

# ROCKS AND CAVEMEN, CLICKS AND TROLLS: HOW TROLLING MAY BE THE MODERN EQUIVALENT OF ROCK- THROWING

Kyle J. Powell<sup>1\*</sup>, Glenn Geher<sup>2</sup>

<sup>1</sup>Florida State University, Tallahassee, FL, USA

<sup>2</sup>State University of New York at New Paltz, New Paltz, NY, USA

\*[powell@psy.fsu.edu](mailto:powell@psy.fsu.edu)

## ABSTRACT

---

Research on human cooperation has presented many well-evidenced mechanisms for the sustainment of cooperation throughout evolutionary history, including direct and indirect reciprocity, reputation-based incentives, and formal punishment. These mechanisms, however, rely strongly on individuals interacting repeatedly with known others in face-to-face contexts—contexts that differ dramatically from the anonymous, invisible, and asynchronous nature of modern cyberspace (Suler, 2004). As such, examples of mass cooperation that occur in online contexts, such as online trolling—the intentional and malicious disruption of online discourse to generate chaos and provoke responses from powerful individuals and organizations—require different mechanisms to function. The current paper explores whether a lesser-known framework, Bingham and Souza’s (2009) coalitional enforcement model, can help shed light on trolling behavior. By allowing individuals to project punitive threat cheaply and at a distance, online trolling may function as a digital analogue to the rock-throwing coalitions of early human groups. We argue that this perspective offers new insights into both the nature of trolling and the broader dynamics of cooperation and punishment in online spaces.

**Keywords:** evolutionary psychology, trolling, coalitional enforcement, Trolling Pathway, cyberpsychology

The rise of the internet has greatly altered the ways in which human beings interact with the world around them (Twenge, et al., 2019). With one click of a button, an individual can order goods from halfway around the world, get back in touch with someone they have not seen in decades, or brush up on their French language skills. In addition, an individual can also use the internet to disrupt otherwise peaceful discourse by starting meaningless arguments, sending death threats, using inflammatory or hateful language, and sowing online chaos in general. These behaviors are typically described as “trolling,” and the individuals who engage in trolling are typically called “trolls” (Cheng et al., 2017). Trolling is a very complex phenomenon, and it can reduce the overall quality of online discourse in some cases. Indeed, online trolling is a pervasive social phenomenon that has been shown to result in increasingly hostile, aggressive, unwelcoming, and prejudicial online communities (Anderson et al., 2014; Coles & West, 2016; Sanfilippo et al., 2017). Yet, little theoretical work has been done to frame this phenomenon in terms of the larger human experience.

Online trolling, with its advanced interconnected networks of potentially thousands of individuals around the world all working toward the same goals, can be considered an act of mass cooperation (Cook et al., 2018). The current paper proposes trolling is merely a natural extension of Bingham and Souza’s (2009) *principle of coercive threat*, within their model of cooperation known as *coalitional enforcement*. By framing trolling in this way, we are not only offering an explanation for how the phenomenon fits into the larger human world, but we are also offering a new model of trolling from which testable predictions can be explored.

### ***Popular Models of Human Cooperation Cannot Account for Online Trolling***

*Cooperation* is defined as one individual paying a cost for another individual to receive a benefit (Rand & Nowak, 2013). These costs and benefits are measured in terms of *evolutionary fitness*, or how well an individual can survive and reproduce in its environment (Geher, 2014). In a well-mixed population of individuals, defectors will always have higher fitness than cooperators because they are receiving benefits from the cooperators and giving nothing in return. Natural selection, therefore, should reduce the abundance of cooperators in the population. And yet, humans are an extremely cooperative species, even among anonymous strangers in one-off laboratory interactions (Burnham & Johnson, 2005). It must be that some mechanisms exist to cause cooperation to be favored over defection in humans. This section will briefly summarize how some popular models of human cooperation used by researchers cannot fully account for the phenomenon of online trolling. For further explanation of the models of cooperation discussed here, see Appendix A.

The online environment in which trolling takes place does not lend itself particularly well to many popular models of human cooperation. For example, a very simple form of human cooperation, *direct reciprocity*, involves the continued interactions of individuals who know each other and provide help in exchange for receiving help later (Rand & Nowak, 2013; Trivers, 1971). The online environment, however, tends to be largely anonymous (individuals do not know each other), invisible (individuals largely do not interact face-to-face), and asynchronous (interactions do not occur in real-time) (Suler, 2004). Therefore, in the online environment, where individuals may not know each other, be able to interact face-to-face, and there is no guarantee of ever interacting with the same person twice, direct reciprocity becomes extremely challenging.

The online environment also presents a problem for models of human cooperation where reputation building is emphasized, such as in *indirect reciprocity* and *moralistic aggression*, two models that focus on the importance of having a reputation as a good cooperator (McGuire, 1991; Nowak & Sigmund, 2005; Rand & Nowak, 2013). In an online environment that is anonymous, invisible, and asynchronous, building a reputation and knowing the reputation of others may be very challenging.

Additionally, consider *punishment*, which occurs when an individual incurs a fitness cost to themselves to inflict a fitness cost on a defector to make them more likely to cooperate in the future (Raihani et al., 2012). Perhaps online trolling is simply a form of collective punishment. However, there is a key point to be made here: Online trolls do not have to incur a fitness cost to themselves when they engage in online trolling. Because online trolls can be anonymous and invisible, and they can engage in trolling from anywhere in the world, online trolling is a very low-cost, low-risk way to get others to change their behavior (Howard, et al., 2019; Ortiz, 2020). For this reason, online trolling does not fit the model of using punishment to ensure cooperation.

What is needed to help explain the phenomenon of online trolling is a model of cooperation that does not necessarily rely on reciprocity or reputation. What is needed is a model that describes large groups of people banding together to enforce cooperation from a distance at a very low cost to themselves. We describe such a model, Bingham and Souza's model of coalitional enforcement, in detail in the next section.

### ***Coalitional Enforcement: The Rock-Throwing Ape***

The mechanism of mass human cooperation that is of primary interest in this paper is coalitional enforcement, and it was proposed by Bingham and Souza (2009). Coalitional enforcement is a model that describes how human beings, throughout history, have enforced cooperation through low-cost, long-range means, such as by throwing rocks, using bows and arrows, and using firearms. We argue this kind of enforcement is exactly what occurs in the online environment today with the phenomenon of online trolling. For this reason, the coalitional enforcement model is uniquely suited to help explain online trolling in a way that alternative models cannot. This section will describe the coalitional enforcement model in more detail.

The model begins with a basic premise. Human beings are unique from all other animals for only one major reason: our anatomy has evolved such that we can throw small objects with extreme accuracy and power, and this ability allows us to project coercive threat over a long distance. Natural environments contain finite resources (Meadows et al., 1972; Meadows & Randers, 2012). Many organisms and species living in these environments compete for these resources (Ayala, 1972). Bingham and Souza (2009) suggest that projecting threat from a distance gives human beings the ability to successfully manage this competition for resources. This uniquely human rock-throwing ability, combined with strong social cooperation tendencies (see Baumeister & Leary, 1995), allowed early human beings to police themselves; if a larger, more physically imposing individual attempted to cheat and take more resources than the other members of the group, they could be pelted with rocks. A collection of individuals that uses this strategy of attacking from a distance is referred to as a *coalition* (Bingham, 2000), and the first coalitions are estimated to have contained approximately 150 individuals (Dunbar, 1992).

