do many things, but if they do not have any impact directly on the process, then they are not engaged in the process" (Kwiatkowski, 2017, p. 48).

Access to the decision-making table
Young people consider their involvement meaningful when they have access to the negotiating table. Decision-making at COPs is performed by the parties of the UNFCCC, i.e. state representatives. Young people, alongside many other NSAs, do not have an official seat at the table. One participant in Kwiatkowski's (2017) qualitative research sug-

gests that “physical access could be guaranteed through the inclusion of Youth Delegates in the official countries delegations” (p. 48). Physical access to decision-making processes is a crucial element of meaningful youth involvement because it represents power.

Influence beyond COPs & networking Lastly, many participants emphasize that meaningful involvement is broader than being able to contribute meaningfully in the COP negotiations: it should produce influence beyond the official negotiations and establish a social network. The ability to exercise influence beyond COP meetings is closely linked to legitimacy and authority. Important, international events, like the COP, create ideal networking opportunities between countries, actor groups and generations. A precondition, therefore, is that the youth are recognized as equal constituents.

7.2 Elements of meaningful involvement

Based on the conceptual understanding of meaningful participation and the empirical data collected by Kwiatkowski (2017), it appears that ‘to be participating meaningfully’ is characterized by several elements. The underlying concept is *recognition*. However, there seem to be two different kinds of recognition in this context: recognition as equal partners (Kwiatkowski, 2017) and recognition as authoritative (Nasiritousi, Hjerpe & Linnér, 2016).

While there is a clear difference between equality and authority, in the instance of public participation, the authority of one actor requires a sense of equality with others. NSAs cannot be authoritative if considered marginal in status. Thus, this research defines the concept of recognition to encompass both equal and authoritative forms of recognition for agents. Recognition is a prerequisite for the transition from actor to agent (Thew, 2018). Once one has been recognized, it is critical whether one has *influence* on policy, decision-making, and agenda-setting at the conference as well as beyond it (e.g. Kwiatkowski, 2017). Meaningful participation depends on the power relations between actors and hence, on the level of *empowerment* marginal groups experience. Power and empowerment are closely examined in the conceptual frameworks of participation, both in the ladder and pyramid approaches, and thus, depend on the degree of participation. Therefore, one could argue that a conceptual understanding of meaningful participation starts with the degree and type of participation. This thesis suggests the following elements of meaningful involvement as part of a codependent relationship illustrated in a triangular model: recognition, influence, and empowerment (Figure 6).

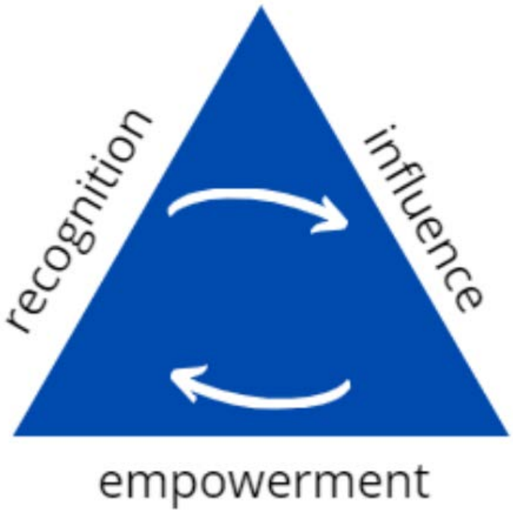


Figure 6: Co-dependent triangle of meaningful involvement

8 Case analysis: youth participation in global climate talks

The above literature review on participation, power, interests, and meaningful involvement leads to the following propositions as an answer to the research question: *how meaningfully involved are young people in GCG?* Genuine meaningful involvement can only be achieved through transformative participation. Lower forms of participation gradually reduce the meaningfulness of the engagement; hence nominal participation does not result in meaningful involvement. The following section first briefly presents how young people are currently participating in global climate talks. Second, it discusses general interests of youth participation from both youth and institutional/adult perspectives. Third, it analyzes youth participation in GCG based on White’s (1996) typology of interest.

8.1 How are young people participating at global climate talks?

Young people have shown their interest in GCG already in the early days of UNFCCC climate negotiations in the 1990s (Fisher, 2016; Orr, 2016; Thew, 2016). However, it was not until 2005 in Montreal, Canada, at COP11 of the UNFCCC that youth delegates in the international climate negotiations started organizing themselves by establishing the International Youth Delegation (International Youth Climate Movement, 2020). Over the course of two decades, youth organizations have established themselves into groups with similar perspectives and objectives, building the basis for becoming a constituency in the UNFCCC. In 2009, youth officially gained constituency status as YOUNGO which opened an array of systematic communication channels with the secretariat and the Parties (UNFCCC, 2010). Today, the International Youth Delegation is called the International Youth Climate Movement (IYCM) because it has expanded its participation far beyond youth delegations, such as in the forms of capacity building, training and education. Its newest, and to the public most visible form is the climate strike movement (Dirth, 2019).

After the stalemate of Copenhagen, when NSA participation in the UNFCCC was highly uncertain (Kuyper, Linnér & Schroeder, 2018; Orr, 2016), many young people became aware of the sense of urgency in the climate discussion with decisions potentially jeopardizing their future and the future of their children. Youth acknowledged their responsibility to act and expressed their dissent through climate activism (Escobar, 2015; O'Brien, Selboe & Hayward, 2018). Fridays for Future strikes started in 2018 with the 15 year old Greta Thunberg, and have since generated a mass mobilization of young activists across the globe (Fisher, 2016; O'Brien, Selboe & Hayward, 2018) with a multitude of narratives emphasizing the immediate threat to people and their well-being (Dirth, 2019).

The general approach of the decision-making parties, as well as of the observing NSAs on the inside of the conference, is deliberative and consensus-based participation (Kwiatkowski, 2017). This approach is built on the principles of deliberative democracy in which equal participation and non-confrontational communication are used to strengthen relationships between parties (Dryzek, 2005). By refraining from the use of coercion, threat, and *power over* (cf. Hayward & Lukes, 2008), deliberative democracy theoretically leads to the most just outcomes through rational consensus-seeking (Young, 2002; as cited in Kwiatkowski, 2017, p. 15). In practice, however, structural inequalities based on economic power prevail; thus, privilege and opportunity are not equally available to all (Young, 1990).

The agonistic approach, on the other hand, is based on agonistic pluralism, which considers public disruption a positive tool to address political conflicts (Mouffe, 1999). Agonistic pluralism believes that structural inequalities have a profound impact on democratic processes that marginalize weaker parties and hence feed into injustice (Young, 2003; as cited in Kwiatkowski, 2017, p. 16). Mouffe (1999) argues that an agonistic approach "acknowledges the real nature of its frontiers and recognizes the forms of exclusion that they embody, instead of trying to disguise them under the veil of rationality or morality" (p. 757). However, there is also a "conflictual consensus" between parties that follow previously agreed upon procedural guidelines (Hansen & Sonnichsen, 2014; as cited in Kwiatkowski, 2017, p. 16). Conflictual consensus "is based on a general conformity to a set of ethical and political principles but includes a disagreement amongst the opponents about the interpretation of those principles" (Kwiatkowski, 2017, p. 16). On the outside, young people, and civil society at large often choose to participate with "disruptive, agonistic strategies", i.e. challenging the status quo through activism, particularly when addressing emotional or controversial topics that are avoided during deliberations on the inside (Kwiatkowski, 2017).

In the context of global climate talks, the deliberative path involves young people engaging in lobbying, policy work and youth delegations. The agonistic path, in contrast, involves youth activism inside and outside of the conferences (Kwiatkowski, 2017), such as the school strikes for climate, climate marches and boycotts. Deliberative strategies are generally more meaningful than agonistic ones because they recognize young people as equal and valuable constituents. Participation through deliberative means gives the youth a platform that has the potential to influence decision-making and can therefore be very empowering. Agonistic strategies, on the other hand, hinder youth from being recognized and hence disempowerment and little influence are likely to follow. However, intermediate paths are often more easily accessible and therefore preferred. Kwiatkowski (2017) argues that traditional paths are commonly combined with communications through media, such as social media activism that mobilizes people and calls for action with the use of specific hashtags. Combinations of strategies can allow young people to be disruptive and challenge the status quo, while at the same time being able to achieve recognition, be empowered and eventually influence decision-making. Many youth organizations refer to a three-pillar approach of youth engagement (Figure 7): (1) activism, (2) media and communications, and (3) policy and lobbying (Kwiatkowski, 2017).

activism
media &
communications
policy &
lobbying

Figure 7: Three-pillar approach as suggested by Kwiatkowski (2017)

More specifically, based on participant observation at the 40th Intersessional of the UNFCCC Subsidiary Bodies (SB40), Thew (2018) identifies six modes of NSA participation, of which *conference access, side-events, exhibits, plenary interventions* and *high-level meetings* can be placed in the deliberative path. *Actions* on the other hand are demonstrations performed by YOUNGO; if they have a positive focus they can be deliberative. However, when they have a negative focus, seek to challenge the status quo and attract attention, as they often do, they are part of agonistic strategies.

8.2 Interests in GCG participation: youth & institutions

Young people, like all other constituencies, have various objectives in participating in GCG. From literature on NGO participation in international climate change negotiations (e.g. Betzold, 2013), it can be deduced that the *logic of influence* (Schmitter & Streek, 1999) plays a prominent role. Kwiatkowski (2017) finds in her interviews that many young people participate in global climate governance to influence decision-making and ultimately to create political change. Many negotiators prefer to settle agreements for young people as opposed to including them in the negotiations (Thew, 2018). This sort of mindset is extremely patronizing and ignorant of the interests of the youth; without participating (or at least trying to participate) in the policy-making processes, youth has no power to bring ideas to the table, no power to influence, no power to create positive change. Young people today will be tomorrow's leaders (e.g. Percy-Smith & Thomas, 2010), therefore, they should be given a seat at the negotiating table. It is their future, after all, that is being discussed and that will be impacted by current decision-making. Young people today bring a creative mindset, an astounding capability to mobilize support to an audience much bigger than only their peers, and perhaps most importantly they bring new perspectives. Climate change has been an issue for many decades and it has been discussed over and over again, yet we are still not moving fast enough to a sustainable solution; young people may be able to turn the page by bringing some fresh wind into the room, and by working together with those already sitting at the negotiating table.

Moreover, it is often argued that youth participation enhances personal development and aids the acquisition of social skills and experience (Checkoway, 2011; O'Donoghue, Kirshner & McLaughlin, 2002). Hence, it can be assumed that many young people seek to participate to gain insight and experience, and to expand their knowledge – in order to achieve this, youth needs to feel empowered. However, one might ask: why not participate on a local, more small-scale level rather than in a global governance setting if this is the main objective? If the logic of influence holds, many young individuals may perceive the influence to be greater on a global level. Lastly, besides gaining experience, it can be assumed that young people have an interest in creating and expanding their social and professional networks. Building a network of people and hence, enhancing their social capital can eventually help young people to have a more significant influence (Nasiritousi, Hjerpe, Linnér, 2016) through, for example, collaborations and lobbying efforts from an insider position (Thew, 2018). A prerequisite for enhancing youth's social capital is to be recognized, which is, besides influence and empowerment, the third crucial element of meaningful involvement (cf. section 7.2). These are, however, all assumptions derived from the general discourse on NSAs, and further research needs to explore them concretely on young population groups.

While youth themselves benefit from being actively involved in GCG, local and global systems also have much to gain from including the youth. The main objective for supporting and encouraging youth participation lies in the "perceived ineffectiveness of past methods" (Yunita, Soraya & Maryudi, 2018, p. 53; Abelson et al., 2003). Greater public involvement is associated with higher legitimacy (Cornwall, 2008) and better, more democratic and sustainable solutions (Checkoway, 2011; Yunita, Soraya & Maryudi, 2018). Involving young individuals of civil society somewhat expands this logic. Another objective to support youth involvement could be taken from findings that link meaningful youth involvement with increasing organizational development, greater commitment and more effectiveness (Zeldin, McDaniel, Topitzes & Calvert, 2000; as cited in O'Donoghue, Kirshner & McLaughlin, 2002, p. 18). "Individuals participate, organizations develop, and communities change" (Checkoway, 2011, p. 343). Checkoway (2011) argues that the highest potential for empowerment occurs when all three levels, individual, organizational, and community level are involved (cf. Schulz, Israel, Zimmerman & Checkoway, 1995). Hence, one could argue that the more diverse the individuals in age, cultural and societal background, the more diverse the developments and solutions are at the organizational and community level. If the improvement of policies and climate solutions is the goal, then including young people in political processes would arguably be a logical step to work towards such a goal.

8.3 Analysis: forms of participation

This section analyzes youth participation in GCG based on White's (1996) typology of interests: nominal, instrumental, representative, and transformative participation, as defined in chapter 6.

Nominal participation

The UN Action Plan *Agenda 21*, which was produced at the Earth Summit in Rio de Janeiro in 1992, states in Chapter 25 (as cited in UNFCCC, 2010):

> It is imperative that youth from all parts of the world participate actively in all relevant levels of decision-making processes because it affects their lives today and has implications for their futures. In addition to their intellectual contribution and their ability to mobilize support, they bring unique perspectives that need to be taken into account. (p. 5)

This sounds very promising and is still applicable today, nearly thirty years later. However, it does not truly describe the realities of youth participation in the UNFCCC, not in 1992 and not today. First, *youth*, as defined by the UN General Assembly, refers to persons between 15 and 24 years of age (UNFCCC, 2010). However, according to the *Guidelines for the participation of representatives of non-governmental organizations at meetings of the UNFCCC*, representatives shall be older than 18 (UNFCCC, 2010). Thus, in fact, the call for active participation in decision-making processes applies to young adults and not to younger youths and children. This limited conception of youth stops a majority of young people from having a platform to be heard in decision-making processes. Moreover, empirical research on youth participation in global climate talks shows that many negotiators consider it crucial to reach a global climate agreement *for* young people and generations to come, but do not deem youth to be a necessary part of the process (Thew, 2018). In such instances, young people are not recognized and cannot become agents to exercise agency meaningfully. Mainly, they are believed not to hold cognitive power based on their age and societal status. These dynamics point towards *nominal* participation forms and resemble Hart's (1992) steps (1) *manipulation* and (2) *decoration*.

More concretely, plenary interventions (one of the modes of NSA participation defined by Thew, 2018) are perceived as *nominal* because, as one participant elaborates, "Interventions show our interest but Parties don't listen. The format isn't interactive, they don't even listen to each other's and ours are at the end when everybody has left," (Thew, 2018, p. 379), implying that they are not recognized and do not have any true potential to influence. Because of the lack of recognition, plenary interventions are perceived to offer little opportunity to exercise agency and hence, do not provide space for meaningful participation (as defined in section 7.2). Another mode of participation is side-events. However, while side-events are good opportunities for agency and are found to be highly valued amongst NSAs (Schroeder and Lovell, 2012), they require high material power, and are thus not equally available to all constituents (Thew, 2018), a fact that epitomizes the ignorance of NSA heterogeneity in global governance literature.

At SB40, YOUNGO hosted one side-event, but Thew's (2018) research shows that out of all her interviewees only youth participants and those working within Article 6[1] negotiations had attended. This suggests a "lack of recognition of YOUNGO's Cognitive Power outside of Article 6 policy" (Thew, 2018, p. 383). Side-events can thus be understood as *nominal* forms of participation for youth. Theoretically, the institution provides the opportunity for meaningful participation, but either there are structural barriers that do not allow youth to organize an event (which, according to the agonistic approach to participation, marginalizes weaker parties and feeds into injustice, as discussed in section 8.1), or participants and decision-makers do not consider it a priority to attend a side-event organized by youth. In either scenario, whether lacking resources constrain meaningful participation or sufficient resources yield neither influence nor any sort of recognition, youth participation is meaningless here.

Instrumental participation

Often the argument is brought forward that today's youth are tomorrow's leaders and should, therefore, be included in political processes (e.g. Percy-Smith & Thomas, 2010). This may be true, but the interest in involving youth in such instances is often instrumental; youth's voices that show support for an institution's policies are merely gathered together to lend it a stronger, more democratic position (Checkoway, 2011; Cornwall, 2008; Yunita, Soraya & Maryudi, 2018) without true youth involvement. One negotiator claims, "An innovative youth movement will always have an impact, we expect new ideas from them" (Thew, 2018, p. 378). If the value of young people's cognitive power (as defined in chapter 5) is indeed recognized, then such participation could be *representative* or even *transformative*. However, if it is simply to demonstrate that youth involvement is valued, while in reality, the youth get no true opportunity to have an impact or to be empowered, then such participation is *instrumental* to the image of the institution and does not provide a platform for meaningful participation of the youth.

Instrumental participation comes at a cost for youth. An example thereof are actions, i.e. demonstrations inside the negotiation halls with the aim to attract attention from decision-makers and media (Thew, 2018). Often actions are negatively perceived by powerholders, particularly when constituents feel there is no other effective mode of participation available to them. Young people then tend to focus on their social and symbolic powers (as defined in chapter 5) to question the legitimacy of the institution (Thew, 2018). However, it is cognitive power, i.e. giving concrete and focused policy suggestions, that is perceived much more positively from decision-makers than utilizing symbolic and social powers (Thew, 2018). "Sometimes youth can be too aggressive, they should push the limits creatively, not just say you need to think about the future," states one negotiator (Thew, 2018, p. 380), implying that youth tend to use symbolic over cognitive powers and therefore, fail to gain recognition. Nonetheless, youth attempt to participate in actions and to make representative claims – not only on the inside of the negotiation halls but also outside. In such instances, young individuals perceive their own agency as weak and their participation as *instrumental*. They are physically present but feel that there is no space for meaningful action on their part. "We need to do something drastic. There is a system problem. This is a flawed process", states one of the youth participants from Thew's (2018) research (p. 381). Ultimately, youth resort to social and symbolic powers and 'disruptive' participatory strategies (as discussed in section 8.1), attempting to challenge the UNFCCC's legitimacy and potentially creating enough public momentum to disrupt the negotiations with the help of the media (Thew, 2018). However, this runs the risk of losing credibility (Thew, 2018).

NSAs fulfill a role as 'watchdogs' (e.g. Betzold, 2013; Thew, 2018). A walk-out, as done in Warsaw at COP19, for example, is an attempt to raise symbolic powers and is believed to lead to a loss, or at least questioning, of the UNFCCC's legitimacy. However, negotiations continue even without youth, or NGOs, or often even without specific SAs such as SIDS. The absence of certain NSAs and their actions to evoke attention are noticed by the media, but, as they are generally not ex-

[1] Article 6 of the UNFCCC deals with climate change education, training, awareness-raising, access to information and public participation. It was later rebranded as Action for Climate Empowerment (Thew, 2018).

pected at the negotiation tables because of the way NSA involvement has been handled in the past, this often does not extend to the negotiating tables (Thew, 2018). A walk-out demonstrates that youth perceive their participation as *instrumental*, and that resorting to non-discussion-centred methods is, ironically, their only option to be heard. The other previously discussed frameworks of participation (the ladder framework and the TYPE pyramid) do not explicitly discuss instrumental participation. It could potentially be placed on Hart's (1992) step (3) *tokenism*, but that does not completely capture that this participation comes at a cost for youth and is only a means to achieve their goal because they see no other viable alternative.

Representative participation

At SB40, YOUNGO followed Article 6 discussions because youth felt the conference was particularly open to their involvement and ideas (Thew, 2018, p. 377). Young individuals perceived they had cognitive power that led to recognition in the discussion, which increased their agency. Based on interviews with youth participants and negotiators, Thew (2018) found that when youth felt like they were heard, they were motivated to take on a more active role. In the example of Article 6 negotiations, the propositions made by the youth were not only taken seriously but were also implemented in the final draft. Through the discussions, they developed relationships with Article 6 negotiators (Thew, 2018) and gained leverage power (as defined in chapter 5). Young people are likely to identify this sort of participation as *representative* because they are able to voice their opinion and have leverage over the decision-making process. Some may even have expectations of it becoming *transformative*, as can be deduced from statements of YOUNGO members: "People who work on Article 6 are all friends. They all want the same thing", "I've been following Article 6 as that is something we can act on" (Thew, 2018, p. 377). These statements show that young people feel not only recognized as authoritative and equal partners, but also that they can act and have an influence – they feel empowered. This participation therefore satisfies all three elements of the triangle of meaningful involvement defined in section 7.2, meaning that it is a prime example of meaningful youth participation.

Representative participation aligns with step (6) *adult-initiated, shared decisions with children* on the ladder framework. Step 6 on Hart's (1992) ladder was compared to *pluralistic* participation on top of the TYPE pyramid in section 4.3, which, as suggested by Wong, Zimmerman and Parker (2010), is ideal for both the development and empowerment of young people but also best for the advancement of the community. However, White's (1996) *transformative* participation might come closer to this ideal, pluralistic form than Hart's (1992) step 6 because *transformative* holds empowerment as a determining factor, fueling into meaningful participation (as discussed in section 7.2). Hence, empowerment is what draws the defining distinction between *representative* and *transformative* participation.

Transformative participation

Article 6 negotiations can be perceived as *representative* participation, but if we examine them from another angle that includes empowerment, they are also a prime example of *transformative* youth participation. Youth have used their cognitive power to offer productive input for policy amendments which have been successfully implemented and are recognized by negotiators and decision-makers (Thew, 2018). "In an area where I didn't have a firm view, a good suggestion from YOUNGO about the Doha Work Programme influenced my position" (p. 378) states a negotiator, emphasizing the possible gain for policymakers when including the youth in a meaningful way. (Thew, 2018). Young people perceive the recognition of their cognitive power from decision-makers, which in return empowers them to maintain interest and give focused and innovative policy suggestions. However, as Thew's (2018) research shows, this recognition barely extends to decision-makers who are not directly involved in a specific discussion. According to Thew (2018), policies are negotiated in silos and agency is limited within them. Youth participation in Article 6 negotiations being perceived as *transformative* does not guarantee the same in other policy silos.

Like other frameworks of participation, White's (1996) typology suggests a progression from less to more meaningful participation. This assumption seems to hold, as illustrated in the above analysis, even though the distinctions may blur into each other. However, White (1996) points out that, once contextualized, the quality of involvement becomes more important than its form or degree (also see Cornwall, 2008). "Nominal forms of participation can give citizens a foot in the door if there has been no constructive engagement with them before" (p. 273), whereas "when 'empowerment' boils down to 'do-it-yourself'" (Cornwall, 2008, p. 272), transformative participation results in frustration and becomes rather meaningless as defined in this paper. Thus, a side-event, although perceived to be mostly nominal, is not necessarily meaningless because it might create opportunities for being empowered and recognized, and having influence in the future. This insight highlights the critical need for in-depth observational research on youth participation in GCG because existent frameworks struggle to capture the implications of young people's involvement in global climate talks, specifically after Paris. The Paris Agreement takes a bottom-up approach that specifies the roles of NSAs more concretely and hence, enables and constrains formal and informal NSA participation in new ways (as discussed in 2.1). This context, such as the negotiating policy silo or established relations to other participants and decision-makers, can be more decisive on how meaningful the participation is than the form of participation itself.

What about youth climate strikes? White's (1996) typology assumes that the public is involved within the confinements and knowledge of an authoritative institution. However, in the case of youth participation in GCG, the youth climate strikes have rapidly grown into a form of participation that cannot be ignored. The movement, led and controlled by youth, has created space to address issues they consider important. Thew, Middlemiss and Paavola (2020) show in their research that over time youth have shifted from "emphasizing their own future vulnerability. . . to amplifying the present vulnerability articulated by other stakeholders" (p. 9). This shift was thought to bring about more recognition from policymakers and other NSAs, allowing for more agency and meaningful involvement (as defined in section 7.2) for youth. The youth climate movement, however, lacks adult/institutional consent and guidance (Wong, Zimmerman & Parker, 2010), and thus does not fit within the parameters of White's (1996) typology, but resonates quite neatly with what Wong, Zimmerman and Parker (2010) define as *autonomous* participation, on the right foot of the TYPE pyramid (as discussed in section 4.3). It is a proactive approach that seeks to create new opportunities for policymaking and participation. Hence, the function of this form of autonomous participation is awareness-raising and pressure-building. However, like Wong, Zimmerman and Parker (2010) highlight, youth-led participation runs the risk of leading to frustration and disempowerment when youth have unrealistic expectations and ignore the broader social structures and confinements in which they operate. Moreover, from the perspective of power and the element of recognition, youth climate strikes do not receive recognition from authorities, hence, hindering the transition from actor to agent for young people (as discussed in chapter 5).

9 Conclusion

This thesis draws together different theoretical understandings of participation, power, and interests, and operationalizes *meaningful participation* as a codependent triangle of recognition, influence, and empowerment. Previous work has used the term *meaningful participation* loosely but has lacked a more concrete, definitional understanding. At this stage, it is challenging to assess meaningful youth participation in GCG in detail because of the following reasons. First, empirical research on youth participation in GCG is scarce, and this thesis relies significantly on data from Thew (2018). Second, meaningful youth involvement cannot be generalized as policies are negotiated in silos (Thew, 2018); what holds in one policy area can be entirely different in another. Nonetheless, this research established a conceptual and definitional base to start thinking about the role of youth in global climate talks.

