We are Legion:

Hacktivism as a Product of Deindividuation, Power, and Social Injustice

by

Jessica Bodford

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Approved April 2015 by the
Graduate Supervisory Committee:

Virginia S. Y. Kwan, Chair
Paulo Shakarian
Bradley Adame

ARIZONA STATE UNIVERSITY

May 2015

ABSTRACT

The current study examines the role that context plays in hackers' perceptions of the risks and payoffs characterizing a hacktivist attack. Hacktivism (i.e., hacking to convey a moral, ethical, or social justice message) is examined through a general game theoretic framework as a product of costs and benefits, as well as the contextual cues that may sway hackers' estimations of each. In two pilot studies, a bottom-up approach is utilized to identify the key motives underlying (1) past attacks affiliated with a major hacktivist group, Anonymous, and (2) popular slogans utilized by Anonymous in its communication with members, targets, and broader society. Three themes emerge from these analyses, namely: (1) the prevalence of first-person plural pronouns (i.e., *we*, *our*) in Anonymous slogans; (2) the prevalence of language inducing status or power; and (3) the importance of social injustice in triggering Anonymous activity. The present research therefore examines whether these three contextual factors activate participants' (1) sense of deindividuation, or the loss of an individual's personal self in the context of a group or collective; and (2) motive for self-serving power or society-serving social justice. Results suggest that participants' estimations of attack likelihood stemmed solely from expected payoffs, rather than their interplay with subjective risks. As expected, the use of *we* language led to a decrease in subjective risks, possibly due to primed effects of deindividuation. In line with game theory, the joint appearance of both power and justice motives resulted in (1) lower subjective risks, (2) higher payoffs, and (3) higher attack likelihood overall. Implications for policymakers and the understanding and prevention of hacktivism are discussed, as are the possible ramifications of deindividuation and power for the broader population of Internet users around the world.

“Anonymity itself does not seem to be a social ill.
Rather, the state of anonymity seems to encourage
whatever potentials are most prominent at the moment,
whether for good or for ill. When we are anonymous
we are free to be aggressive or to give affection,
whichever expresses most fully our feelings at the time.

There is liberation in anonymity.”

—Gergen, Gergen, & Barton (1973, p. 130)

TABLE OF CONTENTS

Page

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

We are Legion: Hacktivism as a Product of Deindividuation, Power, and Social Injustice

In the days following the controversial and possibly racially driven murder of Michael Brown on August 9, 2014, riots began to emerge in the streets of Ferguson, Missouri. Activists seeking to spread the news of this social injustice traveled from neighboring cities and states, but not all activists were physically present at these protests. Indeed, representatives of the hacker—or more specifically, hacktivist—group Anonymous took to online forums, video feeds, and news channels, posting live video and protest updates for the world to see (Rogers, 2014). This movement, coined Operation Ferguson, escalated over the following days as Ferguson police responded to the riots with increased violence. By August 12, Anonymous members had begun to directly target the phone lines and associated websites of Ferguson City Hall. But while many viewed such actions as signs of support for Michael Brown, many "suspect[ed] their motives and question[ed] their behavior" (Wines & Fitzsimmons, 2014), bringing to question the reasons underlying Anonymous's actions. To date, no clear answers have emerged (Rogers, 2014).

Central Problem

A *cyber-attack* is defined as an attempt to damage, destroy, or gain illegal access to a computer network or system. Within the broader family of cyber-attacks falls *hacking*, or the act of gaining access to—and control over—third-party computer systems (Antón et al., 2003). The intentions underlying hacking are, by nature, unclear: Although it is possible that hackers wish to cause grave harm, severe economic losses, or destruction of critical infrastructure, it is equally possible that they wish to do nothing more than exact revenge upon their targets (Manion & Goodrum, 2000).

Hacktivism, which forms the focus of this work, shares many of the methods observed in hacking[1]. However, its name evinces a clear distinction that sets it apart from other forms of cyber-attacks: As Jordan and Taylor (2004) aptly state, "hacktivism is activism gone electric" (p. 3)—it is the emergence of protest in cyberspace to make a statement that is often morally, politically, ethically, or legally driven (Colesky & Van Niekerk, 2012; Manion & Goodrum, 2000). Cyber-attacks in their general form are estimated to cost the U.S. government hundreds of billions of dollars per year (Krawczyk, 2014), far more than annual government expenditures following natural disasters. Hacktivist attacks, on the other hand, incur costs not only to governments but also quite extensively to corporations, organizations, and other frequent targets. Ponemon Institute (2012) estimates that companies lose $22,000 for every minute that their website is taken offline due to a DDos (Distributed Denial of Service) attack, a commonly used method among key hacktivist groups to make a statement against particular groups of choice (Mansfield-Devine, 2011). Security company Symantec estimates that, across all forms of hacktivism against both governments and industries, hacktivist attacks cost approximately $114 billion each year (Albanesius, 2011).

To be sure, such attacks should not be carried out lightly. Although hacktivism entails pursuing a cause of personal or moral import, it is also accompanied by substantial risk should the perpetrators be caught. In the past, apprehended members of hacktivism groups such as Anonymous and Lulzsec have faced charges of a criminal nature, including fines up to $183,000 and prison sentences of up to ten years (Pilkington, 2013;

[1] Unlike hacking, few methods utilized in hacktivist attacks entail intrusion into—and control over—another's computer system. Distributed Denial of Service attacks, for example, are often used to take a target's website offline without directly or indirectly

Vincent, 2013). Thus, prior to taking part in a hacktivist attack it is imperative that, in line with game theory, the member weigh both end goals (hereafter referred to as *payoffs*) and consequences (*risks*) if caught by defenders[2] (Dixit & Nalebuff, 2008; Liu, 2005). Game theory, first founded under the name *minimax theorem* in the late 1920s (von Neumann, 1928), originally stated that within a two-person, zero-sum game with only limited possible moves, one can mathematically calculate the strategies that would guarantee the optimal payoffs for each player given the other player's actions. In the near century that followed, the minimax theorem has been broadened and extended to apply to a wide range of strategic decision-making scenarios that collectively fall under a game theoretic framework.

In its most general sense, this framework holds that rational decision makers should carefully weigh all known consequences of their decisions before acting. In the present study, game theory can be thought of as a strategy by which an individual chooses an action (here, to attack or not to attack) based on available information (risks and payoffs involved). Because neither risks nor payoffs are typically known prior to taking part in an attack, they are both subjective in nature—that is, the hacktivist must estimate the probability and magnitude (e.g., severity) of both risks and payoffs, and base his or her final decision of attack on these estimations.

Scope of the Present Research

[2] Although we will not directly address the role of the defender (i.e., the party tasked with defending a system from attack) within this study, we acknowledge the importance of defenders' actions in preparation for—or retaliation against—hacking attempts. For the purposes of this work, we will focus exclusively on the decision-making strategies of the hacker alone.

The present research posits that these estimations of subjective risks and payoffs do not remain static in the face of a pending hacktivist attack. Instead, we expect the immediate context to play a crucial role in tailoring such estimations to best fit the situation at hand. Before discussing examples of immediate contextual factors, it is fitting to first discuss the nature and purpose of experimental priming in psychological research.

Priming is a phenomenon by which exposure to one stimulus impacts a response to another, later stimulus. Research on semantic priming in the early 1970s suggested that participants were quicker at recognizing words when they directly followed semantically similar words (e.g., “butter” following “bread”, but not “monkey”; Meyer & Schvaneveldt, 1971). Furthermore, priming is more implicit than it is overt: Participants are unaware of the impact a prime may have on subsequent behavior (Schröder & Thagard, 2013). Even so, a vast body of research across many areas of psychology suggests that the effects of exposure to a prime—be it textual, visual, or aural—can be both salient and long-lasting (Sumner & Samuel, 2007; Tulving, Schacter, & Stark, 1982). When exposure to different priming conditions is randomized, it is assumed that individual differences are left to chance; as such, differences in responses across conditions are attributed to the effect of the randomly presented prime, rather than to extraneous variables.

A large body of psychological research has illustrated how contextual cues exert powerful influences on human cognition and behavior. Environmental primes can activate related—but dormant—concepts and, in turn, impact perceivers’ judgments and decisions (Bargh & Chartrand, 1999; Higgins, 1989; Wyer & Srull, 1989). Subtle cues have yielded powerful effects in a wide variety of contexts, from auditors making fraud-risk judgments (Hackenbrack, 1992) to managers making hiring decisions (Highhouse, 1997), from Wall Street professionals making investment decisions (Alter & Kwan, 2009) to general estimates of personal cancer risk (Kwan et al., 2012) and disease threats (White, Johnson, & Kwan, 2014).

In the present research, we seek to identify contextual factors—here, motives stemming from the nature of the attack itself—that impact hackers' subjective risks and payoffs and, therefore, their likelihood of carrying out a hacktivist attack. For decades, psychological research has explored the motives underlying human cognition, action, and behavior and the ways in which context influences these outcomes (Churchill, 1991; Newcomb, 1950; Perrin, 1923). Similarly, we posit that the motives precipitating a hacktivist attack play an important role in hackers' subjective risks and payoffs. Limited past research on the role of motives in cyber-attacks has primarily focused on two elements: (1) hacking as a general form of cyber-attack, rather than hacktivism in particular (Turgeman-Goldschmidt, 2005; Voiskounsky & Smyslova, 2003); and, more specifically, (2) psychosocial determinants of hackers' decisions to carry out an attack. These psychosocial determinants are indicative of the hacker him- or herself, such as the motives of disgruntled or angered employees to exact revenge, seek monetary rewards, or pursue similar goals by hacking a target company's information system (Greitzer et al., 2010; Greitzer et al., 2012).

Hacktivism and Anonymous

The term *hacktivism* was first coined in 1998 from members of the hacker organization cDc (Cult of the Dead Cow), who sought to share their sentiments against the perceived injustice surrounding the Tiananmen Square Massacre nine years prior

(Vegh, 2003). Whereas activists might share such sentiments through sit-ins, marches, and other embodied forms of protest, the developing network of the World Wide Web made it possible for cDc members to carry out such protests online, reaching a broader audience in a fraction of the time required for in-person demonstrations. This key difference between activism and hacktivism is no small matter: Although both forms of protest hamper the productivity of their targets, individuals who partake in such movements in person put themselves at a high degree of risk of public identification and arrest. Hacktivists, on the other hand, are better able to ensure anonymity and avoid detection. They are, in essence, given the ability to use cyberspace as a shield.

In the years that followed, hacktivism has evolved into a distinct, publicly recognized form of protest, commonly portrayed by hacktivists and the media alike as watchdogs that regularly perform valued functions for society (Rogers, 1999). Out of this movement rose the international hacktivist network Anonymous, recognized by *TIME* Magazine as one of the 100 most influential people in the world (Gellman, 2012). Anonymous, originally composed of members of the European hacker group Chaos Computer Club (Shakarian, Shakarian, & Ruef, 2013), states that it professes no ideology or creed (Hai-Jew, 2013); instead, it is constructed around a set of distinct principles and beliefs that have, since their inception in the mid-2000s, come to define them. More specifically, Anonymous states that its actions advance Article 19 of the United Nations Declaration of Human Rights (Dahan, 2013):

> Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive, and impart

> information and ideas through any media and regardless of frontiers (United Nations, 2007).

As such, these principles have been closely tied with Anonymous activity over the years, such that in the face of social injustice or restrictions of basic human rights (e.g., freedom of speech, information), members of Anonymous are likely to rise to action. Due to the dispersed and largely unidentifiable nature of this group, such actions are primarily coordinated through online forums such as 4chan.org, which guarantees full anonymity to its users, including the exclusion of user IP addresses to all but system administrators (Stryker, 2012). In the wake of a trigger event such as the shooting of Michael Brown in Ferguson, one or more members of Anonymous would post a call to action on 4chan with two key elements of information: (1) details of the trigger events precipitating the attack, and (2) details regarding how members should join in this attack (Massa, 2011). *Responses to these calls may range from dozens to tens of thousands of members depending on the trigger event and necessary actions, which begs the question of what factors underlying these events and actions are most likely to mobilize Anonymous members* (Brown, 2014; Phillips, 2013).

It might seem that a key determinant of response volume to Anonymous calls to action is the nature of the social injustice precipitating the attack; however, additional conjectures among hacktivist scholars suggest that members may be motivated by a desire for power, knowledge, challenge, and recognition (Hai-Jew, 2013), motives that suggest more a self-focused, rather than society-focused, end goal. Due to the heterogeneous nature of Anonymous and its many members, it seems doubtful that all Anonymous members across age, ethnicity, gender, and way of life are motivated to join in hacktivist attacks for the same reasons. In light of this predicted heterogeneity, it was necessary to utilize a bottom-up approach to identify several key underlying motives that capture past Anonymous attacks from its inception in 2005 to the present day.

Pilot Study 1

In a first pilot study, we examined a diverse sample of Anonymous-affiliated hacktivist attacks to glean and code for the key motives characterizing each. More specifically, we took into consideration both (1) the trigger events precipitating the attacks (e.g., political action perceived as social injustice) and (2) the methods utilized during the attacks (e.g., more playful, publicly visible actions may connote sensation-seeking motives[3]).

Materials. Although several branches of research have addressed specific case studies of hacktivist activities, it was important to gain an understanding of these case studies that was up-to-date at the time of data collection (i.e., and not limited by lengthy publication cycles), and accessible enough that various information sources could lend an overarching and unbiased understanding of the events that took place in each case study, rather than depend on a single, and potentially biased, source of information for each event. As such, Wikipedia served as a necessary tool to identify an updated timeline of Anonymous-associated events that were detailed by a diverse range of users. Cases were selected based on two key criteria: (1) detail surrounding each trigger event, and (2) heterogeneity of targets and causes within each event.

[3] Examples of methods that may imply sensation-seeking motives, rather than more sinister attacks against targets, include calling in mass orders for pizzas, taxis, and SWAT teams against a target such as the Church of Scientology in 2008 (Coleman, 2014).

First, the explanation of trigger events precipitating the attacks must have been detailed enough to suggest at least one reason underlying the attacks that followed. Similarly, the nature of the attacks must be discussed, even if only briefly. For example, “support[ing] a civil movement against corruption in India” (Timeline of events associated with Anonymous, n.d.) alone does not provide details of the aims or nature of the civil movement (e.g., retaliatory, suggesting a self-protection motive, versus an offensive measure, suggesting power against the government), nor to the nature of the corruption (e.g., a single person, suggesting third-party punishment, versus a government group, suggesting social injustice). Furthermore, no mention of the ways in which Anonymous has supported this movement is made, which leaves the reader unclear as the nature and magnitude of the attacks.

Second, some attacks were listed separately by target, even if the prior trigger events and timespan were identical. For example, members of Anonymous carried out attacks against a number of corporations and countries based on their reactions to—and censorship of—Wikileaks; similarly, the multiyear Occupy Wall Street movement has involved a variety of police forces and organizations at distinct, protest-related time points around the country. These examples were, for the sake of creating a detailed case study for each set of events, combined into single case studies.