The crux of this argument lies in the fact that throwing rocks from a distance is easier and cheaper in terms of cost to survival than fighting the cheater up close, and a group of individuals can all throw rocks at the same time, maximizing the threat and boosting the fitness of all members of the rock-throwing coalition. Under these conditions, fewer individuals will be able to get away with cheating, and fewer individuals will be willing to try to cheat, for fear of violent reprisal. There is a distinct fitness cost associated with cheating under these conditions. This idea is known as the *principle of coercive threat*, and the use of this coercive threat to police coalitional groups is known as *coalitional enforcement* (Bingham, 2000). The application of coercive threat to solve the conflicts of interest problem (a process that only human beings are capable of) is argued to be directly responsible for the founding of the first civilizations, the building of archaic societies, and, ultimately, the development of democracy and the modern

state. It is worth noting that Bickerton and Szathmáry (2011) have made a similar argument for the influence of rock-throwing ability on the development of modern civilization.

There is evidence that modern hunter-gatherer societies continue to use thrown projectiles for violence. A survey of 57 hunter-gatherer societies concluded that 12% of these societies frequently use simple thrown projectiles to engage in interpersonal violence (Hrnčič, 2023). For example, the native Niuean people of the island of Niue in Polynesia have been recorded using throwing stones as one of their primary weapons as recently as the last several centuries (Isaac & Isaac, 2011). Indeed, there are many instances in recorded history of European colonists encountering hunter-gatherer peoples armed with simple thrown projectiles, such as rocks (see Isaac, 1987, for examples of these instances). The use of rock-throwing by these groups specifically appears to have declined over the last several centuries, with mechanically powered weapons such as the bow and arrow becoming more common (Whittaker, 2013). Some attribute this decline to colonialism efforts and interactions between hunter-gatherers and more technologically sophisticated groups (Hrnčič, 2023; Lombardo & Deaner, 2018).

Now, in today's always-online world, it appears as if there is a new way to project threat over a distance: *online trolling*, or the intentional and malicious disruption of online discourse to generate chaos and stir up responses from powerful individuals and organizations. The remainder of this paper will describe the phenomenon of online trolling in more detail and argue that it functions as a modern extension of distance-based coalitional coercive threat; a form of low-cost, distributed threat projection that serves to disrupt and influence powerful actors in digital environments.

### ***Trolling Versus Cyberbullying***

A very common distinction made by researchers is the distinction between trolling and cyberbullying. Most researchers agree that these two phenomena are separate, with some overlap (March & Steele, 2020). Essentially, there are three primary differences between trolling and cyberbullying: First, trolling is typically conducted by a group, and cyberbullying is typically conducted by a single individual (Cook et al., 2018; Cheng et al., 2017). Second, trolls usually target individuals that they do not know personally, such as politicians or celebrities, and cyberbullies usually target individuals they do know personally, such as someone with whom they attend school (Cheng et al., 2017). Third, trolling attacks tend to be more general and less personal; a troll would target a politician's policies instead of their manner of dress, for example (Cheng et al., 2017). In cyberbullying, however, attacks tend to be highly personal, with cyberbullies specifically calling out personal attributes of their targets (Cheng et al., 2017). See Table 1 for a review of these differences. Note that these are just the most common qualities of these two phenomena; not every single trolling attack or cyberbullying incident will possess these qualities exactly. Importantly, the current framework focuses specifically on online trolling and not on cyberbullying.

**Table 1**  
*Key Differences Between Trolling and Cyberbullying in the Most Typical Case of Both Phenomena*

| Phenomenon    | Perpetrator<br>Characteristics     | Victim<br>Characteristics                            | Nature of the Attack            |
|---------------|------------------------------------|------------------------------------------------------|---------------------------------|
| Trolling      | Usually conducted by a group       | Victim is usually not known personally by the troll  | Attacks tend to be very general |
| Cyberbullying | Usually conducted by an individual | Victim is usually known personally by the cyberbully | Attacks tend to be very general |

### ***Trolls and Their Characteristics***

The word *troll* originates from an Old Norse word meaning *fiend* or *demon* (Zarenti & Katsadoros, 2023). The word choice is seemingly appropriate, given the large number of studies that have correlated trolling behaviors with high levels of Dark Tetrad personality traits (narcissism, Machiavellianism, psychopathy, and sadism) (Sest & March, 2017; March & Steele, 2020; Gylfason et al., 2021; Powell, 2024). This relationship is so prevalent in existing research that some have started simply assuming its existence and searching for mediators or moderators that affect the relationship (Masui, 2019). Only one study from those reviewed failed to find a direct positive correlation between trolling behaviors and Dark Tetrad traits (Brubaker et al., 2021).

With few exceptions, existing studies have found that men are more likely to engage in trolling behaviors than are women (Sest & March, 2017; March & Steele, 2020; Powell, 2024). In addition, men were more likely to perceive trolling behaviors as functional, meaning that they felt the behaviors were serving a purpose beyond mere amusement (Brubaker et al., 2021). Combining this finding with the above findings on the Dark Tetrad, one can glean that many individuals who participate in trolling behaviors are malevolent men who troll for reasons other than entertaining themselves. This assertion will become important later.

Trolls are more likely to act in groups than alone, according to existing research. In fact, even individuals who admitted to engaging in trolling behaviors themselves discussed how trolling is a group activity, and anyone who goes at it on their own is likely to be stigmatized (Sanfilippo et al., 2018). Trolling as a group activity is so well-documented that some researchers have started positing the existence of an unofficial *trolling network*, with the implication being that trolls are typically aware of other trolls occupying the same regions of cyberspace, and these trolls can team up and troll the same targets (Cook et al., 2018).

But why is trolling as a group so common? The answer to this question is critical to the argument being made throughout this current paper. Trolls gather in groups to better exert their power (Howard, et al., 2019; Ortiz, 2020). Or, as Bingham and Souza (2009) would put it, the trolls gather in groups because it gives them the ability to exercise greater coercive threat. In other words, a group of trolls in the trolling network is the cyberspace equivalent of a coalition, banding together to reap the fitness benefits of coalitional enforcement. In forming these online coalitions, the trolls can leverage the characteristics of online interactions to their advantage. For example, online trolls do not need to know each other personally or even be active online at the same time to effectively apply coercive threat. This ability to work together from anywhere in the world at anytime can make these online coalitions much more powerful in projecting threat than ancient coalitions were. This comparison between the ancient rock-

throwing coalitions and modern trolling groups is a fundamental link between the theoretical work of Bingham and Souza (2009) and existing research on trolling that will serve as the basis of a new theoretical framework explaining trolling behavior and why individuals engage in the behavior.