Meaningful youth participation in GCG can be understood as a progression from *nominal* to *transformative* participation, as suggested by White (1996). This research found that nominal youth participation (at a cost and for display), such as in plenary interventions or side-events, does not allow for meaningful involvement. However, transformative (empowered) youth involvement, such as in the Article 6 negotiations, does result in meaningful involvement. These findings resonate with the proposed answers to the research question: *how meaningfully involved are young people in GCG?* in the beginning of chapter 8. However, one must carefully consider the context and circumstances of the participation because these can be more decisive factors in determining whether the participation is meaningful than the participation form itself. Nominal/non-participatory forms can offer an opportunity for setting the base of meaningful involvement, whereas transformative/participatory forms can quickly end up as do-it-yourself projects, resulting in frustration rather than empowerment (cf. Cornwall, 2008).

If participation lacks institutional consent and guidance (because it is youth-led and youth-driven) and thus, lacks recognition of authority, youth are actors who cannot transform into agents. Therefore, youth strikes at global climate talks – which are outside, disruptive participatory strategies, and which rely heavily on symbolic powers – cannot result in meaningful involvement as understood in this thesis. However, symbolic and social powers, as commonly employed in climate activism, are understood to bring about a loss or questioning of the UNFCCC's legitimacy. Participatory forms on the inside, on the other hand, can lead to meaningful participation if youth is recognized, empowered to contribute, and able to provide concrete policy advice (cognitive powers). Cognitive powers are found to lead to recognition and to provide space to exercise agency, as opposed to symbolic and social powers. Thus, there seems to be an opposition between different power sources: symbolic and social powers challenge the legitimacy of UNFCCC multilateralism and cognitive powers allow for meaningful youth participation, implying that if participation is meaningful it will not question the legitimacy of the institution.

As this thesis has shown, current frameworks do not allow for a thorough understanding of youth participation in GCG. There is a great need for empirical research to explore youth's actions, interests, and effects on decision-making in global climate talks further. In recent years, the question has transformed from 'should youth be involved, and why' to 'how can youth be involved productively and meaningfully'. Hence, research needs to direct focus towards this new question of youth involvement in GCG in order to understand ongoing youth participation as well as how youth can be involved more meaningfully in the future, to the advantage of young people, institutions, and the overarching aims of GCG. Moreover, future research should assess the extent to which the findings of linearity and meaningfulness of participation hold, and how different power sources and dynamics define youth agency. The effects of the current Coronavirus crisis, which heavily alter day-to-day business of the UNFCCC (COP26 in Glasgow originally scheduled for November 2020 has been postponed) as well as of the IYCM (strikes have been moved online, for example), will be intriguing to study in terms of meaningful youth participation, as well as overall NSA involvement in global governance.

While youth climate strikes have not been the main focus of this research, their power to mobilize, their highly informed and curious minds next to their endurance and willingness to 'show up' for the health and wellbeing of our planet, today's and future generations is a highly noteworthy development in the field of youth participation. Youth strikes are neither truly new nor revolutionary, but there is undeniably something that gives the youth today a unique character and that could provide them with a rare opportunity to evolve our thinking not only of climate diplomacy but also regarding the inclusion of 'non-conventional' actors in all aspects of governance. Surely, there are issues regarding representation and accountability of such actors that need considerate attention and further exploration into the extent young people can be included more meaningfully within the current structures of global (climate) governance. Nonetheless, this literature review has hopefully instigated some sort of rethinking and reevaluation of how we perceive 'the youth' as an actor, a valuable resource, a partner, and a solution to many of our challenges today.

10 References

Abbott, K. W. (2012). The transnational regime complex for climate change. *Environmental and Planning C: Government and Policy, 30*(4), 571-590.

Abelson, J., Forest, P-G., Eyles, J., Smith, P., Martin, E., & Gauvin, F-P. (2003). Deliberations about deliberative methods: issues in the design and evaluation of public participation processes. *Social Science & Medicine, 57*(2), 239-251.

Adams, A. (2007). *The role of youth skills development in the transition to work: a global review*. Washington DC: The World Bank.

Albin, C. (1999). Can NGOs enhance the effectiveness of international negotiation? *Interna-*

tional Negotiation, 4(3), 371-387.

Andonova, L., Betsill, M., & Bulkeley, H. (2009). Transnational climate governance. *Global Environmental Politics, 9*(2), 52-73.

Ansell, N. (2005). Children, youth and development. London: Routledge.

Arnstein, S. (1969). A ladder of citizen participation. *Journal of the American Institute of Planners, 35*(4), 216–224.

Bersaglio, B., Enns, C., & Kepe, T. (2015). Youth under construction: the United Nations' representations of youth in the global conversation on the post-2015 development agenda. *Canadian Journal of Development Studies, 36*(1), 57-71.

Betsill, M. (2008). Environmental NGOs and the Kyoto protocol negotiations: 1995-1997. In M. Betsill & E. Corell (Eds.), *NGO diplomacy: The influence of nongovernmental organizations in international environmental negotiations* (pp. 43-66). Boston: MIT Press.

Betzold, C. (2013). Business insiders and environmental outsiders? Advocacy strategies in international climate change negotiations. *Interest Groups and Advocacy, 2*(3), 302-322.

Binderkrantz, A.S. (2005) Interest group strategies: navigating between privileged access and strategies of pressure. *Political Studies 53*(4), 694–715.

Braun, B., Schindler, S., & Wille, T. (2019). Rethinking agency in international relations: performativity, performance and actor-networks. *Journal of International Relations and Development, 22,* 787-807.

Brenton, A. (2013). Great powers' in climate politics. *Climate Policy, 13*(5), 541-546.

Bulkeley, H., Andonova, L., Bäckstrand, K., Betsill, M., Compagnon, D., Duffy, R., . . . VanDeveer, S. (2012). Governing climate change transnationally: assessing the evidence from a database of sixty initiatives. *Environment and Planning C, 30*(4), 591–612.

Bulkeley, H., & Newell, P. (2015). *Governing climate change*. Abingdon: Routledge.

Camino, L. A. (2000). Youth–adult partnerships: Entering new territory in community work and research. *Applied Developmental Science, 4*(1), 11–20.

Checkoway, B. (1998). Involving young people in neighborhood development. *Children and Youth Services Review, 20*, 765−795.

Checkoway, B. (2011). What is youth participation? *Children and Youth Services Review, 33*, 340-345.

Cornwall, A. (2008). Unpacking 'participation': models, meanings and practices. *Community Development Journal, 43*(3), 269–283.

Connelly, J., Smith, G., Benson, D., & Saunders, C. (2012). *Politics and the environment: theory to practice*. New York, NY: Routledge.

Cumiskey, L., Hoang, T., Suzuki, S., Pettigrew, C., & Herrgård, M. M. (2015). Youth participation at the third UN World Conference on Disaster Risk Reduction. *International Journal of Disaster Risk Science, 6*, 150-163.

De Burca, G., Keohane, R. O., & Sabel, C. (2014). Global experimentalist governance. *British Journal of Political Science, 44*(3), 477-486.

Dellas, E., Pattberg, P., & Betsill, M. (2011). Agency in earth system governance: refining a research agenda. *International Environmental Agreements. 11*(1), 85-98.

Dirth, E. (2019). *Processes for just future-making: recommendations for responding to the demands of the Fridays for Future movement* (Policy brief 2019-9). Potsdam: Institute for Advanced Sustainability Studies.

Dryzek, J. S. (2005). Deliberative democracy in divided societies: alternatives to agonism and analgesia. *Political Theory, 33*(2), 218–242.

Ekman, J., & Amnå, E. (2012). Political participation and civic engagement: towards a new typology. *Human Affairs, 22*, 283–300.

Escobar, A. (2015). Degrowth, postdevelopment and transitions: a preliminary conversation. *Sustainability Science, 10*(3), 451-462.

Falkner, R. (2010). Business and global climate governance: A neo-pluralist perspective. In M. Ougaard & A. Leander (Eds.), *Business and global governance* (pp. 99–117). London: Routledge.

Falkner, R. (2016). A minilateral solution for global climate change? On bargaining efficiency, club benefits, and international legitimacy. *Perspectives on Politics, 14*(1), 87-101.

Finlay, S. (2010). Carving out meaningful spaces for youth participation and engagement in decision-making. *Youth Studies Australia,* *29*(4), 53–59.

Fisher, S. R. (2016). Life trajectories of youth committing to climate activism. *Environmental Education Research, 22*(2). 229-247.

Freire, P. (1970/2003). *Pedagogy of the oppressed (30th Ed.)*. New York: Continuum.

Giddens, A. (2009). *The politics of climate change*. Cambridge: Polity Press.

Gulbrandsen, L., & Andresen, S. (2004). NGO influence in the implementation of the Kyoto protocol: compliance. *Flexibility Mechanisms, and Sinks, Global Environmental Politics, 4*(4), 54-75.

Hampson, F. O., & Heinbecker, P. (2011). The „new" multilateralism of the twenty-first century. *Global Governance, 17*(3), 299-310.

Hansen, A. D., & Sonnichsen, A. (2014). Radical democracy, agonism and the limits of pluralism: an interview with Chantal Mouffe. *Distinktion: Scandinavian Journal of Social Theory, 15*(3), 263–270.

Hart, R. A. (1992). *Children´s participation - from tokenism to citizenship*. Florence: UNICEF International Child Development Centre.

Hayward. C., & Lukes, S. (2008). Nobody to shoot? Power, structure, and agency: a dialogue. *Journal of Power, 1*(1), 5-20

Hjerpe, M., & Linnér, B-O. (2010). Function of COP side-events in climate-change governance. *Climate Policy, 10*(2), 167-180.

Hjerpe, M., & Nasiritousi, N. (2015). Views on alternative forums for effectively tackling climate change. *Nature Climate Change, 5*, 864-867.

Hoffmann, M. J. (2011). *Climate governance at the crossroads: experimenting with a global response after Kyoto*. New York: Oxford University Press.

International Youth Climate Movement (2020). The IYCM [online]. Retrieved from https://youthclimatemovement.wordpress.com/the-international-youth-climate-movement/ on 15 April 2020.

Jagers, S., & Stripple, J. (2003). Climate governance beyond the state. *Global Governance, 9*(3), 385-399.

Jordan, A. J., Huitema, D., Hildén, M., Van Asselt, H., Rayner, T. J., Schoenefeld, J. J., ... Boasson, E. L. (2015). Emergence of polycentric climate governance and its future prospects. *Nature Climate Change, 5*, 977-982.

Kara, N. (2013). Beyond tokenism: participatory evaluation processes and meaningful youth involvement in decision-making. *Children Youth and Environments, 17*(2), 563–580.

Keck, M., & Sikkink, K. (1999). Transnational advocacy networks in international and regional politics. *International Social Science Journal, 51*(159), 89-101.

Kleres, J., & Wettergren, Å. (2017). Fear, hope, anger, and guilt in climate activism. *Social Movements Studies, 16*(5), 507-519.

Kollman, K. (1998). *Outside lobbying: public opinion and interest group strategies*. Princeton, NJ: Princeton University Press.

Kuyper, J. W., Linnér, B-O., & Schroeder, H. (2018). Non-state actors in hybrid global climate governance: justice, legitimacy, and effectiveness in a post-Paris era. *WIREs Climate Change, 9*, 1-18.

Kwiatkowski, L. (2017). *Paths to meaningful youth involvement at the international climate change negotiations: lessons from COP22 in Marrakesh* (Unpublished Master's thesis). Uppsala University, Uppsala, Sweden.

Kwon, S. A. (2019). The politics of global youth participation. *Journal of Youth Studies, 22*(7), 926-940.

Larson, R., Walker, K., & Pearce, N. (2005). A comparison of youth-driven and adult-driven youth programs: balancing inputs from youth and adults. J*ournal of Community Psychology, 33*(1), 57–74.

Lisowski, M. (2005). How NGOs use their facilitative negotiating power and bargaining assets to affect international environmental negotiations. *Diplomacy and Statecraft, 16*, 361-383.

Mitra, D., Serriere, S., & Kirshner, B. (2014). Youth participation in U.S. contexts: student voice without a national mandate. *Children & Society, 28*, 292-304.

Mouffe, C. (1999). Deliberative democracy or agonistic pluralism? *Social Research, 66*(3), 745-758.

Määttä, M., & Aaltonen, S. (2016), Between rights and obligations – rethinking youth participation at the margins. *International Journal of Sociology and Social Policy, 36*(3/4), 157-172.

Naím, M. (2009). Minilateralism: The magic number to get real international action. *Foreign Policy, 173*. Retrieved from https://foreignpolicy.com/2009/06/21/minilateralism/ on 27 February 2020.

Nasiritousi, N., Hjerpe, M., & Linnér, B-O. (2016). The role of non-state actors in climate change governance: understanding agency through governance profiles. *International Environmental Agreements, 16*, 109-126.

Nasiritousi, N., & Linnér, B-O. (2016). Open or closed meetings? Explaining nonstate actor involvement in the international climate change negotiations. *International Environmental Agreements, 16*, 127-144.

Newell, P. (2000). *Climate for change: non-state actors and the global politics of the greenhouse*. Cambridge: Cambridge University Press.

Newell, P. (2005). Climate for change? Civil society and the politics of global warming. *Global Civil Society, 6*, 1-31.

O'Brien, K., Selboe, E., & Hayward, B. M. (2018). Exploring youth activism on climate change: dutiful, disruptive, and dangerous dissent. *Ecology & Society, 23*(3), 42-54.

O'Donoghue, J., Kirshner, B., & McLaughlin, M. (2003). Introduction: moving youth participation forward. *New Directions for Youth Development, 2002*(96), 15-26.

Ojala, M. (2012). Regulating worry, promoting hope: how do children, adolescents and young adults cope with climate change? *International Journal of Environmentalism and Science Education, 7*(4), 537-561.

Ojala, M. (2013). Coping with climate change among adolescents: implications for subjective well-being and environmental engagement. *Sustainability, 5*(5), 2191-2209.

Okereke, C., & Coventry, P. (2016). Climate justice and the international regime: before, during and after Paris. *WIREs Climate Change, 7*, 834-851.

Orr, S. K. (2016). Institutional control and climate change activism at COP 21 in Paris. *Global Environmental Politics, 16*(3), 23-30.

Percy-Smith, B., & Thomas, N. (2010). *A handbook of children and young people's participation: perspectives from theory and practice*. London: Routledge.

Phiri, S. (2019). *Youth participation in politics: the case of Zambian university students*. Pennsylvania: IGI Global.

Pretty, J. (1995). Participatory learning for sustainable agriculture. *World Development, 23*(8), 1247–1263.

Price, R. H. (1990). Wither participation and empowerment? *American Journal of Community Psychology, 18*(1), 163–167.

Rietig, K. (2016). The power of strategy: environmental NGO influence in international climate negotiations. *Global Governance, 22*, 269-288.

Rosenau, J. N. (1995). Governance in the twenty-first century. *Global Governance, 1*(1), 13-43.

Schmitter, P.C., & Streeck, W. (1999). *The organization of business interests: studying the associative action of business in advanced industrial societies*. Discussion Paper 99/1. Cologne, Germany: Max Planck-Institut für Gesellschaftsforschung.

Schroeder, H. (2010). Agency in international climate negotiations: The case of indigenous peoples and avoided deforestation. *International Environmental Agreements, 10*(4), 317–332.

Schroeder, H., & Lovell, H. (2012). The role of non-nation-state actors and side events in international climate negotiations. *Climate Policy, 12*(1), 23-37.

Schulz, A. J., Israel, B. A., Zimmerman, M. A., & Checkoway, B. N. (1995). Empowerment as a multi-level construct: perceived control at the individual, organizational and community levels. *Health Education Research, 10*, 309−327.

Shier, H. (2001). Pathways to participation: openings, opportunities, and obligations. *Children and Society, 15*, 107–117.

Sukarieh, M. (2012). From terrorists to revolutionaries: the emergence of 'youth' in the Arab World and the discourse of globalization. *Interface, 4*(2), 424-437.

Sukarieh, M., & Tannock, S. (2008). In the best interests of youth or neoliberalism? The World Bank and the new global youth empowerment project. *Journal of Youth Studies 11*(3), 301–312.

Theis, J. (2007). Performance, responsibility and political decision-making: child and youth participation in Southeast Asia, East Asia and the Pacific. *Children, Youth and Environments, 17*(1), 1-13.

Thew, H. (2018). Youth participation and agency in the United Nations Framework Convention on Climate Change. *International Environmental Agreements, 18*, 396-389.

Thew, H., Middlemiss, L., & Paavola, J. (2020). "Youth is not a political position": exploring justice claims-making in the UN climate change negotiations. *Global Environmental Change, 61*, 1-10.

Treseder, P. (1997). *Empowering children and young people: promoting involvement in decision-making*. London: Save the Children.

UNDP (2014). *Enhancing youth political participation throughout the electoral cycle: a good practice guide. United Nations Development Program*. Retrieved from https://www.undp.org/content/dam/undp/library/Democratic%20Governance/Electoral%20Systems%20and%20Processes/ENG_UN-Youth_Guide-LR.pdf on 7 April 2020.

UNESCO (2020). By youth, with youth, for youth [online]. Retrieved from https://en.unesco.org/youth on 11 April 2020.

UNFCCC (2010). *Youth participation in the UNFCCC negotiation process: The United Nations, young people, and climate change* [online]. Retrieved from https://unfccc.int/files/cooperation_and_support/education_and_outreach/youth/application/pdf/youth_participation_in_the_unfccc_negotiations.pdf on 26 February 2020.

UNFCCC (2019). Side events and exhibits of COP 25. Retrieved from https://unfccc.int/process-and-meetings/conferences/un-climate-change-conference-december-2019/events/side-events-and-exhibits-at-cop-25#:~:text=Official%20side%20events%20are%20a,for%20meeting%20the%20climate%20challenge on 31 October 2020.

United Nations Environment Programme (2019). *Emissions gap report 2019*. Nairobi: UNEP.

Victor, D. G. (2006). Toward effective international cooperation on climate change: numbers, interests and institutions. *Global Environmental Politics, 6*(3), 90-103.

Villa-Torres, L., & Svanemyr, J. (2015). Ensuring youth's right to participation and promotion of youth leadership in the development of sexual and reproductive health policies and programs. *Journal of Adolescent Health, 56*(1), S51–S57.

Vormedal, I. (2008). The influence of business and industry NGOs in the negotiation of the Kyoto mechanisms: the case of carbon capture and storage in the CDM. *Global Environmental Politics, 8*(4), 36–65.

Ward, S., & Parker, M. (2013). The voice of youth: atmosphere in positive youth development program. *Physical Education and Sport Pedagogy*, 18(5), 100-114.

White, S. C. (1996). Depoliticising development: the uses and abuses of participation. *Development in Practice, 6*(1), 6–15.

Wong, N. T., Zimmerman, M. A., & Parker, E. A. (2010). A typology of youth participation and empowerment for child and adolescent health promotion. *American Journal of Community Psychology, 46*(1), 100–114.

Young, I. M. (1990). *Justice and the politics of difference*. Princeton: Princeton University Press.

Young, I. M. (2002). *Democracy and justice, inclusion and democracy*. Oxford University Press.

Young, I. M. (2003). Activist challenges to deliberative democracy. In J. S. Fishkin & P. Laslett (Eds.), *Debating Deliberative Democracy*. Malden: Blackwell Publishing Ltd.

Yunita, S. A. W., Soraya, E., & Maryudi, A. (2018). "We are just cheerleaders": youth's views on their participation in international forest-related decision-making for a. *Forest Policy and Economics, 88*, 52-58.

Zeldin, S., McDaniel, A. K., Topitzes, D., & Calvert, M. (2000). *Youth in decision-making: a study of the impacts of youth on adults and organizations*. Chevy Chase, MD: National 4-H Council.

Humanities

Why Are We Not Doing The Right Thing?

Towards an Action-Oriented Moral Education to Remedy Moral Akrasia

Patrīcija Keiša

Supervisor
Mariëtte Willemsen (AUC)
Reader
Emma Cohen de Lara (AUC)

Photographer: Mirthe van Veen

Abstract

Most academic and public discussion about morality revolves around deliberating on the right moral position to take. Yet, our focus on this complex issue overshadows the fact that even having a clear sense of what the right thing to do is, we often fail to enact it in our everyday lives. This gap between our moral belief and action, called akrasia, poses a problem because it hinders our fulfillment as human beings, and because morality is inherently prescriptive – holding a moral belief requires that we strive to align our behavior to it. Following several scholars who argue that our widespread moral akrasia needs to be addressed through social institutions such as education, I suggest that reducing the gap between our moral beliefs and action should be the focus of moral education. After analyzing various reasons for akratic behavior and providing contemporary real life examples of it, I argue that the educational remedy for it lies in everyday moral reflection within an emotionally supportive community. Furthermore, I suggest more particular moral practice oriented educational activities to counter specific causes of akrasia.

Keywords and phrases: *akrasia, moral behavior, moral education, moral reflection, community*

Acknowledgements

Thanks to my supervisor Mariëtte Willemsen for her gentle guidance and encouragement.

Thanks to Hyunsuh Kim, Mélanie Butterati, Myriam Bellamine, Nika Peršić, Tatia Eleni for their openhearted and meaningful contributions.

Thanks to Gabrielė Plukaitė for their invaluable support and attentive peer review.

Thanks to all the friends who assured me that my work matters and that they are curious to read it.

Contents

> “Video meliora proboque,
> deteriora sequor.”
> “I see the better way and
> approve it, but I follow the
> worse way.”
> —Ovid,
> *Metamorphoses*

1 Introduction

When we think of ethics, much of our attention is directed towards inquiring about the moral principles and values we should adopt – in other words, we concern ourselves with figuring out what the right thing to do is. Often, this is not an easy question to answer and, hence, we take it to be our main concern when it comes to morality. However, there is something that appears to consistently slip our attention: even when we seem to know, think, feel, sense the right thing to do, we often do not follow through with it in our everyday lives. We value freedom, yet we buy clothing that is produced by people working under conditions akin to slavery. We value nature, but we fail to adopt an environmentally friendly lifestyle. We value kindness and community, yet we find ourselves being harsh and self-oriented in our daily interactions.

This gap between belief and action is designated by the Greek term akrasia ($\alpha\kappa\rho\alpha\sigma\iota\alpha$), which signifies “not being strong enough” or “a weakness of the will” (Hare 77). If something is in a condition of weakness, we might like to strengthen it. In fact, it appears to be common knowledge that we wish to minimize this gap between our belief and action – think of expressions like “practice what you preach” or famous quotes such as “the only thing necessary for the triumph of evil is for good men to do nothing”[1]. Still, it does seem that we, the “good men”, often truly do nothing and frequently even act in exact opposition to our values and beliefs. Moreover, our failure to act morally is “typically not episodic” — rather, we commit it on a daily basis (Rorty 99). Observing our widespread akratic behavior in the light of the current environmental crisis, philosopher Lisa Kretz calls for an urgent “shift in theoretical and pedagogical approaches toward ethics that inspire moral action” (“Climate Change” 23). She invites educators and ethicists “to dialogue more widely about teaching *being* ethical” to bridge our moral knowledge and behavior (“Teaching Being Ethical” 167). In a similar vein, Heesoon Bai argues that “environmental education needs to be taken up as a moral education” in order to reclaim our moral agency (311). In a discussion of animal ethics and akratic meat consumption, Elisa Aaltola holds that in order to remedy akrasia, we need to rethink social structures, including education, “in a direction, where they support cultivation towards virtuous forms of life” (“Problem of Akrasia” 127). It is in answer to these contemporary scholars’ invitations that I will build on their knowledge and seek an educational remedy to the gap between our moral belief and action.

I will begin by laying theoretical grounding for akrasia as a phenomenon and discussing why it poses a problem. Then, I will describe common reasons for akratic action and give examples of such behavior. Here, I will refer to a Socratic conversation[2] I organized for the purposes of this research, in which we reflected on the question ‘Why am I not doing what I think is the right thing to do?’ Six AUC students, including myself, took part in the conversation on 3 April 2020. It was facilitated by my thesis supervisor, Mariëtte Willemsen. We followed the Socratic method as described by Hannah Marije Altorf, where conversation relies on experience to avoid distanced hypotheticals and where participants question each other while trying to imagine themselves in the other’s position (5, 8). Each participant had prepared a specific personal example of their own akratic action, which was analysed and reflected upon together with the group. I had initially chosen this method to expand my insight into moral akrasia outside of personal experience and literature, but it has gained larger significance as I advanced my research because the way it fosters a focus on real life situations, empathy[3] and shared reflection is in line with the educational remedies to akrasia that I offer in Chapter 4.

This Socratic conversation was helpful for bring-

[1] This quote is commonly attributed to Edmund Burke, but the direct source is ambiguous (and is besides my point here).