This collection process yielded 25 distinct case studies spanning nine years (i.e., mid-2005 to October, 2014). These studies were summarized into blurbs of approximately similar length (*M* = 82.12 words, *SD* = 26.92 words) based on facts obtained and verified through multiple news and media sources. Case studies are displayed in Appendix A.1. It is worthy of note that these 25 case studies may not be fully representative of past Anonymous activity. The ease with which details about each case study could be tracked and noted should be directly indicative of its publicity, whereas lesser known Anonymous actions may have evaded our detection. If this is the case, there may be third variables at play that partially determine whether an Anonymous attack will grow more public, such as the nature of the target (e.g., government, well-known corporation) or the attack (e.g., scandalous or emotionally evoking material; Berger & Milkman, 2012). The threat of seeking and including less publicized events, however, is falsely attributing a lesser-known act to Anonymous without evidence of the group's involvement (e.g., instead of a separate hacker group spurred by ulterior, non-hacktivist motives). Instead, each of the 25 case studies examined were confirmed through online accounts or interviews to have been directly attributed to Anonymous.

Coding. Six trained undergraduate research assistants separately examined each case study and reported themes and motives that appeared to underlie each event. These responses were compiled by similarity (e.g., status with power motives, vigilantism with third-party punishment) into six key motives that fall into two key groups of rationales: personal significance (i.e., self-focused) and societal significance (society-focused). These motives are shown in Table 1.

Table 1. *Personally and societally significant motives*

Personal Significance	**Societal Significance**
Self-protection	Third-party punishment
Status, power	Social injustice
Sensation-seeking	Economic redistribution

We worked with our trained coders to define, through extant literature and the case studies before them, each of these six motives, which were subsequently outlined for use in future coding. We define *self-protection* as a desire to protect oneself from harm or negative consequences, often taken as a reactionary measure against personal threat. *Status, power* can be described as the seeking of power for oneself (i.e., to enhance one's own self-worth) or among a group of others. *Sensation-seeking* denotes a desire for experiences and feelings that are varied, novel, complex, and intense, or a willingness to take risks out of boredom or curiosity. *Third-party punishment* describes a third party's desire to punish a person or identifiable group of people for violating social norms (i.e., vigilantism, "playing God"). Unlike our social injustice motive, this motive implies a strict morality, sinning, or criminality component. *Social injustice*, on the other hand, is defined as a desire to correct unfairness or injustice on a societal level, perhaps to maintain personal rights. Unlike third-party punishment, social injustice holds a strict legal or unjust component. Lastly, *economic redistribution* entails a desire to redistribute wealth or resources to bring disadvantaged groups to equal standing with a separate, more privileged population.

Coders then used this list of definitions, a more extensive version of which is found in Appendix A.2, to classify each of the 25 case studies along 9-point Likert-type scales for each of the six motives. They were then asked to identify the single motive that most strongly defined each case study, the results of which are displayed in Table 2. Coders reached a high mean intraclass correlation coefficient[4] of $\alpha = 0.925$ ($F_{\text{mean}}[5,25] =$

[4] The intraclass correlation coefficient (ICC) captures the reliability of ratings across multiple coders. Each case study was rated by the same collection of raters. To calculate

17.974, $p < .001$), with α values ranging from 0.779 to 0.977 and 76 percent of case studies garnering an α greater than or equal to 0.90.

Table 2. *Prevalence and strength of motives in Anonymous case studies and slogans*

Motive	**Case Studies**		**Slogans**	
	Absolute	*M (SD)*	*Absolute*	*M (SD)*
Self-protection	2	3.04 (2.46)	1	5.06 (2.89)
Status, power	0	5.60 (1.99)	8	7.37 (1.97)
Sensation-seeking	0	3.75 (2.27)	1	3.92 (2.87)
Third-party punishment	6	6.44 (2.54)	0	3.44 (2.77)
Social injustice	17	7.45 (2.32)	1	4.20 (2.90)
Economic redistribution	1	1.93 (2.12)	0	2.12 (2.32)

Of these results, three motives demonstrated particular strength and prevalence across case studies: social injustice (17 case studies, $M = 7.45$, $SD = 2.32$), third-party punishment (6 case studies, $M = 6.44$, $SD = 2.54$), and status/power (0 case studies[5]; $M = 5.60$, $SD = 1.99$). Respectively, these three were significantly stronger than the remaining three motives, $11.25 \leq t(146) \leq 22.33$ status/power, $19.44 \leq t(148) \leq 29.00$ third-party punishment, and $12.92 \leq t(148) \leq 21.68$ social injustice, collective $p < .001$. The results of our first pilot study therefore suggest that various self- and society-focused motives may indeed have precipitated past Anonymous activities. If we were to focus particularly

ICC, we utilized a consistency model in which systematic differences between raters (e.g., if one rater consistently rates one degree lower than another rater) are held constant. An α of 1.0 indicates perfect agreement across raters for the same case study.

[5] Although no case studies were described predominately as motivated by status/power, this motive remained the next highest motive by mean rating, and by a statistically significant degree of separation (see Table 2).

on the single strongest and most prevalent motive of each of these foci (i.e., self and society), we would be left with a desire (1) for status and power and (2) to correct cases of social injustice as key motives underlying Anonymous activities.

We plotted the coded values of each of the six motives as a function of time to observe whether any temporal trends existed across our 25 case studies. The resulting graph, which is depicted in Figure 1, demonstrates varying polynomial trends by motive: Social injustice and economic redistribution appear to increase in strength over time; sensation-seeking and third-party punishment appear to decrease; and self-protection and status, power seem to increase temporarily before returning to previous levels of prevalence within our sample of case studies. On average, these polynomial trends explain 14.48 percent of observed variance, with R^2 values ranging from .079 (sensation-seeking) to .256 (third-party punishment).

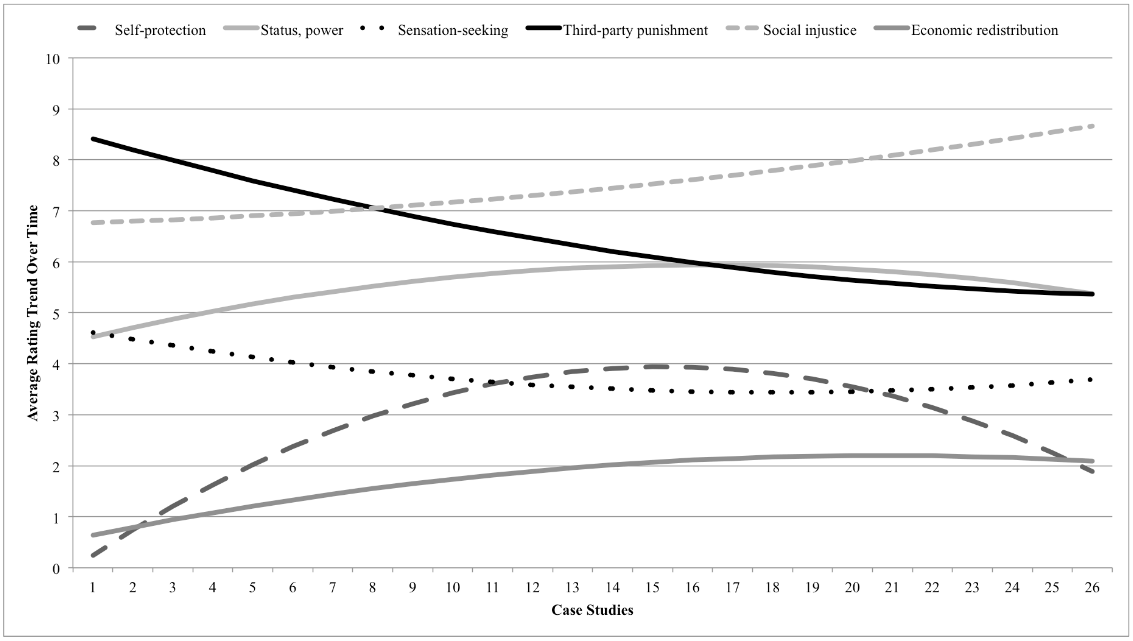


Figure 1. Polynomial trends in motive strength over time

In our first pilot study, we posited that Anonymous members are mobilized by contextual factors surrounding (1) the trigger events that precipitate their actions, and (2) the nature of their actions that follow. It would, however, be imprudent to assume that these are the only sources of influence in Anonymous activity; rather, it seems more likely that Anonymous members actively seek—rather than only passively respond—to convey their message to others. As we have explained, Anonymous calls to action appeal to members, targets, and broader society through the group's self-proclaimed principles, which embody a desire to correct social injustice and wrongdoing. As such, it was next imperative to examine the motives that Anonymous itself appears to communicate to the world, both within its borders (ingroup members) and beyond (outgroup members).

Pilot Study 2

In a second pilot study, we examined the slogans Anonymous most often uses to broadcast its message, as well as the motives that precipitate their actions. Anonymous primarily communicates through 4chan and similar forum channels, but also through Twitter feeds, websites, and YouTube channels that are widely accessible to the public. Furthermore, Anonymous communicates its presence through two key media: (1) iconography, such as the widely recognized headless man in a suit (Coleman, 2014), and (2) slogans, or sayings, commonly associated with the videos, Tweets, and text messages associated with Anonymous. Examples of such slogans include "We are Legion," "We do not forgive, we do not forget," and "Expect us" (Shakarian, Shakarian, & Ruef, 2013).

Materials. As a starting point, we chose to examine text slogans associated with Anonymous rather than affiliated iconography or symbols due to their inherently multivariate nature as primes. More specifically, symbols boast a complexity that may have introduced a number of confounds within our pilot test depending on a wide range of interpretations, varying degrees of intricacy or detail across different symbols, and coders' own personal and cultural backgrounds, which could color their perceptions of visual stimuli. As such, Anonymous-associated slogans were exclusively utilized in our second pilot study.

Slogans were identified through three key sources of information related to Anonymous: a book chapter centered around its growth and presence as a major hacktivist group ("Cyber attacks by nonstate hacking groups: The Case of Anonymous and its affiliates"; Shakarian, Shakarian, & Ruef, 2013), a documentary focusing on the workings and beliefs of Anonymous as a collective (*We are legion: The Story of the hacktivists*; Knappenberger, 2012), and a novel offering an anthropological perspective of Anonymous as a culture (*Hacker, hoaxer, whistleblower, spy: The Many faces of Anonymous*; Coleman, 2014). Eleven key slogans were identified that were explicitly mentioned in at least two of the three sources, indicating popularity and prevalence of these slogans beyond a single case study or trigger event. These slogans are listed in Appendix A.3.

Coding. The same six trained undergraduate coders who took part in Pilot Study 1 examined each of the 11 key slogans, and were asked to focus particularly on word choice and use of capitalization and punctuation (i.e., to emphasize importance of a word or phrase). Slogans were scored on a 9-point Likert-type scale along each of the six key motives previously identified, namely: self-protection, status/power, sensation-seeking, third-party punishment, social injustice, and economic redistribution. Coders then identified the single motive that each slogan seemed to embody most strongly.

Results of these analyses are displayed in Table 2. Coders reached a strong mean intraclass correlation coefficient of α = 0.868 ($F_{\text{mean}}[5,25] = 13.930$, $p < .001$), with α values ranging from 0.756 to 0.986.

Of the six motives identified in Pilot Study 1, status/power emerged most strongly and most prevalently throughout the 11 slogans assessed (8 slogans; $M = 7.37$, $SD = 1.97$). Indeed, this single motive described slogans significantly more strongly than the remaining five motives, $9.47 \leq t(64) \leq 21.54$, $p < .001$. Examples of status/power-coded slogans include “The people should not be afraid of their government. The government should be afraid of its people” and “You want to see Anonymous rise up? Try to shut down its message. Then you’ll see what Anonymous can do.” Of the society-focused motives (i.e., third-party punishment, social injustice, economic redistribution), again social injustice emerged as the strongest and most prevalent motive of the three (1 slogan; $M = 4.20$, $SD = 2.90$), $2.12 \leq t(64) \leq 5.79$, $.001 \leq p \leq .038$. Examples of social injustice slogans include “We stand for freedom of speech, the power of the people.” It would therefore seem that Anonymous, through text-based messages targeted toward both its ingroup (e.g., members, supporters) and its outgroup (society, targets of attacks, government), seeks to utilize slogans that either prime a sense of power or a concern for societal justice.

In addition, coders noted a clear theme that arose throughout a majority of these 11 Anonymous slogans: the frequent use of the first-person plural pronoun *we*. Indeed, a linguistic analysis of these slogans using the Linguistic Inquiry and Word Count software (LIWC; Pennebaker, Booth, & Francis, 2007) indicated a particularly high prevalence of social words (e.g., *we*, *our*), with a rate of 16.07 compared with a national average of

9.50 in personal texts and 8.00 in formal texts. Self-references (e.g., *I*, *me*, *my*; rate = 8.04), on the other hand, did not differ significantly from personal texts (11.40) or formal texts (4.20). Full results from this brief analysis are displayed in Table 3, which indicates the following: the seven primary dimensions of the LIWC analysis, the collective score obtained by our 11 Anonymous slogans, the average scores of typical personal and formal texts (as per Pennebaker, Booth, & Francis, 2007), and the number of words the LIWC software assesses to arrive at the scores depicted. For example, LIWC analyzes text samples for 730 words related to cognitive processes, including words pertaining to insight, causation, discrepancy, certainty, and inhibition.

Table 3. *Linguistic analysis of 11 Anonymous slogans*

LIWC Dimension	**Slogans**	**Avg. Personal Texts**	**Avg. Formal Texts**	**# Words**
Self-references	8.04	11.40	4.20	12
Social words	16.07	9.50	8.00	455
Positive emotions	1.79	2.70	2.60	406
Negative emotions	3.57	2.60	1.60	499
Cognitive words	10.71	7.80	5.40	730
Articles	8.04	5.00	7.20	3
Large words	13.39	13.10	19.60	≥ 6 letters

Of these dimensions, self-references include *I*, *my*, *me*, and *mine*; social words include *us*, *we*, *they*, and *our*; positive words include *love*, *nice*, and *sweet*; negative words include *hurt*, *ugly*, and *hate*; cognitive words include *cause*, *know*, and *ought*, and articles include *a*, *an*, and *the*.

Table 3 suggests that the use of social words (i.e., *we*, *our*) in our 11 Anonymous slogans is approximately twice that of average personal and formal texts—a more substantial increase than is evident in any of the remaining six dimensions. Within social psychological research, the use of first-person plural pronouns such as *we* has accompanied an array of findings related to two key frameworks of thought: (1) the collective or social self, as opposed to the strictly personal self, as delineated through Social Identity Theory; and (2) the loss of the personal self in the context of a group or collective, known as *deindividuation*.