### ***Better Weapons Result in Increased Threat***

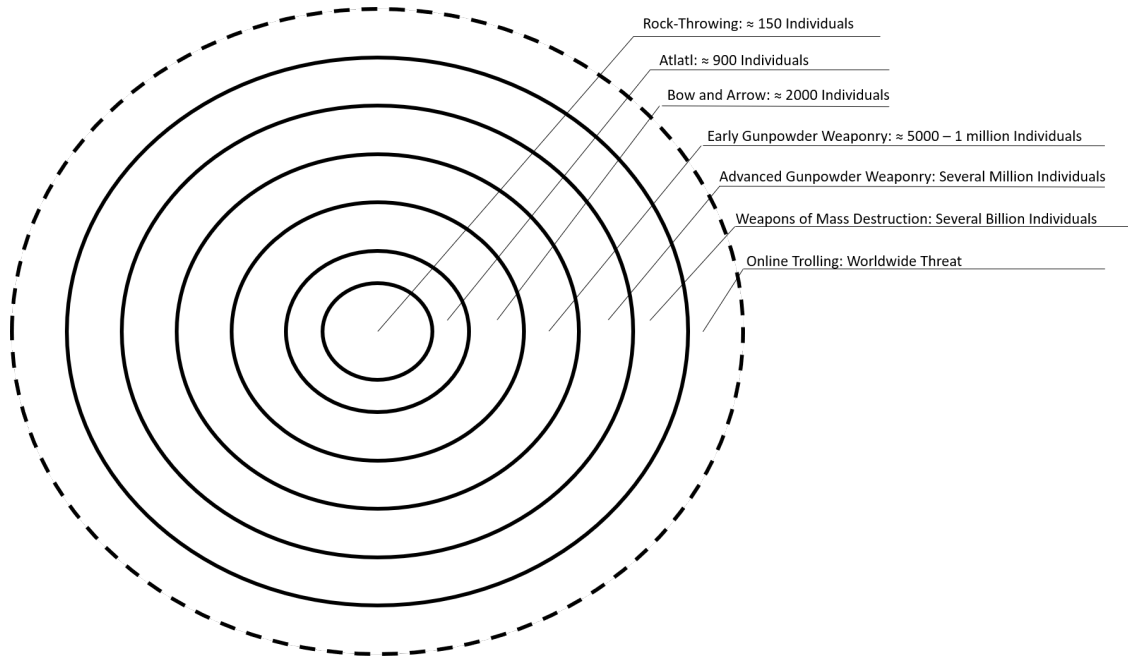
The following model, seen in Figure 1, shows how, in Bingham and Souza's (2009) view, better weapons are better able to project coercive threat. As the model shows, many better weapons have been developed over the course of human history. The projection of threat to keep people from cheating the system is clearly threaded through this history, and, as more powerful, longer-range weapons were developed, human coalitions saw a rapid subsequent increase in size. This effect is directly because longer-range weapons can threaten more people at once, and, thus, more people will choose to cooperate rather than cheat and suffer potential fitness costs.

The unbroken circles in the model reveal the limited nature of the weapons discussed by Bingham and Souza (2009). Even weapons of mass destruction, such as atomic weapons, have their limitations when it comes to threatening their use and deploying them. Online trolling, however, represented by the largest, dashed circle, does not suffer from the limitations the other weapons in the model do. Anyone with internet access can become an online troll, and as the reach of the internet becomes wider and wider, the ability of online trolls to threaten a large portion of the world's population becomes a very real possibility. As of October 2024, it is estimated that 5.52 billion people (or 67.5 percent of the human population) have internet access, and this number is estimated to grow to 7.30 billion by 2029 (DataReportal et al., 2024; Statista, 2023a). Even regions that would not be traditionally considered Western, Educated, Industrialized, Rich, and Democratic (WEIRD, see Henrich et al., 2010) contain hundreds of millions of internet users in some cases (Statista, 2023b). The spread of the internet and the accessibility of online trolling is only going to grow in the next few years. As it does, the ability to project threat through the internet will grow alongside it.

In essence, online trolling has the potential to become the most powerful "weapon" in history, one that can be easily leveraged by a group, utilized over and over in all types of contexts without risk, and capable of forcing even very powerful individuals to change their behavior (Sanfilippo et al., 2018; Cook et al., 2018; Mathews & Goerzen, 2019).

**Figure 1**

Representation, based on the work of Bingham and Souza (2009) and put together by the authors, of increased number of individuals threatened, and, concomitantly, human group size, as weapon range and quality improved.



### ***Online Trolling as a Mechanism of Coalitional Enforcement***

The typical troll is a malevolent male working with a group to disrupt online discourse to achieve some aim beyond their own amusement (Sest & March, 2017; March & Steele, 2020; Gylfason et al., 2021). This aim typically includes spreading some political or social ideology, distracting people from current issues, and attacking powerful individuals (politicians, celebrities, billionaires, etc...) and organizations to bring about some change to the status quo (Mathews & Goerzen, 2019). What these coalitions of trolls are really doing, whether they are aware of it or not, is projecting coercive threat over the victims of their trolling attacks.

The politicians, celebrities, billionaires, and other powerful people of today's world can be thought of as the modern analogue of a larger, more physically imposing individual in ancestral times. One can argue they both have strengths that afford them more resources than the average person. In ancestral times, this would have been literal physical strength, with stronger individuals possessing the capacity to hoard more food, mates, and other resources. Today, the strengths possessed by the powerful include money, fame, status, and government sanctioned authority over others, each one allowing an individual access to more resources than individuals not in possession of these things. Note that both of these situations fall under the classic definition of *power* in social psychology: asymmetric control over valued resources in a social relationship (Galinsky et al., 2015; Magee & Galinsky, 2008).

The trolls in cyberspace can be thought of as the modern equivalent of the physically weaker members of the coalition that relied on their ability to accurately throw rocks to project coercive threat over their more physically powerful conspecifics. Throwing rocks, and thus projecting coercive threat from a distance, was an effective way to police physically powerful conspecifics in the coalition because of the relative ease and low cost to survival the activity posed. Online trolling, with its use of insults and other vitriol instead of rocks, is effective at

projecting coercive threat over a potentially limitless distance for the exact same reasons. It is possible that being a part of the cooperative trolling network confers similar fitness benefits to being a part of an ancient rock-throwing coalition.

Gone are the days when throwing rocks is an acceptable or effective way to send a message of coercive threat. Today, it makes a lot more sense for a coalition of individuals (likely a coalition without access to a large number of resources) to gather together in cyberspace and coordinate trolling attacks on powerful individuals (individuals with access to a large number of resources). The intent of these attacks is typically to bring about some reaction from the victim; the desired reaction depends largely on who the victim is and the nature of the trolling attack. In this way, trolling is often described as being context-dependent.