[2] The role of this conversation was to serve as an inspiration, a collective brainstorm for my research. Hence, it is not my aim to use it for detailed qualitative data analysis, but rather to propose it as a method of philosophical inquiry within the humanities.

[3] Here I mean empathy in the common sense of the word, see Altorf for an extended discussion of its meaning within her analysis of Socratic conversation.

ing in contemporary, first person accounts of akrasia, pointing to the particular difficulties we face in enacting our moral beliefs and affirming akratic action both as an individual responsibility and as a social issue. The shared reflection is the basis on which I will continually employ the pronoun "we" when talking about akrasia – it is not them, some weak, immoral, anonymous people who fail to act morally. Indeed, most of us are akratic. In fact, much of my inspiration for this research comes from my own frustration with my failures to bridge my behavior to my moral beliefs. I am in absolute agreement with Aaltola who writes, "instead of moral blame, the phenomena [of akrasia] deserve the type of attention that may lessen their hold on us" ("Meat Paradox" 2). My aim is not to shame and scold us for our failures, but rather to take a melioristic viewpoint, which maintains that we can make the world better through human effort (Bromhall 44). It is with such a mindset that I will seek an educational approach that could minimize akrasia and foster moral action.

I will first overview the subject of moral education as it is currently conceived, and suggest that a focus on moral action could help 21st century moral education escape the relativist, secularist, individualist reproaches that target it, because any moral ideal requires moral action as its goal. To continue, I will discuss akrasia itself as an important learning opportunity and I will proceed by outlining aspects that an action-oriented moral education should account for as well as by offering suggestions for particular learning activities. However, my goal here will not be to provide a ready-made educational policy for a particular institution. Rather, my focus on moral education and the pedagogical pointers that I will provide are to highlight akrasia as a social issue that needs to be addressed and solved in communion. That being said, throughout my work I will recognize and try to reconcile the intrinsic conflict between personal responsibility and social influences that is at play in our failures to act in accordance with our moral beliefs.

2 Akrasia

2.1 Defining Moral Akrasia

Akrasia designates a theory-action gap: it is "the disposition to act contrary to one's own considered judgment about what it is best to do" (Steward). Not all instances of akrasia are necessarily moral: for example, I can hold it best for my own good not to eat an extra piece of cake, yet I fail to follow through and I eat it anyway.[4] Notwithstanding, often the theory on akrasia addresses cases that are, in fact, related to morality – the failure to do what one holds to be the right thing to do. This can represent both an action that directly goes against our moral judgement or a lack of action where our morality would require us to act.

One of the earliest defining discussions on akrasia is found in Plato's dialogue *Protagoras*. There, Socrates suggests that true akrasia is actually impossible because "one cannot choose a course of action which one knows full well to be less good than some alternative known to be available" (Steward). Such an assertion implies that moral beliefs are necessarily motivating – it is sufficient to have a sincere belief about what is morally right in order to be moved to do the morally right thing. This ethical position is called motivational judgement internalism (Rosati, Burch ch.1). The opposing view to internalism is externalism, which holds that, while moral belief can motivate, there is no necessary connection between moral belief and the motivation to act morally – in other words, acting according to one's moral beliefs is contingent (Rosati, Burch ch.1). Both of these positions, according to Burch, present an essential truth – internalism underlines the close and, indeed, generally expected link between moral judgement and action, while externalism accounts for the actuality that moral judgements do not always move us to act in accordance with them.

There is much to be said about the existence of akrasia and its implications for human rationality, judgment and motivation. From the perspective of analytic philosophy, it is of interest to examine the logical soundness and rationality of the phenomenon (Davidson qtd. in Chappell 92; Audi 528). For cognitive sciences, it can be a fruitful object

[4] It is arguable whether such (or any) act of akrasia is truly morally neutral. For example, by eating that extra piece of cake, I might normalize the indulgence of sweets for people around me, and this might ultimately compromise the health of my community. Admittedly, going so far in everyday moral considerations can potentially have a paralyzing effect, making an individual think she can do nothing without bringing about morally damaging consequences. That being said, I make this comment simply to emphasize that the moral implications of many, seemingly neutral, everyday actions are often overlooked, so potentially many, if not all, cases of akrasia may be considered moral in nature.

of analysis to better the understanding of the human mind (Kauppinen). I, however, aspire to take a more pragmatic approach. Instead of pondering how akrasia can possibly happen, I wish to inquire why it does happen and how it could be remedied. It is important here to note an assumption that underlies both the way akrasia is commonly defined and the discussion of it that is to be pursued in this paper. Moral akrasia is the failure to act upon one's own *personal* moral belief. We have the tendency to associate morality first and foremost with the rules imposed by the society, the church, the family and other external structures. However, in the highly secular, relativist and individualist postmodern Western world of the 21st century,[5] there is a high level of skepticism toward any such moral authority. When these external structures implicitly or explicitly appeal to their dictated standard of morality, we tend to view them as *moralizing* and, in many cases, we refuse to act in accordance with their standards. These situations do not constitute cases of akrasia, because the moral beliefs that we fail to follow with action are not our own. This is not to say that the morality of, for example, a religion cannot be internalized as the morality of an individual, but by this internalization it then becomes one's personal morality. In sum, both in the larger context of discussing akrasia and in the more particular context of our time, moral views5 are conceived of as a predominantly personal matter. In fact, that is what makes moral akrasia such a puzzling problem in the first place. We ourselves judge something to be the right thing to do, yet we fail to do it.

2.2 Akrasia as a Problem for Morality

The study of ethics is essentially a study of how we ought to live (Shafer-Landau 1). This prescriptive nature of moral judgments is why the theory-action gap poses an issue; as Richard Hare puts it, "no one can say that there is no problem here, unless he denies that it is the function of moral judgements to guide conduct" (70). Hare also notes that this failure to live up to one's moral ideals is "perhaps the central difficulty of the moral life" (72).

In her essay "Are moral considerations overriding?", Philippa Foot quotes Professor D. Z. Philipps's statement that "moral considerations are, for the man who cares for them, the most important of all considerations" (181). Failing to act in accordance with that which is arguably the most important undeniably poses a problem. However, Foot argues against Philipps's thesis by providing examples of situations in which people do care about morality, but whose moral considerations are overridden by other factors, for instance, financial matters (184-6). One may be concerned by the detrimental effects of climate change, but agree to work a well-paid job funded by the fossil fuel industry. Overall, Foot is saying that it often happens that we care about morality while failing to take moral considerations as having the utmost importance in our lives. In this way, Foot's stance is rather realistic and descriptive, while Philipps's claim is idealist and prescriptive. However, ultimately, their positions are not necessarily in any true opposition. The two authors bring to the fore the same question – why do we simultaneously care about morality, but end up treating it as secondary, and consequently fail to act in the way we think we ought to act?

Here, it is relevant to note that the premise assumes that we care about morality, thus, this is not the case of an amoralist "who seemingly makes moral judgments, while remaining utterly indifferent" or of a person who refuses to engage in any moral considerations at all (Rosati). We *do* care and, according to Philipps, we feel remorse if we give in to other things that contradict our morality (qtd. in Foot 182). The experience of moral failure can yield strong negative feelings, as it often creates "a sense of having failed ourselves and a need to re-examine who we are" (Athanassoulis 351). Germain Grisez and Russell Shaw deepen this argument, asserting that,

> "Ought" in the moral sphere points ultimately to an ideal of the fullest possible personhood and the richest possible community. The moral "ought" is a kind of verbal road sign directing travelers to their full humanity realized through freedom of self-determination. (99)

This view is in line with a long psychological tradition, which holds that "identity means being true to oneself in action" and "identity is rooted in the very core of one's being" (Aquino & Reed 1427). If morality is our own prescription to a fuller personhood, to fulfilling our identity and to nurturing our

[5] This depiction of "the Western world" is, of course, a generalization – but it is the world with these general social characteristics that I am discussing and addressing.

relation to others, and we are routinely failing to follow through with it, undeniably, there is something going wrong. This gives us sufficient motivation to seek a solution to minimize this failure and to align our behavior with our morality, in other words, to promote the opposite of akrasia – enkrasia.

2.3 Reasons for and Examples of Moral Akrasia

Historically, in Western philosophy, there are two main interpretations of the inner conflict at the origin of akrasia: that of rational failing and that of moral failing. For Aristotle, it is the former; acting akratically means that one's passions have overtaken one's reason. Acting in accordance with one's better judgement of morality, that is, acting enkratically, is then a matter of resisting passions (Bromhall 28). Aristotle metaphorically describes akrates – people who act akractically – in two ways: "first, it is as if the incontinent person is drunk, intoxicated by desires or habits that push him to commit acts he knows to be wrong" and "second, akrasia is likened to an illness such as epilepsy, with temporary, disabilitating seizures" (Aaltola "Meat Paradox" 6).

This "rational failing" interpretation certainly presents a truth about moral akrasia – often, it can seem that our choice not to do the right thing is taken as if in a state of intoxication, not genuinely under our conscious control. Moreover, it can happen that acting in accordance with our moral standards presents less pleasure and comfort than acting akratically. But, as summarized by Aaltola, Aristotle distinguishes two types of pleasure: "Whereas alien pleasures are base and antagonistic toward virtue ..., proper pleasures derive from virtuous activity; they enable us to flourish and fulfil our telos[6] ("Meat Paradox" 7). The pleasure of acting morally is, thus, more fulfilling than other (oftentimes, bodily, short-lasting) pleasures and should be preferentially sought out. Yet it is also admittedly more difficult to achieve – it requires more effort – and that is why we tend to fall into akrasia. This interpretation of akrasia puts great emphasis on personal responsibility and self-control, but does view enkrasia as a truly achievable goal.

The second view of akrasia is that of moral failure. This perspective is held by early Christian thinkers, such as St. Paul and St. Augustine. For them, akrasia represents "conflict between the spirit and the flesh" (Bromhall 30). However, in contrary to Aristotle's view, here the mastery over one's passions (of the flesh) is fundamentally impossible to achieve:

> Akratic action is an intractable element of the Christian experience; it is a consequence of the Fall and the corrupting effect it had on human nature. So long as one is attempting to act contrary to the corrupted longings of human nature, one will struggle. (Bromhall 30)

Hence, striving to live in accordance with our "innermost spirit" is inevitably difficult and can never truly be satisfied (30). Although such a view at first appears quite discouraging to our attempts towards a moral life, it is also liberating, perhaps even soothing, in that it admits akrasia to be a universal human struggle rather than a particular failure of an individual. Kyle Bromhall takes this early Christian idea of struggle, strips it from the association with shame and sinfulness, and interprets it through the doctrine of meliorism of the psychologist and philosopher William James. Meliorism holds that "the world can be made better through human effort" (44). Thus, the perspective presented by Bromhall allows for the viewing of akrasia as a necessary and legitimate struggle, but also provides hope towards improving our ability to act morally. It underlines the universality of the issue and avoids narratives of embarrassment and failure, while maintaining the potential for personal responsibility and growth.

It is this melioristic position on akrasia I wish to side with throughout this exploration and on which I will particularly elaborate when discussing possible remedies to akrasia in Chapter 4. I will now proceed to outline and exemplify the diverse causes of akrasia. This list, of course, is not exhaustive, but rather a general overview of everyday akratic actions.[7] Moreover, while I choose particular cases to represent a specific cause, in reality, several causes are

[6] Telos (Greek: $\tau\epsilon\lambda$ος) - ’inherent purpose’.

[7] This means that I am omitting the causes of episodic akrasia, where, for example, occasionally we engage in moral self-licensing (Mullen and Monin) and avoid moral action or commit immoral acts because we feel like we are generally morally good people and "deserve" not to do the right thing from time to time. I am also excluding akrasia for its own sake, as Augustine famously confesses to stealing pears in his childhood: "My pleasure lay not in the pears; it lay in the evil deed itself" (2.4).

often at play simultaneously, which contributes to the complexity of akrasia as a phenomenon.

2.3.1 Desire, Impulsivity and Comfort

In the spirit of the Aristotelian view of akrasia, desire and impulsivity, or what he calls “alien pleasures”, are common reasons why we fail to do the right thing. These are often hedonistic and consumerist inclinations, as cleverly epitomized by Aaltola, “I want dairy ice-cream, even if it comes from maltreated cows” (“Meat Paradox” 7). Of course, for this to be akratic, it presumes that the agent holds a belief that animal welfare matters – as many of us do. Distance – both temporal and geographical – also plays a role here. Our long-term wish might be to be a virtuous person who does not cause harm to animals, but our short-term desire is to enjoy ice cream. This short-term desire promises immediate satisfaction so we succumb to it. Similarly, the cows that are maltreated are ‘somewhere far away’; in fact, we will never see them, regardless of whether they are happy or suffering, but the ice cream is right here and we will directly enjoy it. However, Aaltola’s example presents akrasia as an explicit thought process, whereas this is not always the case. Oftentimes, we can impulsively act upon the hedonistic desire for the oh-so-tasty ice cream without even deliberating cow welfare. Then, the realization of akrasia can come as a regretful afterthought of “I shouldn’t have,” almost as in a hungover. It is also possible that this realization does not actually even cross our mind or, perhaps, we make sure that it does not. Aaltola discusses at length the phenomenon of meat eater’s akrasia, where “one both loves and eats animals” (“Meat Paradox” 2). In these cases, people employ many different strategies to avoid acknowledging their own akrasia in order to continue enjoying their steak. This is a serious issue that I will discuss more thoroughly in Chapter 4. Another related reason for akrasia is that moral action often requires a sacrifice of comfort. As opposed to fulfilling a desire, which motivates the active pursuit of an akratic act, there are also situations where akrasia is the comfortable status quo, whereas doing the right thing requires an active, troublesome effort. This is something that we established as a common point of agreement in the Socratic student conversation I organized for this research. Everyday situations where we choose comfort over moral action include, for example, buying fruit packed in excessive plastic packaging simply because that is what the most convenient grocery store sells, not defending someone from bullying because we do not want to get involved in the conflict ourselves, eating meat because it is more easily accessible than a vegetarian meal. Overall, it seems that moral action requires additional effort and, hence, a sacrifice of our usual comfort. This means it is also relevant to discuss our habitual state of being.

2.3.2 Habits and Self-control

Just as the Greek and Latin roots of ethics and morals stem from habits and customs (Arendt 5), much of our akrasia lies in habit. It is often the case that our habits are not attuned to our moral beliefs, yet are major driving forces for our behavior. As Bromhall describes, “a habituated action will have a strong degree of motivational force behind that action, merely by virtue of being habituated, regardless of the wishes of the agent at the time” and, moreover, “a habituated action suppresses actions contrary to that habit” (33). Thus, even if we are aware of our akrasia, it can be difficult and emotionally draining to fight against a bad habit. In our Socratic conversation, a participant shared her experience[8]:

> I believe it is morally wrong to use other people as a means to alleviate your own anger, and that we have a moral duty to process our emotions in a way that does minimum harm to others around us. However, in situations in which I am irritated, frustrated, or overwhelmed, I tend to get angry at people who have done nothing wrong to me. A specific example is with an ex-boyfriend from high school. I was stressed out by school-related problems, and when he asked what was wrong, I could hear myself screaming at him, insulting, mocking him, and being mean for no reason. It happened a few times, and it felt as if I could not hold it back, even though my “moral” brain was disgusted that I was behaving this way.

[8] This is a direct quote from the personal examples that the participants were asked to prepare and write down prior to the conversation

Her story aptly illustrates Bromhall's theory that habits are enacted "regardless of the wishes of the agent" (33). It also captures the feelings of frustration and helplessness that akrasia can cause. Overcoming a habit seems to require hard work and self-control, and it might be difficult to pinpoint where to start or exactly what this work entails. Bromhall argues that habit is exactly at the root of akrasia: "Akratic actions happen because the agent has not trained herself properly and/or has not manipulated her environment to the extent required to bring about successful resistance of her habituated actions" (41). Aaltola notes that our hedonistic and consumerist inclinations are also a matter of habit that require self-control to be changed. She writes:

> Most of us know the practical dilemmas of believing x, and yet wanting to do something that violates x, whether this is in relation to overindulgence on food or drink, protracted idleness or flying to holiday destinations in the era of climate change — moreover, most of us are probably capable of recognizing the lack of self-discipline underlying this phenomenon. (5)

Ultimately, most of these habitual actions take place mindlessly, almost automatically. A helpful theory for understanding this is psychologist's Daniel Kahneman's explanation that "human behavior is governed by two distinct systems. 'System 1' is automatic and unconscious, whilst 'system 2' includes conscious deliberation" (Aaltola "Meat Paradox" 2). Habits are part of system 1 responses; acting upon them is quick and requires little effort. Going against habits would require employing system 2, but this is not an easy task because system 2 "is inherently lazy and seeks to conserve energy whenever possible" and thus we easily "default to system one responses" (Bromhall 47). Hence, being akratic often means choosing the path of least adversity and failing to work towards enkratic habits. Time also appears to be a relevant variable: in situations where we lack the time for considerate deliberation and need to proceed quickly, we tend to choose the habitual akratic option.

2.3.3 Lacking Moral Reflection and Phronesis

While we affirm holding general moral beliefs, we can fail to reflect on them on a daily basis and to recognize the moral implications of our everyday actions. For example, one can judge it morally right to cause the least possible suffering to other human beings. Yet, it does not mean that she will readily recognize the emotional pain she causes to her parents by regularly being impatient and brusque with them. Nor does it mean that she will realize that buying a new dress in a fast-fashion store perpetuates the suffering of the mistreated workers in sweatshops. In this way, akrasia can be caused by a lack of moral reflection and failing to recognize the larger moral considerations of everyday action.

Nevertheless, even when engaging in active moral reflection and having a strong belief in our values, we can often find ourselves not knowing exactly how to act in a way that would represent our moral beliefs. This is well captured by the Greek concept of phronesis, which is the "ability to see what virtue requires" in particular situations and is "developed through practice and experience" (Athanassoulis 348). In our Socratic conversation, a student recounted the experience of seeing a woman being verbally harassed on the street and feeling the urge to help, but failing to intervene because she did not know what to do. Similar situations are fairly common and can also represent a conflict of different moral considerations. Trying to help someone in danger can mean potentially putting ourselves in danger too. It is true that morality often concerns a tension between our own and others' well-being. In most cases, this tension is rather abstract – think of the ice-cream-making maltreated cows we will never meet. However, in a situation where we may physically put ourselves in danger to help someone, this tension becomes very direct and tangible. It is a highly charged moral situation when our physical well-being is directly placed against someone else's. In any case, even if we do decide that helping is the right choice, taking action would require deliberating about what to do, gathering the courage to act instead of doing nothing, and putting in the effort to act instead of saving energy and remaining passive. Moreover, this was clearly a time sensitive situation, where she needed to take action quickly. Once we look at how many layers there are that essentially act as barriers to action, it is no wonder that we often do not behave in accordance with our moral intuition. A lack of phronesis can also stop us in more calm and subtle situations, where we have time to deliberate and not much more is at stake, but we simply lack the experience to act. As Athanassoulis rightly notes, “wanting to be sympathetic and knowing what to say to a grieving person are two different things” (351).

2.3.4 Disengagement and Lack of Empowerment

Our lack of moral action can also come from the feeling that morality is simply too demanding. We might feel that there is too much suffering and too much wrong in the world for us to alleviate it. Thus, we experience a lack of hope and do not believe that our actions matter. Consequently, we feel a loss of moral agency and refuse to take moral responsibility. To elaborate on this, Lisa Kretz refers to three forms of “emotion-overloading” as identified by psychologist Martin Hoffman: over-arousal, fatigue, and indifference. In sum, these stages follow one another as we are exposed to situations where empathizing with the ones suffering causes such distress that it tires us out and we eventually completely disengage, emotionally and morally (“Emotional responsibility” 343). Kretz takes a note of everyday examples of this disengagement: many of us have become numb to the homeless people we pass on the street; similarly, the media exposes us to a “barrage of negative and violent images of extreme suffering” on a daily basis has desensitized us so much that images of war or of animal slaughterhouses fail to trigger any emotional engagement (343). We seem to shrug it off, as if saying: “Yes, ideally we would like to help, but what can we do? Nothing, it seems.” Another common example of this type of disengagement is when one refuses to take environmental action, such as to stop consuming animal products, because she does not believe her individual behavior would have a significant impact on reducing the animal farming industry’s greenhouse gas emissions.

2.3.5 Social Script Dictates Otherwise

While our moral akrasia undeniably concerns our individual beliefs, identities, habits and behavior, akratic action is largely a social phenomenon: “moral psychological aspects of individual life cannot be reduced to the individual, but rather find origins from the surrounding socio-political arrangements” (Aaltola “Problem of Akrasia” 126). Amelie Rorty writes, “Just as a disposition to chronic bronchitis may indicate a toxic environment, so individual akrasia may indicate social disorder” (649). There is a plethora of examples of this social disorder or, in other words, of the social conditions that create fruitful ground for moral akrasia. In fact, it seems that very few cases of akratic action take place without there also being a fairly obvious social aspect involved. For example, we find ourselves not helping a woman being harassed on the street because that is simply not the common thing to do – we rarely see other people get involved in such situations and, in this particular situation nobody else is helping. We buy goods wrapped in unnecessary plastic and consume meat because that is the normalized, easy thing to do. The way our society is designed promotes certain habits of consumption, behavior, interaction and attitude, which we have internalized despite the fact that they are largely contradictory to our personal moral beliefs. These have become part of our easily accessible system of thinking, as previously discussed. Akratic habits rooted in social norms have additional grounding and become especially difficult to change. Moreover, when discussing this with my peers in the Socratic conversation, we came to the realization that we often lack the practical example of how to act otherwise, and it takes focused creative engagement to find an implementable alternative course of action. As in the case of a woman harassed on the street, we might truly not know what to do since we so rarely see people step up in such situations.

Another aspect of the social component is that going “against the flow” and acting in accordance to our morality often has a social cost. A common case is an example from school, where somebody is being bullied and we have the moral intuition to defend them, but we do not follow it because we fear being bullied ourselves. Similar fear of social consequence is widespread. Research into the reasons vegetarians often violate their diets found that “adhering to a vegetarian diet with absolute adherence would lead one to feel alienated, socially withdrawn, rude, burdensome, wasteful, and socially awkward” (Rosenfeld and Tomiyama 8). This is because meat consumption is such a deeply ingrained norm that not only society at large, but also friends and family are simply not ready to accept this choice of their loved ones. Social consequences need not even be so drastic and negatively defined. For example, I might buy a beautiful dress from a fast fashion store simply because I, without much reflection, know that people will praise me more for looking pretty in my dress than for having a strong moral stance against the fast fashion industry. However, here it is also true that if I try to 'wear' my moral stance and tell others about why I do not buy such dresses, I risk that they could see it as a personal reproach to their moral goodness and, as a defense, hold a negative attitude towards me – more on that in Chapter 4.

Moreover, Rorty notes that our socio-economic conditions not only condone action that turns out to be akratic to our personal morality, they actually promote akrasia itself by holding contradictory values. She writes:

> Social institutions and economic systems encourage and foster the very actions that they also condemn. While promoting habits of cooperation, they also reward radical independence; while condemning aggression, they also praise "aggressive initiative." While admiring selfless devotion, they also reward canny self-interest. Except in extreme cases, rewards and sanctions do not form a clear and guiding pattern. (Rorty 653)

This idea is also clearly illustrated by how widespread is the paradoxical behavior of both loving and eating animals. Lisa Kretz reminds that often, without us being aware of it, psychological research lies behind manipulative marketing strategies we are exposed to daily: "hyper-consumerist identities are meant to ensure insatiable desires for the accumulation of ever-changing commodities and services" ("Climate Change" 18). For instance, there are studies that explore the psychological phenomenon of moral self-licensing and suggest that marketers "incorporate words such as virtues, ethics and/or noble, when describing attributes of their brand in advertising" (Geiger-Oneto and Minton 2530) in order to let the consumers' moral guard down and induce them to purchase luxury goods they would have otherwise refused. Thus, it appears crystal clear that external forces play a large role in our akratic behavior.

3 Moral Education

3.1 On The Possibility and Necessity of Moral Education

The notion of moral education is changing as the view on morality shifts in our society. As a consequence, nowadays, a first point of disagreement concerning moral education is not about the method or content, but about the possibility and necessity of such education in itself. For example, one objection is that "the idea of teaching anything involves the passing on of expertise" and that an expertise in morality seems implausible (Gingell & Winch 147-8). Another argument is that education involves assessment and any assessment of morality appears highly problematic to implement (148). While these objections rest on extremely inflexible views on what constitutes education,[9] they do serve to highlight the general modern and postmodern attitudes of scientism and moral relativism. Scientism engenders an "aversion to anything not measurable" (Purpel 310) and morality is perceived as belonging to a highly subjective, immeasurable realm (Purpel 310). With the rise of individualism and secularization, morality has become a deeply personal matter and there seems to be no space for a moral authority in society (Smith et al. qtd. in Sarid 245). Yet, the general idea of a "moral education", traditionally associated with religious schooling, seems to imply exactly that – a moral authority and, by extension, moral indoctrination. Thus, already in the 1960s, the term "moral education" is described as having an "archaic ring" and a puritanical connotation (Kohlberg qtd. in Sarid 245).

