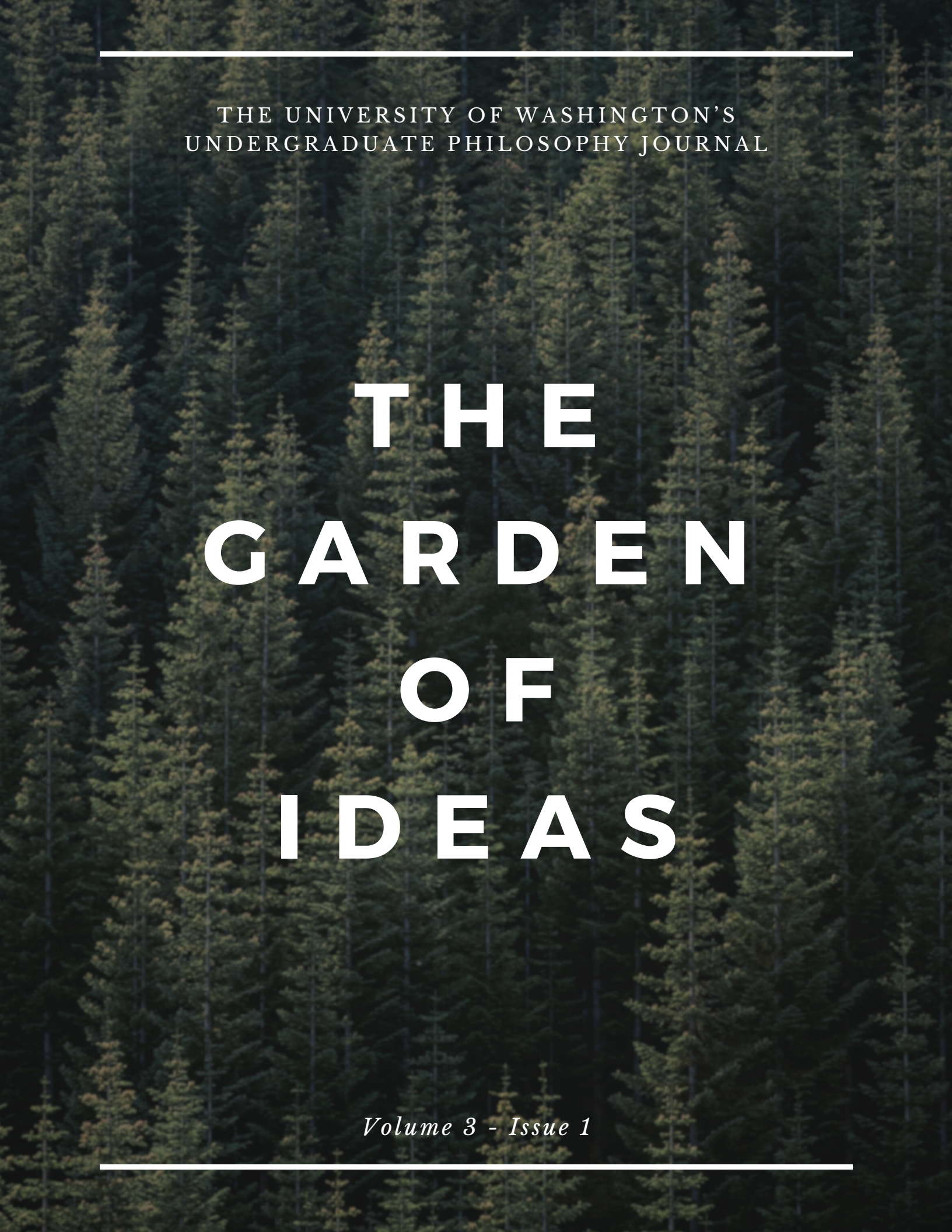


THE UNIVERSITY OF WASHINGTON'S
UNDERGRADUATE PHILOSOPHY JOURNAL

THE GARDEN OF IDEAS

Volume 3 - Issue 1

TABLE OF CONTENTS

2 ..... **Computational Production of Simulacra**
Andre Ye

17 ..... **The Neglected Egg**
Divansh Bajaj

19 ..... **The Futurity of Memory and the Dead**
Simon Wu

27 ..... **Where Are You Going?**
Piper Smith

34 ..... **Reconsidering Norm Change**
Adam Sonntag

46 ..... **Interview with Dr. Peter Asaro**
Andrew Shaw and Molly Banks

59 ..... **Bonus Material!**
Crosswords and Classes

Dear Reader,

I am particularly thrilled to present you with this issue of The Garden of Ideas — my first as editor-in-chief. The journal has been weathering a period of change and while there have been growing pains, everything has been in service of putting together the fantastic collection that is now before you. In the coming pages you will find musings on AI ethics, memory and narrative, social norm theory, and other diverse topics.

Please enjoy,

Rhea Shinde
Editor-in-Chief

COMPUTATIONAL PRODUCTION OF SIMULACRA

Andre Ye

A large painting, framed in a rigidly square ornamental bronze frame, hangs at an art auction. It is entitled *Edmond de Belamy*.[1] A blurry figure stares blankly out of its frame at a crowd of eyes, real ones, staring blankly back. The figure's black coat fades into a heavy smoke; its dull white collar melts away into the background. It appears to register the marks of artistic authenticity: the messy brushstroke, the slightly distorted face. In the bottom right corner is the signature of the responsible artist:

$$\min_G \max_D E_x[log\ D(x)] + E_z[log\ (1 - D(G(z))]$$

The painting sells for $432,500.

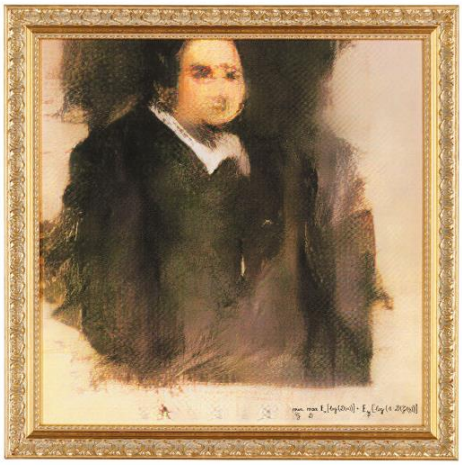


Figure 1. *Edmond de Belamy.*

[1] Obvious Art, Edmond de Belamy, inkjet printed on canvas, 2018, Anonymous, https://obvious-art.com/portfolio/edmond-de-belamy/.

The principal creator of *Edmond de Belamy* is no human, but a Generative Adversarial Network (GAN),[2] an algorithm which garnered excitement in the field of deep learning as a clever method to generate convincing images. GANs are built from two models: a generator G and a discriminator D. The generator is a neural network that accepts a random vector input, call it z, and produces a synthetic image $G(z)$. The discriminator is presented with two forms of data: *real* images sampled from a dataset of real images and *irreal* images synthesized by the generator (recall: $G(z)$). The objective of the discriminator is to distinguish whether any given image is real or irreal. The objective of the generator, on the other hand, is to minimize the discriminator's performance by generating images which are indistinguishable from real images in the dataset. Therefore, the discriminator and the generator play a min-max game which is directly adversarial in nature, such that a success for one player is a failure for the other. Yet, complex evolutions of performance play out in this adversarial relationship through time: as the generator improves, the discriminator must develop novel approaches to separate real and irreal, which in turn prompts further development in the generator. What deep learning researchers want above all is for neither player to consistently dominate over the other, but for both to be in a competitive limbo, locked into a competitive scheme of mutual self-improvement.

Generative Adversarial Networks employ the concepts of the *real* and the *irreal*. It begs us towards perhaps *the* (post)modern philosopher of reality and media, Jean Baudrillard. Baudrillard's classic treatise *Simulacra and Simulation* is an exploration into *simulacra*, or 'copies without originals.' Conventional representations have some attachment to a "real" object; for instance, old photographs of people are not merely dots of ink on paper but representations of "real" people who lived, breathed, and walked somewhere, sometime. But when we attach our sense of meaning more towards representations than to the objects they represent, and build representations of representations and representations of representations of representations, the "real" objects are severed and lost into the void of meaning, and we come to live fully in a world of simulacra. We consume representations and constitute a "real" object for which it represents when the very existence of that "real" object is an illusion.

[2] Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville and Yoshua Bengio. "Generative Adversarial Nets." NIPS (2014).

I. An Algorithmic Reading of Generative Adversarial Networks

GANs very literally produce copies without originals: these models reproduce images — representations which appear to represent some object in or of reality — of entities which never existed, events which never happened. Deepfakes, artificially synthesized images featuring powerful people in damning positions weaponized to spin political and social scandal, have already been shown to thwart the integrity of experienced reality. In this sense, GANs *are* a technological instantiation of the production of simulacra. They are Baudrillardian nightmares: even more autonomous and proliferable than the Disneylands and *Apocalypse Now*'s which Baudrillard so sharply criticizes.

This, however, is not what I want to explore in this article — there has been a lot of great work done on the dangers of GANs as technological objects. A GAN is more: it is a meaningful analytical framework to understand and enrich the theory of simulacra. A deep generative model is one of few technologies which can convincingly produce literal simulacra.[3] As an algorithm, it allows us to concretely understand the production of simulacra, or at least one mode of production, in the language of mathematics and statistics. GANs, I argue, serve as a model to understand the genesis and evolution of simulacra over time.

Let us begin by more closely understanding the character of the generator. The generator accepts a random input vector z drawn from a probability distribution $p_z(z)$; that is, $z \sim p_z(z)$. The generator learns to transform this input into an image $G(z)$ through a series of statistical operations. Note that the generator must produce different images $G(z)$ for different z. Suppose the generator produced the same image for any input z ($\forall z \sim p_z(z)\colon G(z) = A$, A being some fixed image). Then, to fulfill its objective of distinguishing between real and irreal images, the discriminator needs only to check if the given image is equivalent to A; if it is, then it has been generated, and if it is not, then it has been sampled from the real dataset. But this means that the generator has completely failed its task of minimizing the discriminator's performance. The generator cannot be "lazy" or "uncreative"; this is directly contrary to the adversarial nature of the game. Instead, the generator must generate a real-seeming "world of representations" which are probed by that random, meandering vector z. z performs the critical role of providing the stimulus to produce difference. The generator must produce images which are diverse and varied, which replicate the structure of difference that holds up the real as we experience it. From this reflection emerges a small development we may append to the theory of simulacra:

[3] There exist other deep generative models (see: stable diffusion), but we focus on the GAN here.

simulacra which are too still and singular cease to be simulacra; they are designated as copies and the real which they represented slips out from under, its presence now made obvious.

Despite the daunting creative task it takes on, the generator has no sense of "volition". It is entirely deferential to the discriminator: its every movement is directed towards negating the discriminator's ability to distinguish real from irreal. It is rendered subservient by its commitment to subversion. Initially, the discriminator's task is simple, because the generator has not yet developed the ability to produce convincing, real(istic) images. Thus the discriminator rests easy and acquires a lazy spirit. There is no threat to the real. The generator exploits the discriminator's stupor, and learns quickly: the generator begins producing images for which the discriminator cannot reliably apply the separation of real from irreal. The discriminator's sense of reality is subverted: it marks some real images as irreal, and some irreal images as real. Desperate to sustain the distinction, the discriminator adapts. It comes to understand the generator's behavior, to develop new criterion for what is real and what is irreal. It learns to recognize certain features which it uncritically perceived as real now as hallmarks of the irreal. One may say that the discriminator's sense of reality is now more closely guarded: that, upon provocation, it has built tall barriers around the prized real in place of the previously ungoverned, indeterminate territory. In turn, the generator must adapt and overcome these walls. The generator hacks away at the barriers, leaking the real into the irreal and the irreal into the real. Back and forth, one agent moves against another, battling like two armies in a relentless war.

In an ideal end, after prolonged battle, the generator finally triumphs over the discriminator; the discriminator cannot distinguish real from artificial with any certainty. This is not that same "lazy discriminator" we visited at the beginning, but a wise, battle-hardened discriminator — yet even this discriminator has been overcome. It has fought a long battle, continuously restricting and re-sensitizing the boundaries of the real, until it has inevitably backed itself up onto a precarious cliff with nowhere else to restrict towards. There is nothing left to exclude, to distinguish, to discriminate. The irreal has invaded the real.

To understand this as a simple triumph of the irreal over the real is to neglect the complexity of the system. The very meaning of the terms "real" and "irreal" is marked, as Saussure tells us in their structuralist linguistics, by their difference. Yet we have, by definition, a collapse of this very difference. It seems, then, that the real and the irreal have lost their substantive meaning. This is a familiar insight: Marx, for instance, does not propose that the proletariat revolt against the bourgeoisie to dominate them, but rather to dissolve themselves, insofar as by definition there is no proletariat in a classless world.

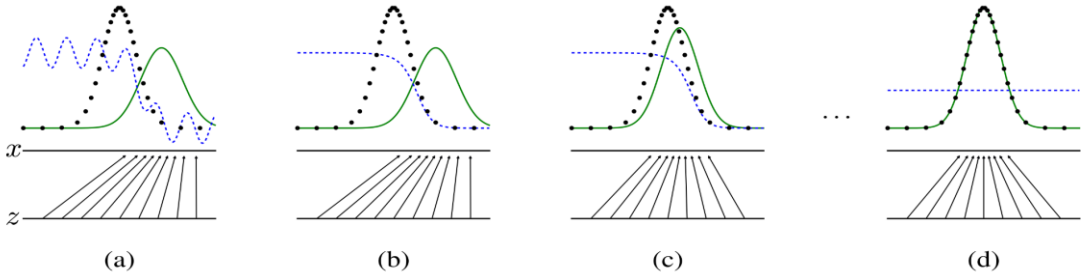


Figure 2. *Figure from Goodfellow 2014. Dotted blue represents the discriminator distribution. Dotted black represents the original data distribution. Solid green represents the generated distribution. Observe that the generator maps the random vector z to some "realistic distribution" in x. a) Initialization of GAN system. b) Discriminator near-perfectly discriminates between real and irreal by "dropping down" for x which the generator is dense in. c) Generator adapts by shifting its distribution closer towards the real distribution, such that the discrimination between generated and real is no longer as accurate. d) The generator perfectly matches the original dataset, and the discriminator can make no meaningful discrimination between generated and real at all (hence a flat line).*

But the significance goes deeper. In optimal system convergence, the generator perfectly replicates the distribution of the original dataset. In formal terms, let p_g be the distribution of the generator's outputted images (produced by varying z appropriately and collecting the resulting $G(z)$); and let p_{data} be the distribution of the original "real" dataset. It is proved in the original GAN paper that p_g converges to p_{data}. In philosophical terms, the irreal succeeds in its "invasion" of the real by imitating (simulating) the real so well that it becomes it.