Two illustrative examples reveal the context-dependent nature of trolling attacks. First, former republican governor of Texas, Rick Perry, became the target of trolling on the social media site “Facebook” in 2012 after a speech wherein he implied lawmakers know more about regulating women’s reproductive health than doctors and the women themselves (Sanfilippo et al., 2017). The trolls, some of whom were identified as physicians and women’s health specialists, mockingly sought health advice from Perry to expose the ludicrous nature of Perry’s public statements (Sanfilippo et al., 2017). Another example is the case of Anita Sarkeesian and the wide-ranging trolling attacks that came to be known as *Gamergate*.

Anita Sarkeesian is a Canadian-American feminist scholar, journalist, and activist who, in 2012, began a series of online videos and blog posts that spoke out against the perceived inherent misogyny of the video game industry (Burgess et al., 2017). In response, Sarkeesian received almost daily death threats, rape threats, bombing threats, and one of her speaking engagements was threatened with mass violence, resulting in the event being cancelled (Burgess et al., 2017). In this case, the trolls that went after Sarkeesian went to great lengths to show their distaste for her views and bury her under a mountain of violent hatred. And, interestingly, the trolls were at least partially successful in this endeavor; by forcing the host of her speaking event to cancel because of the shooting threats, the trolls deprived Sarkeesian of an opportunity to share her views with a captive audience.

Breaking these examples into their component parts will be illuminating. Both examples begin with some powerful individual making some disliked statements. The online responses, then, are hateful, insulting, and intended to trigger a reaction. The powerful individual in question then must decide how to respond to the trolls. These two examples diverge in one interesting way: the trolling against Perry was largely seen as socially acceptable, while the trolling against Sarkeesian was largely seen as socially unacceptable (Sanfilippo et al., 2017; Burgess et al., 2017). However, this distinction, we argue, is not relevant to the bigger picture. Both coalitions of trolls, those that went after Perry and those that went after Sarkeesian, ultimately had the same goal: to exert coercive threat and produce a reaction in their targets. Here is the elegance of framing trolling within Bingham and Souza’s larger theory. Most trolls, regardless of political beliefs, social beliefs, favored methods of trolling, and favored targets, have the same goals: to exert coercive threat over individuals who are or who are perceived to be more powerful than themselves to disrupt the status quo and bring about a desirable response from their targets. In ancestral times, when a larger, more powerful individual presented themselves and tried to take more than they earned, a coalition of physically weaker individuals would band together and inflict a fitness cost on them by throwing rocks. Today, when a more powerful individual with access to more resources presents themselves and does something disliked, a coalition of weaker individuals without access to those resources may band together to troll them, inflicting a potential fitness cost as well. This is the way of the modern struggle for balance.

## ***The Trolling Pathway***

By and large, we expect a typical trolling attack to follow a pattern similar to that seen in Figure 2.

Essentially, a trolling attack begins when the victim of the attack (usually an individual in a position of power or perceived to be in a position of power) says or does something disliked. This description is intentionally vague because, as the examples discussed above reveal, the victim can be just about anyone, of any political and social orientation; all that matters is the trolling network has a negative reaction to the victim's actions and mobilizes an attack. This attack is simply labeled the *Trolling Behavior*.

There are essentially two types of trolling behavior, which we have labeled *Functional* and *Nonfunctional*. Nonfunctional trolling is trolling for fun, and, while there may be some future work that looks into this idea further, this particular model focuses exclusively on functional trolling. Functional trolling, as we have defined it, is trolling with the intention to project coercive threat, and is where the remainder of this analysis will focus.

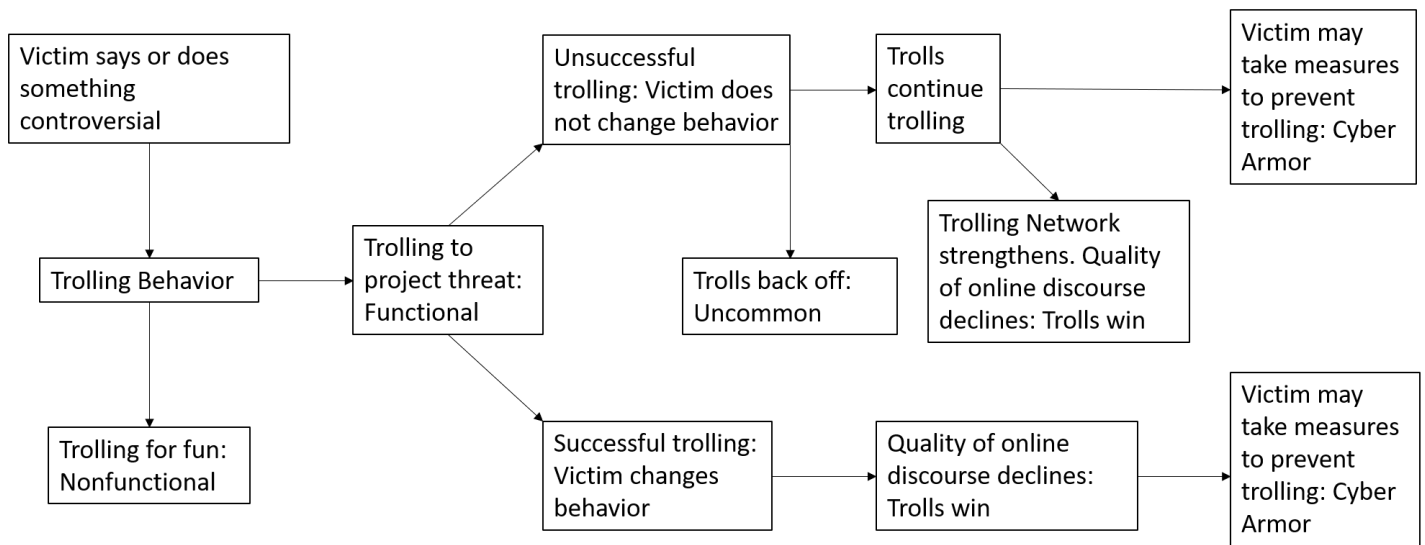
Successful trolling occurs when the victim changes their behavior in response to the trolling. This change could take the form of everything from cancelling a speaking event to rolling back disliked policy decisions, depending on who the victim is. The trolls win in this case, they successfully project coercive threat, and the quality of the online discourse surrounding the topic in question will decline as a result. This decline in online discourse occurs when online trolls start taking over forums, comment sections of websites, and social media profiles of the victim (Coles & West, 2016; Sanfilippo et al., 2017). The presence of many trolls in these spaces can make these spaces more hostile, unwelcome to newcomers, prohibit the development of online communities, and even encourage the normativity of aggressive online behavior (Coles & West, 2016). There is also evidence that high trolling activity in an online space can result in increased polarization and allow for the development and maintenance of prejudicial attitudes (Anderson et al., 2014). These factors reduce the overall quality of the online space and make it harder for people to have discussions about current issues. After a successful trolling attack, the victim may then take measures to prevent future trolling attacks; they may build what we call *cyber armor* (discussed in more detail below).