Deindividuation and Social Identity Theory

A *social identity* describes the segment of an individual's self-concept that is derived from his or her membership in a social group (Turner & Oakes, 1986). Social Identity Theory holds that an individual is not limited solely to a personal self, but rather to several selves that correspond to various group memberships (Tajfel & Turner, 1979). As such, in different social contexts an individual may think and behave on two distinct self-construals, or "levels of self" (Turner, 1987)—namely, their personal self (e.g., derived from factors unique to the individual) and the self associated with group membership, or their ingroup. This distinction may, therefore, be summarized as a self-definition based upon *we*, rather than *I*.

Past research has explored the influence of self-construal priming on cognition, in which experimentally altering the salience of an individual's personal or group self may lead individuals to adopt individualistic (*I*-focused) or collective (*we*-focused) mindsets and values (Gardner, Gabriel, & Lee, 1999). As such, the substitution of *we* in place of *I* may have profound consequences on the ways in which individuals construe a given situation. An example of one such consequence is *deindividuation*, in which the salience of one's personal self is eclipsed by the aims and principles of one's larger group. This change in salience may lead a group member to adopt new values, behaviors, and social norms, freeing them from prior social restraints and making them more likely to exhibit behaviors that they might normally inhibit (Festinger, Pepitone, & Newcomb, 1952; Diener, 1979; Zimbardo, 1971).

Philip Zimbardo was among the first social psychologists to develop a model of deindividuation (Zimbardo, 1969), which he later tested in his well-known but controversial Stanford Prison Experiment. Over the course of six days, university students who were randomly assigned to adopt the roles of mock prison guards who began to subject mock prisoners (i.e., other university students) to extreme authoritarian measures and, in some cases, psychological torture (1971). Zimbardo held that this rapid adoption of a new and group-defined self superseded participants' previously held personal morals and values. As such, he states in his model of deindividuation that certain social and contextual triggers can lead to a temporary state of suspended personal identity, a theory that in decades since has been used to explain vandalism, graffiti, and gang behaviors (2004).

With the increasing prevalence of personal computers and the World Wide Web, however, researchers began to adapt this model of deindividuation to fit a society in which complete anonymity was, and is, becoming more easily possible. The Social Identity model of Deindividuation Effects (SIDE; Lea & Spears, 1991) has since become a primary model of deindividuation within the realm of computer-mediated communication for its emphasis on anonymity and reduced identifiability through the Internet (Chan, 2010; Lee, 2007; Postmes et al., 2001; Spears et al., 2002). Whereas previous models of deindividuation (e.g., Diener, 1979; Zimbardo 1969) emphasized social contexts in the physical world, the SIDE model has stressed the strength of social context in cyberspace, theorizing that anonymity online may change the relative salience of personal and group identity and thereby impact subsequent behavior (Postmes, Spears, & Lea, 1998).

Impact on subjective risks. In the present research, we expect the anonymity inherent in Anonymous activities to shift the salience of members' personal selves into a sense of self more closely tied with Anonymous as a social group, albeit a geographically dispersed one. In so doing, the prevalent use of *we* in Anonymous slogans may prime ingroup members—as well as members of their outgroup, such as society members or potential targets—with a sense of a wide-reaching, cohesive, and inherently faceless collective.

As has been previously stated, the risks and payoffs involved in a given cyber-attack should, as per the theoretical framework of game theory, be an important consideration when deciding whether to carry out such an attack. Because hackers are rarely informed of the exact risks and payoffs involved, however, these values must be estimated and are therefore subjective in nature. We posit that these estimations are conditional upon contextual factors that may sway perceptions of the riskiness or benefits inherent in a pending hacktivist attack.

Referring specifically to the use of *we* so common in Anonymous calls to action, we expect that exposure to *we*, as opposed to *you*, language may shift a member's salience of personal self to that of his or her group self, resulting in a dispersion effect whereby the risks characterizing a hacktivist attack are carried by all members of the group, rather than by the single individual who must bear those risks alone. As such, we predict that the use of *we* results in a deindividuation effect largely driven by group identity salience and the anonymity inherent in most Anonymous communication channels. This deindividuation primes a dispersed sense of risk—that is, the individual considers him- or herself to be just another face in the crowd, and therefore less easily identifiable—and therefore perceives a lower sense of personal risk should he or she choose to carry out the attack.

On the other hand, the use of *you* language in Anonymous calls to action should result in the opposite effect, in which a member's personal identity salience grows stronger, and his or her salience of group identity temporarily wanes. Should this be the case, a frequent use of *you* should focus the subjective risks involved in a hacktivist attack solely on the individual, leading to an overestimation of these risks in comparison.

Impact on subjective payoffs. Alongside varied risks, hacktivism also entails an end goal acquired only after the successful completion of the necessary attacks. These end goals, or payoffs, are also subjective in nature and must therefore be estimated prior to engaging in the attack.

Just as we expect the use of *we* to lead to a perceived distribution of risk, it may also emphasize the wider array of benefits that the broader ingroup—here, Anonymous and the message it is fighting to convey—may possess as a collective. Here, it is important to distinguish between *you*, being the payoffs that only the individual contemplating joining an attack will possess, and *we*, being the payoffs that both the individual and the broader group will jointly possess should the attack prove successful. Because individual payoffs are a baseline, the addition of payoffs for one's ingroup may serve as an additional advantage.

As such, we expect the prominent use of *we* in Anonymous calls to action to make more salient the varied payoffs experienced by both the individual and the larger group, therefore leading to an increase in subjective payoffs overall.

Motives: Power and Justice

In the early 1940s, psychologist Abraham Maslow posited in his well-known paper "A Theory of human motivation" that we—as humans—require a basic set of needs that, once obtained, allow us to seek further needs that are less necessary but still desirable during the lifespan (Maslow, 1943). More specifically, he states that only after physiological and safety needs are met (i.e., food, water, health; personal and financial security) can we seek needs related to love and belonging, followed by esteem-relevant needs, and ultimately the desire to self-actualize by realizing one's full and unique potential in life.

Power motives and self-enhancement. More recently, evolutionary social psychologists have revisited Maslow's original hierarchy of needs to incorporate the importance of immediate situational threats and opportunities that may lead an individual to place higher import on motives related to the context at hand (Kenrick et al., 2010). Within this renovation of Maslow's hierarchy (i.e., the fundamental motives) is a particular focus on Maslow's concept of self-actualization, which the authors state is largely subsumed within a fundamental human desire for status and power. This *power motive* describes the common human search for others' respect and esteem, which—once obtained—allows the individual to influence the thoughts and behaviors of others (Hofer & Chasiotis, 2011). Past research suggests that individuals primed with power motives often exhibit behaviors more likely to garner additional respect and admiration from others, including increased aggressive and risk-taking behaviors (Anderson & Galinsky, 2006; Griskevicius et al., 2009; Griskevicius & Kenrick, 2013).

Power motives are, in this sense, a strategy toward self-enhancement. *Self-enhancement* describes the motivation to feel positively about oneself, thereby maintaining satisfactory levels of self-esteem. Past research suggests that human desires for self-enhancement grow more salient in situations of risk, threat, or competition (Kwan, Kuang, & Zhao, 2008), even to such a high degree that one's self-views become overly, and inaccurately, positive (Beauregard & Dunning, 1998; Crocker et al., 1987). A wealth of extant research supports that self-enhancement motives are more powerful than any competing self-serving motive, including self-verification (i.e., the desire to be perceived according to firmly held values; Swann, 1983) and self-assessment (i.e., the desire to assess and identify the aspects most important to one's own identity; Trope, 1983); stated simply, self-enhancement appears to be a "cornerstone" motive throughout life (Sedikides & Gregg, 2008).

Impact on subjective risks. Not only have power motives been linked with increased risk-taking behaviors, but they may also play a role in risk estimation. Research on power priming suggests that those in a high-power mindset are unrealistically optimistic in their perceptions of risk, and therefore more likely to demonstrate risk-taking behaviors (Anderson & Galinsky, 2006). The *optimism bias* describes the phenomenon by which, when placed in a threatening situation, an individual is optimistically biased to believe that he or she is less at risk than others in the same situation (Helweg-Larsen & Shepperd, 2001). Research on this decision-making fallacy suggests that one factor that may precipitate the optimism bias is self-enhancement, which causes individuals to foresee the outcome of a risky situation to be more optimistic in their favor, sacrificing realistic estimations of risk for an estimation that affirms the individual's need for high esteem (Shepperd et al., 2002).

We would therefore expect that, in line with research on the influence of power motives on risk estimation, the use of Anonymous slogans that convey power and status may lead members to underestimate risk. Hacktivists who may feel powerless in their present situation—either due to, or in addition to, the trigger events leading to the call to action (e.g., Internet censorship, government corruption)—may respond readily to suggestions of power; as such, participants who view a power-inducing slogan may be more likely to perceive a risky situation as less risky, and subsequently be more likely to consider joining in the attack.

Impact on subjective payoffs. It has been stated that, as per Maslow's hierarchy of needs and the more recent evolutionary psychological fundamental motives, humans are inherently self-interested. Only after one's own physiological and safety needs have been met should considerations for others—for example, kin or an individual's broader ingroup—grow more salient. Regarding power in particular, research suggests that individuals primed with power display an increased focus on what they stand to gain from their actions (e.g., in a game of blackjack; Galinsky, Gruenfeld, & Magee, 2003;

Lammers, Stapel, & Galinsky, 2010). We might, therefore, expect that participants primed with power motives will perceive the payoffs stemming from a successful hacktivist attack to be much higher because the primary person benefitting is the individual him- or herself. As such, we predict that power motives will highlight an individual's personal benefits, thereby leading that individual to overestimate the payoffs involved in carrying out a hacktivist attack.

Justice motives and Social Justice Theory. Relative Deprivation Theory describes the phenomenon by which individuals are motivated to correct examples of perceived injustice on three key levels: (1) personal injustice, or one's own deprivation compared with other members of their own group; (2) fraternal injustice, or their ingroup's deprivation compared with outgroups; and (3) third-party injustice, or the deprivation of unaffiliated others or society as a whole (Crosby, 1976; Jennings, 1991). But whereas Relative Deprivation Theory posits that human judgments of injustice are derived from social comparison (i.e., person to person, group to group), research on the *justice motive* supports a more selfless, moralistic approach toward perceptions of injustice (Greenberg, 1987).

Since the early 1980s, political psychologists have argued that justice is a preeminent concern among humans and is therefore an expression of a fundamental requirement of society (Lerner, 1975). Social Justice Theory therefore holds that, similar to Relative Deprivation Theory, humans are motivated to correct injustice even as a third-party—that is, instances of unfairness that do not directly affect individuals themselves; but in addition, it states that the origins of this innate desire for social justice stem from a preference for a society in which rules are followed, corruption is punished, and virtuosity is justly rewarded (Lerner, 1977), particularly when group membership or placement within society is made salient (i.e., the relational model of justice; Tyler, 1994).

Impact on subjective risks. Relative Deprivation Theory posits that in the face of social injustice, individuals will be motivated to correct the unfairness at hand, undergoing either legal or illegal collective action to do so. However, later attempts to experimentally test this supposition found the opposite: Even when primed with high levels of social injustice, participants' willingness to engage in legal or illegal collective action to right perceived wrongs remained unaffected (Martin, Brickman, & Murray, 1984). The authors contended that one possibility for this null finding was participants' perceived personal costs in engaging in collective action, particularly if illegal in nature; however, this possibility has not yet been empirically tested. Because mass reactions against social injustice may draw the attention of authorities seeking to quell such uprisings, individuals may perceive collective action to be a particularly risky endeavor. If this is indeed the case, we might expect primes of social injustice to increase participants' perceptions of risk.

Impact on subjective payoffs. A second hypothesis stemming from Social Justice Theory posits that people seek justice as a by-product of purely self-interested motives—more specifically, it is easier to maximize one's own outcomes and acquire desired resources in a society that is just and fair (Lerner, 2003). Should this hypothesis hold, we might expect that a social injustice prime would increase the salience of an individual's own personal benefits, thereby highlighting the payoffs resulting from a just solution to an unjust situation. As such, we predict the presence of Anonymous slogans priming justice motives to lead participants to perceive greater payoffs for self-interested purposes than they might in the absence of such a prime.

Research Question

To revisit the primary goal of the present study, we seek to explore the following question: First, how do contextual cues (i.e., language and slogans used by Anonymous) activate (1) senses of deindividuation and (2) power or justice motives? Second, does the activation of these two factors influence hackers' perceptions of risks and payoffs, and thus their perceived likelihood of carrying out a hacktivist attack? In equation form, we will explore the possibility that:

$$P(\mathrm{A}) = f(\mathrm{R_S}) + f(\mathrm{P_S}) + f(\mathrm{R_S P_S}) + \ldots \tag{1}$$

whereby $P(\mathrm{A})$ signifies probability of attack (i.e., attack likelihood), which is a function of subjective risks ($\mathrm{R_S}$), subjective payoffs ($\mathrm{P_S}$), and the interaction of the two ($\mathrm{R_S P_S}$), along with any additional factors not discussed in the present study.

Hypotheses

We organized our overarching research question into four key hypotheses:

H1A: *Deindividuation and risks*. We predict that the frequent use of *we* in Anonymous calls to action will prime participants to feel deindividuated among the perceived collective of Anonymous, therefore perceiving risk to be distributed evenly among other respondents to the call, and therefore lower to the participants themselves. We expect *you* language, on the other hand, to have the opposite effect, singling out the participant and creating a sense that they alone face the risks underlying the attack.

H1B: *Deindividuation and payoffs*. The prominent use of *we*—as a conjunction of the individual and the larger group—in Anonymous calls to action may increase the salience of payoffs that both the participant and the group stand to gain, therefore leading to an increase in subjective payoffs overall. *You* language, however, may highlight only personal payoffs. Although it is possible that we might heighten individuals' self-interest by priming a self-focused pronoun, it remains the case that *we*, here, benefits both the self and the larger group. Therefore, due to the singular nature of *you* in the absence of group payoffs, we expect the use of *we* to be more powerful.

H2A: *Motives and risks*. In line with research on the influence of the optimism bias and power motives on risk estimation, we expect that Anonymous slogans conveying power and status may lead participants to estimate lower risk than when this motive is not activated.

As per Relative Deprivation Theory, we predict that individuals will perceive a hacktivist attack to be a more highly risky endeavor when primed with social injustice, due to a perceived increase in the personal costs involved in engaging in illegal collective action.

We expect that participants in the combined power and justice motive condition will fall in the middle of these two groups on risk perception, due to a hypothesized decrease in the face of power motives and a hypothesized increase following exposure to justice motive primes.

H2B: *Motives and payoffs*. Because the primary person benefitting from a successful attack is the individual him- or herself (i.e., personal benefits), we expect that participants primed with power motives will perceive attack payoffs to be much higher.

As per Social Justice Theory, we predict that a social injustice prime will also increase the salience of an individual's own personal benefits, thereby highlighting the payoffs that would result from restoring justice to a situation or society; however, we expect the extent of this increase in payoffs to be of a lesser degree than that of a power motive prime, which is regarded as one of the most influential motives yet studied in psychological research.