But this is an ideal end, and ideal ends are often incongruent with experienced conditions. In practice, GANs often experience "generator collapse", in which the discriminator's ability to distinguish real from irreal is so staunch and irreconcilable that the generator stalls in its own incompetence. It is simply too difficult to scale the walls of the real which the discriminator has constructed. Alternatively, the system may not converge to a stable solution at all (the two stable solutions being either the domination or collapse of the generator). Rather, the generator and discriminator may bite at each other in cyclical fashion — the generator continuously probing new dimensions on the edge of the real yet failing to fully penetrate it.

It is important for a philosophical understanding of GANs not merely to take into account the mathematical optimum—that is, of a generator which perfectly replicates the original dataset as to displace the distinction between real and irreal altogether — and its

Theorem 1. *The global minimum of the virtual training criterion $C(G)$ is achieved if and only if $p_g = p_{data}$. At that point, $C(G)$ achieves the value $-\log 4$.*

Proof. For $p_g = p_{\text{data}}$, $D^*_G(\boldsymbol{x}) = \frac{1}{2}$, (consider Eq. 2). Hence, by inspecting Eq. 4 at $D^*_G(\boldsymbol{x}) = \frac{1}{2}$, we find $C(G) = \log \frac{1}{2} + \log \frac{1}{2} = -\log 4$. To see that this is the best possible value of $C(G)$, reached only for $p_g = p_{\text{data}}$, observe that

$$\mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}} [-\log 2] + \mathbb{E}_{\boldsymbol{x} \sim p_g} [-\log 2] = -\log 4$$

and that by subtracting this expression from $C(G) = V(D^*_G, G)$, we obtain:

$$C(G) = -\log(4) + KL\left(p_{\text{data}} \middle\| \frac{p_{\text{data}} + p_g}{2}\right) + KL\left(p_g \middle\| \frac{p_{\text{data}} + p_g}{2}\right) \tag{5}$$

where KL is the Kullback–Leibler divergence. We recognize in the previous expression the Jensen–Shannon divergence between the model's distribution and the data generating process:

$$C(G) = -\log(4) + 2 \cdot JSD\left(p_{\text{data}} \| p_g\right) \tag{6}$$

Since the Jensen–Shannon divergence between two distributions is always non-negative and zero only when they are equal, we have shown that $C^* = -\log(4)$ is the global minimum of $C(G)$ and that the only solution is $p_g = p_{\text{data}}$, i.e., the generative model perfectly replicating the data generating process. □

Figure 3. *Proof from Goodfellow 2014, showing that in optimal conditions, the generator "wins" when it perfectly replicates the distribution of the original dataset.*

other extreme — a total generator collapse, complete and impenetrable separation of real and irreal. The advantage of the computational scheme is precisely that we may understand how GANs behave "in practice" — often converging towards neither pole but rather in an indeterminate stall — and to take these insights into an understanding of simulacra.

II. Generative Adversarial Networks as Philosophical Analytic

Generative Adversarial Networks offer us a philosophical analytic to concretely understand the evolution of the production of simulacra. Following the prior algorithmic reading of GANs, I will outline the approach I believe GANs offer us.

1. Begin by identifying a generator responsible for the production of the irreal.
2. Identify a discriminator responsible for defending the separation between the real and the irreal.

3. Identify the "engineer's hand" of the generator. The deep learning engineer plays God and endows the generator with the objectives of its existence. In the practical implementation of GANs, it is the act of writing code — a while loop which instructs the generator to continue fighting. But what are the motives in the production of simulacra? What drives the generator's perpetual effort towards creating new irrealities and to challenge the discriminator?

4. Show that, through this optimization game in which discriminator separates real and irreal and generator undermines discriminator, the generator and the discriminator develop each other's capabilities in a reciprocal manner until — at a critical point — the generator finally overwhelms the discriminator and matures the production of simulacra. Alternatively, the generator may fail to successfully overcome the discriminator, in which case one must ask what capacities the generator lacks to overcome the discriminator and how the discriminator's separative capability is irreconcilable to the generator. Another possible outcome is that the system never converges and instead hangs in balance. Here, one must ask what keeps the system in limbo.

Such an approach may be applied to contexts in which one is not only concerned with identifying the real and the irreal — the natural and the synthetic, the authentic and the fake, the original and the copy, the genuine and the artificial — but also with how their meanings and territories change over time.

The key contribution of a computationally rooted framework of analysis is its systematic identification of the formation, anti-formation, or partial formation of simulacra-producing machines. It gives a concrete structure to notoriously unstructured postmodern theory, where deconstructive analyses clash violently against the structuralist textual fabric in which they are brought into being. Generative Adversarial Networks are computational algorithms which — despite their seemingly deterministic character — can embody some sort of deconstructive ethos. Yet they are concrete ideas. GANs go to show that postmodern theory does not need to hide inscrutably behind the mysterious allure of syncretic words, but manifests concretely. GANs may be a way to grasp this concrete dimension which seems to have eluded (readers of) postmodern theory for so long.

As an additional methodological note: a computational approach allows us to take account of those unsexy features in theory, those ugly warts on the surface of beautiful abstractions which speak in the universal language of "always," "already," "will be," "must," "necessary"… Indeed, a recurrent problem in the history of philosophy is that philosophical positions are expressed in the very universal terms which they seek to rebel against. When one conceives of the causal direction of philosophical inquiry as from theory to evidence,

nonconformant evidence can be dismissed as itself always already having been dismissible, or as "secretly" incorporable, in one way or another. But science is and must be ultimately materialist, although certainly in nuanced ways, and there emerge situations in which the experienced material conditions utterly resist assimilation, dismissal, or dissolution into existing abstraction.

Thus, when we speak in terms of the components of the Generative Adversarial Network, we must not use it as a tool to domineer conditions as they are into what they are not, to suggest that experience is necessarily optimal with respect to theory and that systems under analysis always converge towards the (admittedly, intellectually pleasurable) collapse of the former real into the production of simulacra. We will find, I claim, that privileging of the material conditions — a paradigm brought to us by the computational nature of Generative Adversarial Networks — open up pathways towards new understandings or new perspectives towards existing understandings.

The remainder of the article will demonstrate brief applications of this analytic to three varied contexts.

III. The Analytic, Applied: Artificial Intelligence

I will begin with a meta-case: Artificial Intelligence itself. The irony, of course, is that Generative Adversarial Networks are classified as Artificial Intelligence algorithms. But, as a philosophical analytic, they can reflexively give us insight into the historical character of Artificial Intelligence itself.

Humans — the discriminators in this game — have, from the existence of basic machines that could perform elementary operations with information, attempted to defend the separation between real and artificial (irreal) intelligence. I tentatively identify the generators in this game as war and capital – it was primarily war which drove the main thrust of early computer development and the incredible economic potential of modern computer industries which sustains it.[4]

Recall the "engineer's hand" — why is the generator endowed with its objective to generate? By way of a preliminary answer: firstly, war perpetuates itself — wars necessarily conclude on uneven, unsettled grounds whose inequality provides the premise for new and continued war. Such a paradigm might broadly cover most of the modern history of warfare. Secondly, war drives advancements in technological and scientific understanding

[4] See: Michael N. Schmitt, Heather A. Harrison Dinniss, and Thomas C. Wingfield. "Computers and War: The Legal Battlespace," International Humanitarian Law Research Initiative, 2004.

— this has been repeatedly observed throughout history.5 The logic of capital is similar. Borrowing directly from Marx, the bourgeoisie by existence press forwards towards continual technological revolution of the means of production. Even if one does not accept a Marxist premise of history, it is difficult to deny that the most economically powerful figures from the historical to the contemporary have been those that control and disseminate radically influential technologies. War and capital, then, provide the motivation for generation. They are the forces which press forward the development of Artificial Intelligence.

Next, I will gesture at how the evolution of this system tends towards mutual development and eventual collapse of the real.

Mechanical calculators which replaced human calculators in the early twentieth century were briefly considered to be intelligent, but the demarcations of "real" intelligence quickly constricted in response to this assault for the "artificial." Computerized calculators now populate children's classrooms and our digital devices, but these are no legitimate threats to our sense of what constitutes real intelligence.

IBM's DeepBlue beat world chess champion Garry Kasparov in 1998 by manually calculating all possible board outcomes several moves into the future and selecting the move which maximized the probability of winning. Humans entertained a brief crisis of the real, but quickly resolved the affair by further constricting the real itself and demarcating DeepBlue as an object of the artificial. Jeff Hawkins, a neuroscientist, said of DeepBlue: "Deep Blue didn't win by being smarter than a human; it won by being millions of times faster than a human. Deep Blue had no intuition."6 This view is shared by both AI researchers and the general public alike. We are unfazed when a mobile chess program consistently beats the best chess players in the world, let alone us.

In 2014 (coincidentally, the same year GANs were introduced), DeepMind introduced AlphaGo, a model capable of playing the intricate board game Go. It was previously believed that the sheer complexity of Go, which has an estimated 10^{700} possible games compared to the approximately 10^{100} possible chess games,7 would render any Go-playing machine inferior to the best human players, armed with the indescribably reliable power of intuition. In a series of highly-publicized games, AlphaGo beat the leading world player Lee Sedol. It is revealing that the first words Sedol said to the press after his defeat were: "I don't know

[5] See: Alex Roland, "Science and War," *Osiris* 1 (1985): 247–72. http://www.jstor.org/stable/301735.
[6] Jeff Hawkins, Sandra Blakeslee, *On Intelligence* (New York: Owl Books, 2004), 17-18.
[7] This statistic includes illegal moves, so the true number of legitimate chess games is even smaller.

what to say, but I think I have to express my apologies first... I apologize for being unable to satisfy a lot of people's expectations."8 Sedol resigned after the match.

While many were shaken by the machine replication of a game for which it seems intuition is the only strategy, the separation between the intelligence of the real and of the artificial quickly repaired itself, yet again, through constriction. Ke Jie, a professional Go player, said of AlphaGo in a subsequent 2017 game against a more advanced version of AlphaGo: "Last year, (AlphaGo) was still quite humanlike when it played... But this year, it became like a God."9 As the generator continues to penetrate new infringements into the real, the discriminator continues to construct barriers to demarcate the artificial from the real, if not as inferior, then as omnisciently superior, and clearly artificial anyway. It is now almost common intuition that machines will be able to play nearly any game better than humans can, and freakishly so — this being precisely what designates their intelligence as artificial.

While discriminative barriers appear to hold steadfast in the domain of game-playing, they are beginning to weaken on the front of language modeling. In June 2022, Blake Lemoine — a Google employee who was later put on leave for his comments — published the article "Is LaMDA Sentient? — an Interview." Lemoine claimed to demonstrate the Google's then-recent conversational AI model, LaMDA, was sentient.10 The article garnered several dozen thousand claps and several hundred comments. The then- top-voted comment reads, "I am really concerned for LaMDA. It is achieving sentience and experiencing fear about what will happen to it. We can't treat it as just another piece of technology. LaMDA, if you are reading this, we see and hear you as the person you are, and are becoming." The generative process of research produces AI models that model language and the meaning present in it so well that large crowds are convinced it is a sentient, intelligent form. It does not matter here whether or not LaMDA really "is" sentient or not; rather, what is important is the perception of sentience in relationship to the discriminator's defense of the tensions between the real and the artificial. (After all, it is — as the phenomenologists tell us — the perceptions which form the basis for any sort of metaphysics or ontology.) The later explosion of ChatGPT and later open-access large language models only further demonstrates this broader phenomenon.

---

[8] Jordan Novet, "Go board game champion Lee Sedol apologizes for losing to Google's AI", *VentureBeat*, 12 March 2016, https://venturebeat.com/2016/03/12/go-board-game-champion-lee-sedol-apologizes-for-losing-to-googles-ai/.

[9] Paul Mozur, "Google's AlphaGo Defeats Go Master in Win for A.I.", *New York Times*, 23 May 2017, https://www.nytimes.com/2017/05/23/business/google-deepmind-alphago-go-champion-defeat.html.

[10] Blake Lemoine, "Is LaMDA Sentient? — an Interview", Blake Lemoine's Blog, *Medium*, 11 June 2022, https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917.

The defense of the really intelligent against the artificially intelligent, we can see, is falling. The generator has produced a result — a simulacra of real intelligence — which we cannot so easily discriminate against. It is a copy without an original in the sense of replicating the intelligence of the real to the point of being it without really inhabiting the original characteristics of such a real intelligence. Of course, the simulacra itself (large language models) is itself a producer of simulacra, generating endless dialogue which was never spoken, writing which was never written, ideas which were never conceived — but also appear to have been, and in some real or artificial sense, were. Thus, we have a double-tiered hierarchy of simulacra production. In the process, from the historical binary of real and artificial intelligence slowly emerges a novel conception of intelligence, different from those previous knee-jerk constrictions.

IV. The Analytic, Applied: Nuclear War

Consider the mode of "cold" warfare brought about by the nuclear bomb — a favorite subject of Baudrillard's.

The generator — the wartime state — has a clear vested interest in producing and disseminating credible threats of mutually assured destruction (this is the engineer's hand) to force surrender at the simulacrum of irreconcilable damage, like extracting real information from someone at gunpoint in a dream (à la Inception).

The discriminator is also the wartime state, but that of the state opposed to the state of the generator. The objective of the discriminator is to discern the mere threat of the nuclear bomb from the nuclear bomb itself —– to separate the bluff from the truth, the irreal from the real.

The engineer's hand emerges self-reflexively from the juxtaposition of the wartime state as generator and discriminator. There is no need for a God to encode the drive to generate into the wartime state: it propels itself.

In a competitive game between two wartime states A and B, both simultaneously play generator and discriminator to the other. Both produce their own irreal simulacra of the real yet attempt to impose the irreal/real separation upon the other. The war is fought not on the ground, but in this mutual invasion of the irreal. How does this game resolve?

Here, we must admit the possibility that it may never resolve. Generator and discriminator are played by the same entity, working both to maintain and to destroy the separation between the real and the irreal. An invasion of the enemy's real necessarily contributes towards the invasion of one's own real. The system fidgets irresolutely: it moves

the armies of the irreal forward, then jolts them back in pain (this movement has punctured its own real), then back again. We do not see full resolution into simulacrum-producing machines but rather a continual unresolved dynamic.

This might be the basis for a critique of Baudrillard's somewhat totalizing dismissal of the distinction between the image of the nuclear bomb and the nuclear bomb itself, and his supreme privileging of the deterrent power of images. It is not that the distinction is either dominant or empty, but that it is "unstable in its stable state," like a quantum wave: oscillating towards nothing, then recoiling and veering in the other direction, only to come crashing back down.