Nevertheless, plenty of scholars resist the doing away with moral education altogether. Many of them are critical of the term 'moral education' itself and accuse it of unnecessary compartmentalization. The reproach is that it promotes "the dualistic notion that there are two kinds of education, one of which is moral and the other where "moral"

[9] Kretz is highly critical of such a view; siding with Freire's ideas expressed in Education of the Oppressed, she objects to "the neoliberalization of education, which encourages a bank-model of "learning" wherein "right answers" are memorized and regurgitated for credit" ("Teaching Being Ethical" 152). From both my academic research and my experience as a student, I strongly endorse this criticism.

is irrelevant" (Purpel 311). As noted by Nel Noddings, Daniel Callahan, Roger Straughan and others (an observation that seems perhaps too truistic even to necessitate scholarly citation) moral issues are essential to human life and permeate all aspects of it. Straughan points out that the amount of time dedicated to moral matters in education is disproportionate to the frequency at which people are confronted with moral questions in their lifetime (120). An optional ethics course, as cleverly argued by Sharaf Rehman, connotes the quite absurd idea that the ethical aspects of life are also optional (15). Noddings takes an even stronger stance, asserting that "the primary aim of all educative effort is the nurturance of the ethical ideal" (173). The overarching claim is that moral concerns are ever-present and important, and, by refusing to address morality in education directly and seriously, we would ignore an essential aspect of life. There remains the question of how to properly address moral education in our time.

3.2 Common Methods

An alternative to the traditionally conceived "indoctrination" path of teaching *right and wrong* is a method based in neutrality, which accommodates pluralism and aims for value clarification (Gracia 9-10). Such an approach is better suited to (post-)modern liberal societies as it caters to individualism and moral relativism, steering each moral learner to "get in touch with his own values, to bring them to the surface, and to reflect upon them" (Purpel and Ryan qtd. in Straughan 16). However, a fundamental point of criticism for this method is that it, just like moral indoctrination, fosters a lack of critical discussion and its neutral pedagogy allows for the indiscriminate promotion of, for example, anti-democratic attitudes (Gracia 8, 10; Purpel 310; Straughan 17). This is a reminder that the liberal democracy is not, in fact, a value-free society and, for instance, the acceptance of other people's freedom and equality is not a matter of mere personal preference. This presents another complexity that any contemporary conception of moral education needs to accommodate.[10]

In an attempt to bring in more argumentative discussion and find a path between neutrality and indoctrination, Gracia suggests a *third* method of moral education, a Socratic or deliberative approach (11).[11] This method aims "to enrich the analysis of the question at stake, trying to increase the wisdom of the decisions to be taken" and holds that "everyone should be able to give an account of their own value choices, despite how difficult this could be" (12). While this emphasis on discussion, reflection and decision-making is laudable, Gracia does not show why the Socratic method would be exclusive with the neutral-deemed values clarification approach – it is potentially an improvement to it. At any rate, group deliberation and value clarification are aspects that merit more consideration and I will return to them in Chapter 4.

There is something the moral-educational approaches addressed here so far and many of those discussed in literature have in common, for example, in the overview Straughan offers in *Can We Teach Children to Be Good?*, – they present an almost exclusive focus on teaching moral judgement and reasoning. Moral action is an implicitly desired goal but is largely left out of proposed approaches to moral education. Kretz observes that the dominant pedagogy is based on the knowledge-attitude behavior model, which "assumes sharing knowledge inevitably leads to behavior change" ("Emotional Responsibility" 346). This model is related to Lawrence Kohlberg's influential cognitive-development theory of morality, of which "the central tenet ... is that the sophistication of a person's moral reasoning predicts his or her moral behavior" (Aquino & Reed 1423). Kretz is critical of this model, pointing out that "a number of empirical studies falsify" its premises and that the connection between cognition and action is purely hypothetical (346). This is also what Straughan concludes in his overview:

> children may be taught a great deal about morality without being taught to be moral agents; they may fail to use the information and the skills they have acquired, when faced with a real-life moral decision, or they may fail to act upon the moral judgements they have formed. (110)

[10] This topic merits a discussion of its own and lies outside of the scope of this research project.

[11] Gracia's approach to the Socratic method differs from the one described by Altorf, which we used for the puposes of our student conversation. The former interprets it more loosely, in a general sense drawing from Socratic dialogues as presented by Plato, whereas the latter presents a particular step by-step method of facilitating conversation.

It is clear that this gap between moral judgement and moral action is by no means a problem relegated to childhood. In fact, children might be somewhat excused, especially young ones, since morality is oftentimes proposed to them as a set of entirely external rules and values, if not by the school then by their parents. If they fail to see appeal and refuse to obey the moral authority of their school or parents – as pointed out by Ariel Sarid, such refusal is part of a crucial developmental stage (257) – it means they have not internalized this morality and likely do not have a moral identity of their own yet. Thus, children who have been “morally educated” but do not act in accordance with what they have been taught are not necessarily akratic. The real conundrum is posed by all ages of adults who have formed a somewhat stable conception of what they personally believe to be right and wrong, yet fail to act in accordance with it. Ideally, of course, as children we would already start to develop our own morality and simultaneously learn to enact it. In the face of widespread akrasia, I take the claim that “we teach who we are, and that’s the problem” (Cohen qtd. in Bai 325) to be of crucial importance – if we adults are chronically akratic, how can we begin to teach children to be otherwise? We have to start working from our already present akrasia and dedicate serious attention to methods of fostering moral action. Moreover, I will discuss in Chapter 4 the idea that, in view of the virtue educational benefits of failure as proposed by Nafsika Athanassoulis, akrasia itself can play an important role in moral learning. Here, I would also like to address briefly the target-audience and the context of the educational solutions I am embarking to discuss. My approach is in line with that of Athanassoulis:

> Virtue education is not contained, it ranges from the home, to the community, to the school. Nor is it limited to a particular period of one’s life. Rather we are developing our moral characters through-out our lives . . . and in different educational contexts. (359)

Hence, moral education as I envisage it is not reserved to specific educational institutions or age groups.

3.3 Moral Action and Enkrasia as a Virtue

If, as discussed in the section on the problematic nature of akrasia in Chapter 2, since the moral “points ultimately to an ideal of the fullest possible personhood and the richest possible community”, it is in our best interest to act morally (Grisez & Shaw 99). Yet, the gap between moral belief and action is a commonplace and eternal problem that has been discussed for centuries by moral philosophers and others. Consequently, this issue merits serious attention from any form of moral education. In fact, moral action is something that is to be desired and condoned by any kind of moral stance, since morality by definition is action-oriented. Thus, the promotion of action that matches moral beliefs should emerge as a higher, universal goal of moral education. Here, I would like to propose a perspective from virtue ethics, which I deem illuminating for addressing moral akrasia and its opposite, our emerging goal, moral enkrasia.

Contemporary virtue ethicists delineate a separate group of virtues of higher order, also called structural virtues or master virtues (Steutel 129; Szutta 2.1; Hofmann et al. 286). These differ from other “lower order” virtues in that they are not concerned with the pursuit of particular goals, but with the way in which one ought to pursue goals in general (Steutel 129). These are are described as “a matter of personal psychic strength — ability and willingness to govern one’s behavior in accordance with values, commitments, and ends one is for” (Adams qtd. Szutta 2.1). Such virtues include perseverance, consistency, courage, patience, temperance and endurance. As noted by Jan Steutel, they largely have to do with self-control and “can be regarded as corrective of contrary inclinations” (131). These virtues of higher order are meant to “help us overcome various limitations, whether psychological (e.g., laziness or the lack of self-confidence) or situational (e.g., adversities of life)” (Szutta 2.1). Overall, it is clear these are the virtues for striving to minimize akratic action and promoting behavior that follows one’s beliefs. Cultivating these virtues is important for cultivating enkrasia. Thus, I suggest this should be of direct concern to moral education.

An apparent objection to this proposal is that the higher order virtues are not necessarily moral virtues in themselves – they are instrumental. As cleverly noted by Natasza Szutta, “a terrorist may also need patience and self-control” (2.1). Promoting that everyone acts according to their personal morality might also lead to more actions that are generally considered immoral in our society.[12] However, an emphasis on action and, thus, on instrumental virtues is not to signal that practicing moral reflection, deliberation and judgement are unimportant. Rather, it underscores that actions are a crucial next step to assent moral reflection as meaningful. Hence, these virtues are not sufficient, but constitute a necessary condition for moral action. For example, we need patience and self control in order to work against a habit that we ourselves deem immoral and to develop a better habit in its place.

Ultimately, I deem that the virtue ethics perspective of the higher order virtues is helpful in highlighting the practical nature of morality and turning moral education towards the cultivation of moral action. That being said, there is a lot to be discussed about the preconditions and practicalities of fostering these virtues.

4 Education towards Enkrasia

4.1 Akrasia as a Starting Point

Akrasia is a universal phenomenon and, admittedly, it might be somewhat quixotic to genuinely envision a world where it would be entirely eliminated. As asserted by Bromhall, it is only human to act akratically and, in line with early Christian thought, “so long as you are alive, you will struggle against certain tendencies of action and thought, and sometimes you will fail” (38). Nevertheless, I have shown that this gap between moral belief and action does pose a problem for the development of our full personhood, the integrity of our identity, our community and our relationship to the world. But this problem can be viewed as an opportunity. Addressing akrasia and striving to minimize it presents an opportunity for personal and social growth (Callard 172). Bromhall argues that akrasia is “a key component to meliorism” because it “provides the agent with valuable information that cannot be gleaned from anything else and acting on that information grounds the agent’s belief that improvement is possible through increased effort” (44). Nafsika Athanassoulis writes extensively on the positive role failure can have in moral education. She distinguishes between constructive failures that “lead to a positive lesson” and destructive failures that do not result in such a lesson and “may even have a damaging effect on the agent” (352). In the case of a constructive failure, “the person acknowledges the failure, may feel a variety of negative emotions associated with it and may take steps to make amends for the failure, but the end result is positive as the agent has learnt something of value” and, ultimately, has taken a step towards a more virtuous existence (351). A destructive failure, on the other hand, can have a detrimental effect on the agent’s self-esteem, lead to disempowerment and cause a total loss of interest in the original goal or ideal (351-2). However, it appears that the real difference between these two types of failure is a difference in attitude and approach. As Athanassoulis rightly notes, it might be “advisable to do away with the language of failure” (356). This term does carry a heavy negative connotation, but the growth-oriented approach in question presents an opportunity and incentive to redefine our relationship with it. Failure here could be redefined as an observation of “I am not living my life in the way I deem to be right” that leads to an inquiry upon the steps to improvement.

4.1.1 Resistance to Bringing Morality into Thought and Conversation

As previously established, a precondition for moral akrasia is caring about morality. That is, in order to observe a gap between belief and action, we need to have belief in the first place. And, indeed, few of us would admit to not having any moral beliefs or not caring about morality as such. Those who do are excused from this discussion. However, caring about morality does not necessarily mean that it makes it to the forefront of our everyday thought and interaction. In fact, oftentimes we feel resistant to this topic. Even the mention of the words “moral”, “value”, “virtue” can be seen as carrying a “moralizing” tone and make us attribute a

[12] This is a more complex point that I cannot afford to discuss at length here. I am highly skeptical of the assertion that a human being given a supportive environment to flourish and express herself would become a terrorist. This is the kind of moral educational community which I am asking for in Chapter 4. It is in this space that we should cultivate patience and self-control, and strive away from the alienation and distress that lead to such extreme behaviors as terrorism.

lecturing “holier than thou” attitude to the speaker. Moreover, as observed by Julia Minson and Benoît Monin:

> While societies may differ on what it means to be moral, they agree that it is good to be so. Yet anecdotal evidence suggests that overtly moral behavior can elicit annoyance and ridicule rather than admiration and respect. Common terms such as “do-gooder,” “goody-goody,” or “goody two-shoes” capture this negative attitude. (200)

This might be connected to the unpleasant realization that “the right way to live”, while it embodies values, rules and virtues we agree with, is something to which we rarely adhere. Research shows that we tend to hold negative attitudes towards those who are better at living up to moral standards than we are (Monin et al.; Zane et al.; Rothenberger). These studies suggest that this response represents a defense reaction to a threat on our positive self-perception: the enkratic action of others reminds us that we also have the opportunity to act morally, yet we do not take it, and, consequently, this can make us question our own moral goodness (Monin et al. 89-90). It is evident that moral considerations do not occupy the forefront of our everyday lives if it takes other people’s moral action to *remind* us of the possibility of moral action and to make us reflect on the gap between our own moral beliefs and behavior. This means that it is not unusual for us to ignore our moral akrasia; we generally avoid thinking about it or dissociate from it. A typical case of this is the meat eater’s akrasia, where one holds “belief in the value of animal wellbeing and life” yet consumes meat all the same (Aaltola, “Meat Paradox” 3). In order to deal with this cognitive dissonance, people often dissociate or choose strategic ignorance, wherein they mentally “disconnect” meat from the idea of living animal or avoid engaging with the “information available on animal minds, suffering or welfare issues” (2-3). Meat-eaters are also observed to express resentment towards vegetarians simply because they anticipate the mere fact of another person’s vegetarianism to be a personal moral reproach (Minson and Monin; Rothenberger). This resentment being part of the defense mechanism to cope with dissonance, it seems that akratic meat eaters go to great lengths not to recognize the gap between their belief and action. Returning to Athanassoulis’s account, this presents a case of “destructive failure,” where moral failure is not acknowledged and thus cannot lead to an attempt at betterment. A reason for this can be found in the fear that stems from perceiving any failure as a purely negative experience and a threat to one’s self-worth instead of seeing it as an opportunity to learn, persevere and improve.

Moreover, the negative feelings towards more morally enkratic people point to an experience of competition and a feeling of comparative inadequacy. The fact that others are good seems to simply imply that we are bad and, in order to reaffirm ourselves, we feel that we must find fault in them.

This is not a surprising pattern of thought considering the highly individualistic and competitive neoliberal society in which we live. Consequently, as Bai notes, a large part of our behavior is oriented towards invalidating others’ experience, which we do “by not respectfully and sensitively receiving and acknowledging another’s expression of experience”, for example, by “denying, ignoring, dismissing, invalidating, trivialising and/or ridiculing it” (317). Minson and Monin’s study establishes a causal link between the meat eaters’ derogatory evaluation of vegetarians and their anticipated “threat of being morally judged and found wanting” by the vegetarians (205). Thus, it seems we judge out of the fear of being judged. This fear of judgement and of discovering our shortcomings is equally represented in our internal relationship to ourselves. Awareness of the gap between our belief and action can be irritating and bring forth feelings of shame and embarrassment (Callard 172; Athanassoulis 351). Consequently, as in the example of akratic meat eaters, we often alienate ourselves from our own experience of akrasia by doing all we can to avoid recognizing it, which, ultimately, jeopardizes our potential for moral agency.

4.2 A Shared Space of Moral Reflection and Acceptance

In order to take steps towards moral action and take moral ownership of our life, we need a space to acknowledge our akrasia and seriously reflect on it. To start perceiving akrasia as a learning process and not as an irreparable disempowering failure, we need to establish and feel that:

- Akrasia is normal and human;

- We need not be judged and condemned for it;
- We are not in competition with others for moral goodness.

These points are other-related and, thus, this is not a task to be undertaken in solitude – it needs to happen in conversation with others. Bai writes, “we are fundamentally intersubjective beings, how we are received, understood and treated by others matters crucially to our sense of self and our reality” (317). Moreover, since akrasia is not solely an individual but a social problem, the solution to it should also be sought out in a social context (Aaltola “Problem of Akrasia” 126). A major task within this social context is to discuss akrasia itself as well as to acknowledge both our general and particular akratic tendencies. For this inquiry to take place, we need to pursue “an intentional turn from isolation to empathetic connection” (Zajonc qtd. in Kaufman and Murray 107). This sharing space has to also welcome all emotions associated with akrasia – feelings of shame, frustration, anxiety and anger that we often suppress within ourselves and shy away from. Bai notes that even though there is increasing awareness of the importance of emotion in learning, including moral learning, “emotions, especially dark or negative ones . . . are usually seen as irrelevance and distraction, if not hindrance, to the development of intellect, and therefore are kept out of school as much as possible” (319). She notes that the same view is accepted “at home and at work” (319), leaving little space for healthy expression and fostering emotional suppression instead. Kretz argues that emotions are an essential component for moral engagement and, therefore, we need to attend to them in order to move towards moral behavior (“Emotional Responsibility” 345).

What logically follows from a truly open, emotionally engaged reflection on akrasia is the active evaluation of our beliefs and actions. Bai writes:

> Unless we come to an explicit realisation that we are the way we are, in our thought patterns, feelings, habits and actions, due to prior conditioning and programming in alienation that we have received from our culture and family of origin, we tend to live and act out of alienated consciousness, and in turn spread, unconsciously, more alienation. (319)

Such contemplation of the self and its surroundings is necessary for “fully taking responsibility as moral agents” and becoming able to make a change (319). Bai views this as a healing process in which “we need to engage . . . collectively through relationship building and community development” (320). This suggestion is in line with contemplative pedagogy, which presents “a transformative educational practice favor[ing] approaches to teaching and learning that encourage self-actualization” (Kaufman & Murray 101). The goal for this contemplative space is to foster three dimensions of intersubjectivity:

- Experiences of attunement with self, others, and the object of shared attention;
- Shared emotions and ideas;
- Affective engagement (Kaufman & Murray 115).

If we participate in this introspective process together with others, we can move towards the realization that, instead of judgement or resentment, we can give and receive compassion. Instead of “pay[ing] attention only long enough to develop counter arguments”, we can work towards “a deep, openhearted, unjudging reception of the other” and, by extension, of ourselves (O’Reilly qtd. in Kaufman & Murray 108). Thus, in place of fearing failure and avoiding facing our shortcomings, we can know and show our flawed selves, and strive to grow.

In order to establish such a transparent, accepting, compassionate, communal contemplation space, we need to recognize the opportunity for community that goes largely unnoticed in institutional education settings, because emotional distance is systematically encouraged and both students and teachers are asked to “leav[e] [their] heart at the classroom door” (Rosales). Then, emerging from this open-hearted moral contemplation space, moral engagement and action oriented educational activities can be pursued. I will now dedicate more attention to this primary moral self-contemplation and then proceed to outline further remedies to particular causes of akrasia.

4.3 Value Clarification and Recognizing Everyday Morality

First, there is more to be said about this process of introspection. If we want to act morally, we need to engage seriously with morality in thought. Here, I suggest a return to the value clarification[13] method, where one is encouraged to "get in touch with his own values, to bring them to the surface, and to reflect upon them" (Purpel & Ryan qtd. in Straughan 16). But many practical value clarification exercises only refer to morality in abstraction. These tasks require the moral learner to create personal value inventories, rank values, compare values, make either-or choices etc. (Kirschenbaum) in a way that risks being too theoretical. Such an approach enables us to recognize our moral identity explicitly, but does not necessarily bridge the distance between said identity and our behavior, allowing for the "loving and eating animals" type of moral disconnect to go unnoticed. To correct this, we need to remind ourselves of the core nature of morality, which is to guide our action. Thus, in an action-oriented moral education model, any clarification exercise should include linking values to real life behavior and thinking about how we reflect them in our lives.[14] This can equally go in the other direction – we can look at our day, week, month, year and think about what moral considerations were at stake, and what moral values we did and did not enact. This type of reflection would foster a moral engagement with our everyday life. Such an exercise is in line with Hannah Arendt's suggestion about morality:

> Could the activity of thinking as such, the habit of examining whatever happens to come to pass or to attract attention, regardless of results and specific content, could this activity be among the conditions that make men abstain from evil-doing or even actually "condition" them against it? (5)

In short, there is a strong suggestion that "lack of attention ... can feed akrasia" (Aaltola "Meat Paradox" 8). But "the act of learning to pay attention is not a switch that one may easily flip on; rather, it is a contemplative process that requires practice and perseverance" (Kaufman & Murray 104). Moreover, truly paying attention and engaging with this world morally can be hurtful. On a personal level, it forces us to face our akrasia explicitly and, ultimately, to face the gap between who we are and who we would like to be. This is an incredibly vulnerable position, and dwelling on it can also become highly unproductive, because the akratic realization itself sounds like an attack – "why are you not doing what you should?" It only makes sense that we would routinely protect ourselves from this question, since addressing it risks making us feel weak, disempowered, like 'failures'. This is where the aspects of community and constructive failure are of importance. First, if we feel weak and disempowered, we ultimately fear that we are worse than others and that they will judge and reject us. In order to foster the courage to acknowledge our moral shortcoming and deal with the feelings they cause, we need others – a community – to whom we can admit all this and who will reassure us of their acceptance and support despite, or perhaps even because of it. Moreover, they will likely share the same struggle and will, in turn, require reassurance of our support. To continue, this acknowledgment of akrasia is not aimed at instilling guilt about our failures, but encourages us to look at these failures with curiosity in order to consider the obstacles that keep us from doing the right thing and contemplate how they could be overcome This is actually a direct exercise suggested by the value clarification method, called "removing barriers to action." It requires one to set a goal, think of what is hindering the fulfillment of this goal, and then suggest how this barrier to action could be removed (Kirschenbaum 105). In other words, to strive towards an enkratic life, we first need to follow the ancient invitation to self-knowledge and to the examination of our lives (Aaltola "Meat Paradox" 5). The list of

[13] Here again, I am omitting discussion about the difference between moral values, beliefs, judgements and virtues. All of these, to different degrees, can form our moral identity and, for the purposes of this research, a detailed discussion of them is unnecessary. Consequently, value clarification here is an umbrella term for the clarification of any components of our moral identity.

[14] This is not to say that the values clarification method does not have action as its goal. Kirschenbaum notes that this is a misconception brought on by the term "clarification"; he writes, "becoming clear about one's values . . . is certainly a major part of values clarification. But values clarification also means acting on one's values, bringing those values to fruition in one's life" (13). Thus, it is nothing revolutionary to suggest that the recognition of our values must be directed towards their actualization in our life. By calling for an action-focused approach, I am simply putting an urgent emphasis on reflecting on and aiming at the fruition of those values in the context of widespread akrasia.

the causes of moral akrasia I provide in Chapter 2 is not merely a theoretical compartmentalization, it is the type of inventory each one of us would benefit from realizing about our own lives. And, while each of us has an undeniably unique combination of personality, motivations, desires, thoughts and external influences, I deem that much of our akrasia is shared and that we can (and should) look for solutions together.

4.4 Empowerment, Practical Action and Creativity

As discussed in Chapter 2, a significant reason why we are akratic is that we do not believe that we can effect moral change in this world. Reflecting on her ethics class, Kretz writes, "When we look at a plethora of daunting current moral issues I've heard students lament 'This is so depressing'" ("Teaching Being Ethical" 162). This overwhelming feeling of despair can cause us to feel powerless and withdraw from any attempt at moral action. The world is full of wrongdoings and suffering, and it seems we can do nothing significant about it. To counter this, Kretz underlines the importance of hope and suggests practical exercises for fostering it. One is to find a person whom we admire morally and to bring their story to discussion with others, so that it can "serve as evidence that significant positive changes are not only possible but have also already happened" (163). The use of moral exemplars is a long employed method in most models of moral education (Noddings & Slote 354). However, here we need to recall the defensive reactions we might experience towards people who seem more moral than we are. A way to deal with this resistance would be to learn about the struggles these people experienced on their way to moral action and how they overcame them. Moreover, in relating to these moral exemplars, we should strive to do the hard work of shifting our attention from 'what they are and what we are not' to 'what they are and what we could become'. Indeed, the study by Zane et al. suggests that negative attitudes towards morally admirable people decrease when we "have a second opportunity to act ethically after initially ignoring" it (337). What is more, it could be of importance that these moral exemplars are close to us geographically, historically and culturally – the life of a kind neighbor might hold more tangible inspiration than that of Martin Luther King Jr. or Mother Theresa. We might equally strive to recognize, praise and discuss morally admirable action in our direct peers.