We expect that joint exposure to both power and justice motive conditions will yield an additive impact on payoffs, such that participants will perceive the highest payoffs when exposed to both motive primes.

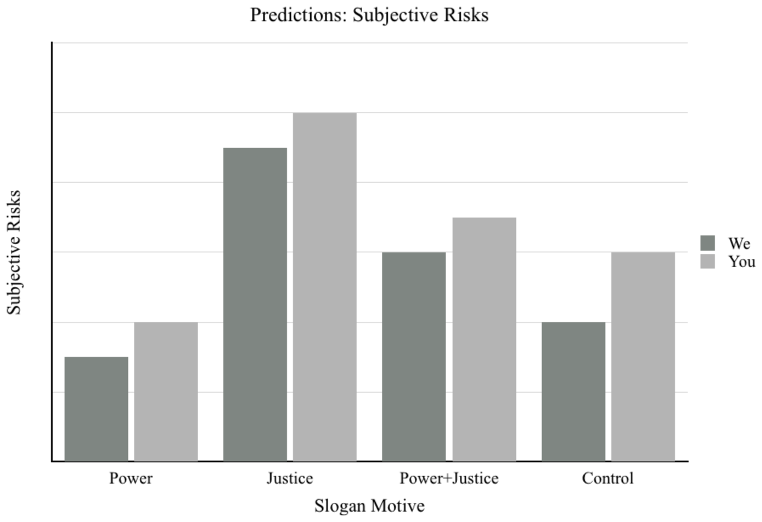


Figure 2. Hypothesized subjective risks by condition

We predict that participants randomly assigned to the control condition will maintain their predisposed sense of subjective payoffs, such that we expect control participants to perceive fewer payoffs than participants in either the justice or power motive conditions.

H3A: *Additive effects on risks*. We

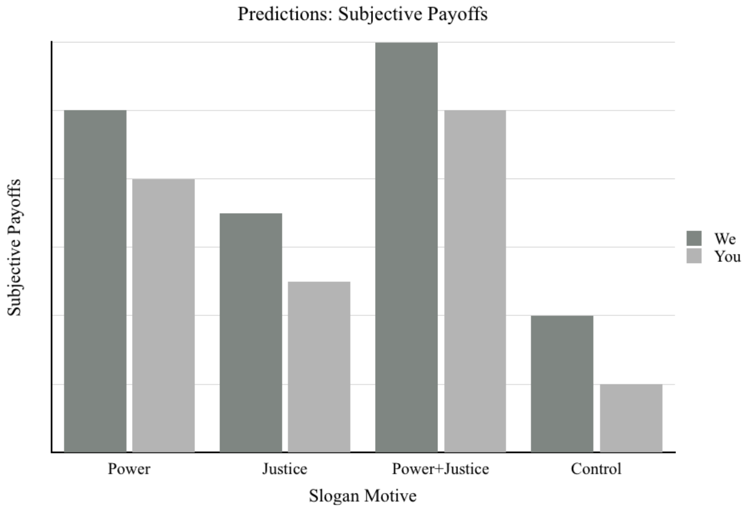


Figure 3. Hypothesized subjective payoffs by condition

expect that the effects of our individual manipulations (i.e., deindividuation and motive) may have an additive effect when combined, such that the *we* (deindividuation) condition and the power motive condition—both hypothesized to yield low subjective risks compared to other conditions—should, combined, yield particularly low subjective risks. As such, we further hypothesize that the condition yielding the highest perceived risks will be the *you*/individuation and justice motive condition, whereas participants in the joint (power and justice) and control conditions—that is, the absence of a motive prime—will fall between these two risk estimations, regardless of deindividuation condition (*we* or *you*). Figure 2 depicts Hypothesis 3A in graphical form.

H3B: *Additive effects on payoffs*. Similarly, we expect the two conditions that we hypothesize to have the lowest subjective payoffs—namely, the *you*/individuation condition and control condition (i.e., no motive)—to jointly yield the lowest perceived payoffs overall. Priming of *we* language and power motives, on the other hand, should yield high subjective payoffs, whereas the justice motive condition—regardless of deindividuation condition—will fall between these two payoff estimations. We predict that the highest perceived payoffs, regardless of deindividuation

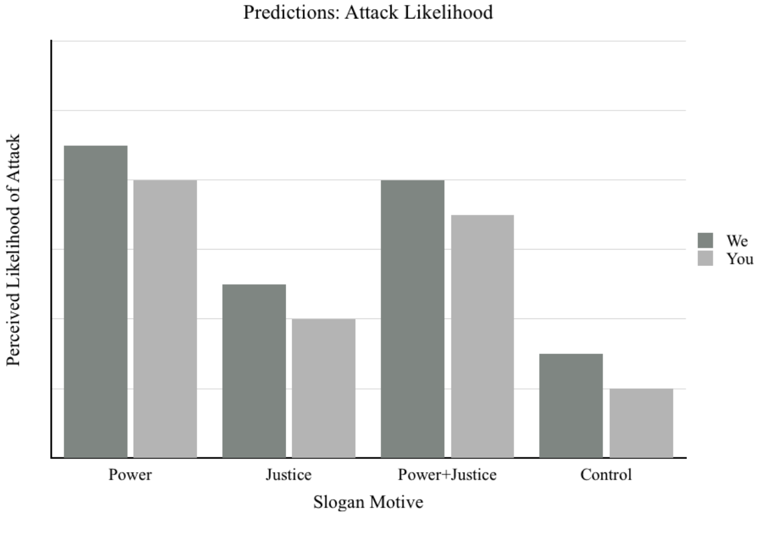


Figure 4. Hypothesized attack likelihood by condition

condition, will stem from participants in the joint motive condition, whose payoff estimation should take into account the increased payoffs that we expect to follow exposure to both power and justice motive primes. H3B is displayed graphically in Figure 3.

H4: *Attack likelihood*. We expect participants to form mental ratios of the risks and payoffs inherent in a given hacktivism scenario, and that this ratio will help guide their overall perceived likelihood of attack; that is, as illustrated in Equation 1, we expect participants to consider an attack more likely as risks decrease and payoffs increase. On the other hand, an attack should be most likely when risks are low and payoffs are high. Thus, building off Hypotheses 3A and 3B, Figure 4 summarizes our fourth hypothesis in graphical form.

Table 4. *Summary of primary hypotheses by condition*

Estimation	**Deindividuation**	**Motive**	**Additive Interaction**
Risks	H1A: We < You	H2A: P < C < J+P < J	H3A: We lower; P lowest, J highest
Payoffs	H1B: You < We	H2B: C < J < P < J+P	H3B: We higher; C lowest, J+P highest
P(Attack)	H4: You/We + C < You/We + J < You/We + J+P < You/We + P		

C = Control (no motive); *J* = Justice motive; *P* = Power motive; *J+P* = Joint motives

Collectively, these hypotheses are summarized in Table 4. We acknowledge that the combined effects of our two conditions may not be additive in nature, but rather dynamic—that is, multiplicative—such that the power of one condition may change depending on its joint appearance with another condition. Following data collection, we will examine whether our interaction effects imply any trends that we did not directly hypothesize.

Method

Participants

Undergraduate students enrolled at Arizona State University were invited to participate in an online experiment that was delivered through Qualtrics online survey software. Participants were recruited if they were currently majoring in computer science, computer science engineering, informatics, information systems, and related disciplines. A list of venues that were used to identify and recruit participants can be found in Appendix B.1. Our target participant audience comprised students from computer and information science backgrounds due to the similar overlap in demographic variables with many Anonymous members and other hacker groups. More specifically, Anonymous members, hacktivists, and general hackers are particularly likely to be male, approximately college-aged, of slightly above-average digital literacy, and of a diverse range of races and ethnicities (Yar, 2013).

It is worthy of note that participants were not intended to be of highly above-average digital literacy. Research suggests that those who engage in hacking activities at the level of the present study (i.e., not mass-scale or state-affiliated cyber-attacks) are likely to be of only slightly above-average digital literacy due to the fact that highly skilled hackers can obtain better payoffs by engaging in more rewarding forms of cyber-attacks (Shim, Allodi, & Massacci, 2012). As such, we expected the range of computer science (and related) students, regardless of grade-point average or year in college, to hold sufficient knowledge about technology to fit the demographic build of those we might expect to use 4chan and similar Anonymous communication channels, recognize and consider partaking in Anonymous calls to action, and consider themselves "Anons" (i.e., members).

In return for their participation, each completed survey was submitted as a raffle ticket for one of six Roku 3 streaming media players[6]. This experiment was available until the beginning of the week during which the funding agency providing participant incentives required proof of raffle prize receipt, at which point the experiment was closed from further participation. We witnessed a high participant attrition rate throughout this period, in which approximately 16 percent of participants who opened the experiment did not proceed beyond the first page, which presented a consent statement required by the IRB and outlined the topic and expected length (i.e., 15 minutes) of the current study. As such, by the time we closed the experiment 644 individuals had opened the study, whereas only 542 had answered at least one item.

These 102 participants were excluded from analyses, along with 68 participants who failed a catch question placed toward the end of the study (i.e., *If you are reading this, please select 100*), yielding a final sample size of 474.[7] Among these students, we

[6] The Roku 3 was chosen to accommodate requirements for tangible, non-monetary participant rewards totaling less than $100 each from the Jumpstart Research Grant for graduate research at Arizona State University. As one of CNET's top-rated and least expensive gadgets of 2014, the Roku 3 was expected to be a desirable method of participant recruitment.

[7] Boxplots of each of our three key analyses (i.e., the impacts of our manipulations on subjective risk, subjective payoff, and estimated attack likelihood) yielded no evidence of extreme cases or outliers within our sample, with the exception of a single participant in one instance, who reported a level of subjective risk more than two standard deviations below the sample mean in the *you*-justice condition. For all remaining analyses, this participant answered within a standard range of responses. All boxplots can be found in Figures 5, 6, and 7 in Appendix C.

observed a mean age of 22.60 (*SD* = 3.36) and modal age of 19, 43.9 percent of whom were female. 63.8 percent of our sample identified as White, 17.1 percent as Asian or Middle-Eastern, and 62.0 percent hailed from middle or upper-middle classes of an annual household income of at least $50,000.

Basic Design

The present study adopted a 2 × 4 between-subjects analysis of variance design (ANOVA), whereby our deindividuation manipulation was either *we* or *you* (two levels) and our motive manipulation could be one of four possibilities: power motive, justice motive, joint motives (i.e., both power and justice), or control (three levels). Participants randomly assigned to the control condition saw neither power motive- nor justice motive-inducing Anonymous slogans. There were, therefore, eight possible conditions to which participants were randomly assigned through a block randomization command in Qualtrics. These eight conditions are outlined in Table 5 for purposes of clarity.

Table 5. *Summary of eight experimental conditions*

	We (Deindividuation)	**You (Individuation)**
Power motive	We + Power motive	You + Power motive
Justice motive	We + Justice motive	You + Justice motive
Power + Justice motive	We + both motives	You + both motives
Control	We + Control (no motive)	You + Control (no motive)

Procedure. Through the channels listed in Appendix B.1, participants were invited to take part in an online experiment. Participants first read a scenario that aimed to induce them into the frame of mind of a student who happens across an Anonymous call to action, which was randomly manipulated as per our eight conditions. Participants then answered a series of questions regarding perceived risks, payoffs, overall attack likelihood, covariates, and demographic variables.

Materials

Guided visualization scenario. Each participant viewed a short, guided visualization scenario asking him or her to adopt the mindset of a person who was similar in age range, university affiliation and major, and gender[8]. Guided visualization (also known as guided imagery) techniques were first empirically used in clinical psychology in the mid-1980s to reduce anxiety and enhance work performance by imagining oneself in a situation that is less physically or psychologically threatening (Ayres & Hopf, 1985; 1990; 1992). In the decades since, other areas of psychology have adopted guided visualization as a strategy to increase the vividness and clarity of participants' mental imagery through text-based prompts. We utilized this third-party, guided visualization approach—that is, one in which participants consider the actions of an imagined other—to avoid the possibility that participants would withhold information or opinions due to conflicting interests, morals, or values. For example, some participants might have felt uneasy stating that they, in a particular situation, would take part in a hacktivist attack that in real life may lead to fines and incarceration; also, and beyond concerns of legal safety, participants who viewed hacking as an immoral activity regardless of the situation may have reflected only these anti-hacking views in their responses.

[8] We manipulated gender of this third-person individual by first asking the participant's gender, and redirecting to a scenario based on the participant's response (e.g., females to a scenario involving a female subject).

Limited past research on cyber-attacks has utilized a scenario-based approach toward understanding risk estimation, particularly in the area of insider threat (i.e., employees' intent to carry out a cyber-attack on his or her own employer's information system). In particular, Greitzer and colleagues (2010) constructed a scenario following an employee, Adam, who is disgruntled with his job and colleagues, and begins to contemplate an attack on his company's information system. Participants were asked to rank variations of this scenario by the degree of risk they perceived Adam to pose against his company (e.g., a past history of anger management issues might imply greater risk). Two teams of researchers utilized statistical modeling to assess computer-predicted risk across a number of scenarios in which hacker familiarity and access to an information system, as well as employee use of malware-infested websites, are taken into account (Chinchani et al., 2004; Espenschied, 2012). Sinclair and Smith (2008) developed three scenarios that vary by an organization's number of employees, technological support, organizational change (e.g., turnovers, corporate merger), management structure (i.e., degree of hierarchy), and number of office locations. The researchers then consulted extant literature regarding each of these factors to estimate risk to an organization's information system in case of insider threat. Similarly, Pfleeger and Caputo (2012) developed a set of cyber-attack scenarios that they presented to industry and government representatives to analyze in terms of riskiness and possible solutions.

It is, therefore, apparent that across a diverse range of methods in a restricted area of research, the use of scenario-based approaches toward cyber-attack risk estimation may be beneficial in the absence of participants who overtly identify as hackers, hacktivists, or disgruntled employees. As such, we utilized a scenario-based approach that guided participants into the mindset of an imagined third party who was considering engaging in an Anonymous-affiliated hacktivist attack. This scenario is displayed in Appendix B.2.

Call to action. The scenario went on to explain that while Googling up-to-date news stories on the current Net Neutrality controversy, Jordan (i.e., a gender-neutral name, with pronouns altered to convey either a male or female subject depending on the participant's gender) stumbles across a page containing a call to action that directly opposes Net Neutrality. We used a single call to action that varied only along our two factors—namely, deindividuation and motives. Because the most prominent motive underlying past case studies of Anonymous activity was social injustice (see Pilot Study 1), we composed a call stemming from a trigger event that was socially unjust and also relevant at the time of data collection—namely, Net Neutrality. The language of the call closely followed that of an Anonymous-released video available through YouTube, which used a computer-generated voice to read the call to action (AnonymousOfficial24, 2014).