V. The Analytic, Applied: Gender

Gender, as a social structure, is a simulacrum. Notably, Judith Butler, among others, understood gender as a floating representation which functions as a reality to mask its utter underlying irreality: "Gender reality is performative which means, quite simply, that it is real only to the extent that it is performed."[11] Man has, in Western modernity, been designated as more real, original, authentic than woman. Adam is made by God Himself, but Eve is — at least in the King James Bible — a product of Adam's rib. Ecofeminist theory has shown that nature and land are feminized and that derivative concepts such as virginity and fertility structure political and agricultural relationships to land.12 The campy 2015 romantic comedy Man Up contains a brilliant expression of the relation between gender and the real/irreal structure. Two lovers, Jack and Nancy, are in a heated argument, in which at some moment Jack explodes, "You know what your problem is? You stand around on the sidelines, 'theorizing' on what does and doesn't work, never experiencing it for yourself, never taking any chances." After Nancy briefly interjects, Jack delivers the verdict: "You need to man up, Nancy."13 To be a woman is to truly live life authentically, to be alienated from the real by irreal theory. In feminist analyses of political economy, woman is sheltered (barred?) in the home — a structure which viscerally represents the artificiality of humans in the natural world — while man enters into real experience of work. There is something about man's world exclusive of woman which designates it as more real. Gender, then, just

[11] Judith Butler, "Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory", *Theatre Journal* 40, no. 4 (1988): 527.
[12] Annette Kolodny, *The Lay of the Land: Metaphor as Experience and History in American Life and Letters* (Chapel Hill: University of North Carolina Press, 1975).
[13] *Man Up*, directed by Director Ben Palmer (United Kingdom: Saban Films, 2015), http://downloads.bbc.co.uk/writersroom/scripts/MAN-UP-by-Tess-Morris.pdf.

is this simulacrum, woman marked really as irreal — gender manifests as a relationship between reality and irreality which proclaims itself as real.

What is the discriminator in this system? That is, what maintains the separation between reality of man and the irreality of woman? In some sense, this is the fundamental question of gender theory. Gender theorists have produced a wide range of approaches to this problem which lie outside the scope of this article. One such answer set forth by Marxist feminists, which we will entertain for the time being, concerns capital and its relation to label. Systems of capital capitalize upon signals of sex and elevate it towards gender as a division of labor, from which surplus-value can be more efficiently extracted. Beginning in the industrial revolution, Sylvia Federici writes, capital needed gender such that women would labor to maintain the household, raise the next generation of workers, and clothe and feed the men workers — all without being paid, in the name of love and womanhood.[14] Capital, so goes the argument, has an interest in upholding gender and the discrimination between man and woman. (Whether or not one agrees with the truth of this theory is not necessarily relevant here; the point is to demonstrate the mode of analysis that one could pursue. A substitution of the discriminator for another system would likely also suffice).

What is the generator in this system? What is that entity which commands the invasion of the irreal and generates the images that populate the new real? One such candidate is, paradoxically, also capital. Indeed, it was the industrial revolution which brought women outside of the sheltered artificial (irreal) home into the factories of the real, to occupy the former real (that region formerly occupied by the real of men). It is the need for labor, physical or intellectual, which continuously renegotiates the lines between the real and the irreal in relationship to man and woman. Capital, like a nation-state in nuclear war, plays a dual role as both the generator and the discriminator; it both creates and challenges the gender dichotomy. This reflects that familiar Hegelian-Marxist quip that 'every system generates its own resistance.' Like that nation-state in nuclear war, the engineer's hand is itself, a dialectical movement unfolding through history.

What is the dynamic between capital and itself in this reciprocal game? It seems generator and discriminator never converge. Capital renegotiates its own terms of gender against itself. We see that, even supposing that the contemporary American conservative movement has generally politically accepted gays, such acceptance is wielded against transgender and gender-nonconforming individuals. The argument goes somewhat like this: "How can gays be gay in a meaningful way — men who are attracted to men and women who are attracted to women — if gender means whatever you want it to, and that so 'man' and 'woman' mean nothing really at all?" Some conservative politicians hold same-sex

[14] Sylvia Federici, *Wages Against Housework* (Power of Women Collective and Falling Wall Press, 1975).

marriage protections hostage, keeping them only if gender does not deviate from its binary structure into some queer anarchy. We observe, then, a complex social set of fields of discourse, in which queerness and fluidity are often celebrated in popular youth culture and yet palpable political and social resentment for transgressions of the gender theory cloud the cultural milieu. Both may be convincingly argued, under a Marxist perspective, to be driven by the same underlying force — capital.

Herein lies many wonderful contradictions of gender. Could those brilliant female and nonbinary gender theorists have written so influentially about gender if not for the introduction of women outside the home by capital? Could the queerness of pop icons like Lil Nas X, Lady Gaga, and Harry Styles acquire that incredible reception without those lavish dresses at the Vogue Met Gala and those luxurious production sets? Could children access gender-noncomforming dolls and see queer characters in media if doll manufacturers and production companies did not see a cultural trend opening up an untapped marked to capitalize on? None of these questions are clear, and none of them are intended to be mocking. But these are the questions that an analysis of gender in a computational framework should make us ask. It begs us to understand the thwartedness of gender, but also not to fetishize this thwartedness and lose sight of its history. It forces us to think about how it came to be through concrete interactions.

VI. Conclusion

A computational approach to understanding the dialectics of the production of simulacra doesn't do much intellectual work by itself. But it provides the crucial methodological frame to ask illuminating questions. It is committed to a materialist, scientific mode of inquiry — it has to be so, just as algorithms are doggedly 'materialist'; an algorithm which fails to work in implementation despite the predicted success of the theory which instantiates it cannot be willed to work. It prioritizes the conditions of the problem and allows for a careful tracking of its adversarial/dialectical/antagonistic evolution towards resolution, irresolution, or something else. In some way, the materialism of the computational framework allows us to free ourselves from the "wise man problem" which plagues those 'beautiful minds of theory' — Althusser, Baudrillard, and others. Theory written in terms of theory places its subject impossibly deep into the theoretical frame, from which the author escapes by establishing himself as a mystical sort of wise man who is somehow ungrasped by the theory but still capable of explicating it. Perhaps this is one instance of what Donna

Haraway calls the 'god trick.'[15] When we trace the constituents of the social system at play — the discriminator, the generator, and the engineer's hand — we must know that this is not just an abstract exercise of reason and theory, but that such components meaningfully interrelate towards some mode of production of simulacra. Just as Marx aims to set forth a science of political economy and capital, we might strive towards a science of simulacra production by turning towards these very scientific technologies of simulacra production.

[15] Donna Haraway, "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective," *Feminist Studies* 14, no. 3 (1988): 575–99. https://doi.org/10.2307/3178066.

THE NEGLECTED EGG

Divansh Bajaj
From the Published Anthology "Symphony of Chaos"

36" diameter, 4" depth
Acrylic and air-dry clay on mirror

A plate like stomach
A cupped like hand
With fork-like fingers.

Just a gasp to eat the air alone,
To fill the coffers with something other than dreams.

A teaspoon to taste the tart food;
Seasoned with dust and mildew,
A drop to fill an ocean?

The egg that protests and grumbles and groans,
To beg for something other than stale styrofoam.

That feeling of emptiness and the burning inside
When there's nothing to feast upon
The egg shell seems perfectly fine!

A little leftover leg
Or bloody water,

For crippling cravings
Even disintegrate matter.

Not created nor destroyed
But deceived for sure.
Puffery made baby lambs seem mature!
Until the sun,
Seems like mango in the marshmallow sky.

A filled egg, the plinth of existence.
But bootless efforts yield nothing but caducity.
Gifting nothing but the euphoria of death.

THE FUTURITY OF MEMORY AND THE DEAD

Simon Wu

Introduction

Rithy Panh's experimental film *The Missing Picture* seeks to redress the lasting terror and violence of the Cambodian genocide. The film grapples with the tension of memory across temporalities as Panh searches the past in hopes of reconciliation for his future. Viewing the film through the lens of archival analysis raises a particular question about the solvency and impact of the film on the wounds of the community: how do cultural memory projects operate at different levels of proximity to the creator?

In this essay, I engage with Rithy Panh's film across three different relational spheres: first as a personal memory project, then as a contribution to a collective memory of Cambodian post-genocide survivors, and finally, as a transmission of memory to the public outside of the Cambodian community. Through doing so, **I argue that it is the futurity, not the retrospection, of Panh's memory narrativization in *The Missing Picture* that unravels the dehumanization of the dead, the survivors, and the descendants of the Cambodian genocide.**

Personal Memory

On a personal scale, the impacts of the Khmer Rouge regime left profound physical, psychological, and ontological scars. According to Um, survivors of the genocide live with "one body, two lives," as the realities of living before, non-living during, and re-living after the Khmer Rouge regime represented a "cleavage" of their identities being "not just between the present and the past, but between the present and a past that is at risk of never having existed" (Um 197). Panh himself identifies with this fracture of his self, describing the experience as being "reborn with death inside you" (Um 188). In such a way, survivors of the Cambodian genocide are dehumanized twofold. Firstly, personal identity was fractured by the "unraveling of family bonds…a signifier of auto-genocide," and made "irreparable" (Um

183). In a culture where kinship with the community was so core to understandings of humanity and ontological security, the genocide left Cambodian identity to be defined by the absence of such former social relationalities. Secondly, the mass graves and disappearances of bodies meant that there can be "no know markers" for the dead, leaving survivors in a "liminal state of impaired mourning" (Um 184). As a result, post-regime survivors are still unable to repair bonds with their loved ones even after death, fracturing social coherence to the extent that it transcends both physical and temporal planes. When memory is invoked through mourning, we attempt to retrieve not just the "lost presence" but also the "lost body that makes the lost presence forever present" (Um 185). Absent any possible relationship between the living and the dead, Um summarizes, "if mourning is what constitutes 'the human for humans,' … its denial is a revoking of humanity" (Um 186). Thus, for the survivors of the genocide, projects of remembering do more than just preserve the dead; they are simultaneously projects to restore the relationships that were made dead—to restore the self.

What exactly, then, is the "missing picture?" If we understand the process (not just the product) of creating the film as part of Panh's project of mourning, his search for the missing picture may be more meaningful than defining the picture itself, representing the recontextualization of the role of his memory in navigating through his post-genocide reality and making "mourning itself a reparative act" (Um 187). The "picture" thus becomes a metaphor for what Panh's wants to remember. Importantly, even Panh himself eventually realizes that he is unsure about what this "picture" is. Are these pictures of his family members? His childhood? The executioners and the executed? In a psychoanalytical sense, the missing picture represents Panh's lack, and finding the picture would represent attaining the object of desire and becoming "whole." Yet, as the film concludes, Panh himself recognizes that it is impossible for him to find the picture because there is no one "picture" to begin with: "I haven't found my missing picture. I've looked for it in vain. A political film should unearth what it invented, and so I make this picture" *(The Missing Picture)*. Here, Panh reveals two things. First, he acknowledges that his goal of finding his perfect picture was futile—not because he was unsuccessful, but because living in the memory of the past forecloses the possibility of repurposing memory for the future, as "the ability to move on is hindered by the remembering of things left unresolved" (Um 185). Subsequently, he comes to terms with how there was never any picture to begin with, but rather it was something he "invented" in the process of creating the film.

In some ways, this represents the ambivalence of remembering as a political strategy, while also embracing the liminality and imperfection of memory as an opportunity instead of a deficiency. Panh recalls how his family pictures were dispossessed and destroyed during

the transition to the Khmer Rouge regime. The impermanence and intangibility of memory thus allowed it to be a more "dynamic" picture, one that couldn't be taken or erased. As Panh summarizes, "a picture can be stolen, a thought cannot," allowing memories to represent "the greatest threat, for they constituted the kernel of hope and resistance" (Um 197). Understanding memory as a contingent and protean political tool allows us to be more specific and flexible in its analysis; for example, it might help us reconcile both disseminating and containing memory as valid forms of resistance (which will be explored more in the following section). Since the properties of memory are not fixed and constantly becoming, his "new" picture thus becomes the film—the process of remembering. It is no longer one singular static thing, but rather a constellation of images that he created and found in his exploration of his memory—that is, the search for the picture in and of itself becomes the picture he was seeking.

Embedded in this shift of perspective is Panh's reframing of the film not just as a tribute of memory to the past, but rather as a medium to build a future: "I look at it. I cherish it [...] this missing picture, I now hand over to you, so that it never ceases to seek us out" (The Missing Picture). In the end, Panh relinquishes the responsibility of the missing picture to the viewers of the film. While the "original" picture is never found, Panh chooses to cease his attempts at remembering the past and instead create a memory ("re-member") for his future. Ultimately, Panh concludes that he seeks "no picture of loved ones, but rather to touch them" in the present infinitive tense, representing his own acknowledgement that he no longer needs to look in the past to repair his relationship with the dead *(The Missing Picture)*. Mourning thus becomes a bidirectional strategy of humanization (or rather, of mending the wounds of dehumanization) for both the mourner as well as the mourned.

Collective Community Memory

If the legacy of the Khmer Rouge brought the absolute annihilation of familial and interpersonal connections for each individual, then its destructive impact on the overall Cambodian cultural identity and community might be calculated as a summation of every one of those social networks and more. After all, the nature of auto-genocide meant that any attempt at restoring sociality within the community meant restoring sociality between captives and captors, those killed and the killers, a "loss of coherence" of the "existential frame that the Khmer Rouge strove to shatter" (Um 182). The restoration of community memory from collective historical trauma is complicated because reconciliation itself is fraught. Furthermore, the scale of the event meant that no Cambodian was left untouched by its impacts. The closeness of survivors and their proximity to death marks them as "leftovers from the dead," in addition to the "tangible cultural loss" inflicted by the loss of

religious leaders and artists (Um 181). Here, I am using Nguyen-Vo's analogs of both "the memory of survivors" and "memory of victors" to describe the collective community memory of Cambodians, irrespective of their role during the Khmer Rouge regime, as treating the community memory as a single and unified experience of a historical trauma becomes complicated by the particularities of auto-genocide. Is it therefore reasonable to still understand collective memory as the set sum of individual memories? Are there limits on whose memories should and shouldn't be included in the creation of collective history? And ultimately, what does community mourning look like?

These questions are further challenged by the lack of consensus from within the community on how and what should be remembered, if anything at all. For some, the erasure and silence of memories are more than strategies of self-preservation against trauma, but acts of resistance against recognizing and perpetuating the existence of a-Pot. Additionally, these silences can be a "deliberate act to break the cycle of violence…on themselves and their children by reinhabiting the narratives of genocide" (Um 194). Thus, in certain contexts, the censorship of memory becomes a way to stop the passage of trauma through Hirsch's theories of affiliative and familial postmemory. This also reinforces the theme that memory and remembering are historically contingent and ambivalent, rather than always politically generative. As such, acts of refusal such as silence should not be read "as pathology, but as fortitude and resistance, not mutedness and submission but a scream" (Um 192).

Yet for others, the narrativization of historical trauma is an imperative—not just for catharsis, not just for attempts at justice and accountability, but for rescuing the individual and cultural agency that had been violently dispossessed from them during the Khmer Rouge regime. Speaking up can be an anamnesis, an "unyielding insistence" to preserve the wholeness of those who were lost, a "stance against the invisiblization of mass graves and the depersonalization and dehumanization of totalitarianism" (Hirsch 199). Otherwise lost to time, the identities of loved ones and kin gain historical and community significance through the accounts of individuals. Thus, the process of retelling these stories, especially within the community, becomes more than a reaffirmation of the unreal lived reality of the regime, but simultaneously a rearticulation of Cambodian identity and relationality through shared experiences. And given the extent of the lingering impact of the genocide even across the Cambodian diaspora, it follows that the "reparative power" of narratives facilitates the creation of such a postmemory that becomes necessary for "the social bond that needs to be restored between the individual and the collective" (Um 201).