Another activity Kretz suggests for counteracting disempowerment is a "solution-focused, service-learning assignment" where students have to engage in or initiate a volunteering project to "morally improve the world in some concrete way" (162-3). She emphasizes that students must choose the project themselves so that it can directly reflect their "own moral beliefs/values/-growth" (163). Although such an assignment can clearly contribute to moral empowerment and foster moral action, I wish to note that we must also fight the tendency to assign morality to a particular enclosed sphere of life – volunteering can accidentally create a feeling that morally right action can be checked off of our to-do list for good and we need not seek it at other moments. Thus, it would be useful to have on-going personal moral projects that we continuously reflect on together with others. Explicitly setting up a project could also help us persevere with our moral goal. For example, one can decide to become vegetarian and track her progress, discoveries and hiccups in this journey. Or, one can wish to treat people with more kindness and compassion. The latter is a simple goal, but can present a lot of practical difficulty. While a component of kindness can be some specific acts of politeness, kindness cannot really be reduced to formulaic steps. It might require reworking our everyday thought patterns that can be full of defense mechanisms such as fear, envy, and anxiety. This means we might need to be creative and look for solutions in sources beyond our usual field of vision. For instance, John Paulson and Lisa Kretz explore the possible use of Buddhist meditation in moral education because the Buddhist tradition "includes techniques and practices for intentionally and systematically cultivating and strengthening the experience and expression of compassion" (326). In particular, they consider the loving-kindness meditation where "the person engaging in [the] practice recites, either aloud or internally, phrases that affirm this aspiration, such as: 'May all beings be free from danger. May they be happy. May they be healthy. May they live with ease'" (326). This type of meditation has been linked to "increased gray matter volume in areas of the brain associated with empathy" and "reduced stress response" (326). However, I would not want to imply that all methods for fostering moral action must be scientifically proven. The force of having an open-hearted community to reflect on moral goals lies in that we can share practical first-hand experience and advice. We can talk about how exactly to help a stranger being harassed on the street or what words to use to show sympathy to a grieving person.

4.5 Habits and Higher Order Virtues

Much of our akrasia lies in habit, and it is not sufficient to come to this realization and then expect a change in our behavior. Changing habits requires self-control. It requires the engagement of our system 2 thinking in order to deliberate about the actions that we otherwise do automatically. This will undeniably require more effort than our usual everyday action. This is where we will need to cultivate our higher order virtues: perseverance, consistency, courage, patience, temperance and endurance. To be ready to put in such continuous effort, we need to maintain a strong motivation, a strong sense of our identity and clear long-term goals. In fact, a general habitual pattern is that we enact short-term over long term goals. For example, we succumb to the short-term satisfaction of buying a new dress over our long-term belief that we should not contribute to the environmental and social harms of the fashion industry. A way to actualize taking control over such habits is to bring our long-term goals to thought explicitly and to ask ourselves what kind of persons we want to be – we need to actualize our moral identity.[15] A study on people who were perceived as moral exemplars by their peers found that their “moral choices were not experienced as self-sacrifice; rather they were manifestations of the exemplar’s moral center of their self” (Kretz “Teaching Being Ethical” 165). In other words, it seems we need to remind ourselves that ‘how we spend our days is, of course, how we spend our lives.’[16] A habit of thinking about this (which can even mean literally attaching this quote to a wall in our room) is crucial to maintain the motivation to change our akratic habits. But then, each akratic habit necessitates its own remedy. If it is a tendency to impulsivity and lashing out at our close ones in place of showing them the kindness we believe they deserve, we might first try to employ such a simple method as counting to ten when we feel a burst of anger. For a more grounded approach, we could benefit from a regular loving kindness meditation. At any rate, a detailed inquiry into habit changing methods is necessary, and likely requires the use of psychological knowledge and approaches like cognitive-behavioral therapy.

There are two important points I wish to bring forth in this discussion about habits and self-control. First, although we consider habits to be everyone’s personal business which, to a large extent, is true – nobody can change my habits but myself – we can benefit from sharing our struggle, our success and generally reflecting on this process together. As Aaltola argues, “reminders of self-control are important, even pivotal in this era of passive hedonism, but they require support from our wider societal settings and institutions” (“Problem of Akrasia” 125). To continue, self-control truly must be accompanied by self-compassion and patience – after all, patience is also a higher order virtue. Undeniably, there is some truth in Plato’s assertion that “each of us must flee away from lack of discipline as quickly as his feet will carry him” (qtd. in Aaltola “Meat Paradox” 5). Yet, we might be better off seeking an attitude that is “corrective to the kind of more-is-better, bigger-is-better, faster-is-better, aggressive and rapacious agency that dominates the world today” (Bai 324). Bai suggests employing the Daoist concept of wu-wei agency, which negates “pushing and straining oneself and others until we are beyond limits and out of balance” (324). In other words, our attempts to become more moral should not stem from submitting ourselves to an authoritative regime of our own making, but from seeking harmonious “unity between self and morality” (Kretz “Teaching Being Ethical” 165). This is also because self-compassion can allow us to know ourselves and genuinely examine the causes of our akrasia, whereas authoritative efforts would risk being directed only at damage control and (yet again) fear of failure.

4.6 Bending the Social Script: Community, Courage and Creativity

There is an important consideration that has been looming in the background of this discussion.

[15] This can also be done through particular exercises that require us to think about our life as a whole. Such exercises include writing a short autobiography, writing one’s own obituary, creating a life inventory and others (Kirschenbaum).

[16] A quote commonly attributed to Annie Dillard.

Let us recall Rorty's strong comparison: "just as a disposition to chronic bronchitis may indicate a toxic environment, so individual akrasia may indicate social disorder" (649). On a similar note, Aaltola writes, "reminders of self-control may not be appealing in the contemporary, consumeristic era, which is marked by the logic of marketing that precisely rests on one's lack of discipline" ("Meat Paradox" 5). As we have seen in Chapter 2, there is clear evidence that marketing strategies are employing psychological research to exploit our weaknesses and foster akratic consumption. Moreover, our social environment has deeply ingrained priorities that are based on competition and self-interest:

> We must fashion ourselves, form our abilities and habits in such a way as to make ourselves employable . . . Our role –our place – in the economy shapes our lives; it determines our security and pleasures and issues in the kind of recognition we receive. (Rorty 654)

So, engaging in the daily moral reflection that I suggest, we might come to the realization that the job we do to gain our living is completely in opposition to our deep moral beliefs. Yet, our livelihood and security largely depend on it. What is more, our social environment also sends mixed messages, presenting a fundamentally akratic space:

> While officially condemning envy as a socially undesirable trait, most societies use, and even induce, envious traits to encourage the development of useful talents and abilities. Market-based, consumer oriented economic systems generate invidious comparisons as a way of increasing consumption. (Rorty 653)

Such a contradictory mindset not only leads us to internalize a constant striving to be better than others and, thus, to fear open communication with them while simultaneously ourselves for this same attitude and ultimately fosters our disengagement from our emotions and moral ideals. Taking this into account, how, then, are we to become more enkratic in face of this abundance of social forces that push us towards akrasia?

To begin, I must reiterate that this explicit acknowledgment of akrasia and the forces that drive it, however painful, is important. Rorty argues that "if [one] were to recognize the extent to which her desires have been manipulated, realizing exactly how socioeconomic policies violate her more general aims, she might be better positioned to check her akrasia" (657). This is also asserted by Athanassoulis: "Key to resisting contrary situational factors is being made aware of their influence." (357) Kretz equally sides with this thought and suggests the strengthening of our moral identity:

> Knowledge of how hyper-consumerist identities are generated, maintained, and then marketed to can enable well-informed approaches to counter such identity construction and behavior in favor of morally grounded identities and behaviors. ("Climate Change" 18)

Moreover, this realization of the social aspects of akrasia allows us to recognise that we are not in it alone and, thus, enables the kind of communal sharing and contemplation that I suggest. Ultimately, together we can take responsibility, develop a support system and look for solutions on both the individual level and the larger social scale. Yet, Rorty rightly warns that "tracing the economic and political sources of akrasia can sometimes also unfortunately deflect individual therapeutic measures" to the extent that we "may self-deceptively disown [our] akratic actions" (657). To such a line of thought Grisez and Shaw answer: "Granted, there are factors beyond our control which powerfully influence our development as human beings, but what of those factors which are within our power to control?" (172) I deem that, especially on the level that concerns our relations to our direct surroundings, there is much we can, in fact, do to become the kind and caring persons most of us would like to be. I do not think I need to refer to literature to remind us of how much warmth and inspiration even a simple act of kindness can radiate. It makes little sense to refuse taking the small steps that are indeed accessible to us. If we refuse to take action to cultivate our moral ideals daily, we are withdrawing from the exercise of our freedom to self-determination[17] (176), lest we readily admit to having no such freedom, in which case, of course, there

[17] As noted by Aaltola, this also resonates with "the existentialist notion of 'bad faith'", where we surrender our freedom to choose, our capacity to use our own reflection, to the hubbub of social custom that tells us what we ought to do, what we ought to be like, and within which we thereby lose our sense of "authenticity"—our capacity to make responsible

is no space for active improvement and this whole discussion collapses.

When it comes to the larger social circumstances, we must remember that “people can be producers as well as products of their environments” (Hannah et al. 674). In the light of this, we must realize that we have the freedom to at least commit ourselves to bettering our social conditions. Kretz’s conception of responsibility is in place here:

> Responsibility as I intend it here is not about identifying who to blame and therefore hold accountable, an approach which enables one to ostensibly absolve oneself of responsibility. Responsibility is to be conceptualized in a forward looking, positive way – it is about taking responsibility through identifying how to help and acting accordingly. (“Climate Change” 16-17)

To continue, she also notes that this social responsibility requires activism, since individual action indeed does not suffice for large-scale reform (17). Therefore, “collective activism is needed to achieve political reform which can encourage or require responsible lifestyles” (17). Moreover, activism also holds the potential to create a strong sense of community and identity, itself engendering a motivating force for a more morally engaged life.

Nevertheless, both activism and any sort of individual moral action require the use of moral courage and creativity. Courage means that we have to be ready to endure the social dangers of standing up for our moral ideals and to insist on them under pressure (Hannah et al. 677). Creativity is necessary so that we can figure out what exactly, in practice, the morally enkratic world we want to live in – the interpersonal relations, the social structures and our attitude towards nature and nonhuman beings – could look like. This creativity is part of the phronesis that we need to employ in our daily attempts to a more morally engaged existence. If we wish to counter the current harmful, over-consumptive, competitive social script, we need to start figuring out what new habits and patterns of relating should replace this.

As a final note, I wish to share a stanza of activist and poet’s Marge Piercy’s poem “The Low Road” through which Kauffman and Murray (102) encouraged collaborative contemplation and co-creativity in their classroom:

> It goes on one at a time,
> it starts when you care
> to act, it starts when you do
> it again after they said no,
> it starts when you say We
> and know who you mean, and each
> day you mean one more.

5 Conclusion

Once we have dedicated in-depth attention to it, akrasia seems to constitute less of a paradox,while still remaining a problem that demands a remedy. Locating particular reasons for the gap between our moral beliefs and daily behavior has allowed us to seek a middle way between a forceful “just do it” mentality that reduces all failure to simple weakness, and a disempowerment that makes a morally integral life seem simply impossible. In sum, I deem that we are facing legitimate difficulties, but by admitting these difficulties and reflecting on them, we have the opportunity to work towards an enkratic life. In part, this certainly requires introspection on an individual basis. Yet, akrasia is also a social phenomenon both in the sense that it is pervasive and that it is, in many ways, enabled by our social structures. Hence, the solution is to be sought socially. Here, in a broad sense, I mean that we need to cultivate shared moral reflection about our daily lives, somewhat in the line of an “action + reflection = transformation” equation (Freire qtd. in Kaufman & Murray 110). In a more particular sense, I hold that it is in educational institutions where such communal contemplation about moral practice has the potential to be pursued, and this is then what we could conceive as “moral education” in contemporary schooling. However, as noted beforehand, I also do not want to imply that there is a part of education where morality matters and that we can ignore it in the remaining parts – such implicit division is exactly what enables us to keep our everyday action distanced from our moral beliefs. Moral education should foster moral reflection in all aspects of education and, by extension, all facets of life.

choices—and become “alienated” from others and ourselves. (“Problem of Akrasia” 135)

Nevertheless, I am again having trouble with strictly delineated terms, because education certainly is not limited to academic institutions. It is hardly a radical idea to assert that we continue learning throughout our lives and outside of schools and universities. The same applies to moral learning. When we think of "moral education", perhaps something like a class full of unruly children comes to mind, where the teacher undertakes to convey morality to them in a transactional manner. However, I deem most adults are in need of a moral education – we lack the practical knowledge of how to live according to our moral ideals and, as I note in Chapter 4, it can be an intimidating and shameful experience to come to terms with our akrasia. Thus, we tend to numb ourselves and become defensive at the mention of morality. The moral education from which we would benefit includes a supportive community that welcomes our frustrations and our emotions to empower us to truly pursue our ideals. But how do we create such a space in a society permeated by individualism, competition and emotional disengagement? This is an undeniably difficult question. For one, we might seek to realize the potential of communities that are already present to us. A moral learning community could begin with a few friends having dinner together and being intimate enough to start a genuine conversation about everyday morality. It could also begin when we encounter somebody whose moral integrity we admire and we muster the courage to inquire about what lies behind their success.

Similarly, institutionalized education has a largely unrealized potential for a supportive community and shared moral reflection. Without really planning for it, perhaps by an intuition, the Socratic conversation I initiated for this research project engendered a tentative step towards realizing this moral community potential that lies also within my university. As we asked ourselves "why are we not doing the right thing?" and, in seeking the answer, discussed personal experiences of akrasia, we were working towards exactly the kind of reflection and open-heartedness I am suggesting is crucial for moral learning. The Socratic method as conceived here, however, puts emphasis on philosophical contemplation in itself rather than for the purpose of seeking practical solutions. Philosophical inquiry into our lives is important in order to begin striving towards moral action – the method can be seen as a first step, or can be adjusted to our goal of enkrasia.

That being said, any such activity requires students and teachers who have the interest and energy for this. Here again, we must acknowledge that in the current system, where we are in a constant rush to prove ourselves and to grow – but to grow to be more employable and get tangible credit rather than to realize our full humanity – there is little space to think about the morality of our lives. How do we create this space? I think, while it is true that radical change cannot be brought on by individual action, any change necessarily starts with an individual who manages to notice, question and bring to the attention of others the akratic contradictions we live in. Moreover, research suggests we have a tendency to underestimate our influence on other people's moral behavior (Bohns et al.). Even by very small conversational steps, we can reevaluate the akrasia we have collectively taken to be the norm. As such, I have done my best to bring this to your attention, dear reader of this essay. I invite you to think about it for yourself and, then, perhaps muster the courage to reflect on it together with others. There is indeed much more to think about than what I have already laid out: there might be more reasons for akrasia than I have noted, the educational methods – formal and informal – I suggest are far from comprehensive, the practical introduction of such moral community in any particular educational institution needs detailed discussion, the important role of the teacher undeniably deserves our attention, the link between activism and enkrasia is to be explored in depth, and, after all, although I have treated it as secondary in order to thoroughly focus on moral behavior, the way we develop moral beliefs in the first place is still of relevance. Let us deliberate about this together.

6 References

Aaltola, Elisa. "The Problem of Akrasia: Moral Cultivation and Socio-Political Resistance." *Philosophy and the Politics of Animal Liberation*, 2016, pp. 117-47, doi:10.1057/978-1-137-52120-0.

—. "The Meat Paradox, Omnivore's Akrasia, and Animal Ethics." *Animals*, vol. 9, no. 12, 2019, pp. 1–16, doi:10.3390/ani9121125.

Altorf, Hannah Marije. "Dialogue and Discussion: Reflections on a Socratic Method."

Arts and Humanities in Higher Education, vol. 18, no. 1, 2019, pp. 60–75, doi:10.1177/1474022216670607.

Aquino, Karl, and Americus Reed. “The Self-Importance of Moral Identity.” *Journal of Personality and Social Psychology*, vol. 83, no. 6, 2002, pp. 1423–40, doi:10.1037/0022-3514.83.6.1423.

Arendt, Hannah. *The Life of The Mind*. Harcourt Brace Jovanovich, 1978, pp. 5-10.

Athanassoulis, Nafsika. “A Positive Role for Failure in Virtue Education.” *Journal of Moral Education*, vol. 46, no. 4, 2017, pp. 347–62, doi:10.1080/03057240.2017.1333409.

Audi, Robert. “The Practical Authority of Normative Beliefs: Toward an Integrated Theory of Practical Rationality.” *Organon* F, vol. 20, no. 4, 2013, pp. 527–45.

Bai, Heesoon. “Reclaiming Our Moral Agency through Healing: A Call to Moral, Social, Environmental Activists.” *Journal of Moral Education*, vol. 41, no. 3, 2012, pp. 311–27, doi:10.1080/03057240.2012.691628.

Bohns, Vanessa K., et al. “Underestimating Our Influence Over Others’ Unethical Behavior and Decisions.” *Personality and Social Psychology Bulletin*, vol. 40, no. 3, SAGE Publications, Mar. 2014, pp. 348–62, doi:10.1177/0146167213511825.

Bromhall, Kyle. “Embodied Akrasia: James on Motivation and Weakness of Will.” *William James Studies*, vol. 14, no. 1, 2018, pp. 26–53, doi:10.2307/26493690.

Burch, Matthew. “Making Sense of Akrasia.” *Phenomenology and the Cognitive Sciences*, vol. 17, no. 5, Phenomenology and the Cognitive Sciences, 2018, pp. 939–71, doi:10.1007/s11097-018-9568-9.

Callahan, Daniel. Ethics Teaching In Higher Education. Plenum Press, 1980, pp. 62-67.

Callard, Agnes. “Akrasia.” *Aspiration: The Agency of Becoming*, Oxford University Press, 2018, pp. 149–76.

Chappell, Timothy D. J. *Aristotle and Augustine on Freedom : Two Theories of Freedom, Voluntary Action and Akrasia*. Macmillian, 1995.

Foot, Philippa. “Are Moral Considerations Overriding?” *Virtues and Vices and Other Essays in Moral Philosophy*, Oxford University Press, 2002.

Geiger-Oneto, Stephanie, and Elizabeth A. Minton. “How Religiosity Influences the Consumption of Luxury Goods: Exploration of the Moral Halo Effect.” *European Journal of Marketing*, vol. 53, no. 12, 2019, pp. 2530–55, doi:10.1108/EJM-01-2018-0016.

Gracia, Diego. “The Mission of Ethics Teaching for the Future.” *International Journal of Ethics Education*, vol. 1, no. January, 2016, pp. 7–13, doi:10.1007/s40889-015-0008-1.

Grisez, Germain, and Russell Shaw. *Beyond the New Morality: The Responsibilities of Freedom*. 3rd ed., University of Notre Dame Press, 2006.

Hannah, Sean T., et al. “Moral Maturation and Moral Conation : A Capacity Approach to Explaining Moral Thought and Action.” *The Academy of Management Review*, vol. 36, no. 4, 2011, pp. 663–85, doi:10.5465/AMR.2011.65554674.

Hare, Richard M. *Freedom and Reason*. Oxford University Press, 1965.

Hofmann, Wilhelm, et al. “Morality and Self-Control: How They Are Intertwined and Where They Differ.” *Current Directions in Psychological Science*, vol. 27, no. 4, 2018, pp. 286–91, doi:10.1177/0963721418759317.

Kauppinen, Antti. “Moral Internalism and the Brain.” *Social Theory and Practice*, vol. 34, no. 1, Florida State University, Jan. 2008, pp. 1–24, doi:10.5840/soctheorpract20083411.

Kirschenbaum, Howard. *Values Clarification in Counseling and Psychotherapy*. Oxford University Press, 2013.

Kraut, Richard. “Aristotle’s Ethics.” *Stanford Encyclopedia of Philosophy*, Center for the Study of Language and Information (CSLI), Stanford University, 2018.

Kretz, Lisa. “Climate Change: Bridging the Theory-Action Gap.” *Ethics and the Environment*, vol. 17, no. 2, 2012, pp. 9–27.

—. “Emotional Responsibility and Teaching Ethics: Student Empowerment.” *Ethics and Education*, vol. 9, no. 3, 2014, pp. 340–55, doi:10.1080/17449642.2014.951555.

—. “Teaching Being Ethical.” *Teaching Ethics*, vol. 1, no. Fall, 2014, pp. 151– 72, doi:10.5840/tej2014111311.

Minson, Julia A., and Benoît Monin. “Do-Gooder Derogation: Disparaging Morally Motivated Minorities to Defuse Anticipated Reproach." *Social Psychological and Personality Science*, vol. 3, no. 2, 2012, pp. 200–07, doi:10.1177/1948550611415695.

Monin, Benoît, et al. "The Rejection of Moral Rebels: Resenting Those Who Do the Right Thing." *Journal of Personality and Social Psychology*, vol. 95, no. 1, 2008, pp. 76–93, doi:10.1037/0022-3514.95.1.76.

Mullen, Elizabeth, and Benoît Monin. "Consistency Versus Licensing Effects of Past Moral Behavior." *Annual Review of Psychology*, vol. 67, no. 1, 2016, pp. 363–85, doi:10.1146/annurev-psych-010213-115120.

Murray, Terry, and Peter Kaufman. "From Me to We: An Experiment in Critical Second-Person Contemplative Pedagogy." *The Intersubjective Turn: Theoretical Approaches to Contemplative Learning and Inquiry Across Disciplines*, edited by Olen Gunnlaugson, 2017, pp. 101–21.

Noddings, Nel. *Caring: A Relational Approach to Ethics and Moral Education*. University of California Press, 2013.

Noddings, Nel, and Michael Slote. "Changing Notions of the Moral and of Moral Education." *The Blackwell Guide to the Philosophy of Education*, edited by Nigel Blake et al., Blackwell Publishing, 2003.

Paulson, John, and Lisa Kretz. "Exploring the Potential Contributions of Mindfulness and Compassion-Based Practices for Enhancing the Teaching of Undergraduate Ethics Courses in Philosophy." *Social Science Journal*, vol. 55, no. 3, Western Social Science Association, 2018, pp. 323–31, doi:10.1016/j.soscij.2017.12.003.

Purpel, David E. "Moral Education: An Idea Whose Time Has Gone." *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, vol. 64, no. 5, 1991, pp. 309–12, doi:10.1080/00098655.1991.9955877.

Rehman, Sharaf N. "Teaching Ethics in an Unethical World." *Annales. Etyka w Życiu Gospodarczym*, vol. 20, no. 4, Wydawnictwo Uniwersytetu Łódzkiego, 2017, pp. 7–18, doi:10.18778/1899-2226.20.4.01.

Rorty, Amélie Oksenberg. "The Social and Political Sources of Akrasia." *Ethics*, vol. 107, no. 4, 1997, pp. 644–57.

Rosales, Janna. "Cultivating minds and hearts." *University Affairs*, 2012, https://www.universityaffairs.ca/features/feature-article/cultivatingminds-and-hearts/.

Rosati, Connie S. "Moral Motivation." *Stanford Encyclopedia of Philosophy*, Center for the Study of Language and Information (CSLI), Stanford University, 2016.

Rosenfeld, Daniel L., and Janet A. Tomiyama. "When Vegetarians Eat Meat: Why Vegetarians Violate Their Diets and How They Feel About Doing So." *Appetite*, vol. 143, Elsevier Ltd, Dec. 2019, p. 1-9, doi:10.1016/j.appet.2019.104417.

Sarid, Ariel. "Between Thick and Thin: Responding to the Crisis of Moral Education." *Journal of Moral Education*, vol. 41, no. 2, 2012, pp. 245–60, doi:10.1080/03057240.2012.678054.

Shafer-Landau, Russ. *The Fundamentals of Ethics*. 3rd ed., Oxford University Press, 2015.

Steutel, Jan. "The Virtues of Will-Power: Self-Control and Deliberation." *Virtue Ethics and Moral Education*, edited by David Carr and Jan Steutel, Routledge, 2005, pp. 130–41.

Steward, Helen. "Akrasia." *Routledge Encyclopedia of Philosophy*. Taylor and Francis, 1998, doi:10.4324/9780415249126-v003-1.

Straughan, Roger. "Can We Teach Children to Be Good? Basic Issues in Moral, Personal and Social Education." *British Journal of Educational Studies*, vol. 32, no. 1, 1984, doi:10.2307/3121134.

Szutta, Natasza. "The Virtues of Will-Power – from a Philosophical & Psychological Perspective." *Ethical Theory and Moral Practice*, Ethical Theory and Moral Practice, 2020, doi:10.1007/s10677-020-10068-1.

Winch, Christopher, and John Gingell. *Key Concepts in the Philosophy of Education*. Routledge, 1999.

Zane, Daniel M., et al. "Do Less Ethical Consumers Denigrate More Ethical Consumers? The Effect of Willful Ignorance on Judgments of Others." *Journal of Consumer Psychology*, vol. 26, no. 3, 2016, pp. 337–49, doi:10.1016/j.jcps.2015.10.002.