The call manipulation, which is displayed in Appendix B.3, was broken into the following set of components: (1) background information regarding the trigger event, (2) first motive-inducing slogan, (3) information regarding the hacktivist attack, and (4) second motive-inducing slogan. Throughout this call, the use of *we* language was altered in half of the conditions to read *you* (e.g., [*We/you*] *must not turn over* [*our/your*] *rights to the highest bidder*); furthermore, two slogans each were chosen to convey either power motives, justice motives, or a combination of the two. Participants in the control condition did not see slogans, but instead a call comprising only background and attack information.

The call was presented in the same format (i.e., background and text color, font and font size) as what might be seen on 4chan and similar communication channels commonly used by Anonymous members. By mimicking this forum-based format, we hoped to emulate the experience of stumbling across an actual call to action while browsing the Internet.

Risks, payoffs, and likelihood. After viewing one of eight possible variations of this call to action (i.e., corresponding with our eight conditions), participants were asked to report (1) subjective risks, separated into probability of being caught and severity of punishment if caught; and (2) subjective payoffs, comprising probability of succeeding and magnitude of payoffs if successful. In this way, we build upon Equation 1 to hypothesize that:

$$P(\mathrm{A}) = f(P[\mathrm{R_S}] \times M[\mathrm{R_S}]) + f(P[\mathrm{P_S}] \times M[\mathrm{P_S}]) + f(\mathrm{R_S P_S}) + \ldots \qquad (2)$$

in which subjective risks and payoffs may be broken down into (1) the probability of a success ($P[\mathrm{P_S}]$) or failure ($P[\mathrm{R_S}]$) occurring, and (2) the magnitude or severity of the benefits ($M[\mathrm{P_S}]$) or detriments ($M[\mathrm{R_S}]$) posed. These items were further delineated into the risks and payoffs posed to Jordan and the broader group of hacktivists to capture self- and collective-focused estimations. We also asked questions capturing how anonymous participants believed Jordan to feel, as well as the extent to which they believed he or she identified with the group of people who will take part in the attack. These two items assessed whether the call manipulation influenced perceived anonymity and group identification. We then presented a single item to capture perceived likelihood of attack, worded in the third-person: *How likely do you think Jordan is to participate in this call to action*? These items are displayed in Appendix B.4.

Randomly assigned condition was coded into three variables indicating (1) We/You language (i.e., dummy-coded[9] to imply deindividuation or individuation condition), (2) Motive (i.e., dummy-coded into one of four conditions based on assigned motive), and (3) overall condition (i.e., 1 through 8, maintained solely for the purpose of examining simple contrasts between conditions).[10]

Our dependent variables of interest were grouped into three overarching categories: risk, payoff, and attack likelihood. Risk variables comprised (1) subjective risk, (2) likelihood of being caught, and (3) severity of punishment if caught; payoff variables included (1) size of payoffs, (2) likelihood of success, (3) benefit to pride, (4) benefit to others, (5) benefit through self-challenge, (6) boost status, (7) fight Net Neutrality, and (8) raise public awareness against Net Neutrality; and our single attack likelihood variable gauged participants' perceived likelihood that the third-person subject would carry out an attack. Our three risk variables were highly correlated (average $r[470] = .518$, $.415 \leq r \leq .694$, $p < .001$), as were our eight payoff variables (average $r[447] = .425$, $.225 \leq r \leq .663$, $p < .001$), displaying high internal consistency ($\alpha_{risk} = .765$ for 3

[9] In linear regression models, predictor variables are typically continuous in nature. When they are not, such as in the instance of our two manipulated factors, dummy coding is used as a method to convey categorical variables in numerical form. In dummy coding, ones and zeros are used to convey necessary information of group membership (e.g., a group coded 0 indicates assignment to one group, whereas 1 indicates the other group).

[10] Conditions 1 through 8 were numbered in order of the motives and language conditions displayed in all graphs throughout this work, namely: all odd-numbered conditions presented *we* language, and the paired increments of motives (e.g., 1 and 2, 5 and 6) were ordered according to the order of our hypotheses and, likewise, our horizontal axes in each graph: power motives, justice motives, joint motives, and control (no motives).

items, α_{payoff} = .856 for 8 items). As such, we formed single composites of each of these variables to yield an Overall Risk and Overall Payoff variable. Henceforth, any mention of Risk or Payoff will refer to these composite variables. The descriptive statistics of these risk and payoff variables, as well as their final composites, are displayed in Table 6. All variables were measured on 0 to 100-point scales, captured in 10-point Likert-type intervals.

Table 6. *Descriptive statistics: Risk and payoff variables*

Variable	***M* (*SD*)**	***N***	**Variable**	***M (SD)***	***N***
Risk	58.73(26.21)	472	Challenge	46.91(30.12)	456
CaughtLikely	46.03(27.42)	469	BoostStatus	41.64(28.39)	446
PunishSevere	57.76(27.15)	473	Fight NN	43.05(29.68)	455
SuccessLikely	33.13(23.92)	460	Raise Aware NN	51.04(29.74)	463
SizePayoffs	33.76(26.30)	457	Risk Overall	54.12(22.16)	473
BenefitPride	58.81(26.50)	464	Payoff Overall	44.06(19.84)	473
BenefitOthers	45.49(29.62)	461			

Covariate: Sensation-seeking. Beyond the power of our manipulations, it is possible that psychological factors play a role in participants' perceived risks, payoffs, and likelihood of attack. One such psychological factor may be sensation-seeking, or the pursuit of exciting, novel, or intense experiences. We believed it possible that individuals high in sensation-seeking would view a new, risky situation such as a prospective hacktivist attack to be alluring, and perhaps a challenge worthy of exploration. Even when asked to adopt the mindset of a third party (i.e., Jordan), individuals who were particularly intrigued by exciting and new situations may have inaccurately estimated others' intentions to pursue the same goals. Furthermore, Zuckerman and colleagues (1993) state that those high in this particular personality trait are more likely to engage in behaviors without considering the negative consequences that might result from their actions.

As such, we assessed excitement-seeking behaviors through the Zuckerman Kuhlman Personality Questionnaire, Impulsive Sensation-Seeking subscale (ImpSS; Zuckerman et al., 1993). The 19-item ImpSS assesses risky activities and quests for complex or intense sensations; however, unlike similar measures of excitement- and sensation-seeking behaviors, the ImpSS does not specify exact situations that may indicate an adventurous lifestyle or tendency toward outdoor activities (e.g., *I would like to go sky-diving*; Costa & McCrae, 1992; Zuckerman, Eysenck, & Eysenck, 1978). These 19 items (α = .919) are displayed in Appendix B.5.

Covariate: Personality. Personality is another psychological factor that we expected to influence our findings beyond the impact of our manipulation. Personality psychologists often assess personality along five key dimensions as per the Neuroticism-Extraversion-Openness Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992). These dimensions, which are displayed in Table 7 alongside their alpha values,[11] have more recently been adapted so that higher scores in any of the five subscales is correlated

[11] Each of the five TIPI dimensions is measured through two items, one of which is reverse-scored. Our measure of internal consistency therefore reflects the degree to which these two items were answered consistently (after reverse-scoring) among our participants.

positively with well-being and similar psychological variables (i.e., Neuroticism was replaced with its inverse, Emotional Stability).

Table 7. *Five personality dimensions*

Dimension	Examples	Internal Consistency
Openness	Unconventional, creative, open-minded	α = .452, $r(474) = .294$, $p < .001$
Conscientiousness	Dependable, organized, attentive	α = .589, $r(473) = .423$, $p < .001$
Extraversion	Outgoing, enthusiastic, unreserved	α = .686, $r(471) = .150$, $p = .001$
Agreeableness	Pleasant, amiable, sympathetic	α = .260, $r(474) = .522$, $p < .001$
Emotional stability	Calm, carefree, not easily upset	α = .660, $r(474) = .498$, $p < .001$

We expected that individuals who were open to new experiences might have been more likely to view a hacktivist call to action as a challenge or exciting possibility, thereby swaying their perceived risks, payoffs, and likelihood of attack. Although the remaining four personality dimensions may also factor into the present study, we had no particular expectations regarding their influence. In the present study, we utilized the Ten-Item Personality Inventory (TIPI), which shortens the broader NEO-FFI into a ten-item scale demonstrating adequate test-retest reliability, convergent validity, and discriminant validity compared with the original NEO-FFI (Gosling, Rentfrow, & Swann, Jr., 2003). These items are displayed in Appendix B.6.

Descriptive statistics for all variables of interest, including covariates, are displayed in Table 8. With the exception of our personality covariate measures (i.e., the five TIPI dimensions) and Impulsive Sensation-Seeking scale (ImpSS), all variables were measured on a 0 to 100-point scale. TIPI and ImpSS variables were measured and compiled on 1 to 11-point scales, which were displayed to participants on the same 10-

point Likert-type interval schedule. Variables marked “MC” signify our manipulation check items.

An intercorrelation matrix of all variables of interest displayed in Tables 7 and 8 is shown in Table 9, found in Appendix C.

Table 8. *Descriptive statistics: Variables of interest*

Variable	***M* (*SD*)**	***N***	**Variable**	***M (SD)***	***N***
AttackLikely	43.93(22.48)	471	Agreeableness	7.30(1.90)	471
PeerPrevalence	27.16(23.05)	454	EmotionalStability	6.95(2.35)	474
ImpSS	5.26(1.75)	474	MC:Empowered	65.86(22.80)	473
Openness	7.49(2.01)	474	MC:SocialInjustice	67.03(23.68)	471
Conscientiousness	7.92(2.02)	473	MC:Anonymity	58.69(27.08)	473
Extraversion	5.28(2.40)	474	MC:GroupIdentify	62.48(22.78)	472

Demographics. We ended the experiment with a series of standard items concerning participant demographics, namely: (1) race/ethnicity, (2) household income, and (3) ESL (i.e., English as a Second Language) status. To protect anonymity, participants were redirected to a separate survey to submit a personally identifiable raffle entry for one of six Roku 3 streaming media players. These demographic items are shown in Appendix B.7.

Results

Hypotheses 1A, 2A, and 3A

To address Hypotheses 1A, 2A, and 3A, we ran an ANOVA in which *we*/*you* language and motives (i.e., as a variable indicating one of four possible motive conditions) were entered as random factors predicting subjective risk, and found a significant two-way interaction of these factors, $F(3,465) = 4.107$, $p = .007$, $\eta_p^2 = .026$. These results, which are displayed in Figure 8, demonstrate a nonsignificantly higher level of subjective risk among participants in the *you*/individuation condition compared with those who were randomly assigned to the *we*/deindividuation condition, with the exception of the control condition, in which this trend is reversed: the highest risk is instead reported by participants in the *we*, rather than the *you*, condition. When primed with power or justice motives separately, participants display the sharpest increase in risk between *we* and *you* conditions; however, we see almost no change in subjective risk when these two motives are paired within the same condition (i.e., joint motive condition of both justice and power motives). In other words, we observe an interaction effect in which the *we*/deindividuation prime predicts a decrease in risk compared with *you*/individuation conditions only when participants are primed with power or justice motives separately. When primed with both at the same time, there is no impact of language on subjective risk; and in the absence of a motive prime (that is, when we make neither power nor justice motives salient), we see a reversal in these trends. Instead, in the control condition we see a marked increase in subjective risk compared with our remaining three motive conditions.

There was no main effect of *we*/*you* language ($F[1,3.021] = 1.262$, $p = .343$, ns) or motive on subjective risk ($F[3,3] = .374$, $p = .796$, ns), therefore withholding evidence of Hypotheses 1A and 2A. However, we do see evidence of Hypothesis 3A in which the *we*-power condition yields the lowest overall subjective risk, and the *you*-justice condition yields a markedly higher (although not the highest, as we originally predicted) level of risk. Collectively, these results can be compared against our original predicted findings in Figure 2. Table 10 in Appendix C displays a table of all means and their significance values with respect to other conditions.

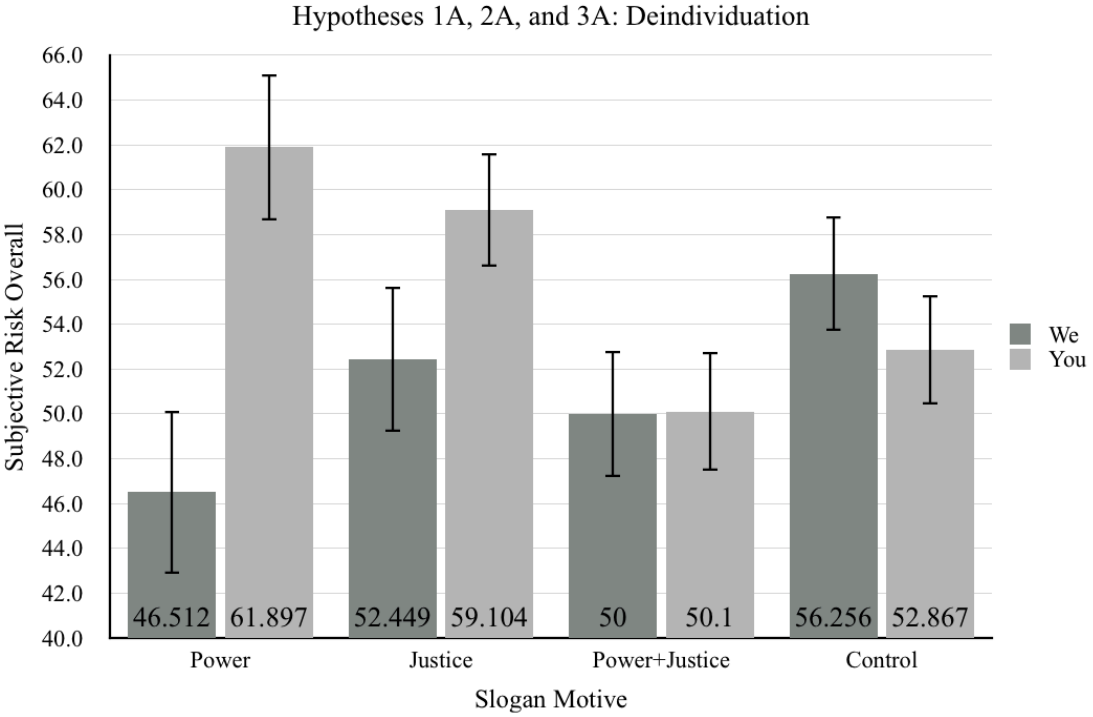


Figure 8. Results of Hypotheses 1A, 2A, and 3A, the impact of *we*/*you* language and motive on subjective risk

To gain a better understanding of how our two manipulated factors impacted subjective risk in isolation of one another, we further probed this ANOVA by examining simple contrasts outlining our predicted effects. For example, we assessed whether a main effect of *we* versus *you* language existed in isolation from the variance explained by our motives conditions. To do so, we ran a general linear model in which we dummy-coded *we* (-1, indicating lower predicted levels of risk) and *you* conditions (1, indicating higher levels) dichotomously. Supporting Hypothesis 1A, we found that when assessing the sole influence of *we*/*you* language on subjective risk (i.e., with the influence of motives removed from the equation), the use of *we* language predicted significantly lower risk compared with participants assigned to *you* conditions, $F(1,465) = 5.127$, $p = .024$, $\eta_p^2 = .011$. It would therefore appear that the variance explained by the interaction of language and motives together, which proved significant in our univariate ANOVA, obscured the significant impact of *we*/*you* language alone on subjective risk.