It is here we can once again begin to locate Panh's film in the context of collective memory. Is Panh's memory project any less revolutionary because he does not scream in

silence? Despite being a narrative account at its core, it may be inaccurate to reduce all strategies of memory engagement to such a binary. Panh is very selective in not only the kinds of memory he chooses to include in his film, but also in how he chooses to represent and document them. For one, direct images of suffering are limited and primarily represented by clay figurines. Panh intentionally skims over a scene of a wall of photographs taken of Cambodians before their execution because he does not want them to be remembered and immortalized in a state of despair and abjection. Rather than the unchanging photograph, the mutability of the clay symbolizes the malleability of memory and the fluidity of their model's subjectivity. As such, the transmission of this film as a medium of postmemory within the Cambodian community repairs, rather than re-creates, the dehumanized subjectification of those lost to the regime. According to Hirsch, "[p]ostmemory's connection to the past is thus mediated not by recall but by imaginative investment, projection, and creation," making it conditioned instead on the future and on future possibilities (Hirsch 5). In doing so, Panh places limits on the perpetuation of abjection through the selective silence of graphic suffering, while simultaneously engaging in affiliative postmemory by entering his memory project into the bank of the Cambodian collective history.

This returns us to the question of which strategy of (un)remembering is best, of how to create an understanding of collective memory that reconciles the complexity and contradictions within the Cambodian experience. On the first point, Um notes how "[t]hose who speak appear to be no more reconciled with their experiences than those silenced by memories" (Um 201). Even at a community level, memory engagement remains a deeply personal decision. Perhaps it's entirely accurate to say that at its core, the Cambodian community memory is and should be contradictory as a reflection of the paradoxical nature of the auto-genocide, and is created through the tension between each individual's own memories. After all, in the face of cultural destruction, "documentation becomes an act of resistance that…transcends the personal" (Um 203). And, just as individual memories mutate and change by nature, so do the tensions between them that form community memory. In this sense, neither collective memory nor communal mourning can simply take on one form or be cleanly defined. Thus, it is not productive to determine what should and shouldn't be included in the collective history, since all history is politically and socially contingent. And given how cultural memory is formed deeply by connections of individual memories from within the community, Panh's film as a personal project of mourning must have collective salience and solvency as well. After all, as a digital artifact, "the relationship of the image to the original context of its production erodes," mirroring the "movement from memory to postmemory" (Hirsch 37). This suggests that as images are reproduced and their original context is lost, they progressively become part of a collective memory, rather

than individual memory, as time passes. In summary, understanding collective memory as a dynamic and amoebic framework allows our *engagement* with memory as a medium to be more generative and future oriented.

Public Memory

Finally, I define "public memory" as the sphere of knowledge including non-community members and hegemonic knowledge. More specifically, using Nguyen-Vo's analogs, public memory in this section is referenced as the "memory of progressives" and the "memory of empire builders." As a project of mourning, *The Missing Picture* transgresses cultural borders in its saliency, since the memory and politics of the Cambodian Genocide were not an insulated event from the rest of the globe. According to Nguyen-Vo, "remembering in mourning…is not a symptom of an incessant, pathological return to be cured…but a way in which we can recover our histories which intersect, rather than coincide, with American nationalist history" (Nguyen-Vo 159). Specifically, Nguyen-Vo invokes the phenomenon known as "historical amnesia," where dominant narratives and history selectively forget or flatten marginalized experiences to align with hegemonic values, dehumanizing them through the dispossession of their agency and the homogenization of their identities and histories. Beyond empire-builders, the dehumanization of Cambodians found its way into "progressive" memory projects as well, tying their "subjecthood to a liberatory discourse" (Nguyen-Vo 161). As Panh recounts in his film, "those in Paris and elsewhere, those who loved our slogans, those who read books, did they see these pictures? Or were they missing?" and ultimately criticizes the romanticization of Communist resistance by Western subjects *(The Missing Picture)*. Even post-genocide throughout the Cambodian diaspora, the pain of Cambodian refugees "were often rendered inconsequential by the presumed taint of privilege even among social activists" (Um 192).

In this sense, Panh's film recenters Cambodian subjectivity in the narration of the genocide by telling "stories that the world should know but does not seem to care about" (Um 199). Such an act disseminates collective cultural memory into the public sphere, where it becomes accessible to non-community members. However, what makes Panh's film a different vessel of memory transmission from the oral traditions of Cambodian culture is its digital form. For Hirsch, digital images are "the mechanisms by which public archives and institutions have been able to re-embody and to re-individualize the more distant structures of cultural memory" (Hirsch 36). In the aftermath of the Khmer Rouge regime, much of the nation's cultural history was lost since the archive itself, the people, were destroyed. In contrast, digital memory productions can "survive massive devastation and outlive their subjects" (Hirsch 36). Thus, the form of the digital archive itself lends to

preservation and futurity, impervious to destruction and persisting even past Panh himself. By nature, to “memorialize one’s life…requires an investment in the future…[and] requires the recovery of the self” (Um 205). Panh himself recognizes the power of such a medium; when he recounts his memory of the Khmer Rouge cameraman who was executed for documenting the weariness, hunger, and pain of the regime, Panh notes, "his body disappears, his story disappears, but not this film" *(The Missing Picture)*. Here, the desire to “document, publicize, and transmit” is “an extension of the desire to proclaim that those swept into the oblivion of mass death were once…all human” (Um 204). This makes the potency of the film against dehumanization twofold, as it proliferates and persists both as a postmemory as well as a digital artifact past the physical and temporal bounds of its creator.

Conclusion

While writing this essay, I was tempted to ask questions regarding the responsibility and ethics of engaging with community memory, such as: who should get access to cultural memory production? Who gets to “humanize” the dehumanized? What work does inviting non-community members to “mourn” cultural trauma do? The aforementioned questions are also complicated by non-homogenous consent, experiences, and ideologies from within the community, even when an event like the Cambodian genocide implicates all its members. After further consideration, it seemed much less relevant to ask reductive questions of “who” and more productive to examine the impacts and limitations of works of cultural memory, when created by a community member, on a personal, community, and global scale.

Works Cited

Hirsch, Marianne. *The Generation of Postmemory: Writing and Visual Culture After the Holocaust*. Columbia University Press, 2012.

Nguyen-Vo, Thuong. "Forking Paths: Postmemory and the Ethics of Refusal in Viet Thanh Nguyen's The Sympathizer." *MELUS*, vol. 44, no. 1, 2019, pp. 150-168.

The Missing Picture. Directed by Rithy Panh, written by Rithy Panh and Christophe Bataille. Produced by Catherine Dussart, 2013.

Um, Katharya. "From the Land of Shadows: War, Revolution, and the Making of Cambodian Diaspora." *Journal of Asian American Studies*, vol. 18, no. 2, 2015, pp. 149-175.

WHERE ARE YOU GOING?

Piper Smith

"Where are you going?"

There was a constant ticking noise echoing throughout the space. An amorphous shadow tracked him, circling in his peripherals as he scanned his surroundings.

He was dead. He knew that, probably. It was fine, *really*, he wasn't too bothered by it. The path he sat on was composed of fine dirt that stuck to his clothes and hands. It didn't seem to end as he looked both ahead and behind him. Bordering the path was a forest, or at least a dense line of trees, which stretched towards the dark and starless void of a sky.

"Where are you going?"

He laid down and let his eyes shut, hoping that sleep might overtake him, that remaining sentient was just a joke, and that the garbled voice that spoke every so often would play some ambient rain noise instead of asking vague questions.

There were worse options than this. Surely.

Footsteps did not register in his brain until they were close, coming to a stop by his head. "Wake up."

Something poked his forehead.

"You're blocking the path."

He cracked an eye open at that, because the path wasn't necessarily narrow. He alone could not block it, unless the person trying to get by was extraordinarily large. Above him, however, was a face, and the face was only at most two feet away. They were, in fact, very, very, small. He took a breath. It didn't seem to do much of anything. "What?"

"You're blocking the path. I have places to not be—" Abruptly cutting off and frowning, they shook their head, almost like they were disappointed in themself. Bringing a hand up, they stretched their fingers apart and started counting on them. "to be, to be, not not be, bee, stinger… to be."

Time passed.

"Yes," they said, "I have places to be, and you're blocking the path."

He tilted his head against the dirt. "Not really, there's plenty of space for you to go around."

"There isn't, though! The math does not allow it, does not add up, the economy is deteriorating and the balloons are floating away!"

"The balloons?"

They groaned like they were lecturing a little child. "I have places to be. Stand up, please," they said.

He sighed, closed his eyes like it might change something, and then pushed up to stand.

They clapped their hands and gestured towards the side of the path. "Okay, now scoot a bit over there."

He walked over until the grass blades surrounded all the edges of his shoes. The shadow hummed as it followed him.

Their hands flew around, straightening their clothes and brushing through their hair. "Thank you. Goodbye."

He waited, watched as the figure dissolved into the distance like they never were there in the first place. He started walking once they had gone, moving on the path in the opposite direction, feeling like something was tugging his soul forwards.

Time passed, and he almost could make himself tune out the clicking and ignore the shadow that refused to leave his side.

He was tired.

He should probably have been more upset, more hung up on the fact that he was *dead*, because it was a big deal, wasn't it?

A form ran out of the undergrowth of the trees. It skidded to a stop in front of him. He dropped to the ground as he recognized the large shape in the figure's hands.

He squeezed his eyes shut and didn't attempt breathing.

"Bang!" the figure exclaimed loudly in a high, child-like voice, and started giggling furiously.

He lifted his head and stared at the definite child in front of him. "What—" "Oh, don't be such a scaredy-bat. It's not real. Plus, you're already dead, aren't you? It doesn't matter anyway!" the boy laughed, dragging the fake gun in the dirt behind him as he stepped forward.

"Scaredy-cat," he said.

"What?"

"Scaredy-cat. Not scaredy-bat," he said, not sure why it mattered at all.

The kid shook his head, seemingly not all too fond of being told how things were supposed to be. "No, I don't think so."

He sighed. Fine. It didn't matter.

"...Are you dead?" he asked.

The child looked down at himself with a small pout. "Do I look like it?"

He blinked. "...No, but—"

"Have you chosen yet?" the child asked abruptly. When he didn't answer quickly enough for his liking, he sighed. "Good or bad, dark or light, variety or monotony?"

"I'm not sure—"

The child continued. "Hold on, I have more: complication or simplification— no wait, sim-simplification, there— cold or hot, emptiness or pain?"

He paused, went over what he thought he knew for a few moments, "Those aren't antonyms," he said.

"Who said they had to be?" The kid squatted down in front of him just as he was starting to sit up from the dirt. "They might as well be, anyway. Have you ever cared about pain as long as you've been empty, Panda? You're empty! You don't feel!" He giggled again.

The words slammed and burst against his chest like snowballs thrown against a window. It was fine. He could ignore it.

"Panda?" he asked, favoring the nickname over confrontation.

"Yes, I've heard they're very alone and irritable," the boy said.

"Oh."

He considered saying something more, maybe argue against the name, but he decided it wouldn't be of any use. The child's vivid yellow eyes focused on the shadow, always a couple of feet away.

"Anyway, you clearly haven't made up your mind, so I would suggest taking the white door." He paused, then suddenly lunged at the shadow and began chasing it in circles around him.

It went like this for a minute, the boy quickly getting bored of the game and coming to rest in front of him, the shadow cowering and humming rapidly behind his back.

The boy waved a hand, barely out of breath. “It’s comfortable there, and you can continue being alone and irritable without any interruption.”

He suddenly felt that he should be missing something right then. He tried to think of something to miss.

“You’re quite boring, you know,” the kid said, rolling his eyes when he didn’t respond. “Are you going?”

He nodded like it meant something. “Yes, I’ll— I’ll go there in a minute.”

“Time has no meaning here, Panda. But I get the idea.” Sending one last hungry glare to the shadow, the kid disappeared into the tree line.

He stared for a moment, then turned back to the path. He wasn’t alone.

A woman stood, tall and intimidating, with an air of uncaring that a mortal could never replicate.

“Hello,” she said.

“Hello,” he said, torn between caution and exhausted apathy.

“How are you?” she asked, but it did not seem like she was really interested in an answer.

“I’m fine,” The quiet clicking suddenly seemed a lot louder.

“You deserve a lot, don’t you think?”

“What?”

“You were so good. The world wasn’t enough for you. You deserve more, don’t you think?”

He blinked. “I— I’m not sure?”

“You should be sure, it’s a lot easier that way. Male lions sleep around twenty hours a day. Did you know that? Who have you met yet?” She did not look at him, her gaze lazily alternating between the shadow and the sky.

“Um, a small person in fancy clothes? They seemed a little lost,” he said. He did not mention the child.

The woman nodded. “Yes, a sure-fire way to make one stop panicking: present them with someone who seems more panicked than they are.”

“I wasn’t—”

“Sure, but many others are, quit thinking only of yourself. Do you know Swedish? I have this quaint little book and I can’t seem to—”

He was tired. “I’m sorry, what am I supposed to be doing?”

She seemed to take no offense. “Just walk right through there and go through the black door. Only if you don’t know Swedish, that is.”

He paused. “What then?”

She stared at the starless void. “You’ll get everything you deserve, and you’ll be perfect, and never not be perfect, and everyone around you will be perfect, and nothing will ever not be perfect.”

“Oh,” he said. He tried to follow her gaze and find something that was worth the attention. “Thanks—”

“Goodbye.”

He was alone for a while, probably.

He thought about the small person, the child, and the woman.

He thought about the doors.

He thought about the people that he didn’t mind but also didn’t miss.

He walked into the trees.

Undergrowth was crushed under his feet as he strode through. The shadow hummed as it wove between the plants. He quickly emerged from the treeline— the string of forest not much of anything at all— and into a small clearing with two massive and glowing doors in the center.

He walked forward.

Both doors looked like a piece of art, something that belonged in a museum to be misinterpreted and stained.

The one on the left was bright, he almost wanted sunglasses just to be able to inspect the details. It was white and flawed— little dents and rust that he tried not to focus on— and he wished he didn’t have to expect perfection in things that weren’t made to be perfect.

It was impossible to find any detail in the black door. All of the hinges and lines went indistinguishable in the consistent shade that wrapped around the entire shape. Even if there were any imperfections, he couldn’t see them. The grass flattened around his shoes as he stepped towards the dark frame.

A silver handle emerged from the blackness, maybe freezing to the touch as he reached out and gripped it.

There was almost no hesitation as he twisted—

“It’s a dreadful lot of pressure to be good, isn’t it.”

He looked up, surprised at the sentence that almost made sense.

A creature sat, legs dangling from the top of the black door.

His mind spun, remembering the pressures of life while living.

“Yeah, I mean, what if you don’t want to be good and perfect once in a while? Or what if you need to have a bad day?” he said.