Humanities

A Move Towards Visibility

Representations of Working-Class Indigenous Women in Classic and Contemporary Mexican Cinema

Martha Echevarría González

Supervisor
Dr. Christina Buckley (AUC)
Reader
drs. Tina Bastajian (AUC)

Photographer: Che Spraos Romain

Abstract

This paper argues that Mexican national cinema, along with other national arts, has contributed greatly to the construction of a homogenous national imaginary in which marginalized groups, such as working-class indigenous women, are essentialized and stereotyped. Borrowing from Mexican writers such as Octavio Paz and Carlos Monsiváis, I provide an overview of the history of Mexican indigenous women in the national arts to understand their placement in the national imaginary. In terms of film, I first examine the Golden Age of Mexican cinema to explore its construction of indigenous female archetypes in the context of the *indigenismo* movement and its impact on spectators through an analysis of *Maria Candelaria* (1944). With this historical discourse in mind, and from an intersectional feminist perspective, I then examine two contemporary Mexican films, *Roma* (2018) and *La Camarista* (2018), to explore how some recent movies are calling attention to the most marginalized groups of society. Drawing on the work of Latin American film scholars as well as on feminist film theory, I present original close scene analyses to examine the ways in which contemporary films are offering working-class indigenous women the opportunity to reclaim their space in the national imaginary. These close analyses show that there has been a move in contemporary films towards making visible the struggles and realities of working-class indigenous women – as gendered, racialized, and classed citizens – on screen and in today's society. As such, this thesis shows the compelling need to reconstruct Mexico's national imaginary in a more inclusive and heterogeneous form, and the trend that is emerging in this direction.

Keywords and phrases: *Mexican cinema, intersectionality, female archetypes, indigenous women, national imaginary*

Contents

1 Introduction

1.1 Visibility in Contemporary Mexican Cinema

In the last decade, several Mexican filmmakers have worked on projects that document or dramatize real life stories about indigenous people in and outside of their communities. Feature films such as *Roma* (dir. Alfonso Cuarón) and *La Camarista [The Chambermaid]* (dir. Lila Avilés), which both premiered in 2018, were made as an attempt to give visibility to one of the most marginalized and ignored groups in Mexico: working-class indigenous women. Like many other contemporary Mexican films[1] which attempt to bring the marginalized to the center, *Roma* and *La Camarista* aim to create a verisimilar representation of female Mexican laborers and thus call attention to a sector of society that is often overlooked in both the arts and in the public sphere. To do so, both films feature indigenous working-class women as protagonists whose struggles seek to expose the realities of the social, economic, racial, and gender inequalities in Mexico City. In an attempt to create realistic and fair representations of these women, both filmmakers have chosen to tell the women's stories as the main plot of the films and to work with actors from a working-class background as protagonists of their own stories.

Roma and *La Camarista* are films that are interested in building a space for new cinematic representations of those women that were hardly visible before: indigenous and mestizo[2] women that have moved into larger cities to work as cleaners, cooks, domestic workers, and artisans. Most importantly, these films explicitly attempt to debunk female and indigenous archetypes that have been constructed and reproduced throughout Mexican film history, particularly since the Golden Age of Mexican cinema (roughly from 1935 to 1960). In order to analyze which conventional paradigms these films are challenging, I will focus on their innovative and disruptive way of representing working-class indigenous women. The close scene analyses of *Roma* and *La Camarista* will look at the relation between these films' cinematic narrative strategies and what feminist authors describe as the 'gaze', the myth of women, and the power relations on screen. Using these terms within the film analysis, I will explore how contemporary films can contribute to constructing heterogeneous and non-essential images of indigenous women in the contemporary national imaginary, and the impact this may have on flesh and blood women off-screen.

Before doing so, this study will present its theoretical framework based mainly on a series of Latin American and feminist film scholars, as well as on the concept of intersectionality. Then, with these theories in mind, it will review the legacy of the Golden Age of Mexican Cinema – also known as Classical Mexican cinema – as a project of nation building by incorporating the ideas of feminist scholars and Mexican intellectuals who discuss the role that women, and particularly indigenous women, have played in Mexican discourse, films, and arts of the 19^{th} and 20^{th} century. Finally, this section will explore the essentialist and simplistic representations of indigenous women in cinema and the national arts, and how these have affected their conditions of visibility in the national imaginary.

After examining the ideological constructions that were made during the Golden Age of Mexican Cinema through a close analysis of *Maria Candelaria*, I will move on to analyze how *Roma* and *La Camarista* magnify marginalized women's visibility today. In this text, the concept of visibility will refer to the extent to which marginalized groups are included in the production of national arts, and thus the extent to which they are incorporated in the national imaginary. Finally, for this study, I assume that the more verisimilar representations of marginalized groups there are in the national arts, the more possibilities these groups have to be included in the national imaginary and consequently, the less difficult it is for them to participate socially, economically, and politically. The more visibility and attention they receive, the more likely it is for

[1] Some examples of recent Mexican films that tell the life and stories of working-class, indigenous, and other marginalized groups are *La Tirisia* (dir. Jorge Pérez Solano, 2014), *Sueño en otro idioma* [I Dream in Another Language] (dir. Ernesto Contreras, 2017), *Lorena, la de pies ligeros* [Lorena, Light-Footed Woman] (dir. Juan Carlos Rulfo. 2019), and *El sueño de Mara'akame* [Mara'akame's dream] (dir. Federico Cecchetti, 2016).

[2] In the Latin American context, mestizo/a refers to a person of mixed European and 'Indian' blood (Encyclopædia Britannica). In the colonial caste system, mestizos held a middle social position, placed under Europeans and above indigenous. Since the physical characteristics that distinguished mestizos from other groups were not always obvious, "mestizaje became as much a cultural identity. . . as a racial identity" (Ching et al. 92-93).

them to receive equal and just treatment from society and the nation.

1.2 Feminism & Intersectionality in Mexican Film Studies

Understanding the different socio-cinematic representations – from the Golden Age compared to those from contemporary times – through an intersectional feminist perspective is essential when examining the role cinema plays in shaping the conditions of visibility for indigenous women of contemporary Mexico. The concept of intersectionality in feminist theory is an attempt to bring diversity into feminism by taking into account the fact that not all women experience inequality in the same way. Intersectionality expresses the idea that each woman lives at the junction of different systems of privilege and oppression; therefore women experience different levels of discrimination in terms of class, race, ethnicity, religion, and so on, which differentiate their experiences of what it is to be a woman. In order to analyze the role that cinema plays in shaping indigenous working-class women's visibility, it is thus essential to look at their representation and their experiences from an intersectional perspective. This study draws on diverse feminist theories of film and visual culture – Claire Johnston and bell hooks, as well as Ana López and Dolores Tierney – to explore the roles that "old" and "new" Mexican cinema play in the construction of marginalized women in the Mexican national imaginary.

In her essay "Women's Cinema as Countercinema" feminist scholar Claire Johnston looks at the unequal evolution of male and female myths in cinema and addresses the way myths of women, the vamp and the straight girl, have operated in Classical Hollywood cinema. Similarly, Ana López and Dolores Tierney explore the myths of women in Classical Mexican cinema through the perspective of gender studies. In 1975, Laura Mulvey published her now seminal essay "Visual Pleasure and Narrative Cinema," where she develops the concept of the male gaze, or the masculine-coded camera, through which Classical Hollywood films reflect and reinforce the unconscious patriarchal binary of male/active and female/passive. Through the male gaze, the ideal (male) spectator derives pleasure via narcissism (identifying with the ego-ideal male protagonist) and voyeurism (identifying with the masculine-coded camera) (Mulvey 837-839). This study is informed by Mulvey's reading of the male gaze, however it focused mainly on the socio-historical background of the spectators rather than on a psychoanalytic interpretative framework.

Drawing on Mulvey's theory of the male gaze, bell hooks introduces the juxtaposed concept of the 'oppositional gaze' which refers to the gaze of those who are not being represented; in her argument, this was the black female gaze (116-117). Since people of color have historically been denied the right to look, hooks argues that an oppositional gaze can become a site of resistance from the dominant white patriarchal power (116). As there is a historical similarity between the power relations of black and white Americans, and the power relations of indigenous Mexicans and creole Mexicans[3], this paper adopts hooks' theory of the oppositional gaze of African Americans and applies it to that of Mexicans of indigenous heritage. For the purpose of this study, the idea of the oppositional gaze will help us to understand indigenous women's place as underrepresented and, therefore, oppositional 'gazers' in both Classic and contemporary Mexican cinema. With the help of these scholars, this thesis will transition from an overview of Golden Age Mexican Cinema to a study of the ways in which contemporary Mexican films are reframing and challenging archetypal representations of marginalized women today.

Following this conceptual framework, this comparative study will incorporate the work of intersectional feminist scholars – again bell hooks and also Claire Johnston – to analyze gendered, classed, and racialized power relations involved in filmic treatments of white and indigenous female and male characters. The work of Latin American film scholars who have already incorporated seminal feminist film theories into their studies of classical Mexican cinema will serve as a foundation upon which this study will build when analyzing contemporary Mexican films and the roles they play in shaping the Mexican national imaginary.

[3] In the Mexican colonial caste system, gachupines (Peninsulares or Spanish-born whites) and creole (Spanish-whites born in Mexico) were at the top of the social hierarchy, while mestizos (indigenous and Spanish mixed) and indigenous followed lower in the social and legal hierarchies.

1.3 Women in The Golden Age of Mexican Cinema: The Construction of a Nation

The Golden Age of Mexican cinema – along with Muralism[4] – is one of the most representative national arts of the 20^{th} century to contribute to the construction of female archetypes and the development of a system of values and beliefs that established the role that men, women, and indigenous people were expected to play in Mexican society. The Golden Age is particularly relevant for two reasons: its magnitude and its influence on its contemporaneous audiences. First, the Golden Age was one of the most prolific periods in the history of Mexican cinema. Between 1932 and 1939, the Mexican film industry produced 236 movies, and in the next ten years, between 1940 and 1949 it increased its productions to a total of 665 feature films (de la Vega Alfaro, 24-25). This cinema grew thanks to the introduction of sound and to the interest of the government in investing in the national arts in order to establish a sense of unifying national identity (de la Vega Alfaro 23).

As the industry flourished, its films' themes and archetypes became a reflection of 20^{th} century Mexico while at the same time contributing to the shape and contour of the national identity (Mraz 92). Part of this project of nation building was the *indigenismo* movement, which offered Mexico a myth of origin free from its colonial past. The purpose of *indigenismo* was to reconstruct the national identity through a romanticized and essentialist picture of indigenous people to "assimilate [them] within the nation state" (Tierney 75). The national cinema, alongside muralist painters, portrayed indigenous people in a way that reflected the notion of a pure and essential Mexican identity, and that prioritized a 'native style' over a 'European style' (Tierney 76-77). However, as Tierney notes, this project was far from reality – most of the underclass indigenous communities were physically, economically, and racially separated from the rest of the nation (74).

For Carlos Monsiváis[5], an important element of the process of nation building is the consumption of popular culture by the masses ("Cultura Popular" 98-99). This echoes Benedict Anderson's idea that imagined communities[6] were first formed thanks to the printed press, and consequently, contemporary ones are formed thanks to the news and media that are consumed en masse. In Mexico, Golden Age cinema was also a means to develop the imagined community and the unified national identity. Therefore, with the hopes of analyzing the past, present, and future constructs of Mexican identities, historians and academics have extensively studied this emblematic era of national filmmaking. Latin American film scholar Dolores Tierney challenges the canonization of the Golden Age cinema and debunks many of the original readings of these films to highlight the ideological and representational contradictions present in this period of filmmaking. Similarly, Latin American film scholar Ana López explores the role of women in popular genres, such as the melodrama and the cabaret film, to map and then interrogate the archetypes of women constructed in Classical Mexican cinema. Other scholars such as Juan Pablo Silva Escobar use these mapped archetypes to understand how the Golden Age cinema has contributed to building a Mexican social imaginary. He argues that films of this period were responsible for elaborating images and ideas of what is conceived as 'typically Mexican', and for inscribing these ideas in the collective consciousness (10). *Maria Candelaria* (dir. Emilio Fernández, 1944), which this study will later explore in greater detail, is a canonized film which constructs a damaging and othering image of indigenous people, and particularly indigenous women.

Movies of the Mexican Golden Age were not so concerned with creating realistic representations of women, but rather with creating archetypes and moralistic characters from whom viewers could learn and with whom they could identify (Monsiváis, "Cultura Popular" 105). Therefore, popular

[4] Mexican Muralism was an art movement and a project of nation building that began in 1920 and lasted until around 1970 with the purpose of reunifying the country after the Mexican Revolution. Its paintings were usually charged with social, political, and historical motifs that aimed at uniting all Mexicans into one common history (Greeley 263-267).

[5] Carlos Monsiváis, belonging to a very active generation of Mexican journalists and writers, became a fundamental figure in the documentation of Mexican values, traditions, and social changes from the late 20^{th} century.

[6] For Anderson, a nation is an imagined community in the sense that it is socially constructed and imagined by individuals, the media, politics, etc. Ian Buchanan states, "it is imagined because the actuality of even the smallest nation exceeds what it is possible for a single person to know – one cannot know every person in a nation, just as one cannot know every aspect of its economy, geography, history, and so forth." (244). Thus, the 'imagined community' is a way for people to abstract their own and other communities.

genres of this period, such as the melodrama and the cabaret film, focused on creating contrasting white female archetypes such as the sacred, well-behaved mother and the sensual mistress, in order to reinforce patriarchal values (López 147-154). However, and even with the presence of the *indigenismo* movement, real indigenous women received little attention in the national cinema, and as scholar Dolores Tierney explains, that scant treatment of the indigenous "often reflects the fantasy of otherness, painting the *indígena[s]* as an exceptional other while suppressing the reasons for [their] social marginalization, backwardness and exploitation." (Tierney 74).

These archetypal representations from Classical Mexican cinema established trends, behaviors and a model for working class and aristocratic families to follow (Monsiváis, "Cultura Popular" 113). These models and archetypes became so embedded in the national imaginary of Mexican audiences and filmmakers that films today still struggle to diverge from them (Silva Escobar 11-12). Contemporary film production has become more independent from the state and therefore offers both films with archetypal representations of female figures (commonly found in mainstream cinema) and films that challenge these figures. Since the academic discussion surrounding contemporary Mexican cinema needs to be updated, this research will analyze two movies produced within the last decade to observe the most recent changes in both cinema and the society that these films reflect.

2 Women in the National Imaginary

2.1 Archetypes and Stereotypes

In 1950, writer and intellectual Octavio Paz wrote *The Labyrinth of Solitude*, a book-length essay in which he attempts to decipher the characteristics, traits, and historical elements that define and shape Mexican identity. It is important to note that an identity in this sense is not an ontological fact, but is rather socially determined and "its meaning is constructed by the people who try to define it." (Ching et al. 7). In this search, Paz draws on colloquialisms and local expressions, and finds that one of the most important elements defining Mexicanness is an identity born from a history of rape and female treason[7] (57-79). Carlos Monsiváis takes Paz's influential essay as a widely accurate historical-cultural interpretation of Mexican society, and notes that "by nature and definition, Mexican culture is a sexist culture" ("Soñadora, coqueta y ardiente" 23). He explains that Mexico is a culture divided by 'feminine' and 'masculine' roles with certain characteristics assigned to each that allow for the perpetuation of a patriarchal ideology ("Soñadora, coqueta y ardiente" 22-23). Similarly, Paz argues that Mexican people always understand the role of women as an instrument or a means – to fulfill men's desires or whatever tasks the law or society assign – but never as an end in itself (57-60). Under these social conditions, diverse realms of artistic expression such as literature, painting, and film have interpreted the position of women in Mexican society.

Monsiváis explores the literature of the 19^{th} century, which was preoccupied with depicting different archetypes of Mexican women, in order to define what he calls *la sensibilidad femenina* [the feminine sensibility] and its opposition, the free woman. The feminine sensibility is morality, inspiration and tenderness, while the free woman is usually incarnated by the prostitute who 'negates' the real femininity (Monsiváis, "De la construcción de la 'sensibilidad femenina' 82-83"). However, while 19^{th} century arts and literature only played around with these archetypes and moral values, mid 20^{th} century Mexican cinema became moralizing and didactic by creating more culture-specific stereotypes of bourgeoise, working-class, and indigenous women. Although the Mexican Revolution (1910) opened many doors for women to participate more actively in society, the national arts of that period did not reflect these changes. Instead, they continued repeating female stereotypes and moralistic ideas (Monsiváis, "Soñadora, coqueta y ardiente" 38).

One of the most effective ways of transmitting a homogenous ideology – in this case the idea of the Mexican – is through archetypes and stereotypes. However, it is worth noting the difference between the former and the latter. Archetypes, as established by psychoanalyst Carl Jung, are "ways

[7] Paz tells the story of Marina or La Malinche who in times of the Spanish conquest was given to Hernán Cortes, the Spanish conquistador, as a wife and slave. Because she gave birth to one of the first mestizos she is seen as a traitor to the Aztec people and as the raped mother of all Mexicans (57-79).

of thinking and acting that derive from the most primitive aspects of our psyche" (Buchanan 25) and that reside in our collective unconscious. Stereotypes, on the other hand, are generalized assumptions about groups and communities which, according to Stuart Hall, "reduce people to a few, simple, essential characteristics," (257) and thus are inherently essentializing and reductionist. The usage of archetypes in stories and narratives is important because it helps readers and audiences identify and relate to the different characters (the hero, the villain, the ruler, etc.). However, as Mary Anna Kidd argues in her study of archetypes and stereotypes in media representation, when archetypes are married to stereotypes it leads to problematic stigmatization of groups, particularly in multicultural societies (26).

Along these lines, then, the question as to what role the moralistic stereotypes from 19^{th} and 20^{th} century Mexican arts played in the construction of indigenous women in the Mexican national imaginary arises. First, it should be noted that the Mexican cinema of the 30s and 40s was strongly unified and controlled by the State, which turned the national cinema into a nation building cultural project (Pick 217). The Mexican government of that time encouraged the cinematographic industry to "participate in the economic and political transformation of the State" (Chávez 120), and, consequently, this industry started dealing with the education of the masses (Monsiváis "Cultura Popular", 118).

Juan Pablo Silva Escobar argues that this cinematic project, especially through location-specific genres – such as *comedia ranchera* – and its applications of excessive stereotypes – for both men and women – contributed greatly to the transformation of the national imaginary (23-24). Additionally, as Monsiváis notes, cinemagoers at the time experienced films as if they were happening in the real world: they would scream in anger, chant or applaud, and, on occasion, even attack the actors who played antagonist roles ("Cultura Popular" 105-106). Therefore, it is fair to assume that audiences would relate the characters' positive and negative traits to the real world and create associations that categorized certain groups under certain traits and characteristics. For example, if a movie stereotyped an indigenous community as lacking education, or an indigenous woman as a typical domestic employer, then audiences were more likely to associate the real life indigenous to a lack of education and poverty. Anecdotes of the time, in addition to analyses such as Silva Escobar's or Monsiváis', show that the stories and characters that these films created had a strong influence on how the spectators learned about and developed a relationship with their environment. Since the stories of the films of the Golden Age of Mexican cinema are mostly about Mexican people living 'Mexican lives', audiences took in an essential way of experiencing and understanding their own *Mexicanness*.

Within this essential understanding of nationality there also exist relatively fixed images of gender. Similarly to the production of *Mexicanness*, imagined archetypes of manliness or femininity were constructed through popular culture. In her study of Classical Mexican cinema, Ana López explores the different female archetypes that the genre of melodrama developed. First, she explains the trope of the 'good' mother, usually portrayed by Sara García. Second, she examines the vamp; a 'bad', haughty, and independent woman, usually portrayed by María Félix (155). For the former, López concludes that these family melodramas and their construction of an asexual, saint mother are still reinforcing the values of a patriarchal society (154). For the latter, López argues that even though this often provides the film with a strong, female character, independent and sexually emancipated, the character type is still built under a patriarchal structure that reflects the dangers of desire for men (156). As López analyzes the construction of archetypes in Mexican cinema, Claire Johnston studies 'the myths of women' in Classical Hollywood films to understand the role that the vamp and the straight girl character play within the narrative. She states that "within a sexist ideology and a male-dominated cinema, woman is presented as what she represents for man [through] myths that transmit and transform the ideology of sexism and render it invisible – when it is made visible it evaporates – and therefore natural." (32-33). Drawing on López's and Johnston's work, then, the creation of female archetypes reflects and practices gendered power relations and naturalizes discourses which construct relations of power between men – that which creates the meaning of woman – and women.

2.2 *Maria Candelaria*: The White Indigenous Woman

This study will now look at one of the representative films of Classical Mexican cinema, *Maria Candelaria*, to explore more specifically the archetypes of indigenous women and the stereotypes that are built around them. This iconic film was part of the project of *indigenismo*, and thus it idealizes the indigenous protagonists as the noble and original inhabitants of Mexico. The film provides a clear example of the relationship between gendered power structures and the construction of archetypes. It also presents a framework with which to explore how racial and class discourses in this context intersect with those of gender.

The story, told from the point of view of a famous painter, is about the life and death of Maria Candelaria (Dolores del Rio), an indigenous woman rejected by her entire community for being the child of a prostitute. In an attempt to protect Maria Candelaria, her fiancé, Lorenzo Rafael (Pedro Armendáriz), steals medicines for her but is caught and jailed. While attempting to earn money as bail for her fiancé, Maria Candelaria decides to model for the famous painter. However, when he asks her to pose naked, she refuses and leaves, after which he finishes the portrait with another woman's naked body. When Maria's community sees her nude portrait, they mob her and tragically stone her to death.

Maria Candelaria was a huge national and international success and has become one of the iconic films of the era. A newspaper article from 1944 praising it for its success at Cannes Film Festival reads it as "a moving love story, in the most beautiful Mexican landscape, with the best actress in national cinema" (qtd. in Avendaño). However, this newspaper also inadvertently displays the racial and class discourse of the 40s by stating that "in this film the soul of our poor Indians beats with all its sadness, stoicism and rare joys." (qtd. in Avendaño). By calling the characters 'our poor Indians', the article shows a sense of ownership of white over indigenous people, as well as an oppressive class and racial discourse, which the film itself reinforces through archetypes and binaries. More recent academic arguments around the film tend to be polarized: some scholars, such as Charles R. Berg, argue that it manages to represent indigenous communities through a positive lens, while others, including Jorge Ayala Blanco and Julia Tuñón, argue that all it does is erase differences and reinforce stereotypes about indigenous people (qtd. in Tierney 74). However, according to Tierney, the issue is not necessarily whether it reinforces or challenges these stereotypes, but rather that the "film's representation of the [indigenous] embodies a hybrid incoherent identity" (Tierney 75). In other words, the way the film constructs indigenous characters is full of contradictions which, as we will explore, end up supporting the racial binary of white as modernity and order, and non-white as backwardness and chaos (Tierney 95).

According to Tierney, the way the protagonists of this film are constructed, lit, and contrasted in comparison to the other characters creates a racial binary. First, it is important to note that both Del Rio and Armendáriz are white bourgeois Mexican actors playing the role of working-class indigenous people in this film. As Tierney notes, throughout the history of cinema, white actors have played universal roles, i.e., "a white actor can be 'raced' by the mise-en-scène to represent a non-white character. . . , but a non-white actor can never play a role that is not racially marked" (86). Part of the process of racializing these white actors is to give them traits and characteristics usually assigned to the 'race' portrayed. In the case of del Rio's Maria Candelaria, the actress was given very little makeup, dressed in simple peasant clothes (which, ironically, were designer made), and made to speak a colloquial Spanish with an exaggeratedly rural accent, which characterizes her as uneducated (Tierney 84, Silva Escobar 24). Del Rio's racialized image as well as her submissive and humble performance around white characters (such as the priest) further reinforces a sense of the indigenous as subordinate (Tierney 85).

On the other hand, even though both del Rio and Armendáriz are 'raced' to serve as indigenous characters, the way they are lit and their role in the film as more progressive than the rest of the community associates them with whiteness and progress. Both the priest and the painter, who are white characters, are portrayed as the 'ideal' future of a nation under construction, while the indigenous community is portrayed as backward and even barbaric for resisting modernity when they stone Maria Candelaria to death for her allegedly progressive attitudes (Tierney 84-90). Consequently, the glaring contradiction of this film is that it presents the indigenous "as both modern Mexico's central couple (Maria Candelaria and Lorenzo Rafael) and as the obstacle to its progress." (Tierney 84).

As previously explored, the films of Classical Mexican cinema were part of a project of nationhood and homogenous identity construction. Therefore, the ways in which the characters in *Maria Candelaria* are constructed and how they relate to each other can help us understand more about the role of women in the national imaginary and its relationship to Mexican national cinema. In order to explore this idea and make a comparison possible between the two contemporary Mexican films in contrast to *Maria Candelaria*, this study looks into three aspects of each film. First, it examines where in the social hierarchy the working-class or indigenous woman is placed in the context of the film. Since Maria Candelaria is characterized as not only a poor woman but also an indigenous one, it is possible to read her character from a perspective of class, race, and gender, both separately and simultaneously. Through an intersectional lens, all of these perspectives (class, race, and gender), establish particular relations of power between Maria Candelaria and her community, her fiancé, or the painter. In this power structure, it is clear that she sits at the bottom of the hierarchy as a poor, indigenous, and 'immoral' woman. Above Maria is the indigenous community, then her indigenous fiancé, and finally at the top stands the white male painter along with the priest. This hierarchy is representative for the distribution of power across Mexico's society of the 20^{th} century.