We ran a second set of contrasts examining the simple effects of motives on subjective risks, and found no significant effect either when examining a stepwise progression of influence (i.e., expecting power motives to be lowest, followed by control and joint conditions, and lastly by justice motives; $F[1,465] = .041$, $p = .840$, ns) or when dichotomizing power and justice conditions alone ($F[1,465] = .273$, $p = .601$, ns). Lastly, we contrasted control conditions (i.e., no motives) against conditions that primed either power motives, justice motives, or both motives, and found a significant difference between the two groups, $F(1,465) = 666.900$, $p < .001$, $\eta_p^2 = .589$. There was, however, no support for Hypothesis 2A.

Hypotheses 1B, 2B, and 3B

To address Hypotheses 1B, 2B, and 3B, we ran an ANOVA using language and motives as random factors predicting subjective payoff. There was no two-way interaction ($F[3,465] = .215$, $p = .886$, ns) or main effect of *we*/*you* language ($F[1,3.412] = .192$, $p = .688$, ns); however, a main effect of motive appeared, $F(3,3) = 8.563$, $p = .056$, $\eta_p^2 = .895$.

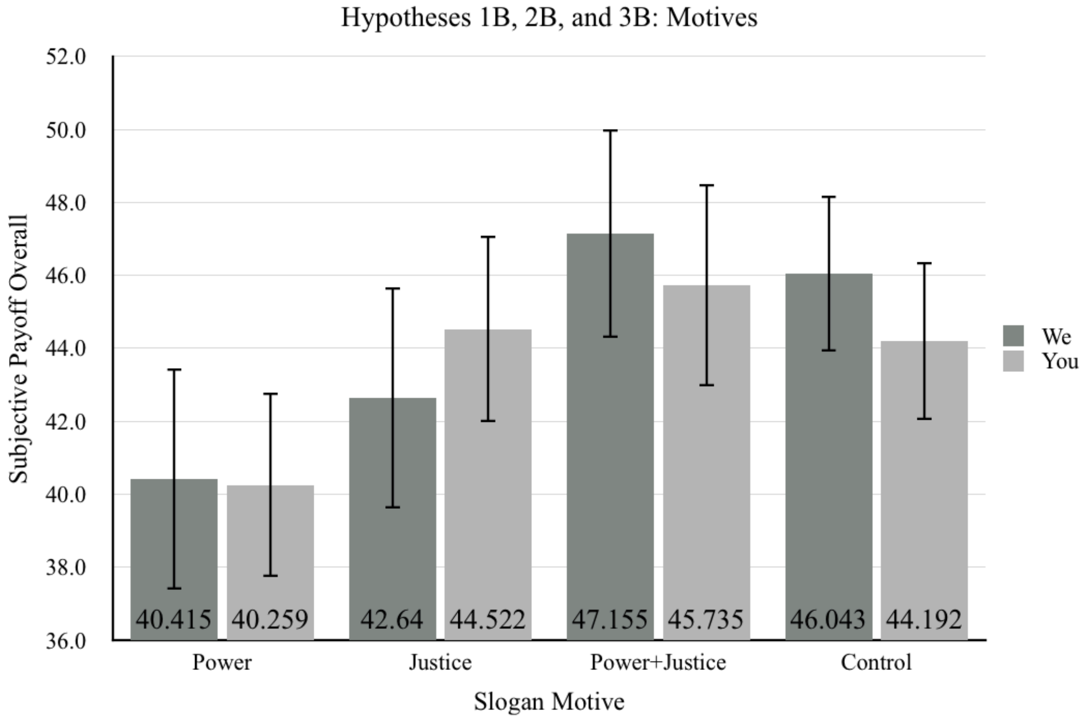


Figure 9. Results of Hypotheses 1B, 2B, and 3B, the impact of *we*/*you* language and motive on subjective payoff

Figure 9 depicts this main effect in graphical form, showing almost parallel trends in subjective payoff across *we* and *you* conditions, but wide variability across our four motive conditions. More specifically, we see that participants primed with power motives report the lowest subjective payoffs—indeed, to a degree that is marginally lower than those in the control group (*t*[465] = 1.853, *p* = .065) and significantly lower than those in the joint motives condition (*t*[465] = 2.209, *p* = .028). Contrary to expectations, participants in the control condition reported higher perceived payoffs on average than those assigned to power or justice motive conditions; however, consistent with Hypotheses 2B and 3B, the joint appearance of both power and justice motives yielded higher perceived payoffs than other motive primes, albeit nonsignificantly when compared with the control group (*t*[465] = .528, *p* = .598, ns).

These findings suggest that when participants are not primed with a specific motive (i.e., control participants), they still perceive a comparably high degree of payoffs—indeed, almost as high as those reported by participants presented jointly with both motives. It is possible that when no particular motive is made salient, participants in the control condition are able to freely supply their own factors to the situation—either motives or other influential factors—that result in a higher degree of perceived payoff overall. These factors might therefore act as a bolstering mechanism in the absence of power- or social injustice-specific primes, pushing participants to view the potential upsides of an otherwise risky situation.

Our original predictions for these three hypotheses are displayed in Figure 3, and a table of all respective means is displayed in Table 11, Appendix C.

We further probed our ANOVA by running simple contrasts that assessed the effects of *we*/*you* language isolated from the effects of motive primes, but found no effect by contrast code, $F(1,465) = .042$, $p = .835$, ns. Similarly, we found no effect when contrasting our motive primes in a stepwise progression (i.e., predicting an increase in perceived payoffs from control to justice motives, power motives, and then to joint motives conditions; $F[1,465] = .011$, $p = .918$, ns) or dichotomously between control and joint motives conditions ($F[1,465] = .279$, $p = .598$, ns). Lastly, we examined differences in payoffs between our motives and control conditions by using dichotomous dummy coding, and found a significant difference between participants who had been primed with one or more motives compared with those who had not, $F(1,465) = 526.804$, $p < .001$, $\eta_p^2 = .531$.

Collectively, we observed no support for Hypothesis 1B and partial support for Hypotheses 2B and 3B, such that the joint appearance of power and justice motives predicted higher subjective payoffs compared with other conditions; however, this increase was only significant when compared with our power motive condition.

Game Theoretic Approach

The primary goal of the present study has been to explore the way in which contextual cues activate senses of deindividuation and motives and, in turn, how the activation of these factors influences hackers' perceptions of risks and payoffs. This approach utilizes a general game theoretic framework in which we expect subjective risks, payoffs, and any interaction of the two to inform participants' perceived likelihood of carrying out a hacktivist attack, which comprised our fourth hypothesis. It was therefore important to examine whether risks and payoffs did, indeed, jointly predict estimated attack likelihood. Building on Equation 1, we first examined this hypothesized relationship by running an OLS (i.e., Ordinary Least Squares) regression in which:

$$\mathrm{AL} = ß_0 + ß_1(\mathrm{RISK}) + ß_2(\mathrm{PAYOFF}) + ß_3(\mathrm{RISK} \times \mathrm{PAYOFF}) \qquad (3)$$

where standardized beta coefficients (ß) represent the number of standard deviations the dependent variable (Y, attack likelihood [AL]) would change for each increase in standard deviation of the predictor variables (X_1, risk [RISK]; X_2, payoff [PAYOFF]; and X_3, the interaction of the two [RISK × PAYOFF]). We centered our two predictor variables prior to forming a multiplicative interaction coefficient to reduce any collinearity between these predictors and their interaction.

The overall regression model was highly significant, $R^2 = .305$, $F(3,465) = 67.915$, $p < .001$, $\eta_p^2 = .180$, suggesting that 30.5 percent of the observed variance in our dependent variable can be explained by participants' subjective risks, subjective payoffs, and the interaction between the two; however, of these three independent variables only subjective payoffs were a significant predictor of attack likelihood (ß = .551, t = 14.218, $p < .001$). There was no two-way interaction (ß = .005, t = .141, p = .888, ns) or main effect of subjective risk (ß = -.040, t = -1.038, p = .300, ns).

Examining our original Equation 1 more closely, we hypothesized in Equation 2 that we could build upon this relationship by delineating subjective risk and payoff into (1) the probability of a success or failure occurring, and (2) the magnitude or severity of the benefits or detriments posed. In other words, our Equation 3 could be adapted so that subjective risk comprises both perceived risk and perceived likelihood of getting caught (i.e., probability of failure), and subjective payoff comprises perceived size of payoffs and likelihood of success. The resulting equation would therefore appear thus:

$$\text{AL} = ß_0 + ß_1(\text{R}) + ß_2(\text{CL}) + ß_3(\text{SL}) + ß_4(\text{SP}) + ß_5(\text{R} \times \text{CL}) + ß_6(\text{SL} \times \text{SP}) \qquad (4)$$

where R indicates risk, CL denotes likelihood of getting caught, SL signifies likelihood of success, and SP stands for size of payoffs. This regression model was significant, R^2 = .230, F(6,441) = 22.006, $p < .001$, η_p^2 = .130; however, only likelihood of success (ß = .344, t = 5.946, $p < .001$) and size of payoffs (ß = .128, t = 2.247, p = .025) were significant predictors of attack likelihood, mirroring our OLS regression model from Equation 3. The impacts of risk, likelihood of getting caught, and both two-way interactions were nonsignificant.

It would therefore appear that, contrary to expectations as per a game theoretic approach to hacking decisions, perceived likelihood of carrying out a hacktivist attack was driven by payoffs, but not risks. This finding suggests that college-aged students expect others (i.e., a third-person subject) to be less than rational decision makers—that is, to weigh benefits, rather than costs, when choosing to carry out a potentially self-threatening action, possibly overestimating others' recklessness or risk-taking tendencies when faced with a threatening scenario.

Hypothesis 4

In Hypothesis 4, we predicted that participants would form mental ratios of the risks and payoffs inherent in a hacktivism scenario and utilize this ratio to guide their overall perceived likelihood of attack.

To test this hypothesis, we ran an ANOVA using *we*/*you* language and motives as random factors predicting attack likelihood, and found neither a two-way interaction (*F*[3,463] = .767, *p* = .513, ns) nor main effects of language (*F*[1,3.114] = 1.276, *p* = .338, ns) or motive (*F*[3,3] = .968, *p* = .511, ns). We further probed this analysis, which is displayed in Figure 10, by running simple contrasts assessing a stepwise progression of strength of each condition on attack likelihood as per our original Hypothesis 4, depicted in Figure 3. As such, we expected an increase from control conditions to justice motives, joint motives, and finally to power motives when assessing the impact of each on attack likelihood. This simple contrast was nonsignificant, *F*(1,465) = .658, *p* = .418, ns. When contrasting participants' responses from the control condition (hypothesized to be lowest in attack likelihood) against those from our power motives condition (hypothesized to be highest), we again found no effect, *F*(1,463) = 1.071, *p* = .301, ns.

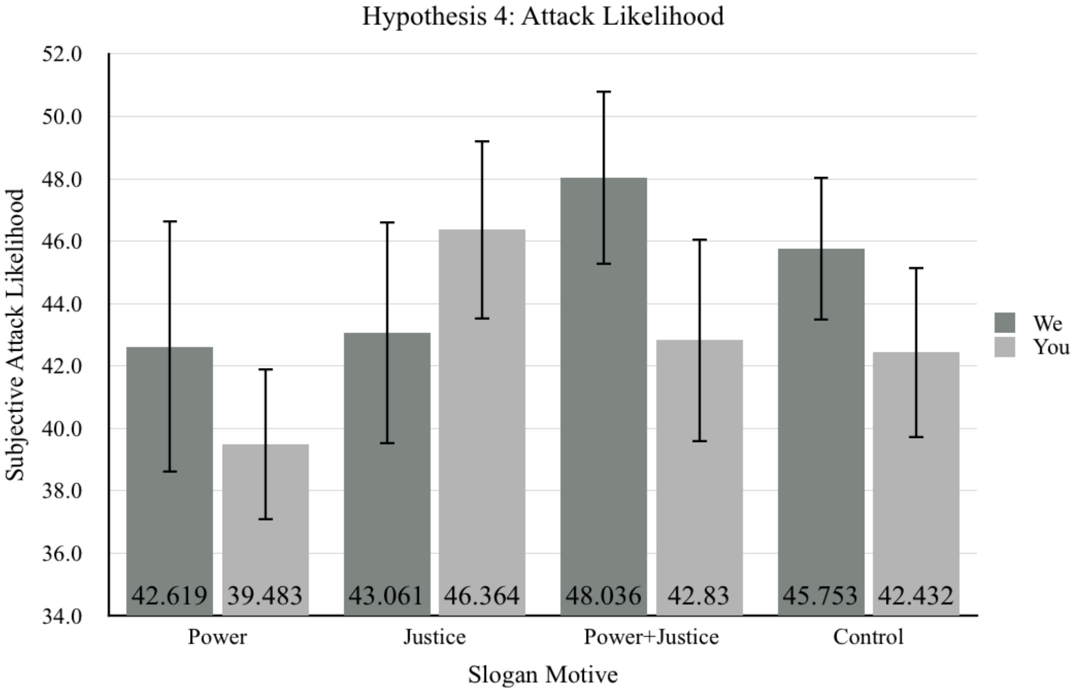


Figure 10. Results of Hypothesis 4, the impact of *we*/*you* language and motive on attack likelihood

Lastly, we assessed whether the difference between the highest condition on attack likelihood (*we*-joint motives) was significantly greater than all other conditions, and found a trending effect (*F*[1,463] = 2.250, *p* = .134, ns). All means and significance values shown in Figure 10 can be found in Table 12 in Appendix C.

Exploratory Analyses

Covariates. To gain a better understanding of our sample, as well as other variables not directly related to our central hypotheses, we ran a series of exploratory analyses. First, we examined our expected covariates for any indication of extraneous relationships between individual differences and our variables of interest. As is shown in Table 9 (variables 4 through 11), the only covariate related to a series of outcome variables was the degree to which hacking behaviors were common among participants' peers (i.e., peer prevalence). Peer prevalence was related to each of our eight payoff variables as well as attack likelihood, our primary dependent variable. Those with hacking peers were also more likely to report during our manipulation checks that they expected Jordan to feel empowered and anonymous, as well as to perceive higher social injustice.

These trends suggest that among participants whose friends carry out hacking behaviors, the risks involved in a pending attack are irrelevant to the subsequent decision to hack; instead, payoffs seem consistently greater when peers have hacked in the past, or continue to do so in the present. There was, however, no moderation effect of peer prevalence or any other covariate on the relationship between risk and attack likelihood ($|.0001| \leq$ interaction $b \leq |.023|$, $.326 \leq p \leq .997$), nor did any but one of our covariates moderate the relationship between payoff and attack likelihood ($|.006| \leq$ interaction $b \leq |.029|$, $.212 \leq p \leq .760$).