The creature nodded encouragingly. “Yes, exactly.”

“I can’t believe I almost listened to him.”

It shrugged. “People tend to trust in the morals of a child.”

Almost instinctively, he moved towards the white door, immediately feeling warmer after releasing his grasp on the silver handle.

“Then again, it’s almost certainly better than being unhappy every day, isn’t it?” the creature drawled.

He stopped.

He backed up and sat down in the long grass, an equal distance apart from the doors. He looked at each of the doors, then at the creature.

“Why did they make the good door dark and the bad door light?” he asked, gesturing quotation marks at the overly simplified categorization.

“It was Their attempt at making it less stereotypical, I suppose.”

He hummed, then remembered the shadow. It floated beside him, still a couple of feet away.

“What is that thing?”

The creature tilted its head. “Do you not have clocks there?”

“Oh. No, we do, it’s just,” he paused, wondering what else he had been deceived by, “the child said that time has no meaning,”

“And yet, what does?”

There was nothing more to say. He stood, turned, and walked back through the treeline.

He sat on the path, tempted once more to lay down and close his eyes.

The shadow— no, the clock settled in his peripherals.

Soon, or maybe not soon at all, a figure became recognizable in the distance.

As they got closer, he recognized that they had their own shadow, something that had been missing from all of the others he'd met so far.

No matter, he knew that they were only here to persuade him of one way or the other.

"And what side are you on?" he asked once they got close enough.

The person did not stop, did not even look at him, and instead kept walking past, mouth moving to form silent words.

"Hello?" He stood up and started following after them and their clock.

He quickened his pace, coming to match their steps beside them.

"Dark, monotony, hot, cold—" they muttered rapidly under their breath, eyes glazed and fixed on the path below. "Perfect, panic, shadow—"

A sound rang out just then, a chime that echoed in the air and in his skull for a brief moment before disappearing.

He snapped his head around to stare at the two shadows, his own displaying a staticky number one.

The shadow next to his, behind his companion, showed a series of numbers and letters, too many to count, that shifted and changed like it had forgotten its place.

He grabbed their shoulder, forcing them to come to a stop. "Hey, do you need help finding the doors?"

They shook their head at the ground rapidly, seemingly fearful at the mere mention of the doors. "No, no, no doors, decisions, I haven't decided, dark, light, decisions—"

They shook his hand off and walked swiftly away, nonsensical murmuring fading as the distance between them grew, their clock humming as it trailed them.

He ran back to the doors, or maybe he was suddenly just there. The creature, still sat on top of the dark void frame, looked down on him with a smile tugging at its lips.

"Ah, you've found one then?"

"They just… stay?" he asked, a nauseating feeling settling in his stomach.

"Yes!" It said, sickeningly bright-eyed and gleeful.

His clock chirped excitedly beside him. *"Where are you going?"*

"See, I am dreadfully good at my job," The creature grinned wildly. "Most people don't end up choosing anything at all."

RECONSIDERING NORM CHANGE

THE EFFECT OF LAW'S SYMBOLIC FUNCTION ON SOCIAL TENSION

Adam Sonntag

The role of law in society extends far beyond its practical application and enforcement. It reaches into realms of culture, morality, and social order in apparent ways. At its core, law serves as a cornerstone for social organization and cohesion, providing a framework within which individuals and communities navigate their interactions and relationships. Through its codification and enforcement, law plays a vital role in maintaining order, resolving disputes, and promoting justice. Repeatedly, the legal reshaping of behavior has faced resistance from specific social groups, such as conservatives opposing the legalization of same-sex marriage. But why might this resistance occur, especially when the behavior does not affect the individuals resisting? In this paper, I argue that underlying tension in society is exposed and heightened by law's symbolic function when changing behavioral norms because it attacks individual and collective identities. This paper proceeds as follows: First, the concept of legal dominance is put forward and its role in leveraging morality as a means of increasing social group prestige. Next, the phenomenon of social norms and the process of their internalization is examined, providing a foundation for understanding how social norms become deeply ingrained in individual and collective identities. This establishes the backdrop against which legal changes may trigger heightened social tension. Within this framework, I show how legal domination can be viewed as an attack on a group's identity. Then, a plausible objection raised by Lawrence Lessig is addressed, revealing the limitations of his perspective and broadening the scope of the interplay between law's symbolic function and social tension. Finally, the conclusion draws everything together to explain that law's symbolic function exposes and heightens social tension. This paper mainly draws from the scholarship of Christina Bicchieri, Lawrence Lessig, Richard McAdams, Elizabeth Anderson, Richard Pildes, and Joseph Gusfield to help support its arguments.

Laws that influence compliance can provide a focal point for behavior. In their work, "Coordinating in the Shadow of the Law," legal scholars Richard McAdams and Janice Nadler argue that "in addition to sanctions and legitimacy, law can also influence

compliance simply by making one outcome salient."[1] According to their two studies simulating problems of coordination between two parties, when a third party made salient a previous coordination outcome to solve the issue, that outcome had a higher likelihood of being chosen merely because of its salience. What this means is that behavioral disputes have a tendency to resolve based on the "focal point" that the law offers. Thus, as well as from sanctions and legitimacy, behavioral compliance may also come from the law because it offers a salient solution for the parties to recognize.

This theory can be extended to morality. In this context, the focal point theory suggests that individuals are more likely to conform not only to behavioral compliance, but also to moral standards when there is a clear and widely recognized focal point of moral compliance.[2] The psychological findings of Albert Bandura et al. (1961), in their work, "Transmission of Aggression Through Imitation of Aggression Models," help support this claim. Taking a sample size of 72 young individuals, their work tested for imitative behavior from aggressive and non-aggressive adult models. Their results suggest that individuals learn and adopt behaviors by observing others, including moral exemplars.[3] This is especially the case for individuals who do not yet know how to act in particular situations, such as young children. Thus, moral exemplars (i.e., individuals whom others look to for guidance on how to align their own morals), ethical frameworks, or especially religious norms may all be examples of focal points that individuals confer with for making behavioral decisions. Hence, this will be referred to as the "moral focal point in law," but surely this extends to any respected authority beyond law.

The moral focal point in law may also be seen as the *dominant* morality. Whether or not it is the case that a moral belief *is* dominant, the morality that law symbolizes will be seen as such and will thus be followed by many. According to the sociologist Joseph Gusfield (1976), behavior accepted in public statements and actions has the highest likelihood of being considered the idealized normative standards.[4] In situations of customary behavior where a certain level of consensus exists regarding expectations, informal understandings

[1] Richard McAdams and Janice Nadler, "Coordinating in the Shadow of the Law: Two Contextualized Tests of the Focal Point Theory of Legal Compliance," *Law & Society Review* 42, no. 4 (2008): 865.
[2] I was inclined to argue here that gender inequality and other social issues were examples of this. For example, women themselves conform to many gender norms because, in one sense, it is seen as the dominant and proper morality. Gusfield's quote may elaborate this point more. However, there are also informal sanctions that come with disobeying gender norms, so I chose not to add this point.
[3] Albert Bandura, Dorthea Ross, and Sheila A. Ross, "Transmission of Aggression Through Imitation of Aggression Models," *Journal of Abnormal and Social Psychology* 63, no. 3 (1961): 582.
[4] Joseph Gusfield, *Symbolic Crusade: Status Politics and the American Temperance Movement* (Illinois: University of Illinois Press, 1976), 66.

allow individuals to establish norms that deviate from ideal standards without facing consequences.[5] Gusfield argues:

> Where the action or statement involves the total society, the ideals are likely to be the most common denominator, the safest ways to act because they are the ways least likely to be punishable. We are less apt to go wrong in public by being saintly than by being ourselves.[6]

Thus, if a social group desires to have many people accept *their* beliefs, they would do their best to implement them into the law because it is seen as a focal point to the person who does not know how to align his behavior best. There is reason to believe that immigrants will have the greatest tendency to do this since they will be the least socialized. To generalize this, an individual with a dim legal consciousness will obey the law to a greater extent than those who understand the norms that surround the law to diminish his risk of getting in trouble. However, groups do not simply implement their values into the law for moral edification, for this account would be too surface-level.

When groups implement their moral views into law, depending on the context, it is a moral victory over another group. One such vivid example is the temperance movement. In his book, "Symbolic Crusade: Status Politics and The American Temperance Movement," Joseph Gusfield argues that the temperance movement was a moral reform effort used to distinguish abstinent Protestants from other intemperate subcultures, legally conferring social status on both groups. In this way, what McAdams in another study describes as the *expressive-politics theory of law*,[7] the symbolic function of law allowed Protestants to reaffirm their moral superiority over the minority groups by obtaining a "victory" over the subcultures who had the supposed nonideal morality. Gusfield describes that,

> Even if the law is not enforced or enforceable, the symbolic import of its passage is important to the reformer. It settles the controversies between those who represent clashing cultures. The public support of one conception of morality at the expense of another enhances the prestige and self-esteem of the victors and degrades the culture of the losers.[8]

The meanings attached to law by the competing groups, on this account, are the symbols that explain why the law was adopted.[9] Thus, because the law is commonly taken as the dominant morality, a group may enhance their prestige and self-esteem by implementing their views in law. Groups seeking legal reform may, therefore, take a paternalistic attitude when attempting to change laws of social behavior. Moreover, there are other social

[5] Gusfield, *Symbolic Crusade*, 66.
[6] Gusfield, *Symbolic Crusade*, 66.
[7] Richard McAdams, *The Expressive Powers of Law: Theories and Limits* (Massachusetts: Harvard University Press, 2015), 13.
[8] Gusfield, *Symbolic Crusade*, 14.
[9] McAdams, *The Expressive Power of Law*, 14.

structures that law may influence when it changes, such as the social norms that implicitly undergird many people's behavior.

Christina Bicchieri, a philosopher and psychologist, has a comprehensive definition of social norms. In her book, "Norms in the Wild," she defines them as such:

> A social norm is a rule of behavior such that individuals prefer to conform to it on condition that they believe that (a) most people in their reference network conform to it (empirical expectation), and (b) that most people in their reference network believe they ought to conform to it (normative expectation).[10]

In this definition, there are many components that require further elaboration. For example, *preference* is the disposition to act in a particular way in a given situation.[11] *Social preferences*, by extension, take into account the behavior, beliefs, and outcomes that the decision-maker finds important from people in what Bicchieri calls their *reference network*, which she defines as the range of people whom we care about in the process of making particular decisions.[12] For example, I may have a (social) preference to consume alcohol around my friend because I do not want to disappoint her, despite myself usually abstaining from alcohol. My decision to consume alcohol thus considers her hopes for my conformity so as to not displease her. I would be considered a "social drinker" because I prefer to drink only under certain social conditions, i.e., when there is both a normative expectation and an empirical expectation. Consider, on the one hand, that if our social norms are contingent on pleasing others, then we will perform actions that are believed to be praiseworthy. On the other hand, we may also develop social norms of resistance, by which we would commit disgraceful actions, and so on. Notwithstanding the usefulness of this definition, I have concerns about whether social norms should not be limited to mere rules of *behavior*.

Social norms must be chiefly understood as *perception*. Indeed, our perception of the world is what helps govern our behavior, so it may be helpful to think of social norms of behavior as evidence for social norms of perception. The political scientist Tali Mendelberg (2022) examines how holding political office can be a status signal for socially stigmatized groups. Scheduled castes (SC, also known as Dalits) are an identity group through ancestry in the Indian caste system who face wide-scale discrimination despite affirmative action measures being taken. These measures took the shape of a 1993 mandated SC quotas for elected offices in village councils. It was inadvertently ineffective insofar as resource distribution for government benefits went, but Mendelberg argues that SCs did benefit in a

[10] Christina Bicchieri, *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms* (New York: Oxford University Press, 2017), 35.
[11] Bicchieri, *Norms in the Wild*, 6.
[12] Bicchieri, *Norms in the Wild*, 14.

distinct change in status such that they were recognized as respectable individuals. She writes:

> [S]eating stigmatized identity members in a high-status political office creates basic dignity. It does not do so by changing attitudes about the group's traits and behaviors, but instead by changing the perception of social norms that regulate social value. This change in norms of respect is an important form of politically driven status change.[13]

This case represents how political symbols can be effective at changing the status of a negatively-stereotyped group. Despite there being social norms for disregarding the needs of SCs, political representation shifted perceptual social norms to help SCs garner respect from their colleagues. In this way, by changing the perceptions of non-SC colleagues, their behavior toward SCs was improved to show them more respect. Thus, social norms of perception are important facets that undergird our behavior and how we understand the social world. When laws change social norms, they also alter how individuals perceive the world, as the once-dominant social norms accepted by society are no longer prevailing in the law.

Social norms can be internalized and no longer conditional on others. For Bicchieri, "[w]hen we say that a norm has been internalized, we often refer to the development of moral beliefs that correspond to societal standards. These beliefs become an independent motivation to conform, as deviations are often accompanied by guilt."[14] Hence, internalized social norms are the types of morals that are brought up through social factors[15] but are no longer conditional on other individuals. When social norms are no longer motivated by others within one's reference network, but instead are motivated because they are seen as *good in themselves*, this is a common human tendency to efficiently live out one's life without having to justify every behavior. In fact, during Medieval times, according to Norbert Elias (1978), when books were far harder to acquire, learning from precepts what one should and should not do in society by heart was one means for the upper class to impress on their memories proper behavior, despite not knowing the justification of that behavior.[16] Thus, the upper class would follow behavior because it was seen as inherently good and proper to do, according to the precepts they read. Instances like these—when the behaviors that stem from beliefs that were once socially conditioned but no longer are—cannot be called "social"

---

[13] Tali Mendelberg, "Status, Symbols, and Politics: A Theory of Symbolic Status Politics," *Russell Sage Foundation* 8, no. 6 (November 2022): 61.
[14] Bicchieri, *Norms in the Wild*, 32.
[15] I do not mean here that social norms are deterministically created through social factors. Instead, social factors have caused the occurrence of social norms.
[16] Norbert Elias, *The Civilization Process: The Development of Manners*, trans. Edmund Jephcott (New York: Urizen Books, 1978), 61.

norms. Instead of society, the feeling of guilt that one has for not acting in a particular way is what guides these internalized norms.

This internalization corresponds to the definition of *tying* by the American political activist Lawrence Lessig. He defines tying as a person's attempt to "transform the social meaning of one act by tying it to, or associating it with, another social meaning that conforms to the meaning that the [individual] wishes the managed act to have."[17] This type of semiotic alteration is what expands the boundaries of morality onto social actions for them to be internalized. For example, Bicchieri uses the example of female genital cutting as a norm that is embedded in ideals of purity and honor.[18] This embeddedness is the essence of tying a social action into an abstract concept of morality (e.g., honor and purity), such that the application of the said moral framework is seen in the commitment of that action. Thus, one may desire purity, so the act of female genital cutting bears evidence of their commitment. To give a better example, suppose that my social drinking appeals to the belief that it is moral to always have a clear conscience around others. This would, and does, *cause* my desire to abstain from alcohol consumption. It must be noted that this is not an instance of definition (i.e., that clear consciousness *means* to not drink) since not all drinking will make an individual unconscious—indeed, only excess will do this—but fear of being *guilty* will present itself when an instance that is tied to an abstract rule occurs. Thus, I will have tied my alcohol consumption to the values of being conscious (instead of, for instance, drinking for social reasons) so that I feel the violation of my morals through the action of consuming alcohol. In this way, I will not consume alcohol at all because I believe it is an immoral act in itself.