The second aspect regards how the working-class or indigenous woman's sexuality is portrayed in the film, and how it is 'gazed at'. Maria Candelaria is sexually objectified by almost all the men around her: the painter, Lorenzo Rafael, and the mestizo store owner, Don Damián. Because the camera angle and perspective reproduces the stare of the male characters, for Mulvey this would be a clear example of the way in which Classical cinema aligns the spectator's active gaze with that of the male character, positioning Maria Candelaria as the passive bearer of the look. However, since Maria Candelaria is not only gendered but also raced and classed, it is important to examine the role that the oppositional gaze plays. As previously explained, the oppositional gaze is the gaze of the underrepresented subject – in this case the indigenous women – which the subject to regain agency through the power of looking. In Maria Candelaria, most viewers are aware that they are watching a movie about an indigenous woman, however, they are not actively aware that what they see is not an indigenous woman but rather a famous white actress in a costume, who is aligned with the other white characters via the lightning and the mise-en-scène. Just like black spectators in Hollywood, indigenous female spectators in Mexico had to "develop a looking relation within a cinematic context that constructs [their] presence as absence, that denies the "body" of the black female so as to perpetuate white supremacy," (hooks 118) alongside a spectatorship that established that the desired woman is white (hooks 118). But *Maria Candelaria* does not truly offer the opportunity for indigenous female spectators to adopt an oppositional gaze. Instead, as Tierney suggests, the director Emilio Fernández places himself in the position of a colonial voyeur, thereby creating a movie that, similar to Eisenstein's *¡Qué viva México!* (1979), "inscribes Mexico within European primitivism" (78, 82), and leaves the women in the film to be gazed at by the white male spectators and characters.

The third aspect is how the working-class indigenous women are constructed in terms of cinematography and portrayed in terms of traits and stereotypes. As explained, Tierney develops the idea that the way the indigenous community in *Maria Candelaria* is depicted is contradictory as some of them – Maria Candelaria and Lorenzo Rafael – are made to look white and act in line with the progressive and noble ideals attributed to whiteness, while the other indiginous characters are made to look of darker skin and given 'barbaric' characteristics including anger and violence. The lighting in the final scene is key to understanding this binary and what it conveys to the audience. As Maria Candelaria runs away from the angry mob that wants to stone her, a dramatic medium close up of her face appears on the screen, beautifully illuminated with a bright key light. This cuts to a shot of an angry indigenous mob who is given barely any light. This makes them appear much darker, both in complexion and in temperament. Tierney notes that, in this scene, "Western notions of white's moral superiority are mobilized" (90), and therefore, as spectators, we are led to understand that the whiter an indigenous woman is, the more noble and modern she will act, while the darker she is the more likely she will act barbaric and dishonest.

While some of the stereotypes in this film, such as portraying the indigenous community as peasants, make historical sense, some others, such as the racial binary of white as modernity and non-white as backwardness, are damaging for the image of these communities in the national imaginary. Also, returning to Johnston's theory of myths, it is important to consider that this film was produced during a male-dominated period of cinema and consequently it presents women, particularly Maria Candelaria, from the perspective of what they represent for men. Stereotyping Maria Candelaria as a submissive and innocent indigenous woman not only essentializes her but also transmits and naturalizes a sexist discourse.

This emblematic film serves as an example of how the melodramas of the Golden Age, which featured female characters, were likely to represent them through myths and archetypes encouraging unequal power-relations between the male and female characters. *Maria Candelaria* therefore serves as an ideal example to compare and contrast to contemporary Mexican films in which working-class indigenous women are also protagonists, such as *Roma* and *La Camarista*. As Silva Escobar states, these stereotypes of the indigenous, and particularly the female indigenous, are part of what contributed to the construction of the national imaginary of *"mexicanidad"* (24). Therefore, it is important to question how recent movies have challenged these images and how much can they contribute to modifying the national imaginary.

3 Contemporary Mexican Cinema

3.1 Post-NAFTA Growth

As the Golden Age of Mexican cinema began its decline in the early 60s, popular filmmakers looked for ways to produce almost anything – regardless of the quality – for quick profits. Mexico's economy collapsed throughout the 70s and 80s, and consequently the government withdrew funds from the cultural and cinematic industries (Maciel 99). Simultaneously, a wave of aspiring, young filmmakers who were particularly inspired by the new European cinema of the time, such as the French New Wave, began an independent and low-budget cinema. In contrast to the Golden Age films, this new Mexican cinema did not attract large audiences until the 1990s when it slowly caught the attention of international filmgoers and investors (Maciel 100-101).

In 1994, the North American Free Trade Agreement[8] went into effect and with it came a wave of consequences, both positive and negative, for the national film industry. According to research carried out by the *Universidad Autónoma de Nuevo León* (UNAL), national film production had decreased greatly by the end of the 1990s due to the economic and political changes brought about by NAFTA, going from a total of 75 Mexican films produced in 1990 to 9 films produced in 1997 and 11 in 1998 (Hinojosa Córdova, Padrón Machorro). However, NAFTA also brought positive effects, such as foreign investment in the local industry and increased access to international audiences and markets. Therefore, even though the total number of productions in the 90s was extremely low, the few films that were produced and internationally marketed became very successful, such as *Como Agua Para Chocolate* (dir. Alfonso Arau, 1992) and *Sexo, Pudor y Lágrimas* (dir. Antonio Serrano, 1999). By the turn of the millennium, the national film industry was ready to take off, thanks to the capital brought in from abroad, and to the aspiring filmmakers that began their training in the 90s, such as Arturo Ripstein, Guillermo del Toro, Alejandro González Iñárritu, and Alfonso Cuarón, among many others who, with their new filmic proposals, began an era known as the New Mexican Cinema.

Through the 2000s, these directors gained fame in both the national and international film industries, in the latter case mainly Hollywood, and became part of a global scene of influential filmmakers. In 2018, Alfonso Cuarón returned to create his most recent project in Mexico after many years of working in Hollywood where he had directed films like *Gravity* (2013) and *Children of Men* (2006). Back in Mexico, his aim was to make a movie that could recreate his childhood memories, and particularly his memories about Lido, the housemaid who had worked for his family while he was growing up. The result was *Roma*, a film set in the Colonia Roma middle-class neighborhood of Mexico City during the 1970s, a time of widespread student protests

[8] NAFTA reduced the trade and investment barriers between Mexico, the United States, and Canada. For the film industry, this resulted in the fall of national film production, monopolization of distributions and exhibitions, and the decrease in attendance and box office (Hinojosa Córdova, Padrón Machorro).

and political violence. The protagonist of the story, Cleo (Yalitza Aparicio), works as a live-in maid in Antonio (Fernando Grediaga) and Sofia's (Marina de Tavira) household, helping them take care of their four children, cooking, and cleaning the house. As the story moves forward and complications arise, Antonio leaves the family and Cleo finds out she is pregnant with the child of Fermin (Jorge Antonio Guerrero), a member of a paramilitary group known as *Los Halcones*.

Soon after the release of *Roma*, director Lila Avilés presented her most recent film *La Camarista*. This movie tells the story of Eve (Gabriela Cartol), a young chambermaid who works in a luxurious hotel in Mexico City. She is a meticulous cleaner, but to everyone else around her, she is considered a lowly maid. Regardless, she enrolls in the hotel's education program for adults and seeks to be promoted from cleaning the 21^{st} floor to the 42^{nd} floor – a meaningful rise through the ranks. This chapter will analyze *Roma* and *La Camarista* as two case studies of contemporary Mexican films that, by addressing the representation of indigenous working-class women, attempt to reshape the image of these women in the Mexican national imaginary.

Throughout cinema's history, both women and men have been portrayed through stereotypes or fixed iconographies so that audiences can easily identify their roles. However, Johnston explains that, due to sexist ideology, the stereotyping of men (what she refers to as myths) "underwent rapid differentiations while the primitive stereotyping of women remained with some modifications" (32). Thus, according to Johnston, the myths through which women are portrayed also transmit and transform the ideology of sexism (32), while the myths portraying men do not limit them, as they change to accommodate changes in reality. In the case of this study, the myths through which working-class women are portrayed expose and explore a location-specific, sexist ideology of Mexico. Using Johnston's concept of myths along with López's study of archetypes, one can compare the construction of indigenous characters (Cleo and Eve to Maria Candelaria) and historicize their origins. *Roma* and *La Camarista* expose how white and male characters in the film treat Cleo and Eve as if they were living manifestations of the archetypes created by films such as *Maria Candelaria*: that of the subordinate, indigenous or lower-class woman who lays at the very bottom of the social hierarchy, and whose sexuality (in some cases) plays an essential role in determining her path.

As previously discussed, Maria Candelaria's racial binary construction as well as her crude relationship with her community suggest that indigenous communities should continue to be segregated unless they are willing to modernize and whiten their lifestyles. As Tierney puts it, since contact between both groups, indigenous and white, brought the death of Maria Candelaria, "rather than the incorporation of the *indígena* within the modernizing state, isolation is the only means to protect indigenous people." (83). Building on this idea, this study will now turn to the development of Cleo and Eve to evaluate how these newer films allow marginalized women to claim a space of visibility within the national imaginary, and thus within the reality of Mexico. To do so, it will follow a similar structure to the analysis of *Maria Candelaria* by exploring three main aspects of the film on both a narrative and cinematic level: firstly, the position of the working-class or indigenous woman in the social hierarchy; secondly, the portrayal of their sexuality, and thirdly; the cinematographic construction and character traits that contribute to their contemporary image in the national imaginary. These three points are relevant not only because they are present in almost all Mexican movies that create an archetype of a working-class woman, but also because each of the selected case studies present a different level of engagement with this archetype.

3.2 *Roma*: A Look into the Daily Life of an Indigenous Maid

Roma is a contemporary, realistic recreation of the 1970s in Mexico City through a realistic lens. Through historical events and cultural references, the film constantly reminds the viewer that it is a recounting of the past. The historical character of *Roma* recognizes and comments on the race, class, and gender discourses of the 70s, but from a critical distance. As this analysis will argue, in order to create this distance, much of the camerawork throughout the film avoids engaging with the drama of the narrative and instead presents dramatic situations through a distant and impartial lens.

The first aspect of *Roma* to explore is the position in which the film places Cleo, as an indigenous woman, in the social hierarchy. As we meet all the characters in Cleo's life, we realize that there is a

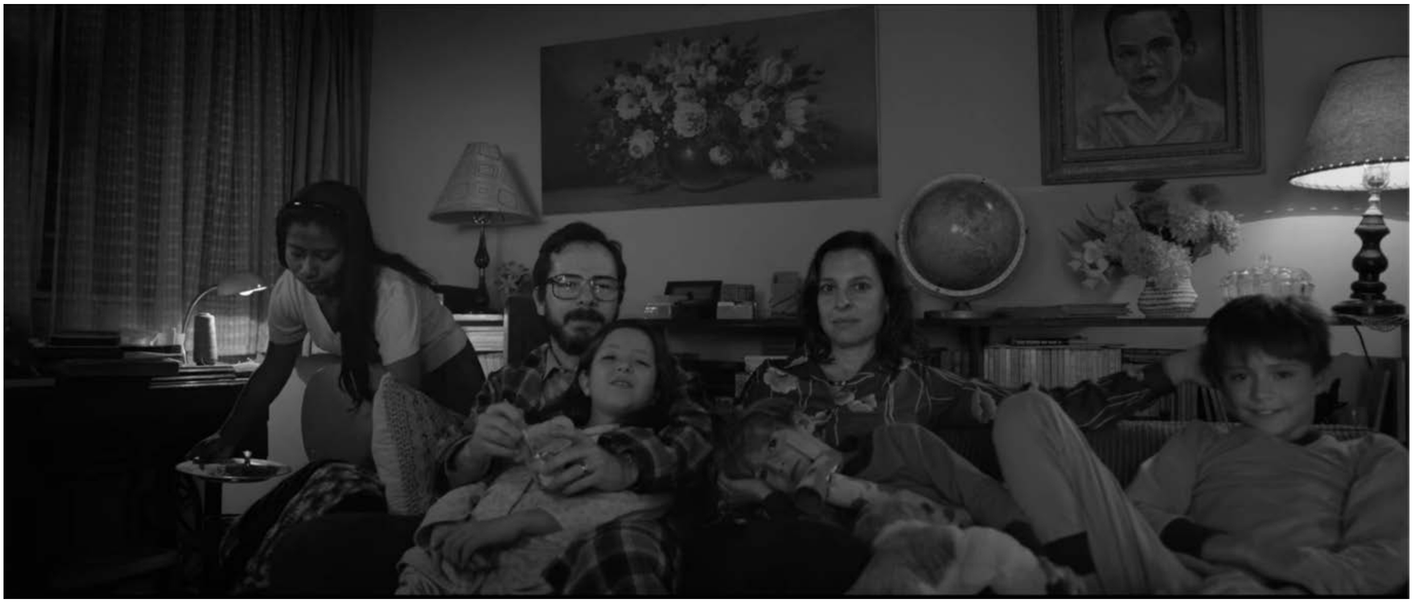


Figure 1: *The position of the characters and the lighting reflect the class and race hierarchy within the film.*

racial as well as a gender and class binary which establishes the different power relations in the film. One of these binaries is the white, middle-class family in opposition to the indigenous, working-class maids (Cleo and Adela, played by Nancy García). The mise-en-scène is telling in this relationship, particularly the lighting of the scene where the family sits to watch a comedy show on TV while Cleo serves them dessert and sits next to them. First, Cleo gives some food to Antonio, the patriarch of the family, and picks up some dirty dishes. This is shot from the perspective of the television that illuminates the faces of the happy family who sit in the foreground of the frame, while the background, where Cleo is walking with the dirty dishes in her hands, is thrown into shadow (see Figure 1). As she walks behind the couch, the camera pans, following her movement. When she reaches the other side of the couch, she sits on the floor next to one of the children who offers her a welcoming hug in the foreground. There, the camera looks at Cleo from the shadowy background where she was standing before, giving a sense that it awaits her return. Soon afterwards, the mother asks Cleo to go and prepare some tea for the father, and as Cleo stands, the camera tracks her walk back into the shadow. The camerawork of this scene, together with the automatism with which Cleo is given orders immediately after a warm moment that invited her to feel like part of the family, exposes the contradictions Cleo experiences as a domestic worker, and the racial, class, and gender hierarchies that stem from the essentialist discourses operating in the Mexican national imaginary. While the mise-en-scène stages the social and racial divide, the camera always follows Cleo as the protagonist. Therefore, the film exposes Cleo's invisibility within its narrative, but makes her visible by constantly bringing her into the foreground of the story.

Cleo also confronts gendered power relations throughout the film, particularly with regard to Fermin, her boyfriend. To explore this relationship, it is important to examine the way in which Cleo's sexuality is portrayed in the film, and how it is gazed by the other characters. After Cleo and Fermin meet and go on a few dates, they go to a hotel room where they have sex. This unconventional sex scene begins with Fermin standing alone and naked in the bathroom, holding a curtain rod. As he walks out of the bathroom, he begins an awkward martial arts performance which Cleo observes, amused, while lying in bed and covering her underwear with the bedsheets. Most of this scene is shot from Cleo's perspective on the bed. Fermin is shot straight on, center punched, and from a medium shot that shows that Cleo is slightly far away from him. The unconventional full frontal shot of a naked male body, as well as the camera positioned from the female perspective, distances the audience from Fermín and guides them to gaze at this man through Cleo's eyes. With shot-reverse-shots of Cleo's giggly and awkward reactions, the spectator and Cleo share a sense of ridicule at Fermin's show of masculinity, as he is trying to display his strength and manly nude body through a strange show, almost as if he was performing an animal courtship ritual. It is also important to note that the "looks" in this sex scene are reversed from a conventional sex scene in two ways: the audience is seeing a nude male body and a (partially) dressed female body, and the audience is invited to gaze at Fermin's body along with Cleo. This rare scene offers a moment for audiences to take the place of the oppositional gaze, almost as if they were sitting next to Cleo and gazing from her perspective; the audience is placed as an indigenous woman watching the hyper-masculine performance of a working-class man.

For hooks, Cleo's oppositional gaze would represent a site of resistance from Fermin's gaze and from the audience's male gaze. Since the spectator is not given an opportunity to gaze at Cleo during a sexual scene, she can use this space to reclaim her agency. In terms of a gendered power-relation, however, this scene, as well as their upcoming encounters, shows clear instances of Fermin trying to impose his dominance over Cleo, sometimes even through threats and violence. However, these scenes where Fermin wants to dominate Cleo also create a sense of disassociation in the spectators by distancing the camerawork from the dramatic action. For example, while they are kissing in the cinema, Fermin finds out that Cleo is pregnant and literally disappears after excusing himself to go to the bathroom. When Cleo goes seeking for Fermin, she finds him in the countryside practicing martial arts. There, his dominance becomes more violent as he insults her and threatens to harm her if she does not leave him alone. However, Fermin's aggression towards Cleo is innovatively shot to keep the spectator at a distance. As they walk in parallel but somewhat far apart from each other (Cleo is clearly trying to keep up with Fermin's pace and his sight), the camera tracks them from a disengaged distance. When Fermín insults and threatens Cleo, there is no reverse shot of Cleo's reaction; we see only her back. Even though there is a heightened tension in the dialogue and in Fermin's body language, the camera stays as disengaged and steady as possible, reminding us that we are distant spectators in a different historical moment. Yet, we are no less affected by what we see on screen.

Finally, during their last meeting, Fermin takes this dominant and violent power-relation even further. As Cleo shops for her baby's crib, the Corpus Christi Massacre[9] of students begins right beneath the shop. Suddenly, Fermin, along with his paramilitary group known as *Los Halcones*, enter the store looking for hidden students, and Fermin points his gun straight at Cleo who silently stares back. As he runs away back to the protests, Cleo's water breaks, causing an early miscarriage. From the moment the armed men enter the shop until Fermin is about to leave, there are no cuts. Instead, there is a slow wide shot pan that follows the students into the shop. As one of the students gets shot, a gun is introduced into the frame as an out-of-focus close up. Steadily, the camera pans to show us that it is Fermin holding a gun and most likely pointing it at Cleo. The lack of fast editing and the slow pan across this very shocking scene again keeps the spectators distanced. The three scenes between Cleo and Fermin are moments full of tension, however the camerawork does not increase the drama but rather creates a critical distance through which the spectators are able to watch from their 21^{st} century perspective. The spectators are invited to understand the lives of indigenous housemaids in 1970s Mexico City and to put Cleo and Fermin's relationship in context, through a historic lens of gendered, racial, and class relations.

The historical character and the distant camerawork throughout *Roma* play an essential role in recognizing the discourses operating in the 1970s. In a slow but climatic scene near the end of the film, Cleo becomes the hero of the story by saving the children from drowning in the ocean while not knowing how to swim herself. The scene finishes with the family and Cleo hugging and crying in relief while sharing how much they love each other. Yet, as soon as they are all back in the city, Cleo immediately goes back to doing house chores and preparing a smoothie for the children. Through these scenes, it is revealed on a narrative level,

[9] The Corpus Christi Massacre, also known as *Halconazo* occurred on June 10^{th}, 1971, the same day that the Catholic Church celebrates the Feast of Corpus Christi. That day, a large group of student demonstrators gathered to protest for better management of education funds and for the end of government repression, among other things. However, as they were marching, a paramilitary group known as *Los Halcones* broke into the protest and triggered one of the most brutal episodes of repression in Mexican history, murdering at least 120 students and injuring hundreds more (Cruz Cárdenas and Mendoza).

that nothing has really changed for Cleo in her social or economic relationship to the family. However, at the level of cinematic form, the drowning scene also shows how the film creates a critical distance for the spectator as opposed to full immersion. As Cleo sits with Pepe (Marco Graf), the youngest of the children, at the back of the beach, she notices that the other two, Paco (Carlos Peralta) and Sofía (Daniela Demesa), are getting too far into the ocean. Scared, she moves quickly towards the shore and, without thinking, runs straight into the water to rescue the two children. The entire scene is filmed in one long shot with a camera that moves sideways tracking Cleo's advancement into deeper waters. From a distance, we see Cleo struggling against the current while the camera itself is surrounded by tall waves and fighting to follow Cleo's movement. Even though the tension is not built through conventional point-of-view and reverse shots, the audience still finds itself swimming against the same heavy current as Cleo and the children are completely engulfed by the intense surround sound of the crashing waves. But while this is one of the most intense moments of the film, we never see them from a close up or from any of their perspectives. Instead, the camera keeps its distance, mirroring Cleo's and the children's struggle in the water without showing us their emotional response until they are out of the water. Then, back in the city and as Cleo goes back to her daily chores, the distance constructed throughout the previous scene allows viewers to recognize Cleo's reality from a critical but still emotional perspective.

To further recognize Cleo's reality within the discourses of the 1970s, it is important to observe how her character is constructed and how this construction contributes to her place as an indigenous and working-class woman in the national imaginary. A key moment raising this question appears in the middle of the film – during the Christmas Holidays of 1970, after Cleo and the family arrive at a countryside ranch where they intend to spend the holidays with some wealthy friends. As we learn, Cleo was already friends with Benita (Clementina Guadarrama), one of the housemaids from this ranch. During New Year's Eve, while Cleo is taking care of some children at the bourgeois party of the house owners, Benita invites her to join her to the 'other' party. Through an establishing shot of the house and the hallways, we see both maids walking from the high-class party on the top floor, to the working-class party in the underground service kitchen. The steep stairs that Cleo and Benita walk down are illuminated primarily at the top, increasing in shadow as they descend, illustrating a binary relationship in which the bright light as well as the white high-classes are on the top floor (as well as at the top of the hierarchy), while the dark and shadowed areas as well as the non-white and indigenous people are on the underground floor (as well as at the bottom of the social hierarchy). As the two friends chat, Benita uses every opportunity she has to point out that Cleo is becoming a city person – too posh for a countryside farmer.

The binary opposition between the scenes in which Cleo is embraced and simultaneously rejected by the family in the city, and the scenes in which she is embraced and rejected by her countryside community convey the sense that Cleo does not really belong in either place. The construction of Cleo's character through these binaries seems to mirror the loss of identity that indigenous people go through when migrating from the countryside to the cities. However, it might also suggest that this loss of identity is transformative. For this, it is important to also examine the production of *Roma* and the choice of hiring non-actress Yalitza Aparicio, an indigenous woman who moved from a small rural community in Oaxaca, Mexico, straight into Mexico City to participate in a huge filmic project. The binary opposition that Cleo experiences in the film is also experienced by the actress herself who was largely rejected and discriminated against by commercial high-class actors in Mexico City. After the success of the film, Aparicio adopted a look that combined Hollywood glamour with traditional indigenous wear, transmitting the idea that she had to go through the same transition as Cleo, from countryside to urban, but successfully constructing a new blend of traits, traditions, and looks. Therefore, when questioning how Cleo's character is constructed and how this contributes to her place in the national imaginary, it is also essential to consider Aparicio's role as an indigenous non-professional actress who went through a similar experience as Cleo in her adaptation from the countryside to the urban setting.