The only covariate to evince a moderating role on the relationship between perceived payoff and attack likelihood was Agreeableness (interaction $b = .050$, $t = 2.123$, $p = .034$), a personality dimension underscoring warmth and amiability. More specifically, as Agreeableness increased, the relationship between payoff and attack likelihood also increased in positivity: Although the average participant reported a higher likelihood of attacking when payoffs were great, agreeable participants (i.e., those that fell at least one standard deviation above the mean in Agreeableness within this sample) were especially likely to do so ($b = .715$, $t = 11.694$, $p < .001$). However, when payoffs were perceived to be substantially lower, the participants most likely to carry out an attack anyway were those who were least agreeable in our sample ($b = .526$, $t = 8.242$, p

< .001). The moderation regression model, which is displayed in Figure 11, was significant, $R^2 = .309$, $F(3,463) = 68.997$, $p < .001$, $\eta_p^2 = .183$.

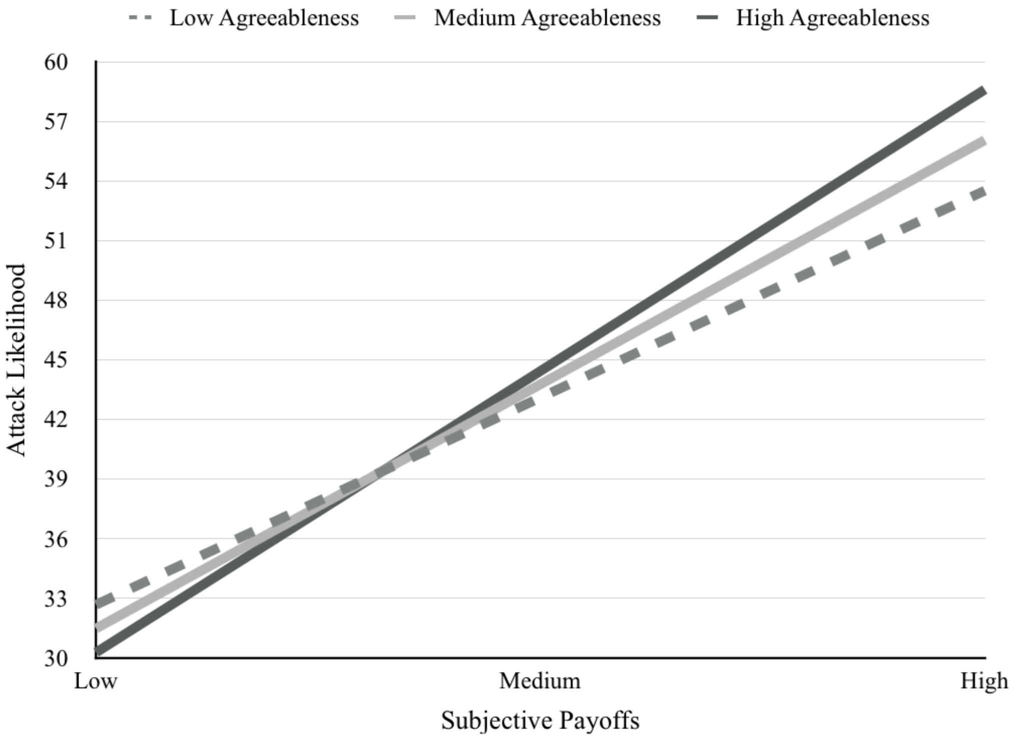


Figure 11. Impact of Agreeableness on perceived payoffs and attack likelihood

Although it is not exactly surprising that disagreeable participants estimated the highest likelihood of carrying out an attack when payoffs were low (i.e., suggesting impulsiveness, informality, and a lack of social concern; Ames & Bianchi, 2008), it is interesting that highly agreeable participants respond most strongly to the influence of payoffs on attack likelihood. It is possible that, because agreeableness signals an increased desire for amiability and empathy, agreeable participants responded most negatively to the potentially detrimental impacts of Net Neutrality on society; indeed, Agreeableness was the only personality dimension correlated with participants' stance on Net Neutrality, demonstrating a particularly dissenting view of this issue (*r*[461] = -.121,

p = .009; all other personality dimensions |.046| ≤ r[464] ≤ |.062|, .186 ≤ p ≤ .321, ns).[12] Agreeable participants were no more likely to view hacking as a possible benefit for others (r[458] = -.012, p = .790, ns) or for society in general (r[470] = .060, p = .194, ns).

Individual differences. To gain a more all-encompassing understanding of the data, we assessed whether there were any differences by gender, ethnicity, or catch question status—that is, whether participants passed or failed our attention check item. The results of these analyses are discussed below.

Risks. When examining the predictive influence of our manipulations on subjective risk, there were no differences in our primary ANOVA results when including all participants, rather than just those who passed our catch question. However, the two-way interaction of *we/you* language and motive held only for male participants (F[3,253] = 3.671, p = .013, η_p^2 = .042), and not among females (F[3,204] = 1.236, p = .298, ns). There were no main effects of language or motives for either gender. These differences, which are displayed in Figure 12, suggest a strong difference in subjective risk particularly among participants assigned to the control (no motives) condition: For females, risk remains low in the absence of a power or justice motive; among males, risk appears to increase markedly.

Furthermore, the interaction effect held for White participants (F[3,315] = 4.151, p = .007, η_p^2 = .038), but not for non-White participants (F[3,140] = .962, p = .413, ns),

[12] Stance on Net Neutrality was unrelated to all other variables of interest, as displayed in Table 8 in Appendix C. It did not emerge as a significant predictor of attack likelihood when assessing Equations 3 and 4, nor did it emerge as a covariate when examining the effects of our manipulations on risk (Hypotheses 1A, 2A, 3A), payoff (1B, 2B, 3B), or attack likelihood (4). It did not moderate the impact of subjective risk or payoff on attack likelihood.

who instead displayed a marginal main effect of motive on overall subjective risk (*F*[3,3] = 7.628, *p* = .065, η_p^2 = .884). Figure 13 depicts these differences, in which *we*/*you* language has less impact on subjective risk among non-White participants compared with motives, whereas a strong interaction effect of language and motives exists among White participants.

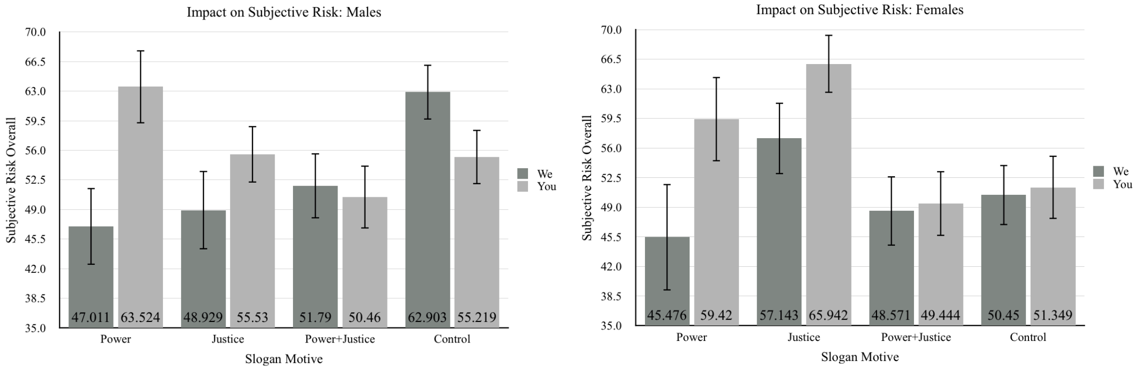


Figure 12. Sex differences of the impact of *we*/*you* language and motives on subjective risk

That non-White participants do not display these same interaction effects—namely, that both pronoun use and motive play an essential role in risk estimation—suggests that *we*/*you* language holds very little influence when paired with power and justice primes. Instead, non-White participants seem to estimate risks most highly when power and social injustice are made more salient—a trend that is completely reversed among White participants. The combined effects of both power and justice motives in our joint condition, however, predicts a strong decrease in subjective risk among our non-White participants.

These findings, though perplexing, may stem from a broader cultural psychological phenomenon in which the use of *you* language primes those of collectivistic background (e.g., Middle Eastern, East Asian, and Latino groups, which comprise a considerable part of our sample) to think not only about themselves, but rather the context that best defines them, which is often a social one. In collectivistic or interdependent societies, concepts of oneself as an individual tend to overlap with close family and friends (Aron, Aron, & Smollan, 1992; Galinsky, Ku, & Wang, 2005), and should this be the case in this sample, it is possible that regardless of pronoun use, both *we* and *you* were priming fundamentally similar self-views.

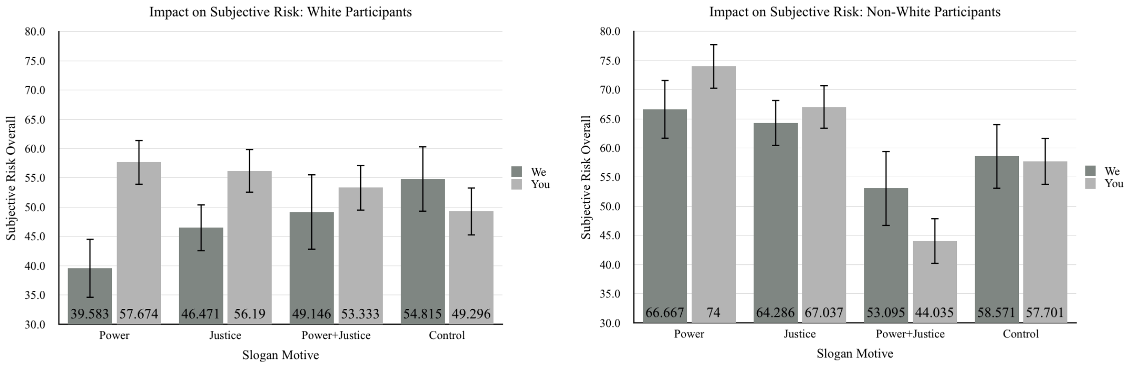


Figure 13. Ethnic differences of the impact of *we*/*you* language and motives on subjective risk

Furthermore, Middle Eastern and East Asian societies are often characterized by high power distance—that is, a strong sense of social hierarchy in society (Bochner & Hesketh, 1994; Hofstede, 1980). Western countries, on the other hand, are characterized by low perceived power distance, in which any changes in power between authority figures and their subordinates is small. As such, by priming power motives and social injustice (i.e., indications that the government and other powerful corporations—which form the upper quartile of many social hierarchies around the world—are dishonest and unjust), we may have unintentionally heightened perceived risk among participants who hail from backgrounds that adhere to these particular cultural values. Such participants might have viewed these primes as disrespectful toward the broader social hierarchy or toward the government, thereby viewing the hypothetical scenario as a more risky endeavor overall.

Payoffs and attack likelihood. When examining the impact of our manipulations on payoffs and attack likelihood separately, we found no significant differences between gender groups, ethnicities, or catch question status.

Game theoretic approach. When building a regression model in which attack likelihood was predicted by risk, payoff, and the interaction of these two factors, we found that only subjective payoffs were a significant predictor of estimated attack likelihood. These effects did not differ by gender, ethnicity,[13] or catch question status. We then examined Equation 4, in which this game theoretic approach is further broken down to comprise only risk, likelihood of getting caught, size of payoffs, and likelihood of success. Our findings, in which only likelihood of success and size of payoffs were significant predictors, were mirrored across genders and regardless of catch question status; however, the importance of the size of payoffs—a significant positive predictor among White participants (ß = .229, t[305] = 3.629, p < .001)—is nonsignificant among non-White participants (ß = -.120, t[139] = -1.006, p = .316, ns), possibly due to a reduction of more than half in sample size between the two groups. At the time of this

[13] Ethnicity was analyzed as a dichotomous variable indicating majority group (i.e., White) and non-majority groups (Non-White) based on small sample sizes among most self-identified racial groups within our sample. For example, although our effects held for most ethnicities except African-American participants (for whom the interaction of risks and payoffs was significant [ß = .537, t = 2.497, p = .026]), several ethnic groups were too small to analyze with confidence (e.g., 2 Native American participants, 5 Middle Eastern participants, 8 participants identifying as *Other*, 17 African-American participants). All remaining analyses of ethnicity in this study divide participants based on majority and non-majority status to maintain necessary sample sizes within each group.

study, we could find no published evidence of ethnic differences in risk estimation or risk-taking behaviors; however, future work may wish to explore whether such differences exist.

Manipulation checks. Toward the end of the experiment, participants were asked to indicate the extent to which they expected Jordan (1) to feel empowered, (2) to perceive the situation as socially unjust, (3) to feel anonymous, and (4) to identify with those who would take part in the attack. The purpose of these four items was to gauge the extent to which our manipulations had the effects that we intended.

More specifically, we hypothesized that feelings of empowerment would be highest among those randomly assigned to the power motives condition; social injustice would be highest among those in the justice motives condition; and anonymity and group identification would be highest among those in the *we*/deindividuation condition. We ran four separate tests of analysis of variance (ANOVA) to predict each manipulation check with two random factors, coded as a *we*/*you* variable and a variable indicating one of four possible motive conditions. Three of these analyses yielded nonsignificant effects (Empowered: $.307 \leq F[3,465] \leq 4.805$, $.115 \leq p \leq .820$; Social injustice: $.376 \leq F[3,463] \leq 1.206$, $.307 \leq p \leq .778$; Anonymity: $.332 \leq F[3,465] = .890$, $.446 \leq p \leq .635$); however, there was a marginally significant main effect of motive on group identification, such that group identification was lowest among participants in the control (i.e., no motive) condition, $F(3,3) = 7.216$, $p = .069$, $\eta_p^2 = .878$.