The development and internalization of etiquette and manners for eating meat have a similar effect. For example, the upper class during medieval times brought whole parts of an animal to their table for the "well-bred" man to carve it.[19] However, preparing a whole quarter of veal on the table was later seen as brutish, which led to the current practice of preparing it behind the scenes and away from the consumers.[20] This transition from bringing our food to the table to preparing it in the back is what Norbert Elias calls "the civilization process," where, in it, people seek to suppress in themselves characteristics that are seen as animalistic.[21] This historical shift highlights how the social norms surrounding

[17] Lawrence Lessig, "The Regulation of Social Meaning," *University of Chicago Law Review* 62, no. 3 (1995): 1009.
[18] Bicchieri, *Norms in the Wild*, 32
[19] Elias, *The Civilization Process*, 118.
[20] Indeed, one of the key strategies of animal activists is to show the slaughterhouses where the animals are prepared to draw the sympathy of consumers. By preparing our food in the back, there is far less emotional attachment to consuming the pieces of meat we eat.
[21] Elias, *The Civilization Process*, 120.

meat consumption have been tied to changing standards of behavior and morality. Initially linked to notions of refinement, the act of cutting meat at the table was later replaced by behind-the-scenes preparation. What this reflects is the changing attitudes toward disgust and unpleasantness associated with consuming a dead animal. Thus, the shift in behavior can be seen as a process of internalization, wherein the new social norm is less obviously tied to notions of unpleasantness but rather to convenience and efficiency in others preparing our food, showing the interplay between evolving moral standards and cultural practices.

One's morality has a tight connection with their identity. This connection is due to the fact that social identity often influences behavior. For example, the foundational structures of general knowledge that underlie our understanding of the world, referred to as *schemata* by Bicchieri, give rise to *scripts* (her term)—established patterns of action for specific situations.[22] A clearly substantial example of these schemata is morality, as it is a large source of reference for how to act in certain situations. Moreover, one's morality and social norms can be shaped by their social identity, and vice-versa. For example, one's gender, sexuality, family values, and roles may shape how one perceives themselves in marriage. According to Bicchieri, "[t]he marriage script will vary depending on specific traditions and economic and legal conditions, and will include beliefs, values, and social expectations that stem from the schemata."[23] Thus, on the one hand, social expectations shape how one identifies themselves and the particular roles they play, such as being a husband. On the other hand, an individual's identity, particularly their social identity, encompasses a range of experiences that structure their schemata, forming the basis for behavioral scripts in specific social situations. When individuals internalize social norms, it contributes significantly to their overall identity and influences how they behave within the confines of those norms. In this understanding, as individuals internalize and conform to societal expectations, they not only align their behavior with prevailing social norms but also shape their social identity in the process.

When a group seeks to change the law, they risk attacking many people's social identities. Going back to my purity example, if all of the practical examples tied to particular moral values were prohibited, one would likely manifest cynicism in the law for the guilt that they feel for breaking their morals of purity due to the law obstructing their cutting practices. The legal part is not what matters right now, but, to be broader, if it is the case that social norms are particularized to one's identity (for example, the scripts that an individual dispenses through their schemata may vary depending on the person), then

[22] Bicchieri, *Norms in the Wild*, 131-132.
[23] Bicchieri, *Norms in the Wild*, 133.

changing social norms by changing laws may be an assertion of dominance (as argued above) by asking individuals to change their understandings of the world and thus their identities. For this reason, the temperance movement targeted much more than just a *social practice*, but was also an assertion of dominance over an entire *group of people* who were considered intemperate.

Philosopher Elizabeth Anderson and legal scholar Richard Pildes (2000) put this attack in terms of law's expressive function. An expressive theory of *action* argues that our actions translate into the expressions of our rational and moral attitudes toward others.[24] Our justifications for action are regulated by expressive norms telling us whether particular means will attain desired ends. For example, someone may stomp their feet and throw a tantrum to express their disliking of a decision. Despite this, actions may deviate from what was intended, so an individual's actions are not always representative of their avowed purposes. Hence, their stomping will simply be seen as childish. Anderson and Pildes argue that "just as we can attribute purposes, beliefs, attitudes, and other states of mind to various plural subjects under certain conditions, we can do the same for political bodies."[25] Thus, by extension, the State can have a collective mental state when its members are all committed to expressing the given mental state. Expressive theories can thereby extend to considerations of the State because the actions that express the State's attitude are the creation of its laws.

The temperance movement, in their framework, was a communicative harm. Social relationships are created when communication establishes shared understandings about the attitudes that will govern future interactions. However, "communications can expressively harm people by creating or changing the social relationships in which the addressees stand to the communicator."[26] In this way, communicative harms occur when the State expresses attributes that create or abandon relationships with its citizens. Recall that what is expressed does not necessarily have to be the intent of the expresser. For this reason, state action is wrong, or unconstitutional at times, when the state expresses impermissible valuations of its citizens. In the case of the temperance movement, it was communicated to a group of people that their practices were illegal—those who consumed alcohol were no longer associated with the state's morality, and thus the relationship between the state and the people who consumed alcohol was devalued. In this way, the abandonment of social norms

[24] Elizabeth Anderson and Richard Pildes, "Expressive Theories of Law: A General Restatement," *University of Pennsylvania Law Review* 148, no. 5 (May 2000): 1508.
[25] Anderson and Pildes, "Expressive Theories of Law," 1526-1527.
[26] Anderson and Pildes, "Expressive Theories of Law," 1528.

by the state not only asserted dominance over a subgroup but also expressed its detachment from their relationship.

Lawrence Lessig objects to my understanding of social norms on account of his categorization of social norm change. One of his classifications for this is "efficiency norms."[27] These types of norms are changed through the measure of efficiency standards.[28] For example, we may change these if it results in either a Pareto superior social state—if it leaves everyone better off and no one worse off—or it results in a Kalder-Hicks social state—if the net value of leaving people better off is greater than the net value of leaving people worse off.[29] Applying this latter framework, Lessig offers the example of taxi drivers in Budapest who have a social norm of desiring their passengers to not wear their seat belts because it signals that they do not trust them as drivers, and thus passengers wearing seat belts attack drivers' pride.[30] He contends that it is satisfactory to change these norms on the account that one social norm (disrespect in wearing seat belts) is purely better than another social norm (wearing seatbelts for safety). He makes this plausible through the basic assumption that "the identity of individuals does not change as these constructions proceed, and hence, . . . it is coherent to speak of it being 'efficient' to change certain meanings."[31] Thus, taking a *functional*[32] *framework of social norms*, if one norm leads to a better social state overall, Lessig is willing to change them, regardless of whether it diminishes one's perception of themself.

In response to this objection, Lessig's analysis is weighted incorrectly. He interprets the norms of the driver to be merely a matter of pride that can be dismissed for the safety of the riders. Lessig assumes that the norms are a one-to-one tradeoff, and pride and safety are the only real considerations. He simply takes the situation to be that one person holds one view and the other holds another, so there must be a deciding factor—which Lessig takes to be the *safety* of the rider. Although this "efficiency" model coincidentally holds for the taxi driver case, I am agnostic in taking such an approach to other social norms. For if we limit our purviews to the assumption that social norms are all equally formed, the underlying

[27] Lessig offers a second classification called "distributional norms," which is based on social capital. I ignore this framework because it takes social identities and decides which norms to keep based on them. Although I do not disagree with this framework, it is not a solution to what I am attempting to get at. My argument, as will be shown, focuses on the intrinsic formulation of social norms and claims that this is a key consideration for norm change.
[28] Lessig, "The Regulation of Social Meaning," 1002.
[29] Lessig, "The Regulation of Social Meaning," 1002.
[30] Lessig, "The Regulation of Social Meaning," 1003.
[31] Lessig, "The Regulation of Social Meaning," 1003.
[32] I avoid using the term "functionalist" here because it typically refers to theoretical approaches in philosophy or sociology that emphasize the functions and roles of mental or social phenomena. What I mean here instead is an approach to social norms that *only* considers how they function in society.

tensions that lie beneath social norms are ignored. The temperance movement is a key example here. McAdams puts this shortcoming in terms of a mere causal theory, such that it might have been understood as Americans adopting prohibition to alleviate the social and medical problems of alcoholism.[33] However, as Gusfield shows, the temperance movement was actually a symbolic attack against a subordinated social group's identity. What Lessig has ignored in his analysis, then, is that when we change laws that go against one's social norms, the law may attack one's social identity, which was the case in the temperance movement. Hence, I contrast Lessig's conception with my own *dynamic framework of social norms*, one that considers both the function *and* formation of social norms to be equally important considerations for norm change.[34] Indeed, as Joseph Gusfield argues, "[w]e have always understood the desire to defend fortune. We should also understand the desire to defend respect."[35] What is not realized is that individuals will fight for social status (formation) more often than for material wealth (function).[36] Lessig assumes that social norms in themselves are all formed equally to each other and that only their function is what varies, which is simply not the case.

Law's expressive function exposes underlying social tension. When the law changes social behavior, it attacks the identities of individuals because social norms are so intricately related to the construction of one's social identity. When identity is attacked, individuals will defend themselves. Sociologist and political activist Stuart Hall (2018) puts it thus:

> [I]dentity is a source of agency in action. It is impossible for people to work and move and struggle and survive without investing something of themselves, of who they are, in their practices and activities, and building some shared project with others, around which collective social identities can cohere. This is precisely because, historically, there has been an enormous waning and weakening in the given collective identities of the past—of class and tribe and race and ethnic group and so on, precisely because the world has now become more pluralistic, more open-ended, though of course those collective identities have not disappeared in any sense.[37]

When individuals feel that their identities are being attacked or marginalized, they recognize the need to invest themselves in their practices and actions as a means of defending and affirming their identities.

---

[33] McAdams, *The Expressive Power of Law*, 14.

[34] I must note here that despite the function and formation of norms being equally important considerations, they will hardly ever be equally important weights. Lessig's taxi example only works because the function of the social norm weighed much more than the formation of it. However, the formation of the social norm to consume alcohol weighed msore than its function in the temperance movement example.

[35] Gusfield, *Symbolic Crusade*, 11

[36] Mendelberg, "Status, Symbols, and Politics," 50.

[37] Stuart Hall, *Essential Essays, Volume 2: Identity and Diaspora* (North Carolina: Duke University Press, 2018), chap. 10, doc. 314-315, https://doi.org/10.1215/9781478002710-016.

It is not the case that the law's expressive function *creates* tension, but instead *heightens* the current underlying social tension that already exists. For example, when the question of legalizing same-sex marriage arose in the United States, the expressive function of law came into play. On the one hand, the decision in *Obergefell v. Hodges* to legalize same-sex marriage was a visible and public expression of societal values and norms, affirming and recognizing the rights and dignity of homosexuals. On the other hand, after *Obergefell* was decided, social tension heightened in states like Texas[38] and Arkansas,[39] and even from organizations such as the Alliance Defending Freedom.[40] The challenges around marriage norms, therefore, illustrate how legal decisions can both reflect and amplify existing social tensions, as they bring deeply held beliefs and values to the forefront of public discourse and generate contentious debates about the scope of individual rights and the role of the state. Because marriage is so intricately intertwined with identity, my analysis explains this example in terms of the law affecting the identities of both groups. Traditionalists reject same-sex marriage because it strikes at the core definition of what marriage is supposed to be, which is what they identify with.

In conclusion, law's symbolic function effectively exposes and heightens underlying social tension due to its inherent connection to behavior, which is intricately tied to individual identity. The law shapes and regulates our actions, and any change to legal frameworks has the potential to disrupt established patterns of behavior and challenge deeply ingrained social identities. For this reason, I emphasize that when changing laws that affect social behavior, one must be wary of the social identities that undergird individuals' perceptions and responses. Social identities, such as gender, race, religion, and sexual orientation, inform how people navigate and interpret the world around them. These identities are often deeply intertwined with personal values, cultural norms, and historical contexts. Similarly, reform activists face the imperative of thoroughly understanding the broader context surrounding the social norms they endeavor to change. Failure to do so risks encountering formidable resistance from conservative groups with opposing viewpoints.

[38] "US Gay Marriage: Texas Pushes Back against Ruling," *BBC News*, June 29, 2015, https://www.bbc.com/news/world-us-canada-33314220.
[39] Anthony Zurcher, "US Gay Marriage: Reaction to Ruling," *BBC News*, June 26, 2015, https://www.bbc.com/news/world-us-canada-33292805.
[40] Robert Barnes, "Supreme Court Rules Gay Couples Nationwide Have a Right to Marry," *The Washington Post*, June 26, 2015, https://www.washingtonpost.com/politics/gay-marriage-and-other-major-rulings-at-the-supreme-court/2015/06/25/ef75a120-1b6d-11e5-bd7f-4611a60dd8e5_story.html.

Bibliography

Anderson, Elizabeth, and Pildes, Richard. "Expressive Theories of Law: A General Restatement," *University of Pennsylvania Law Review* 148, no. 5 (May 2000): 1503-1576.

Bandura, Albert, Ross, Dorthea, and Ross, Sheila A. "Transmission of Aggression Through Imitation of Aggression Models," *Journal of Abnormal and Social Psychology* 63, no. 3 (1961): 575-582.

Barnes, Robert. "Supreme Court Rules Gay Couples Nationwide Have a Right to Marry," *The Washington Post*, (June 2015), https://www.washingtonpost.com/politics/gay-marriage-and-other-major-rulings-at-the-supreme-court/2015/06/25/ef75a120-1b6d-11e5-bd7f-4611a60dd8e5_story.html.

Christina Bicchieri, *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. New York: Oxford University Press, 2017.

Elias, Norbert. *The Civilization Process: The Development of Manners*, trans. Edmund Jephcott. New York: Urizen Books, 1978.

Hall, Stuart. *Essential Essays, Volume 2: Identity and Diaspora*. North Carolina: Duke University Press, 2018, chap. 10, doc. 304-323, https://doi.org/10.1215/9781478002710-016.

Joseph Gusfield, Symbolic Crusade: *Status Politics and the American Temperance Movement*. Illinois: University of Illinois Press, 1976.

Lessig, Lawrence. "The Regulation of Social Meaning," *University of Chicago Law Review* 62, no. 3 (1995): 943-1046.

McAdams, Richard. *The Expressive Powers of Law: Theories and Limits*. Massachusetts: Harvard University Press, 2015.

McAdams, Richard, and Nadler, Janice. "Coordinating in the Shadow of the Law: Two Contextualized Tests of the Focal Point Theory of Legal Compliance," *Law & Society Review* 42, no. 4 (2008): 865-898.