In contrast with the example of Classical cinema, *Roma* breaks with *Maria Candelaria's* proposition that the only way to 'protect' the idealized *indígena* is by isolating them from the moderniz-

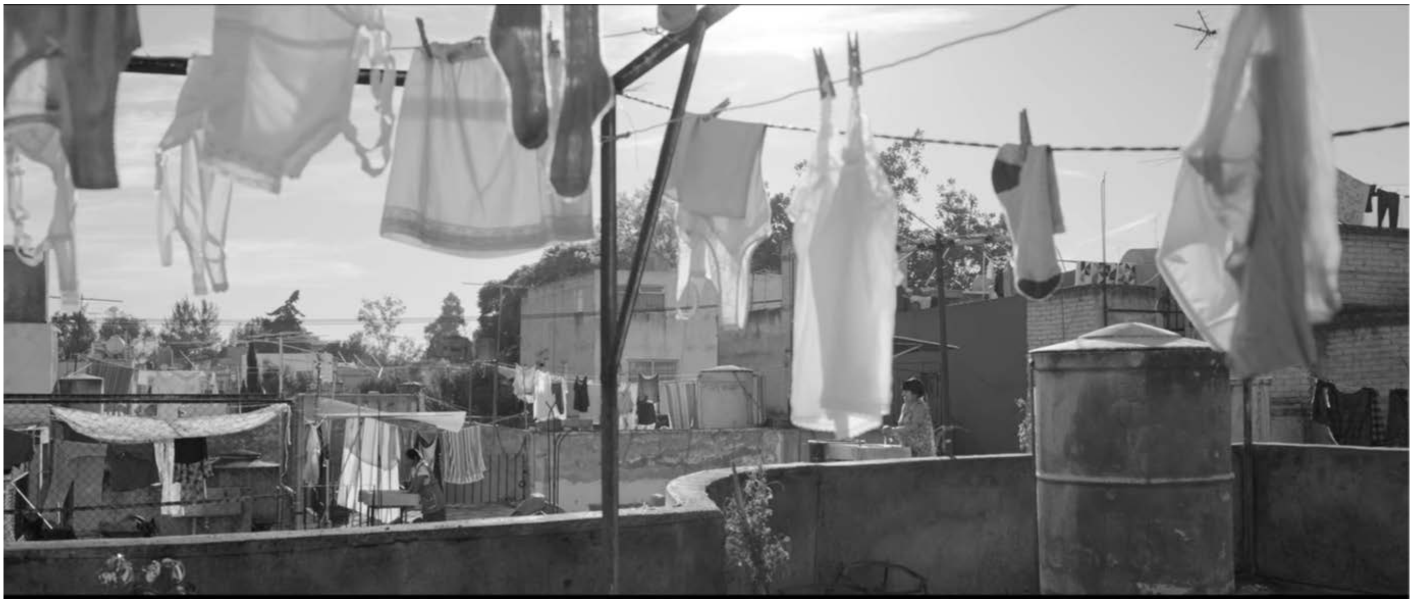


Figure 2: *Wide shot shows several rooftops with many housemaids doing laundry. It speaks to the idea that there are hundreds of other women living the same life as Cleo.*

ing state. Instead, *Roma* makes the working-class indigenous woman visible in the image of the nation, and successfully gives her space for expression as an autonomous person. *Roma* attends to the complexity of Cleo's living situation as an indigenous housemaid economically dependent upon and subjugated to her white employers, and as a young woman dominated by a working-class man. At the same time, *Roma* subtly remarks that Cleo and her story is just one of the innumerable untold and unseen individuals and stories. In an early scene where Cleo is washing clothes on the rooftop with Pepe and Paco playing around her, the camera follows Pepe with a slow pan revealing in the background many more rooftops full of maids who, just as Cleo, are washing clothes by hand and hanging them (see Figure 2). As the scene comes to an end after Cleo lays down to comfort Pepe, the camera slowly pans again to show the activity on the surrounding rooftops. The scene closes by implying that there are hundreds of housemaids like Cleo in Mexico City that are also made invisible by daily classist discourses, and whose struggles and realities need to be seen and heard.

3.3 *La Camarista*: The Life of a Chambermaid in Mexico City

As previously explored, class, race, and gender are so interlinked that belonging to a certain class might directly cause someone to be racialized by others, and vice-versa. In the US and the UK, cultural theorist Stuart Hall explores how black female bodies have often been stereotyped to fall under the category of "mammies", the "prototypical house servants, usually fat, bossy and cantankerous." (251). In other words, by virtue of their black skin, black women are stereotyped and classed within the national imaginary. Something similar happens in the Mexican context because, when working-class women of color are domestic workers, they are racialized and thought of as indigenous. In the case of *La Camarista*, there are no obvious signs that can lead us to assume that Eve is an indigenous woman. However, due to her class, job, and gender, she is associated with the indigenous working-class. Therefore, even though she is technically a mestizo woman, she can be considered as indigenous in the racialized national imaginary. For this reason, the following analysis will examine Eve through this racialized lens when approaching the topic of movement towards visibility for working-class indigenous women.

When released, *La Camarista* resonated with many viewers and film reviewers as a film worth comparing to *Roma*, given that both have similar settings and subject matters (García). *La Camarista* tells the story of Eve, a working-class woman who works as a chambermaid at one of the most luxurious hotels in Mexico City. We follow Eve

– through a motif of unusually close tracking shots – constantly moving from top to bottom of the building: going from the most luxurious bedrooms all the way to the service and laundry rooms in the basement. Eve's main aspirations are to get a red dress that sits abandoned in the lost and found, and to be assigned as the main cleaner of the 42^{nd} floor (the most exclusive of the hotel). The entire movie takes place inside the hotel and its space therefore functions as a microworld. Here, Eve works and cleans, makes friends, showers, sleeps, engages with her sexuality, and lives out her motherhood within these walls, through the absent-presence of Ruben, her four-year-old son, with whom she occasionally gets to speak over telephone calls. This microworld portrays the social, racial, gendered, and economic relations between higher and lower classes (the guests and the workers), and thus it works as a reflection of the same relationships seen in Mexico City as a whole.

To explore where Eve stands in terms of class, race, and gender in relation to the other characters in the film, one must first examine how the social hierarchy is constructed and where Eve is placed within this hierarchy. The verticality of the hotel serves as an analogy for the social hierarchies present in Mexico. The top section of the hotel, where the bedrooms are, corresponds to the wealthy guests; then the middle section, where the lobby and restaurant area is, corresponds to the hotel managers and cooks; and finally, the bottom section, where the service, laundry, and tiny bunk rooms are, corresponds to the cleaners and chambermaids of the hotel. The top area of the hotel is where Eve works, and is responsible for cleaning the 21^{st} floor. This area also holds Eve's dream job: being the main cleaner of the palatial 42^{nd} floor where dignitaries are lodged. Here, Eve interacts with the privileged white guests of the hotel while being ignored by them: from an Orthodox Jew who motions for Eve to push the elevator buttons for him, to a VIP guest who constantly requests ridiculous amounts of toiletry, to an Argentinian woman who gives Eve some money to watch her baby while she takes a shower every morning. All these characters represent the wealthy and mostly white part of (Mexican) society.

Many of these encounters are filmed by constructing a clear vertical line in the middle of the screen that visually separates Eve from the guests. For example, when Eve approaches the Jewish man, she stands on the right side of the screen separated from him by a vertical line created by the elevator. Since we only see the man's back and shoulder, our focus is fully on Eve. Similarly, when Eve goes to the VIP guest's room to bring toiletries, the camera is in the bathroom with Eve in the foreground and the guest sitting on his bed in the background. While they are in separate rooms, the wall emphasizes their distance by creating a clear vertical line down the middle of the screen which separates them even further.

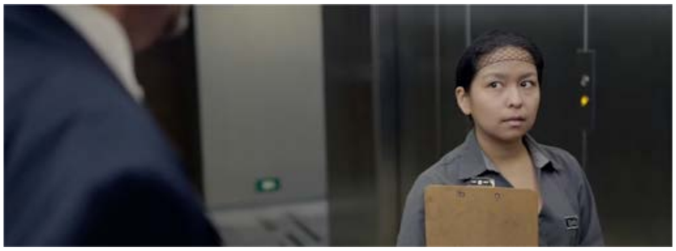


Figure 3: *Eve looking at the Jewish man. A vertical line in the middle of the screen visually separates them*

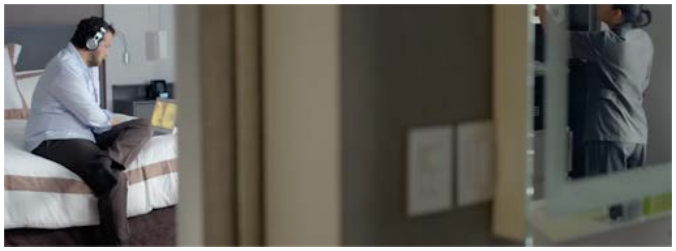


Figure 4: *Eve bringing toiletry of the VIP's guest room. Again, they are visually separated by this vertical line in the middle of the screen created by the wall*

The middle section of the hotel, where the hotel managers, secretaries, and restaurateurs work, represents the middle class of society. The employees of the hotel also maintain a clear hierarchy amongst themselves. Early on in the movie, we see a male chef getting into the service elevator while enjoying some snacks. In this shot, the chef occupies almost all the screen space, barely letting us see that there is an elevator operator next to him (see Figure 5). This brief scene seems insignificant, since we never see this character again. However, a few minutes later in the film, Eve gets into the same elevator discretely eating popcorn, and immediately gets scolded by the female elevator operator who reminds her it is not allowed to eat there. Unlike the previous scene, the elevator operator is the main element, occupying most of the screen space and leaving Eve cropped and out of focus. Clearly, the operator makes a gendered and class distinction between who is allowed to break the rules and who is not. A male chef who has a higher-paid job than a female cleaner can be entitled to more transgressions without punishments within the hierarchy of the hotel workers. The camera position and focus during both scenes visually emphasizes this hierarchy.

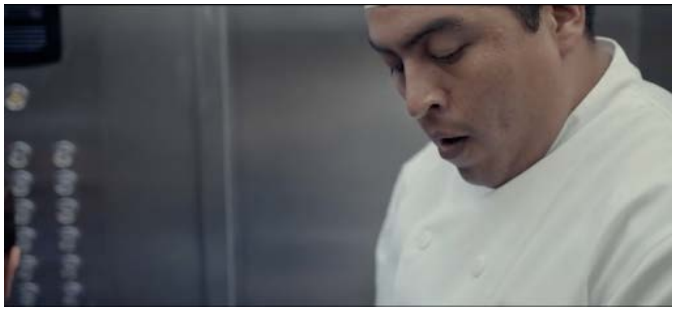


Figure 5: *The composition of the shot – centering the chef and cutting out the elevator operator – reflects hierarchical constructs within the hotel.*

Finally, we find Eve and her coworker chambermaids in the basement of the hotel and in a small classroom where Eve and three other cleaners take elementary school lessons in the early mornings. Eve belongs to this section of the hotel, as well as to the bottom section of society in terms of class, race, and gender. However, as this analysis will later explore, Eve challenges this hierarchy not only by freely exploring all levels of the hotel (even areas restricted only to guests), but also by leaving the confines of the hotel to go onto the rooftop as a brief attempt to escape the hierarchical social and economic system.

Although Eve's gender doesn't necessarily stand out as a factor of oppression, it definitely influences how others around the hotel treat her, the problems she encounters, and how she decides to explore her sexuality within this space. One example of Eve's exploration of her sexuality is in her relationship with the window cleaner with whom Eve often exchanges glances, and who draws hearts on the outer windows of the rooms she is cleaning. One day, after Eve's friend Minitoy challenges her to be more courageous, Eve gives the window cleaner a note with an invitation to meet in a window of the 21^{st} floor where she'll be cleaning. This small step, which in a patriarchal society could be read as a step into female emancipation, leads to an unconventional sex scene. Throughout the scene – a long take with no cuts lasting about three minutes and filmed with a wide angle lens – the camera does not move from its original spot. In the foreground, we see Eve occupying the middle and right side of the screen, and in the background left corner, we see the window cleaner who is at the same time staring at Eve from behind the glass. As Eve notices him, she starts playing around with him, raising and lowering the curtains in a teasing manner. Then, as he keeps on staring, she goes to the edge of bed where she begins to undress. As she sits, she puts lotion on her legs, looks up and slowly and timidly takes off her bra. She lays on her back and as soon as she brings her hand inside her underwear the scene ends.

By being in a fixed and slightly lowered position, the camera suggests that it is a hidden or surveillance-like camera of which the characters are completely unaware. The camera somewhat reflects the setup of the audience in the theater, and thus puts the spectators in a strange and discomforting position in which they are made to question whether they should be watching this private moment through their voyeuristic position. This distanced camera, as well as the discomfort it provokes, also pushes the spectators to distance themselves from the scene and to choose how to gaze at this private moment: either by identifying with the male window cleaner or with Eve's gaze towards the cleaner. However, there is also a distance that does not fully allow for identification with the window cleaner. The man is placed at the very back of the screen in the left corner and slightly out of focus, swinging in and out of the frame and separated from the action by a thick window. The positioning of the window cleaner disassociates the spectators from him and his gaze.

The unconventional camerawork, the distance between the viewers and the male character, the placement of Eve in the middle of the screen, and her fixed stare on the man guide the viewers to identify with Eve's gaze. Since Eve is the oppressed female subject in terms of class, race, and gender, her gaze is another example of hook's 'oppositional gaze'. If viewers choose to follow and identify with her gaze then the scene experience switches from that of a spying camera to a moment of female empowerment and exploration of sexuality. Similarly to *Roma*, and in opposition to *Maria Candelaria*, the protagonist takes control of her sexuality. Eve never actually speaks to the window cleaner, yet she uses this encounter as an opportunity to explore her sexuality without wondering how others may judge her. In terms of sexuality and gender, it seems that Eve is going through a process of emancipation.

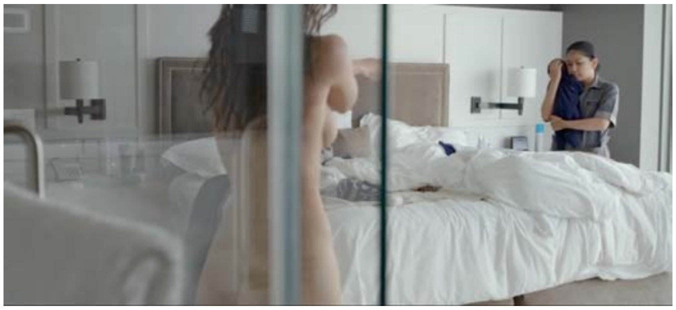


Figure 6: *The composition of the shot - splitter in half by the glass - reflects the social division between the hotel workers and the guests.*

The film also uses the subject of gender to expose the different levels of inequality that women can experience depending on their class and race. To explore this issue, the film builds a relationship between Eve and one of the guests: the Argentinian woman, Romina, who is always in the hotel taking care of her baby while her husband goes to work. The first time they meet, she asks Eve to watch her baby every morning while she takes a quick shower. Eve tries to explain that she is very busy and does not have time to help, but Romina ignores her while thanking her and getting into the shower. Even though Romina does not dismiss Eve as the rest of the guests do – probably because she does not act according to Mexican class structure due to her foreignness – she still makes Eve invisible by often interrupting her. Throughout Romina's shower we see Eve in a tight close-up from which she cannot escape. At the same time however, Eve seems to relax as soon as she holds the baby, who reminds her of her son. When Romina gets out of the shower, the shower door clearly splits the screen in two: on the left side of the screen she stands naked and complaining about not being able to work because of the baby, while on the right side of the screen stands Eve holding the baby as if nostalgic from being separated from her own baby due to her work. Even though both women bond through their motherhood, the visual construction of the scene shows a clear separation between their worlds. Their relationship and the scene construction illustrate the clear binary between rich, white woman and poor, indigenous woman. At the same time, it shows the necessity for intersectionality when trying to understand their struggles as women who are marked distinctly via race and class.

To further explore where Eve is being placed within the national imaginary, it is also important to examine how she is portrayed (characteristics and stereotypes) and constructed (cinematically). Throughout the film, we see Eve as a timid and sometimes fearful character who is overlooked and dominated by almost everyone around her. However, Eve is also a character that challenges many typical assumptions about someone in her position. She studies and educates herself, she is a single mother, she explores her sexuality, and she even goes through the guests' things with a curious spirit (of course knowing that she could get in trouble). Even though Eve often acts submissive in front of her managers and the guests for fear of losing her job, as soon as she is alone, she begins challenging all those classed, racialized, and gendered impositions that oppress her.

Overall, the style of the movie could be described as very claustrophobic. The use of medium close-ups and the lack of establishing shots makes the spectator feel lost and trapped in the space, mirroring how Eve feels inside this microworld. Her initial dreams collapse when she discovers that her friend Minitoy was promoted instead of her to the 42^{nd} floor and that, consequently, she was given the red dress out of pity. As her frustrations and disappointments build up, the scenes get more and more stifling, increasingly making use of tighter close-ups and shaky shots. This sense of entrapment is relieved when Eve heads to the rooftop to take a breath, and the entire space suddenly opens up for her and for the viewers. Through a panoramic wide shot that tilts from Eve up to the open sky, and back down again to Eve, the claustrophobia is broken. However, this is not a fully satisfying sense of relief since we know that this break is only temporary and that she has to go back inside to the microworld that entraps her.

The last two scenes, right after she's back inside from the rooftop, are again entirely shot with tight close-ups and shaky camera, provoking a rough hit back into this claustrophobic world. The viewers are bound to identify entirely with Eve since the camera blurs and cuts off the bodies and faces of everyone else around her. As she makes her way out of the hotel in extreme close up shots, Eve takes another challenging step and leaves the building through the main elevator, rather than the utility elevator, and the main entrance rather than through the back door. This is the first time in the film that Eve walks through the main lobby, since usually only the guests and managers are expected there. Through this action, she challenges the social hierarchy by reminding everyone around her that she exists and that she is no longer willing to accept the position of inferiority that has been imposed upon her. Similarly, by closely following Eve's walk through the main lobby, keeping her at the center of the image and blurring the guests around her, the film is reminding the viewers to recognize all the other cleaners and workers that are often overlooked. The film's final gesture Eve leaving the hotel is symbolic of her joining a society that has ignored her for too long.

On first impression, it could seem that Eve is not so different from Maria Candelaria. She acts submissive in front of others and she is a hardworking and humble maid. However, the way Eve's character is cinematically framed and narratively constructed shows that she is aware of how invisible she is for her surroundings, and that she is willing and able to speak up and to look for recognition. Even though it might seem insignificant for her to explore the 42^{nd} floor, to be in 'only guests' sections, or to walk through the main lobby, these are all signs of Eve's struggle to be seen by those around her.

4 Conclusion

The inherent contradiction at the core of the Mexican national imaginary is that indigenous people are seen as the origin of the Mexican nation while at the same time being society's most marginalized groups. While the *indigenismo* movement tried to assimilate indigenous people into the national imaginary, it failed by erasing them as autonomous people into a politics of whitening, essentializing their identities and creating damaging stereotypes that are still pervasive in mainstream media today. This is shown through an analysis of the film *Maria Candelaria*, that illustrates the portrayal of indigenous women in Mexican film driven by a romanticizing and essentializing national agenda. The broken relationship between white, mestizo Mexicans, and indigenous Mexicans becomes evident in the urban setting, where the latter becomes economically dependent on the former, exacerbating the power imbalance. This dependence creates raced, classed, and gendered power-relations that are often reflected in the national arts and media. Some recent films, such as *Roma* and *La Camarista* have exposed these relations and opened a space for indigenous people to reclaim a space in the national imaginary.

Through a close scene analysis of these two contemporary films, this thesis has shown that there is a growing interest in the stories of working-class indigenous women, one of the most invisible and underrepresented groups in society. Moreover, it argues that the introduction of intersectionality in contemporary discourses allows films to explore new ways of telling the stories of these women and of exposing their struggles without adhering to stereotypes, essentializing their identities, or undermining the possibility for solidarity. By exploring where in the social hierarchy of the narrative these women are placed, the analysis shows that both contemporary films acknowledge the structures of inequalities underlying Mexican society. Then, by inspecting how their sexuality is portrayed and gazed, it demonstrates that both films build characters in the process of recovering their gaze and thereby attempting to emancipate themselves from the patriarchal discourse of female sexuality.

Seeing working-class indigenous women on the screen can be empowering for these marginalized groups, but it can also be a reminder for those in power that these people exist in the same space. On a more tangible level, the recognition of these groups and their fair presence in the national imaginary can be a first step into seeing real life improvements in the living conditions of working-class indigenous women. These attempts to portray varied and complex representations of marginalized women are a tangible movement towards visibility.

5 References

5.1 Works Cited

Anderson, Benedict. *Imagined Communities: Reflections on the Origin and Spread of Na-*

tionalism. London, Verso, 1991.

Avendaño, Reyna. "Dolores del Río y la Trágica Historia de "Maria Candelaria" en Xochimilco." *El Universal*, 14 Dec. 2017, www.eluniversal.com.mx/espectaculos/cine/maria-candelaria-y-la-tragicahistoria-de-dolores-del-rio-en-xochimilco. Accessed 15 May 2020.

Buchanan, Ian. *A Dictionary of Critical Theory*. 1^{st} edition, Oxford University Press, 2010.

Chávez, Daniel. "THE EAGLE AND THE SERPENT ON THE SCREEN: The State as Spectacle in Mexican Cinema." *Latin American Research Review*, vol. 45, no. 3, 2010, pp. 115–41.

Ching, Erik, et al. *Reframing Latin America: A Cultural Theory Reading of the Nineteenth and Twentieth Centuries*. University of Texas Press, Austin, 2007.

Coleman, Arica L. "What's Intersectionality? Let These Scholars Explain the Theory and Its History." *Time*, 29 March 2019, www.time.com/5560575/intersectionality-theory/. Accessed 5 May 2020.

Cruz Cárdenas, Farrah de la, and Damián Mendoza. "Sucedió un Jueves de Corpus Christi." *UNAM Global*, 20 June 2019, www.unamglobal.unam.mx/?p=67251. Accessed 30 May 2020.

Garcia, Lawrence. "The Big Screen: La Camarista." *Film Comment*, May-June 2019, www.filmcomment.com/article/the-big-screen-the-chambermaid/. Accessed 20 May 2020.

Greeley, Robin Adèle. "Witnessing Revolution, Forging a Nation." *Paint the Revolution Mexican Modernism* 1910-1950, edited by Matthew Affron, Philadelphia Museum of Art, 2017, pp. 263-69.

Hall, Stuart. "The Spectacle of the Other." *Representation: Cultural Representations and Signifying Practices*, edited by Stuart Hall, Jessica Evans and Sean Nixon, SAGE Publications Ltd., 1997, pp. 22-79.

Hinojosa Córdova, Lucila and José Antonio Padrón Machorro. "El Cine Mexicano y el TLCAN." *CIENCIA UANL*, vol. 21, no. 89, 2018, doi.org/10.29105/cienciauanl21.89-2, Accessed 27 April 2020.

hooks, bell. "The Oppositional Gaze." *Black Looks: Race and Representation*, Boston South End Press, 1992, pp. 115-31.

Johnston, Claire. "Women's Cinema as Counter-Cinema." *Feminist Film Theory: A Reader*, edited by Sue Thornham, Edinburgh University Press, 1999, pp. 31-40.

Kidd, Mary Anna. "Archetypes, Stereotypes and Media Representation in a Multi-cultural Society." *Procedia – Social and Behavioral Sciences*, vol. 236, 2016, pp. 25-28.

López, Ana M. "Tears and Desire. Women and Melodrama in the 'Old' Mexican Cinema." *Mediating Two Worlds*, edited by John King, Ana M. López, Manuel Alvarado, London British Film Institute, 1993, pp. 147-63.

Maciel, David R. "Serpientes y Escaleras: The Contemporary Cinema of Mexico, 1976 1994." *New Latin American Cinema*, edited by Michael T. Martin, Wayne State University Press, 1997, pp. 94-120.

The Editors of Encyclopædia Britannica. "Mestizo." *Encyclopædia Britannica*, 12 Sep. 2019, www.britannica.com/topic/mestizo, Accessed 2 June 2020.

Monsiváis, Carlos. "De la construcción de la 'sensibilidad feminina'." *Misógino Feminista*, edited by Marta Lamas, Debate Feminista: Oceano, 2013, pp. 81-92.

"El Cine Mexicano." *Bulletin of Latin American Research*, vol. 25, no. 4, 2006, pp. 512-16.

"Notas sobre cultura popular en México" *Latin American Perspectives*, vol. 5, no. 1, 1978 pp. 98-118.

"Soñadora, coqueta y ardiente. Notas sobre sexismo en la literatura mexicana." *Misógino Feminista*, edited by Marta Lamas, Debate Feminista: Oceano, 2013, pp. 21-43.

Mraz, John. "How Real is Reel? Fernando de Fuentes's Revolutionary Trilogy." *Framing Latin American Cinema: Contemporary Critical Perspectives*, edited by Ann Marie Stock, University of Minnesota Press, 1997, pp. 92-118.

Mulvey, Laura. "Visual Pleasure and Narrative Cinema." *Film Theory and Criticism: Introductory Readings*, edited by Leo Braudy and Marshall Cohen, New York, Oxford University Press, 1999, pp. 833-44.

Paz, Octavio. *The Labyrinthe of Solitude: Life and Thought in Mexico*. Translated by Lysander Kemp, Middlesex, Penguin, 1985.

Pick, Zuzana. "Cine y Archivo: Algunas Reflexiones

Sobre la Construcción Visual de la Revolución.” *La Revolución Mexicana en la Literatura y el Cine*, edited by Olivia Díaz Pérez, Florian Gräfe, Friedhelm Schmidt, Madrid: Iberoamericana, 2010, pp. 217-25.

Silva Escobar, Juan Pablo. “La Época de Oro del Cine Mexicano: la Colonización de un Imaginario Social.” *Culturales*, vol. 7, no. 6, 2011, pp. 7-30.

Tierney, Dolores. *Emilio Fernández: Pictures in the Margins*. Manchester University Press, 2012.

Vega Alfaro, Eduardo de la. “El Cine Mexicano Como Fuente y Forma de la Identidad Nacional (1930-1949).” *Nos Vemos en el Cine*, Guadalajara, México, Gobierno de Jalisco, 2007, pp. 17-39.

5.2 Filmography

La Camarista. Directed by Lila Avilés, performance by Gabriela Cartol, Amplitud, 2018.

María Candelaria. Directed by Emilio Fernández, performance by Dolores del Río, Films Mundiales, 1944.

Roma. Directed by Alfonso Cuarón, performance by Yalitza Aparicio, Netflix, 2018.