Unsuccessful trolling occurs when the victim does not change their behavior in response to the trolling. It is possible that the trolls will back down, but what is more likely is they will simply keep on trolling, possibly switching up their tactics and increasing their numbers as they do so. Interestingly, as the trolls continue their trolling behavior, even if the original victim does not change their behavior, they still win, because continuing to troll is likely going to attract more trolls to their cause and strengthen the trolling network. As the trolling network grows in strength, the quality of online discourse will decline, as described above. As the victim continues to deal with the trolling attacks, they may build cyber armor as well to stop it.

The Trolling Pathway model is, to our knowledge, the first attempt to understand online trolling behavior in the context of evolutionary theory. The implications of such a model, and some possible future research directions, are discussed next.

**Figure 2**

*The Trolling Pathway: Shows the most likely steps in the process of trolling to project coercive threat, when the trolling is both successful and unsuccessful.*



### ***Implications of the Current Framework***

An understanding of evolutionary principles can help shed light on our eating habits, education structure, socialization tendencies, and many other aspects of our lives as human beings (Geher et al., 2023). For better or for worse, the use of the internet has become a vital part of many of our lives in the modern world. Modern human beings engage in behaviors in cyberspace with the same passion and alacrity they do in real life space (Tosun & Lajunen, 2009). Thus, those behaviors must be treated with the same care and research vigor as any other behavior, and an application of appropriate evolutionary psychology tools, such as the work of Bingham and Souza (2009), may be the way to accomplish that.

The context-dependent nature of online trolling has made developing theory around the phenomenon challenging (Sanfilippo et al., 2017). However, the suggestion that most trolls have a similar goal, and that goal is evolutionarily driven, may help us better understand how the phenomenon of online trolling fits into the larger human experience. In fact, it may be beneficial to examine this issue, not just from an evolutionary perspective, but from a cyberpsychology perspective as well. As the study of the interaction between technology and human behavior (Ancis, 2020), cyberpsychology is uniquely positioned to help explain online phenomena. In the case of online trolling, it seems that emerging technology has given humans a brand new, superior way to engage in an ancient behavior. What is needed, then, is a novel merger of two fields to approach this topic from two angles: the ancient and the modern. What is needed is an *evolutionary cyberpsychology* approach. Such an approach, in using evolutionary principles to inform human interactions with sophisticated modern technology, could help provide a good starting point for exploring online behaviors such as trolling.

Applying evolutionary principles to help explain the phenomenon of online trolling also gives us a prediction as to the possible responses that common trolling targets may have when they get tired of the constant trolling attacks. According to Bingham and Souza (2009), early methods of projecting coercive threat from a distance, including rock-throwing, the atlatl, and the bow and arrow, became ineffective shortly after the invention of *elite body armor*. Individuals learned they could shield themselves from such attacks, quite literally, and that

made projecting coercive threat with those weapons next to impossible. This invention shifted the coercive power to the coalition with the best body armor, and this change persisted largely until the advent of firearms, which can get through most forms of armor.

The advent of elite body armor, and the subsequent shift in coercive power, is illuminating when one compares these historical shifts to what we are seeing in the modern day. If online trolling really is a new way to project coercive threat, then we should see common trolling targets taking actions to shield themselves from these attacks; we argue this reaction is exactly what is happening.

Take the 45<sup>th</sup> and 47<sup>th</sup> United States president Donald Trump, for example. No doubt an extremely controversial and powerful political figure, Trump was trolled incessantly from the moment he took office for the first time in 2017, not just by Americans, but by citizens in countries around the globe (Fichman, 2022). Trolls attacked Trump on political and ideological grounds, and often mocked the language he used, as well as some of his more notable physical attributes (Fichman, 2022). After leaving the presidency, Trump helped incite an attack on the United States capital to overturn the results of the 2020 election, resulting in him being banned from Twitter (now called “X”) (Conger & Isaac, 2021). In response, Trump founded his very own social media platform, Truth Social, and surrounded himself with political allies (Bhuiyan, 2021). Essentially, Trump constructed a shield for himself, a platform where he could reach only his followers and avoid the masses of trolls.

We argue that instances such as this one, where powerful individuals who are frequently the targets of trolls find ways to shield themselves from trolls on the internet, are going to become more common within the next few years. Elon Musk’s acquisition of the social media platform “Twitter” (and subsequent rebranding to “X”) may be an attempt to build himself a place where he will be safe from trolls (Benton et al., 2022). This shielding from trolls online in the modern-day is analogous to Bingham and Souza’s (2009) elite body armor shielding from projectiles in human history. For this reason, we are calling the concept of shielding oneself from online trolls *cyber armor*.

### ***Future Research Directions***

Our theory suggests that the trolls of today are essentially the rock-throwers of the ancestral past. If this is true, we would expect trolls to possess similar characteristics to the rock-throwers.

It is possible that the rock-throwers of our ancestral past were feeling ostracized from the coalition as a result of being pushed around by the larger, stronger individuals. *Ostracism* occurs when an individual or group is intentionally ignored and excluded by another individual or group (Williams, 2007). Being ostracized may have helped lead them to take control by throwing rocks: in throwing rocks at the more powerful individuals, they found a new coalition to be a part of, and they were able to balance the playing field. In today’s world, ostracism, especially when associated with long-term social estrangements and combined with high dark tetrad traits, can lead to antisocial and revenge-seeking behaviors (Sung et al., 2021). If our ancient ancestors responded to this by throwing rocks because it was the evolutionarily cheap and easy way to take back control and boost fitness, then it makes perfect sense to think that today’s ostracized individuals would band together online and launch trolling attacks on those they perceive to have the power, and they would engage in this behavior because it is the evolutionarily cheap and easy thing to do.

A future study could determine if individuals who are identified as trolls (possibly by giving them the Global Assessment of Internet Trolling Scale, see Buckels et al., 2014) also have a high number of social estrangements and perceived ostracism. In addition, we could use the same predictors in Sung et al. (2021) (impulsivity, borderline tendencies, risk-taking behavior, and life history strategy) to see if these predict trolling behavior, and which is most likely to lead to trolling behavior. Below is a list of predictions for such a study:

- In-line with much of the rest of the research on this topic, we predict that individuals who score higher on a trolling behavior scale will score higher on a dark tetrad scale.
- We predict that individuals who score higher on trolling behavior will also have a faster life history strategy, representing their spending most of their lives with access to few resources.
- We predict that individuals with a higher number of long-term social estrangements will score higher on the trolling behavior scale.
- We predict a moderating effect of the dark tetrad on the relationship between trolling behavior and social estrangements, impulsivity, borderline tendencies, risk-taking behavior, and life history strategy, that is, people who score high on these variables may be more likely to engage in trolling behavior, if they score high on dark tetrad traits as well.

Not only would these predicted findings help find empirical support for the theoretical framework given here, but it would also help provide a clearer picture of the types of personality traits that lead to trolling behavior. Understanding the “trolling personality” from an evolutionary perspective, combined with the understanding that cyberpsychology provides regarding how people interact with technology in general, may be important to future trolling research efforts.