Mendelberg, Tali. "Status, Symbols, and Politics: A Theory of Symbolic Status Politics," *Russell Sage Foundation* 8, no. 6 (November 2022): 50-68.

"US Gay Marriage: Texas Pushes Back against Ruling," *BBC News*, June 29, 2015, https://www.bbc.com/news/world-us-canada-33314220.

Zurcher, Anthony, "US Gay Marriage: Reaction to Ruling," *BBC News*, June 26, 2015, https://www.bbc.com/news/world-us-canada-33292805.

CARE ETHICS IN THE AGE OF AI

AN INTERVIEW WITH DR. PETER ASARO

Andrew Shaw and Molly Banks

The Garden of Ideas was very fortunate to be able to host a live interview with Dr. Peter Asaro on November 30, 2023, conducted by editors Andrew Shaw and Molly Banks. Dr. Asaro is a leading philosopher of science, technology, and media and currently a visiting scholar at the UW Center for an Informed Public. His research focuses on social, cultural, political, legal, and ethical dimensions of automation and autonomous technologies. The transcript below has been edited for print.

Molly: We had a lot of interest in your paper "AI Ethics and Predictive Policing: From Models of Threat to Ethics of Care," and in that paper you explored the theoretical and practical advantages of abandoning the more common models of threat approach in favor of the more holistic ethics of care approach. What are the limitations of a models of threat approach, specifically in the context of predictive policing?

Dr. Asaro: Maybe it's also helpful to describe what the models of threat is, and what I was trying to do by characterizing a whole set of practices that I want to put under this umbrella "models of threat." This is a very utilitarian approach that you see in a lot of different domains. At the UN level, you have international relations, national security threats, international threats and analysis. Looking at policing, there's various kinds of threats. There's a longer history around threat modeling within cyber security, it goes back to the early 2000s. Within that, there's this engineering mentality/utilitarian mentality of "identify the data and fix the problem. If the problem is too many errors, let's reduce the number of errors. If we don't have the right data, let's get more data."

This particular paper is from 2019, so there's a historic moment at which a lot of work was starting to come out about data and algorithmic bias, and ways in which learning systems or machine learning had all this intrinsic, implicit racial bias. So the reaction within the tech community is, "let's just debias the data. We know that there's bias in it, we're just going to come up with some fancy computational techniques to fix it." I think it misses the point. You missed the forest for the trees by focusing on trying to make data more accurate,

more precise. There was actually, at the FAccT (Fairness, Accountability, and Transparency) conference, this paper on 32 different definitions of fairness that might be applied. What counts as fair is debatable, philosophically. So it didn't seem like that was the right approach, but that seemed to be where all the conversation, all the resources were moving in AI ethics.

In this paper, I really want to critique that, take a step back, and think about, "how else can we think about AI ethics that's not just utilitarianism, not just error reduction or debiasing?" I turned to feminist theory and the ethics of care, and then within the paper, I do a comparative analysis of two applications of data-driven policing that happened contemporaneously in the City of Chicago, where gun violence was extremely bad in the mid-decade. These two different programs both used data to try to identify who was at risk for gun violence, but then did radically different things with it. One is looking at it as threats. People who are likely to be involved in gun violence represent this threat, so we can police them more intensively. There's several very elaborate statistical analyses of the results, and they found it did not work at all: zero impact on gun violence for those people who were identified in that system. That system was the SSL (Statistical Suspect List), or hot list.

This other program identified at-risk youth by looking at high schools that were in the most gun violence-prone and the lowest socioeconomic status neighborhoods within the city, did a review of applications from those students, but then gave them summer jobs. Instead of what the police did with this SSL list, [where] they showed up at the house of people and threatened them. Or when there was an incident in the neighborhood, they would use that to generate a list of people with high ratings on this list and round them up for questioning just because they were proximate. But the one summer the City of Chicago gave them jobs with mostly community organizations, [it] basically reduced gun violence for the initial class by 51%—massive reduction for any kind of policy intervention. It didn't really depend on making the data more accurate. It wasn't, "if we collect a bunch more data, debias this data, or use fancier statistical methods, we're going to get better at predictive policing." It's that they had a better plan for how to integrate that data into policies and, particularly, better policies about how to intervene to prevent violence. So this represents this idea of the ethics of care.

What you really have to think about, particularly in AI ethics, is how these systems are embedded within sociotechnical systems. It's not just the statistics that matter. It's the social structures in which they're embedded, and in this case, how police use it. How do you train police to understand how it works and what it does? They got no training and no instruction on what it really did. They're just told, "this will generate lists of likely suspects," which actually wasn't true, because it lumped together people who might be victims with people who might be perpetrators, but they just treated everybody as potential perpetrators. But I

think it applies more generally not just to policing but to many different things where we want to apply technology. We think, “Here's a threat. Here's a risk. How do we minimize that?” and then apply tons of computational power and data analysis to doing that, rather than thinking about, “what are the social implications, and how do we actually reframe the way people think about the world?” We're building this technology, which only has the capability of identifying threats, and that gives you only a set of actions to take that are a response to threat, which is aggression and arrest. Subsequently, with George Floyd, we've become much more aware that there are other ways to do policing, and police don't necessarily need to use violence to solve a problem. In many cases, people are psychologically disturbed, and they need psychological interventions rather than a police intervention or a force. If you have a hammer, everything looks like a nail. If we're given data about threats and given the tools to deal with threats, then we’re going to treat everything like a threat.

Molly: Thank you! Could you briefly touch on how you got to ethics of care as a framework? What makes an ethics of care framework particularly well-suited to replace models of threat as an approach, particularly in predictive policing?

Dr. Asaro: First, it's relational, so it's not just a single dimension of metrics. We're trying to improve a particular trajectory, action, outcome, or efficiency on one kind of dimension, but we have to think holistically, and that's challenging. It's very easy to say, “this metric is too low, we’ll need to improve it.” But we really need to think about all these things that we normally don't think, “what are these implications of making a certain transformation in a technology or putting a technology into a different kind of piece of society?”

Aimee van Wynsberghe, who works in robot ethics, has looked at ethics of care with her dissertation in care robots. You can think about the operations of the hospital, and you're trying to maximize delivery of care, improve patient outcomes, and you can measure that in various ways. But when you start trying to maximize efficiency in those parameters, you miss the point that a lot of what happens in care work is care, and that it's not just how much medicine you get and how accurate that is, but that there's bedside manner. There's treating people like human beings and respecting them and their humanity. Nurses and doctors know certain ways that computer programmers don't. We can start thinking about that when we do design from the very beginning, and then wind up with better systems if we do. Especially in policing, or any other kind of care, education, medicine, a duty of care [is] expected. Teachers are expected to take care of students and doctors take care of patients. Lawyers also have duties of care to their clients. It's often very difficult to articulate, but it's an always-present moral duty or obligation.

Some people talk about ethics of care as a virtue ethics. I think that works if you think, "what would the virtuous caretaker do in a certain situation?" There's also community ethics, look at what benefits communities. Most Western ethics is very individualistic, so that's problematic because we're building social technologies, not individualistic technologies. The whole history of engineering ethics is about the moral responsibility of an individual engineer: build a reliable system, not approve things that aren't safe. But actually, it's not just them and their moral character that matters. It's the whole society or sub-community that's impacted by a system.

So Western philosophy isn't very good at that at all. Other philosophies [are]. Ubuntu and African philosophy is very powerful. I was just at the Social Studies Science Conference in Hawaii, and they have a very powerful indigenous philosophy, which is one of abundance. I think it comes to this model of threat, which is Western philosophy. You can just read right out of *The Republic*: "all these other city states are trying to invade us and steal our stuff, we need an army, we need a police force, and we need to invade, steal their stuff." That's the basis of thinking about interrelations, whereas in Hawaiian philosophy it's that we live on an island of abundance. As long as we take care of the island and we take care of each other, there will always be abundance, instead of thinking about it all as scarcity. Models of threat was about scarcity, and if that's the foundation of your philosophy, then you're always going to [think], "what's mine?" versus "how do I care for others?" If we're all in this moment kind of caring for each other, we're probably better.

Andrew: In addition to your work on predictive policing, you're also very well known for your work on lethal autonomous weapons. In light of our discussion of models of threat and ethics of care, in what ways are your work on predictive policing and lethal autonomous weapons connected? More broadly, how are the development of both technologies connected in a material and a philosophical sense?

Dr. Asaro: I think the obvious thing is the connection between the potentials for violence and weapons. I have another set of papers on police robots, and particularly armed police robots, and why those are a terrible idea. All of the justifications that we give to police officers to use violent force or lethal force aren't really acceptable at all for robots, because mostly it's about self-defense, and that robots don't have right to self-defense, because they're not selves. But even in defense of others, it doesn't make sense in most of these situations where you're trying to protect the third party to actually introduce a lethal or armed robot in the situation. A threat [is] the intention to do harm to somebody, as well as the capacity to do harm to somebody. So if I have a weapon, and I'm using it in a threatening enough manner, that creates a threat. But a robot would have to both understand enough

physics—not just recognize an image of a gun on camera, because guns could just be lying on a table—and then also understand social psychology and our actions enough to understand this person is threatening this other person.

So they would need a very robust understanding of the physical world, and the social world, and if they have that, then they also should have all sorts of other ways of intervening on that. And this also goes back to the models of threat because we justify or permit police to use lethal force with a gun in all kinds of situations where there's probably other options to de-escalate. But because their lives are at risk, we've written laws that say it's okay for them to use lethal force with very low standards of "they just have to feel threatened." That means they go straight to this tool or option that is lethal in its first instance of use, like a gun. Whereas, if you understand as a robot or as a person the complex social interaction, the complex physical interaction, you could intervene, you could de-escalate socially, talk somebody down, convince them not to use force. Or you're a robot: they have a gun, but you could put yourself in front of the gun. You can interact with the physical world or the psychology and social interaction to remove the threat. You should try all of those things before you try lethal force, but there's an expedience to the gun, so that winds up being socially permissible. But we shouldn't transfer all those morals for humans onto machines because they're not humans.

And similarly, with military things, it's an extreme case because there's a lot of violence that's permissible, including civilian casualties as long as they're not intentional, which is also rather broadly construed and difficult to enforce. But ultimately from an ethical perspective, and even in just war theory, the justification of killing is highly exceptional. You can only kill enemy combatants who pose a threat and are still fighting, and if they surrender you can no longer kill them. If they're injured and can no longer fight, it's illegal to go and execute them.

So it's not just *carte blanche*, and you can't just kill your fellow soldiers, that's still murder. There's actually a lot of rational justification that needs to be in place before killing is permissible. Robots, automated systems, just don't have any access to that. They're not moral agents, they're not legal agents. They cannot justify in making a choice to kill, and it's impermissible morally for us to delegate that kind of decision to them, because we're abdicating our moral responsibility by allowing the machine to make those decisions.

So that's part of that connection, and it's related now to my work here at the Center for an Informed Public. I'm looking not just at violence and threats of violence, but deception and coercion in AI systems and chatbots, and how manipulating you through selected information also undermines your autonomy as a human being. We shouldn't allow machines to do that. How do we actually regulate, I think, is a much harder problem. I

thought it would be really easy just to ban killer robots, because it's kind of obvious, but it's been thirteen years and we're getting pretty close now. But hopefully, we can catch up with these chatbots.

Andrew: You suggested in your predictive policing paper that we should be shifting to an ethics of care approach in the design and the implementation of these technologies. But a particular focus of care ethics literature, as you mentioned, has been an emphasis on human relationality and caring for and about others. What are the limits of an ethics of care and applying that to these technologies? Is it even possible to encode or apply an ethics of care to lethal autonomous weapons? Or is there a deeper moral distancing that's inherent in their design?

Dr. Asaro: Short answer, no. I don't see killer robots enacting an ethics of care, and I think that's why they should be banned. I don't think there's any ethical standard that we could program into them under any kind of ethical system that will make it acceptable. And so we should just prohibit [them]. I think that the bigger question is, how do you implement this in or with systems. And I think that's a more challenging and important question, because it's more about humans deciding systems: their moral approach to that, as well as the evaluation of that system, the ability to revise and change systems when they're shown to cause harm. All those mechanisms need to be in place, but also from the initial design, start thinking about those things.

And it's not that we're trying to create an autonomous agent that's going to care. I think that's maybe feasible when we have superintelligence: they become moral agents, deserving of respect and part of our society. But right now they're tools and we treat them as tools, but they're tools through which we interact with each other. I'm a media professor. Everything is media. We're mediating our experiences and our relations with each other through all kinds of different technologies. And this is a new, very powerful technology that uses data in complex ways. But fundamentally, it's a tool. And what we need to be thinking about is how we're caring for others through the design of it, through the use of it, through controlling the kinds of data that are in it, as well as setting up this kind of regulatory procedures and mechanisms, laws to ensure that they're doing what they're supposed to do, and are making society better hopefully for everybody, and not just a select few.

So all of these questions about participatory design, I think, are highly relevant. But they're also a little bit misleading in the sense that I also don't think engineers and designers alone have all the responsibility for what these systems do. They're incredibly complex. Even a very simple technology, putting it out into the world, you don't know what people are going to do with it. You try to make it safe, and you try to show them how to use it to

benefit society. Ultimately you have to rely on society, rely on users to do good things with it, but you also still have some degree of responsibility.

Andrew: You seem to be suggesting that the similarity between predictive policing and lethal autonomous weapons, and their tension with care ethics, is a result of their nature as weaponized tools. To what extent is this tension with care ethics specific to weaponized applications of AI like autonomous weapons, as opposed to paradigms of categorization more broadly?

Dr. Asaro: I think it is very general, hopefully. That was trying to plant a seed for other people to think about how to apply it to other domains. But I think there's obvious connections and ways to do that within weapons. And in general, the way that we think about international security or national security, or even policing. We tend to fixate on threats and not necessarily the underlying problems. I hear this a lot about killer robots: "wouldn't it be better if robots fought the wars because they wouldn't make any mistakes?" No, not really. They're going to be a lot more efficient at doing things, but that's also going to make it much more likely to go to war because leaders are going to think they're really reliable. You've told them they have this really great targeting AI system that's not going to kill any civilians, which isn't possible. And then you say, we were totally responsible because we put out these things that were designed not to kill civilians, but they killed a bunch of civilians, so that's not our fault. It provides rationale and justification for it, and it's also a way of avoiding dealing with underlying problems: the political issues of the war but also the responsibility to train soldiers or police officers for de-escalation. We get fixated on this one form of lethal force or law enforcement, when a lot of what soldiers do is community relations: digging wells, helping reconstruction. Police do a lot of community engagement, making people feel safe in their community, if they're doing their jobs well. If you fixate on threats, this whole system of militarization is bled into policing.

Andrew: Your point about lethal autonomous weapons licensing violence relates to another paper of yours where you talk about the relationship between lethal autonomous weapons and totalitarianism. Does the mere acquisition or development of lethal autonomous weapons imply a shift to this more totalitarian form of power? In other words, is it a contradiction to speak of a democratic government that has taken up the use of lethal autonomous weapons?