Future research could also address specifically cases of successful online trolling. When a target of trolling changes their behavior because of the trolling attack, how will the general public perceive that individual? Will giving into the trolling attack negatively impact an individual’s freely conferred respect and status (i.e., *prestige*; see Maner, 2017)? What will the cyber armor of individuals without the resources of Donald Trump or Elon Musk look like? Current methods of defending against trolling are largely unreliable (see Golf-Papez & Veer, 2017), so what is the next best course of action for the average person? Future research on online trolling, informed by evolutionary and cyberpsychological theory, could help address these questions.

## CONCLUSION

We have presented a theoretical framework describing the phenomenon of online trolling as a natural extension of Bingham and Souza’s (2009) principle of coercive threat. To our knowledge, this framework is the first to explain online trolling behavior using evolutionary principles. Online trolls today group together in coalitions to project coercive threat over a distance in a similar way to our rock-throwing ancestors. The trolls project coercive threat over the internet in an incredibly efficient, evolutionarily cheap, and easy manner. As previously discussed, these online trolling attacks are highly context-specific in their methods and in the response from others to the attack. However, our framework suggests that, generally, the goal of each trolling attack is the same: to bring about some change in the behavior of the target.

Much research still needs to be done to understand the phenomenon of online trolling. This includes further research on the personality of online trolls, their evolutionary characteristics (such as life history strategy), how online trolling affects social hierarchies, and the motivations of individual trolls within the network. It would also be useful to empirically examine the circumstances under which online trolling is viewed as acceptable versus unacceptable. Using evolutionary principles and ancient evolved human tendencies to help inform our understanding of our interactions with sophisticated modern technologies, or an *evolutionary cyberpsychology* approach, may help address these questions in future research and lead to a better understanding of online trolling behavior.

## FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## DECLARATION

We declare that all work herein is original and our own. All authors have reviewed and approved the manuscript for submission, and it is not being considered for submission elsewhere whilst it is under consideration for *Human Ethology*.

## TRANSPARENCY

**Kyle J. Powell:** Writing – original draft (Lead), Writing – review & editing (Lead)

**Glenn Geher:** Supervision (Lead)

## ORCID ID

Kyle J. Powell  <https://orcid.org/0009-0000-2682-2975>

Glenn Geher  <https://orcid.org/0000-0001-7622-7197>

## REFERENCES

- Ancis, J. R. (2020). The Age of Cyberpsychology: An Overview. *Technology, Mind, and Behavior*, 1(1). <https://doi.org/10.1037/tmb0000009>
- Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books.
- Ayala, F. J. (1972). Competition between Species: The diversity of environments in which most organisms live permits the coexistence of many species, even when they compete for the same resources. *American Scientist*, 60(3), 348–357. <http://www.jstor.org/stable/27843135>
- Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The “Nasty Effect:” Online Incivility and Risk Perceptions of Emerging Technologies. *Journal of Computer-Mediated Communication*, 19(3), 373–387. <https://doi.org/10.1111/jcc4.12009>
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117, 497–529. <https://doi.org/10.1037/0033-2909.117.3.497>
- Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin*, 115(2), 243–267. <https://doi.org/10.1037/0033-2909.115.2.243>
- Benton, B., Choi, J. A., Luo, Y., & Green, K. (2022). Hate Speech Spikes on Twitter After Elon Musk Acquires the Platform. *School of Communication and Media, Montclair State University*. <https://digitalcommons.montclair.edu/scom-facpubs/33>
- Bhuiyan, J. (2021, October 20). Donald Trump to launch social media platform called Truth Social. *The Guardian*. <https://www.theguardian.com/us-news/2021/oct/20/donald-trump-truth-social-app-apple>
- Bickerton, D., Szathmáry, E. Confrontational scavenging as a possible source for language and cooperation. *BMC Evol Biol* 11, 261 (2011). <https://doi.org/10.1186/1471-2148-11-261>
- Bingham, P.M. (2000), Human evolution and human history: A complete theory. *Evol. Anthropol.*, 9: 248-257. [https://doi.org/10.1002/1520-6505\(2000\)9:6<248::AID-EVAN1003>3.0.CO;2-X](https://doi.org/10.1002/1520-6505(2000)9:6<248::AID-EVAN1003>3.0.CO;2-X)
- Bingham, P. M., & Souza, J. (2009). *Death from a distance and the birth of a humane universe*. Charleston, SC: BookSurge.

- Brubaker, P. J., Montez, D., & Church, S. H. (2021). The Power of Schadenfreude: Predicting Behaviors and Perceptions of Trolling Among Reddit Users. *Social Media + Society*, 7(2). <https://doi.org/10.1177/20563051211021382>
- Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, 67, 97-102. <https://doi.org/10.1016/j.paid.2014.01.016>
- Burgess, M. C., Byars, F., Sadeghi-Azar, L., & Dill-Shackleford, K. E. (2017). Online Misogyny Targeting Feminist Activism: Anita Sarkeesian and Gamergate. *The Wiley Handbook of Violence and Aggression*, 1-13. <https://doi.org/10.1002/9781119057574.whbva006>
- Burnham, T. C., & Johnson, D. D. (2005). The Biological and Evolutionary Logic of Human Cooperation. *Analyse & Kritik*, 27(1), 113-135. <https://doi.org/10.1515/auk-2005-0107>
- Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017). Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 1217-1230). <https://doi.org/10.1145/2998181.2998213>
- Coles, B. A., & West, M. (2016). Trolling the trolls: Online forum users constructions of the nature and properties of trolling. *Computers in Human Behavior*, 60, 233-244. <https://doi.org/10.1016/j.chb.2016.02.070>
- Conger, K., & Isaac M. (2021, January 8). Twitter Permanently Bans Trump, Capping Online Revolt. *The New York Times*. <https://www.nytimes.com/2021/01/08/technology/twitter-trump-suspended.html>
- Cook, C., Schaafsma, J., & Antheunis, M. (2018). Under the bridge: An in-depth examination of online trolling in the gaming context. *New Media & Society*, 20(9), 3323-3340. <https://doi.org/10.1177/1461444817748578>
- DataReportal, Meltwater, & We Are Social. (2024, October 23). Number of internet and social media users worldwide as of October 2024 (in billions) [Graph]. In *Statista*. Retrieved November 25, 2024, from <https://www.statista.com/statistics/617136/digital-population-worldwide/>
- Dunbar, R. I. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6), 469-493. [https://doi.org/10.1016/0047-2484\(92\)90081-J](https://doi.org/10.1016/0047-2484(92)90081-J)
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137-140. <https://doi.org/10.1038/415137a>
- Fessler, D. M. T., & Haley, K. J. (2003). The Strategy of Affect: Emotions in Human Cooperation. In P. Hammerstein (Ed.), *Genetic and Cultural Evolution of Cooperation* (pp. 7-36). Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/3232.003.0004>
- Fichman, P. (2022). The role of culture and collective intelligence in online global trolling: The case of trolling Trump's inauguration speech. *Information, Communication & Society*, 25(7), 1029-1044. <https://doi.org/10.1080/1369118X.2020.1824006>
- Galinsky, A. D., Rucker, D. D., & Magee, J. C. (2015). Power: Past findings, present considerations, and future directions. In M. Mikulincer, P. R. Shaver, J. A. Simpson, & J. F. Dovidio (Eds.), *APA handbook of personality and social psychology, Vol. 3. Interpersonal relations* (pp. 421-460). American Psychological Association. <https://doi.org/10.1037/14344-016>
- Geher, G. (2014). *Evolutionary Psychology 101*. Springer
- Geher, G., Fritche, M., Goodwine, A., Lombard, J., Longo, K., & Montana, D. (2023). *An Introduction to Positive Evolutionary Psychology*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781009286817>
- Golf-Papez, M., & Veer, E. (2017). Don't feed the trolling: rethinking how online trolling is being defined and combated. *Journal of Marketing Management*, 33(15-16), 1336-1354. <https://doi.org/10.1080/0267257X.2017.1383298>
- Gylfason, H. F., Sveinsdottir, A. H., Vesteinsdóttir, V., & Sigurvinsdottir, R. (2021). Haters Gonna Hate, Trolls Gonna Troll: The Personality Profile of a Facebook Troll. *Int. J. Environ. Res. Public Health* 2021, 18(11), 5722. <https://doi.org/10.3390/ijerph18115722>
- Hamilton, W. D. (1964). The genetical evolution of social behaviour. II. *Journal of Theoretical Biology*, 7(1), 17-52. [https://doi.org/10.1016/0022-5193\(64\)90039-6](https://doi.org/10.1016/0022-5193(64)90039-6)
- Hardy, C. L., & Van Vugt, M. (2006). Nice Guys Finish First: The Competitive Altruism Hypothesis. *Personality and Social Psychology Bulletin*, 32(10), 1402-1413. <https://doi.org/10.1177/0146167206291006>

- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466(7302), 29-29. <https://doi.org/10.1038/466029a>
- Hrnčič, V. (2023). The Use of Wooden Clubs and Throwing Sticks among Recent Foragers. *Human Nature*, 34(1), 122-152. <https://doi.org/10.1007/s12110-023-09445-3>
- Howard, K., Zolnieriek, K. H., Critz, K., Dailey, S., & Ceballos, N. (2019). An examination of psychosocial factors associated with malicious online trolling behaviors. *Personality and Individual Differences*, 149, 309-314. <https://doi.org/10.1016/j.paid.2019.06.020>
- Isaac, B. (1987). Throwing and human evolution. *African Archaeological Review*, 5(1), 3-17. <https://doi.org/10.1007/BF01117078>
- Isaac, B., & Isaac, G. (2011). Unexpected trajectories: A history of Niuean throwing stones. *The Journal of the Polynesian Society*, 120(4), 369-401.
- Izard, C.E. (1977). *Human Emotions*. New York: Plenum.
- Lombardo, M. P., & Deaner, R. O. (2018). Born to Throw: The Ecological Causes that Shaped the Evolution of Throwing In Humans. *The Quarterly Review of Biology*, 93(1), 1-16. <https://doi.org/10.1086/696721>
- Magee, J. C., & Galinsky, A. D. (2008). Social Hierarchy: The Self-Reinforcing Nature of Power and Status. *Academy of Management Annals*, 2(1), 351-398. <https://doi.org/10.5465/19416520802211628>
- Maner, J. K. (2017). Dominance and Prestige: A Tale of Two Hierarchies. *Current Directions in Psychological Science*, 26(6), 526-531. <https://doi.org/10.1177/0963721417714323>
- March, E., & Steele, G. (2020). High Esteem and Hurting Others Online: Trait Sadism Moderates the Relationship Between Self-Esteem and Internet Trolling. *Cyberpsychology, Behavior, and Social Networking*, 23(7), 441-446. <https://doi.org/10.1089/cyber.2019.0652>
- Masui, K. (2019). Loneliness moderates the relationship between Dark Tetrad personality traits and internet trolling. *Personality and Individual Differences*, 150, 109475. <https://doi.org/10.1016/j.paid.2019.06.018>
- Matthews, J., & Goerzen, M. (2019). Black Hat Trolling, White Hat Trolling, and Hacking the Attention Landscape. *Companion Proceedings of The 2019 World Wide Web Conference* (pp. 523-528). <https://doi.org/10.1145/3308560.3317598>
- McGuire, M. T. (1991). Moralistic Aggression and the Sense of Justice. *American Behavioral Scientist*, 34(3), 371-385. <https://doi.org/10.1177/0002764291034003007>
- Meadows, D.H., Meadows, D.L., Randers, J., Behrens, W.W. (1972). *The limits to growth*. Universe Books, New York.
- Meadows, D., & Randers, J. (2012). *The Limits to Growth: The 30-year Update*. Routledge. <https://doi.org/10.4324/9781849775861>
- Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437(7063), 1291-1298. <https://doi.org/10.1038/nature04131>
- Nowak, M. A., Tarnita, C. E., & Antal, T. (2010). Evolutionary dynamics in structured populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537), 19-30. <https://doi.org/10.1098/rstb.2009.0215>
- Ortiz, S. M. (2020). Trolling as a Collective Form of Harassment: An Inductive Study of How Online Users Understand Trolling. *Social Media + Society*, 6(2). <https://doi.org/10.1177/2056305120928512>
- Powell, K. J. (2024). *Understanding the evolutionary characteristics of online trolls (Master's thesis)*. State University of New York at New Paltz.
- Raihani, N. J., Thornton, A., & Bshary, R. (2012). Punishment and cooperation in nature. *Trends in Ecology & Evolution*, 27(5), 288-295. <https://doi.org/10.1016/j.tree.2011.12.004>
- Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, 17(8), 413-425. <https://doi.org/10.1016/j.tics.2013.06.003>
- Roberts, G. (1998). Competitive altruism: from reciprocity to the handicap principle. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 265(1394), 427-431. <https://doi.org/10.1098/rspb.1998.0312>
- Sanfilippo, M. R., Fichman, P., & Yang, S. (2018). Multidimensionality of online trolling behaviors. *The Information Society*, 34(1), 27-39. <https://doi.org/10.1080/01972243.2017.1391911>