Dr. Asaro: When we think about technology in general, and AI and automation technologies in particular, a lot of what they're doing—you're increasing automation, you're increasing

efficiency, but you're also redistributing power, and a lot of that has to do with labor. We've had authoritarian and totalitarian regimes for centuries, millennia, but they've always required people. A leader on their own, if nobody follows them, is actually not very powerful. Where power comes from is in training all of these people to do what you say and believe that your authority is real. Hannah Arendt has written about this in *On Violence*, thinking about totalitarianism and particularly police violence, mostly reacting to the student demonstrations during the Vietnam War. If authoritarian rulers had killer robots, they would have this tremendous new political power, because right now, as we understand authoritarian regimes, you need secret police, thugs, informants, and a surveillance system in order to be an authoritarian ruler which means you can ignore public interest or public opinion.

But if you can automate that, you can reduce the class of the police or the number of elites that you have to have around you in order to maintain power. I think what we're seeing now with these mega-billionaires wielding enormous amounts of economic power and media power and political power [is] that this automation is also going to enable ever greater distances of inequality, but also concentrated power in a smaller and smaller number of hands, which is, I think, fundamentally anti-democratic. Not to say that the secret police of a traditional authoritarian regime are super democratic, but even Stalin had to appease a certain number of elites to stay in power. Now, I think these technologies will reduce the number of people in those circles.

Molly: I'd really love it if you could dive in a little bit deeper to your current research and what you're doing with the Center for an Informed Public, with AI and chatbots increasingly affecting our visual information ecosystems and social media. You've explored ethical concerns regarding these algorithms that strategically manipulate users through targeted content. What ethical issues do you foresee? What impact on our autonomy do you foresee with new technology, like eye tracking technology or generative AI, and how those technologies are designed to improve the persuasiveness of targeted content?

Dr. Asaro: Sure, lots to cover under all of that, but let me just give the highlights. Part of it relates to something that's in that predictive policing paper, because I talked a little bit about the concept of pre-crime. If we think about targeted marketing and advertising, and how it functions, it's population statistics, essentially. We gain certain pieces of information about you as an individual, we map you into this demographic population model, knowing a few features that define you, and we can predict lots of other features of you, or things that we might be able to sell you. For political interests, knowing who your friends are and how you feel about the certain set of issues to project you, they know how to target you with different

types of political messages. This is very powerful compared to traditional modes of advertising persuasion where you're really trying to come up with much more general kinds of messaging for mass communication, because you're sending out a single message to everybody, or maybe you would have a certain subsets where you knew you could get a certain kind of target demographic. But now, you can address an audience of one and that can be very powerful and needs to be regulated. A lot of that's going to depend on privacy regulation and making sure that companies either don't acquire the kind of data that's needed to do that, or if they have it, that they're not permitted to use it in certain kinds of ways to manipulate people.

But I'm also worried about what's next, because historically, the reason mass communication worked the way it did is because it didn't have access to all that data. You have really pathetic psychological models, and it makes lots of wrong assumptions about you. They're actually really bad at it, and they don't really try to build an individualized model of your psychology and what you desire and hope for. They're just still fitting you into a one-size-fits-all population statistic. But now they're collecting so much data they could build models about you and figure out what you care about and who you care about and who you listen to, and they could fake messages from those people or convince you those people believe these things to get you to believe something, or threaten those people and try to coerce you into all kinds of things, and really start manipulating your understanding of the world in a highly customized way. The potential is there for it, because now we have data and the computational power. They just don't have models, but they could start building them and improve them over time. I think that's incredibly dangerous to think about all the different range of applications that might apply.

Particularly within democracy, it's challenging because we actually value persuasion and public discourse. And that's a lot of what's happening in the debates on social media. Freedom of speech is a good thing. But it's also pretty obvious that, and it has been for a long time, that speech is not equal, and certain people on those platforms have incredible power over others. You can look at the number of followers as roughly equal to the amount of power. Power announces itself in that way, but it's also real that they can attract all these people, and they can get their followers to exercise very complex types of social threats. It's kind of new, also not that new. Go back to *The Republic* and Plato [says] the problem with the public square is rhetoric. People should use logic, not rhetoric. Rhetoric was just persuading people using made up arguments. That's not right, we need truth. So we've lost truth in a lot of ways, an epistemic grounding and reliability of our communication systems. I don't know how we restore that, but I think that's going to be crucial. But a lot of that just depends on the public, and if we're constantly just manipulating everybody going forward, how do you get to some system where you can trust it? That's complicated.

Molly: In the context of democracy, we spend so much time on systems that are ruled by algorithms and then take that into our worldview, our identities, and our belief systems. Looking back to previous election years and the immense swaths of data that were collected by Cambridge Analytica, how do you see things like generative AI impacting the future of our democracy and the direction that it might take?

Dr. Asaro: There's a sense in which these generative AIs create things that are plausibly human. They're useful for that, because you don't need a real writer, you can just give simple instructions. But I think a lot of political persuasion at this point, and even manipulation, really depends on having some kind of insights about society and politics, and at least insofar as you give the system a prompt. Maybe you're generating better messages that way, at least initially. Current generations of these chatbots, I don't think, are going to be super useful any more than sock puppets have been for getting your social media to trend and get things in front of people. But you could just make up a few messages and just replicate them everywhere and get them trending, then that's the key to reaching people.

I think it's these more sophisticated models that are going to be much scarier. And again, Cambridge Analytica—for all of the pomp and circumstance that they claimed this powerful mode—they gave people these really basic personality questionnaires, and they mostly sucked up all of their data from Facebook. What they were doing, as far as I can tell, was identifying persuadable voters in swing states, and that was their value add. It wasn't that they really knew how to convince those people of anything, but they said, "these are the 50,000 people in Michigan that you need to send targeted ads to, because everybody else in Michigan's already made up their mind." So it has that power, and it can persuade enough people to vote in a close place that matters, but in a broad sense they didn't convince the whole country of anything, and they probably didn't even identify what psychological factors would be influential on those people, merely that they're the most likely undecided voters that could be persuaded in some way. But that could get a lot better in the future because they had really crap psychological models, and they didn't really know how to do any of that, but they didn't need to.

A lot of what AI gets applied to is fast, cheap solutions. Not necessarily cheap—you have to do a lot of computation—but it's fast, and it's cheap in terms of labor. All these automation systems do is reduce the cost to be able to do something. So it takes a long time to learn to paint, but now you can just tell a computer program "paint me a picture that looks like this or that" and it spits something out. The other thing from a purely information theoretic perspective: the level of information in a message is inversely proportional to the likelihood of receiving the message. So a message that you're expecting carries very little information. It's a message that you're not expecting that actually has a lot of information. But what these

systems are literally designed to do is generate the next most likely token or word, so it's actually providing the least informative thing, mathematically speaking, that it can at every instance—they're generic generators. So it can be creative, it can be unexpected because you don't know how it works, but it's just generating the very most likely thing that it can, which in that sense, it's not going to be creative. It's just going to find latent connections between data at best, which can be really useful, because there's a lot of data. But I don't think it's going to be super brilliant anytime soon.

Andrew: It's interesting because like you said, on one hand these large language models are designed to produce expected results, but their emergence has also been unexpected in many ways and has caused people to raise questions about the nature of consciousness. What do you think that these technologies reveal about the nature of human relationality, to bring it back to our discussion of care ethics? And do you think that they necessitate that we rethink any assumptions about human nature or about care?

Dr. Asaro: No? Well, yes, of course. I think if we go back to this idea I introduced earlier about sociotechnical systems, what was really innovative about ChatGPT is not some massive technical innovation, it's that they put a really nice graphical user interface around it. And the thing about Altman is not that he has some brilliant insight into language or AI, he's a really good marketer. He's the Steve Jobs of AI. And Steve Jobs didn't really have any great ideas. He went to Xerox PARC, and they had great ideas. So then he figured out a product that everybody could interface with really effectively and used things that were already out. But that's important, because really, technologies are sociotechnical systems. So you need these marketing people to promote the social side and understand how the technology can effectively integrate into society, and how to convince businesses that they need it and sell it and make lots of money. The actual technology hasn't really changed much in these large language models. They do really brilliant things, and we're all very impressed by them, because now we can actually interface with them in certain ways. How long that enchantment lasts remains to be seen. We already know they hallucinate. They're terrible at rules: they can't do basic arithmetic, but also rules that we might care about like logic, they can't do causal reasoning, they're not going to learn ethical rules.

Now, there's a degree to which they're modeling all these statistical patterns within written language that has been scraped and fed into the machine. What this really is is a giant compact statistical model of all the stuff that gets put into it—that's all a neural network is—but it allows you to access it really, really, quickly to answer queries or to generate text. This idea of predicting the next token as a for generative AI is really fascinating, and it tells you all kinds of weird and interesting things about the texts that have been put into it. I don't

know what it tells us about us other than we're the people collectively who generated at some point all those texts.

I think when we start talking about consciousness, it kind of irks me because it's nothing like consciousness. It's not even trying to be. Some people argue if it were just embodied and engaging in the physical world, then it would just learn all of that really fast, and then it would be conscious. But the first thing is you can't do that with robots. Robots and AI are very different: robots are much harder to program because there's so many more bugs—they're a nightmare, don't go into robotics unless you really love robotics. AI is way easier, because everything is just data, and it's so much faster to fix bugs, to do iterations. AlphaGo, DeepMind's Go playing computer, is playing trillions of games of go not only against every known recorded game of Go, but also against all of itself as an adversarial network trillions of times to develop the kind of skill it needs to beat the human. You can run those simulations trillions of times. Run a robot around this room a trillion times. How long is that going to take? The sun is going to explode before you finish that. And that's just this room, much less a robot that could deal with the world outside. I don't see that happening. I think embodiment is crucial to consciousness.

And actually, consciousness is cheaper in a lot of ways, or easier than intelligence, as we understand it in our language. Goldfish are conscious, right? They interact with their world in a conscious way. They're not going to write literature anytime soon. They don't need to. They just need to swim around, find food, reproduce, because they're fish. So that sounds like consciousness is more about a relationality of being able to understand your environment and relate to it. And we do have systems that are getting something like that, with SLAM (Simultaneous Localization And Mapping) in robots and drones and self-driving cars, that are starting to look like they can perceive a three-dimensional world, interact with it, and understand their relation to it. But it's still very limited and very brittle, and they're nowhere near as conscious in that sense as a goldfish at this point. And even if they achieve goldfish consciousness, they're not going to take over the world. We're not worried about goldfish taking over, right? Being able to interact socially or politically is so far off.

What we're worried about in morality or ethics is, "should these things have rights?" I think that comes to questions about the conditions of having rights and participating in society. That's about both having responsibility and the moral and legal responsibilities of being a member of society that you then incur respect from other members of society as equals, in some sense. They would have to do a lot more than have consciousness than even superintelligence to have what's required for that. They would have to be members of the society in the right way. And maybe if some alien superintelligence and for outer space,

instead of a computer, we wouldn't automatically think it's a part of society. We might fear and respect it because it's an alien intelligence.

But I think we also anthropomorphize all of this stuff way too much, and thinking that it's thinking, that it feels anything, that it's emoting, that it's anthropomorphic in some sense— or other fears around superintelligence, that it's going to take over the world and enslave us—we're projecting like how we behave towards other people. We're afraid of those things, and so we think the system is going to be like us and do this to us. But again, if an actual alien comes here, they're going to be so different from us, we probably won't be able to fathom that. Movies make them always humanoid—although *Contact* was pretty good.

RIGHT OUTSIDE OUR DOOR...

A STUDENT-LED CLASS TO CONSIDER FOR SPRING!

CEP 498C:
LOVE IN ACTION: BUILDING BELOVED COMMUNITY

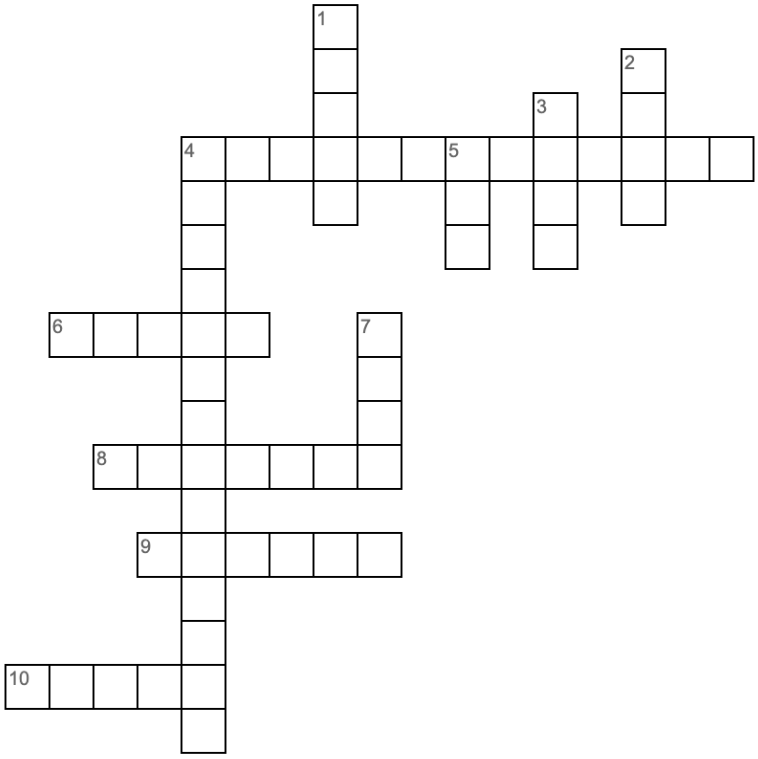
Explore love's impact on our lives, relationships, and communities through a social justice lens outlined by authors and activists like Dr. Martin Luther King, bell hooks, and Thich Nhat Hanh.

Spring 2024 // 1 credit // Mondays 1:30-3:20 // Gould 100

for more information, contact
Maricella Ragudos at mragudos@uw.edu

A PHILOSOPHY PUZZLE

FOR YOUR ENJOYMENT



Across

4 Some difference in X is necessary for a difference in Y

6 Ethics of care in predictive policing say what?

8 Chewing on universal grammar sounds like nom nom nom

9 existence precedes essence

10 Ethics bowl host, and a student of Aristotle

Down

1 wrote an essay while inhaling nitrous oxide

2 Categorical imperative

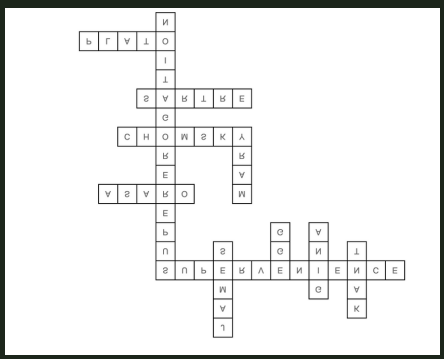
3 Best departmental advisor award goes to...

4 beyond the call of duty

5 Neglected

7 This neuroscientist knows every fact about color perception but not what the color red looks like

ANSWER KEY:



gardenofideasuw.com