- Sanfilippo, P. Yang, S. & Fichman, P. (2017). Managing online trolling: From deviant to social and political trolls. *Proceedings of the 50th Hawai'i international conference on system sciences (HICSS-50)*. Los Alamitos: IEEE Press.
- Sest, N., & March, E. (2017). Constructing the cyber-troll: Psychopathy, sadism, and empathy. *Personality and Individual Differences*, 119, 69-72. <https://doi.org/10.1016/j.paid.2017.06.038>
- Statista. (2023a, July 25). Number of internet users worldwide from 2014 to 2029 (in millions) [Graph]. In Statista. Retrieved November 25, 2024, from <https://www.statista.com/forecasts/1146844/internet-users-in-the-world>
- Statista. (2023b, December 4). Number of internet users worldwide as of 2023, by subregion (in millions) [Graph]. In Statista. Retrieved November 25, 2024, from <https://www.statista.com/statistics/249562/number-of-worldwide-internet-users-by-region/>
- Suler, J. (2004). The Online Disinhibition Effect. *CyberPsychology & Behavior*, 7(3), 321-326. <https://doi.org/10.1089/1094931041291295>
- Sung, A., Geher, G., & Wice, M. (2021). Predictors of Social Estrangements. *EvoS Journal: The Journal of the Evolutionary Studies Consortium*, 9(1), 14-24. <https://doi.org/10.59077/ZLRP9617>
- Tarnita, C. E., Wage, N., & Nowak, M. A. (2011). Multiple strategies in structured populations. *Proceedings of the National Academy of Sciences*, 108(6), 2334-2337. <https://doi.org/10.1073/pnas.1016008108>
- Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two Key Steps in the Evolution of Human Cooperation: The Interdependence Hypothesis. *Current Anthropology*, 53(6), 673-692. <https://doi.org/10.1086/668207>
- Tosun, L. P., & Lajunen, T. (2009). Why Do Young Adults Develop a Passion for Internet Activities? The Associations among Personality, Revealing “True Self” on the Internet, and Passion for the Internet. *CyberPsychology & Behavior*, 12(4), 401-406. <https://doi.org/10.1089/cpb.2009.0006>
- Traulsen, A., & Nowak, M. A. (2006). Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences*, 103(29), 10952-10955. <https://doi.org/10.1073/pnas.0602530103>
- Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, 46(1), 35-57. <https://doi.org/10.1086/406755>
- Twenge, J. M., Spitzberg, B. H., & Campbell, W. K. (2019). Less in-person social interaction with peers among US adolescents in the 21st century and links to loneliness. *Journal of Social and Personal Relationships*, 36(6), 1892-1913. <https://doi.org/10.1177/0265407519836170>
- Whittaker, J. C. (2013). Comparing Atlatls and Bows: Accuracy and Learning Curve. *Ethnoarchaeology*, 5(2), 100-111. <https://doi.org/10.1179/1944289013Z.0000000009>
- Williams, K. D. (2007). Ostracism. *Annu. Rev. Psychol.*, 58(1), 425-452. <https://doi.org/10.1146/annurev.psych.58.110405.085641>
- Wilson, D. S. (1975). A theory of group selection. *Proceedings of the National Academy of Sciences*, 72(1), 143-146. <https://doi.org/10.1073/pnas.72.1.143>
- Zarenti, M., & Katsadoros, G. (2023). Trolls: From Mythology and Folklore to the Internet. *Asian Journal of Language, Literature and Culture Studies*, 6(2), 75-85.

## APPENDIX A

### MECHANISMS OF HUMAN COOPERATION

#### ***Direct Reciprocity***

*Direct reciprocity* occurs when there are repeated interactions between the same two individuals (Rand & Nowak, 2013; Trivers, 1971). If one individual helps another, they can count on receiving help from that person later. For cooperation to evolve, the probability of these two individuals interacting repeatedly over time must be sufficiently high (Axelrod, 1984). In each interaction, the individuals swap paying a fitness cost and gaining a fitness benefit.

#### ***Indirect Reciprocity***

*Indirect reciprocity* occurs when there are repeated interactions within a population and third parties observe those interactions, talk about them, and base their own cooperative decisions on the reputations of those involved (Nowak & Sigmund, 2005; Rand & Nowak, 2013). In this situation, cooperation may be costly, but it leads to the reputation of being a helpful person, which increases chances of receiving help in the future, leading to an overall fitness increase. The *competitive altruism* hypothesis stems from the concept of indirect reciprocity and suggests that individuals may compete to have the best reputation as a helpful person, and they will also compete to secure the most helpful partner (Hardy & Van Vugt, 2006; Roberts, 1998). In this situation, the fitness benefits of having a good reputation outweigh the fitness costs of being cooperative.

#### ***Spatial Selection***

*Spatial selection* occurs when populations become structured such that cooperators can form clusters that allow for increased cooperation even when surrounded by defectors (Nowak et al., 2010; Tarnita et al., 2011). When cooperators are more likely to interact with other cooperators, they earn greater fitness benefits.

#### ***Multilevel Selection***

*Multilevel selection* occurs when there is competition between groups as well as competition within groups (Traulsen & Nowak, 2006; Wilson, 1975). Groups of cooperators will ultimately see increases in fitness over groups of defectors and this process can result in increased selection for cooperation.

#### ***Kin Selection***

*Kin selection* is the basic idea that individuals are generally more inclined to cooperate with those with whom they share genes; genetically related kin (Rand & Nowak, 2013). Kin selection is often tied to the concept of *inclusive fitness*, a specific mathematical method for accounting for fitness effects (Hamilton, 1964).

## OTHER POTENTIAL MECHANISMS OF COOPERATION

### ***The Interdependence Hypothesis***

The *interdependence hypothesis* is a two-step hypothesis for the evolution of human cooperation. In the first step, humans became obligate collaborative foragers such that coordination, sharing, and punishment of free riders was important during food procurement endeavors such as hunting (Tomasello et al., 2012). Since each individual played a role in these endeavors, all group members had a direct interest in the well-being of others; if one individual could not play their part in acquiring food, the total amount of food acquired went down and the whole group suffered. In the second step, the skills and motivations evolved to enhance collaboration in food procurement endeavors were scaled to group life as humans faced competition from other groups (Tomasello et al., 2012). This scaling up resulted in the creation of cultural conventions, norms, and institutions that enforce collaboration and cooperation within groups. Thus, as humans became more interdependent, selection favored cooperation more and more. This hypothesis sets itself apart from others by placing less of an emphasis on altruism and reciprocity (Tomasello et al., 2012).

### ***Moralistic Aggression and Other Emotions that Facilitate Cooperation***

When one individual helps another, and that person fails to reciprocate that help, the original helper and any observers may experience what is known as *moralistic aggression*, or feelings of anger and thoughts of retaliation in response to failure to reciprocate help (McGuire, 1991). Moralistic aggression is believed to occur to some extent in all humans, and is believed to help facilitate cooperation by making those who fail to reciprocate the target of anger, thus ensuring they will be more likely to reciprocate in the future (McGuire, 1991).

Moralistic aggression is not the only emotion believed to have evolved to facilitate cooperation. There have been suggestions made for the role of at least 13 other emotions in the facilitation of cooperation (Fessler & Haley, 2003). Two examples are *guilt*, an aversive feeling that can occur when one inflicts harm on another during a reciprocal relationship, causing them to want to take action to repair that relationship (Baumeister et al., 1994; Izard, 1977), and *righteousness*, a distinct positive emotion that occurs when an individual behaves in a way that promotes cooperation and relationships (Fessler & Haley, 2003). Individuals are motivated to avoid guilt and experience righteousness, making promoting cooperation and avoiding harmful defection important. In this way, these emotions facilitate cooperation.

### ***The Role of Punishment***

*Punishment* is a mechanism by which an individual uses their own resources (i.e., reduces their own fitness) to inflict a fitness cost on a defector to make them more likely to cooperate in the future (Raihani et al., 2012). In humans, punishment can often promote cooperation (Fehr & Gächter, 2000). Punishment can be used in conjunction with any of the mechanisms already discussed to further promote cooperation in human societies.