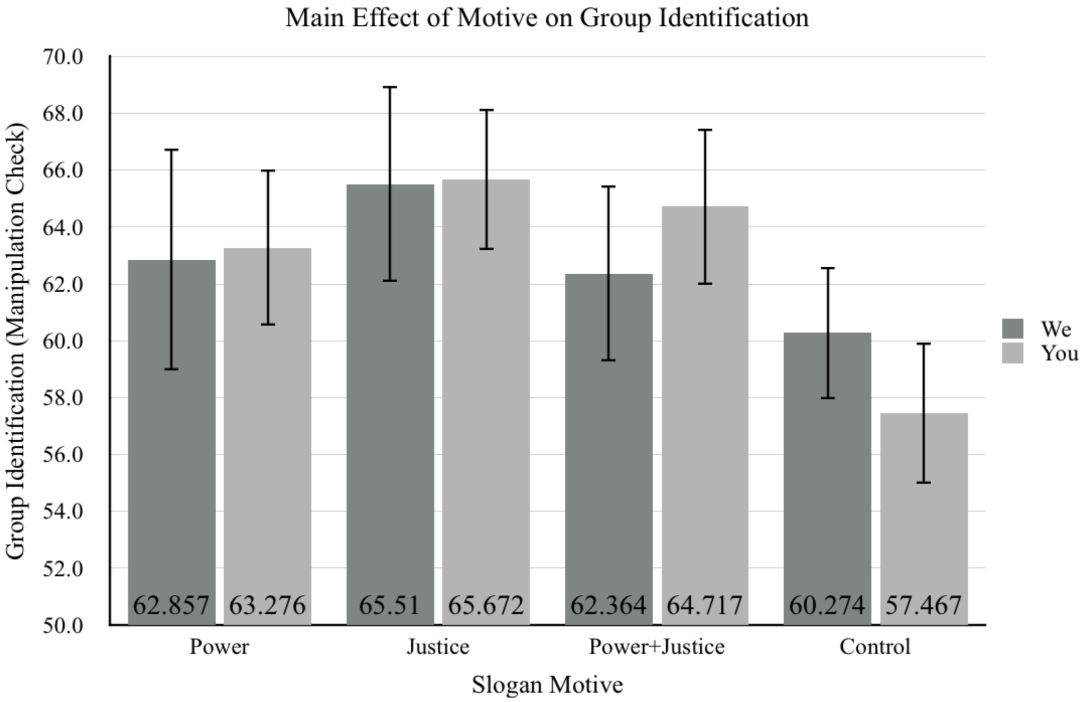


Figure 14. Effect of slogan motive on group identification

This trend is displayed in Figure 14, and suggests that participants who are primed with power motives, justice motives, or both motives simultaneously are more likely to think that the third-person subject (i.e., Jordan) identifies with the group of potential hackers participating in this attack. Participants in the control (no motive) condition were less likely to perceive an increase in group identification.[14]

[14] Stance on Net Neutrality bore no moderating effects on the relationship between Social Injustice, Anonymity, or Group Identification on attack likelihood; however, it yielded a marginally significant moderation effect on the impact of perceived empowerment on attack likelihood ($R^2 = .169$, $F[3,456] = 30.843$, $p < .001$). As views of Net Neutrality grew increasingly negative, the relationship between empowerment and attack likelihood grew increasingly positive; that is, participants who were strongly against Net Neutrality ($b = .478$, $t = 7.943$, $p < .001$) reported higher likelihood of carrying out an attack if they felt particularly empowered (interaction $b = -.0024$, $t = -1.810$, $p = .07$). Participants who supported Net Neutrality also displayed a positive effect between empowerment and attack likelihood ($b = .329$, $t = 5.643$, $p < .001$), but to a less significant degree compared with participants who were against Net Neutrality. In other words, increased empowerment was a stronger predictor of increased attack likelihood among participants who were against Net Neutrality.

Collectively, these primarily nonsignificant effects suggest that our manipulations either did not capture our intended effects, or did not capture the expectedly unconscious impacts of our subtle primes on a level at which participants could knowingly report. An additional factor that must be acknowledged is the time delay between manipulation presentation and manipulation check items, which were shown to participants at the end of the experiment. To gain a better understanding of the trends underlying our manipulation checks, we assessed their relationships with our primary variables of interest and found highly significant positive relationships between perceived payoffs and increased feelings of empowerment (*r*[473] = .414, *p* < .001), social injustice (*r*[470] = .453, *p* < .001), anonymity (*r*[473] = .151, *p* = .001), and group identification (*r*[472] = .340, *p* < .001).

The only manipulation check variable related to perceived risk boasted only a marginally negative correlation: As participants reported higher subjective anonymity, they were more likely to report lower perceived risks (*r*[472] = -.085, *p* = .065), in line with Hypothesis 1A. Overall, all four manipulation check variables were highly positively related to attack likelihood, such that the participants who were more likely to expect Jordan to carry out an attack were also more likely to perceive empowerment (*r*[470] = .404, *p* < .001), high social injustice (*r*[469] = .391, *p* < .001), anonymity (*r*[470] = .234, *p* < .001), and group identification with the potential hackers (*r*[469] = .379, *p* < .001).

These strong correlations suggest that participants who report feeling empowered, perceiving social injustice, feeling anonymous, and identifying with the larger hacker group are also more likely to perceive a high degree of payoffs and likelihood of attack in the impending hacktivism scenario. Because subjective payoffs and attack likelihood are so highly related, we re-examined the relationships between attack likelihood and our manipulation check variables while holding payoffs constant, and again, all correlations emerged highly significant (Empowered: *r*[464] = .231, $p < .001$; Social injustice: *r*[464] = .205, $p < .001$; Anonymity: *r*[464] = .190, $p < .001$; Group identification: *r*[464] = .251, $p < .001$).

That these relationships still prove significant even when controlling for payoff, the most direct predictor of attack likelihood, is noteworthy: It suggests that these four factors, which we predicted would underlie participants' estimations of attack likelihood through our manipulations, are indeed relevant to our primary goals in this study. Because not all of our manipulations produced the same results that we previously expected, there appears to be a disconnect between our intended effects (e.g., that power motives would yield a sense of power) and our actual findings (feeling "empowerment" is completely unrelated to our power motive primes). In future studies, it would be advantageous to work with a series of primes again targeting empowerment, social injustice, anonymity, and group identification, but in different ways than the primes presented in this study: although not all of our intended effects emerged, this key finding suggests that we are, at least, on the right track.

Discussion

Summary of Findings

Risks. In summation, when examining the impact of our manipulations on subjective risk, we found that the impact of power motives and justice motives separately predict higher risks among participants assigned to the *you*/individuation condition, but lower risks among those in the *we*/deindividuation condition. When presented with both motives at once, however, we see a lower degree of subjective risks overall, and no difference whatsoever between the *we* and *you* conditions. This suggests that the effects of power motives and justice motives (i.e., independent of pronoun use) are not additive, as was originally hypothesized; their joint appearance may have an unexpectedly negative impact on subjective risks. As hypothesized, the condition yielding the lowest perceived risk was *we*-power, in which participants were primed with deindividuation and power-inducing motives. When assessing the impact of *we*/*you* language (i.e., isolated from the impact of our four motive conditions) on subjective risks, we found that the use of *we* language predicted significantly lower risk compared with participants in *you* conditions, as was predicted.

Payoffs. There was a main effect of motive on subjective payoffs, such that the lowest payoffs stemmed from participants assigned to our power motive condition, while the greatest perceived payoffs stemmed from those presented with both power and justice motives.

Game theoretic approach. We found that participants' estimations of attack likelihood stemmed solely from payoffs, and not from subjective risks underlying a hacktivist situation. This finding suggests that, contrary to the broader theoretical framework of game theory, we are not rational decision makers. Highly risky decisions may instead be made based solely upon possible benefits, rather than the ratio of benefits to potential costs. Alternatively, this finding might suggest that when asked to predict a third party's action after being presented with a risky (but potentially advantageous) situation, we are more likely to view others as risk-takers or, perhaps, as illogical and payoff-driven.

Attack likelihood. Across our risk- and payoff-specific findings, it appears that, as expected, the use of *we* language leads to a decrease in subjective risks, possibly due to primed effects of deindividuation. The joint appearance of both power and justice motives appears to signal low risks and high payoffs, which—in line with game theory—should predict the highest estimations of attack likelihood. Indeed, the condition yielding the highest estimated attack likelihood comprised *we* language, which signaled lower risk, and joint motives, which signaled both lower risk and higher payoff.

Covariates. We found that although a majority of our covariates were unrelated to our outcome variables of interest, the extent to which hacking is prevalent among one's peers appeared to have a positive impact on estimations of payoff and attack likelihood—that is, participants who might be more familiar with hacking through their peer groups may perceive more payoffs and, therefore, a higher likelihood of carrying out an attack regardless of the risks involved.

In summation, the present findings suggest that when individuals feel deindividuated—that is, an unidentifiable part of a larger group—their estimations of risk decrease, which could be imperative when making risky decisions. Furthermore, it appears that even risky decisions are made based on potential payoffs rather than consequences, suggesting we are not as rational as game theorists and behavioral economists might have us believe. Lastly, it appears that when participants are presented with both power and justice motives (i.e., are made to feel empowered and to perceive high social injustice), the importance of deindividuation on risk disappears—that is, risk estimations are identical regardless of *we* or *you* language—and payoffs greatly increase. Although this effect was not significant within this sample, it would appear that the joint appearance of power and justice motives should—by decreasing risks and increasing payoffs—predict higher attack likelihood in other samples.

Limitations and Future Directions

By employing an experimental design with randomly assigned conditions, we were able to attribute any differences between our randomized groups to the effects of our manipulations. As such, we are able to examine causality, such that participants' estimations of risk, payoff, and attack likelihood stemmed from changes in pronoun usage and motive presentation. Furthermore, our sample size was sufficient in obtaining a necessary statistical power level of 1 – ß = .841.[15]

Although the present findings are encouraging, their potential implications could be further strengthened by replicating these effects in alternative contexts. For example, we observed a relatively high likelihood of attack in this sample (*M* = 43.93 on a 100-point scale, *SD* = 22.48) considering the risks and controversial nature of carrying out such an attack against the government. One possibility for this finding is that in the original guided visualization scenario, participants read that as Jordan read through recent news articles on the Net Neutrality controversy, "s/he finds himself/herself feeling increasingly opposed to Net Neutrality due to the ease with which government branches and Internet providers could exercise control over clients' and citizens' access." We explicitly stated Jordan's reaction to Net Neutrality as a way to communicate to

[15] Post hoc statistical power calculated through G*Power 3.1 with an average effect size of *f* = .16, α error probability = .05, N = 474, df_{num} (numerator degrees of freedom) = 3, and number of groups = 8. Recommended statistical power using an α value of .05 is .80.

participants that Jordan felt a degree of emotional investment in the issue, and was not a removed bystander; this was with the intention of suggesting that although carrying out a hacktivist attack is not a common action among the broader population, there was at least some chance that Jordan might choose to take part. Even so, it is possible that even communicating Jordan's "increasingly" negative stance against Net Neutrality too overtly primed a higher perceived likelihood that she might actually carry out a hacktivist attack, even considering the risks involved. As such, participants may have exhibited some degree of demand characteristics, in which they respond in the way they believe experimenters would like for them to respond.

Furthermore, participants' estimations of Jordan's likelihood of attack may not have mirrored their own decisions if placed in a similar situation, particularly if they considered Jordan to be reckless, risk-taking, or particularly against Net Neutrality as per our guided visualization scenario. Even so, the use of this third-party subject in our prime was also a strength: It was vital that we minimize any chance of social desirability, in which participants are reluctant to report that they, themselves, might carry out an attack due to the illegality and possible immorality of such an action. Even if we were to try to gauge participants' likelihood of carrying out an attack themselves, a concern for truthfulness would still be present, particularly if participants might fear that we—the experimenters—could identify and pursue them for their answers. Just as it might be difficult to estimate the prevalence of domestic violence, illegal downloading, and similar behaviors from self-report measures, so too is it challenging to capture the true psychology of hackers and hacktivists. This potential mismatch between actual and self-reported behaviors is an age-old concern within the field of psychology; however, it must be noted that participants' responses—even (assuming the worst) if not a complete overlap with their true beliefs and intentions—did yield interesting and statistically significant patterns within and across our manipulations.

This potential limitation could be corrected in future research by presenting participants with a scenario in which it is they, rather than a third-party subject, who stumble across a call to action, and they who must decide whether they will carry out an attack. Such estimations should be less subject to bias; however, participants might respond with less variance than we observed in the present sample due to a fear that experimenters could identify and punish participants who openly respond that they would consider carrying out an illegal activity.

Furthermore, participants who view hacking or hacktivism to be an immoral action might respond in the same fashion (i.e., strongly negatively) regardless of the condition with which they are presented, thereby clouding any potential effects underlying our manipulations. More specifically, with an increasing prevalence of legal, and even government-level, sanctions against hacking and hacktivist attacks, news media are more extensively covering stories related to the identification, capture, and arrest of individuals who have participated in such attacks in the past. Documentaries, books, and in-depth news stories are unraveling not only the actions to which the government and large organizations are willing to go to punish hacktivists, but also the severe stakes at play—including prison sentences and fines of tens of thousands of dollars (Pilkington, 2013; Vincent, 2013). These groups, many of which have been targets of hacktivism in the past, also portray hacktivist methods (e.g., Distributed Denial of Service attacks, which have been equated with pressing the 'refresh' button on a webpage hundreds of times per second) as intrusive, immoral, and highly damaging for database systems and web servers, regardless of evidence supporting its relative harmlessness (Coleman, 2014). As such, individuals who have come into contact with these accounts may believe that, for example, taking any stand against Net Neutrality through hacktivist means is unethical, corrupt, or villainous in nature. Whether these participants were presented with increased power motives, highly deindividuating primes, or examples of extreme social injustice would be inconsequential: In all likelihood, they would most likely respond that they would not, under any circumstance, carry out such an attack.

A second limitation that is worthy of note is the nature of our sample or, more specifically, the incentives used to attract members of our sample. In return for their participation, participants were entered into a raffle for one of six prizes, each valued at approximately $100; however, this method proved difficult in attracting participants in a timely fashion. More than 100 participants visited the first page of our online experiment before leaving once more, possibly because they did not consider a chance-based raffle entry to be worth 15 minutes of their time. Rather than offering a set payment or extra credit in classes, we instead attracted participants who were willing to take that chance, or who perhaps were sufficiently intrigued by the subject matter (i.e., Anonymous, hacktivism) that they wished to take part regardless of the nature of our incentives.

We believe that collectively, this research poses a series of theoretically interesting and societally important implications, and it is our hope that future research will further investigate this topic with methodologies that can, perhaps, more accurately examine true behaviors, rather than hypothetical estimations.

Behavioral measures. One study design that could assess these true behaviors might take the form of a fake call to action, manipulated strategically along various factors expected to underlie differences in attack likelihood (here, pronoun use and motive). These calls to action could be placed on various websites conducive toward audiences that might expect such messages, such as 4chan.org, reddit.com, and similar anonymous venues. As was the case in our call to actions, embedded links would purport to direct the reader to, for example, lists of target IP addresses or access to DDOS applications. Instead, these links would redirect to a short survey assessing intent, perceived risks and payoffs, and vague demographic items (e.g., gender, ethnicity, age). Experimenters could then assess the influence of particular manipulations on objective, behavioral measures of participants' intent to follow links and participate in a false hacktivist attack, which would be to a great advantage over self-report or hypothetically worded experiments.

Peer prevalence. An additional direction for future inquiry is to more closely examine the unexpectedly strong influence of peer prevalence on perceived payoffs and attack likelihood. In the present study, we witnessed an effect in which participants whose peers engage in hacking behaviors are more likely to perceive higher payoffs and attack likelihood in a given attack scenario. It is possible that by manipulating the degree to which participants believe it common for their peers to engage in hacking behaviors, participants' perceptions of the payoffs underlying hacktivist behaviors will be directly impacted and, therefore, their perceived likelihood of attack. This phenomenon, by which individuals' thoughts and decisions are influenced by the degree to which a particular behavior is perceived as common in a social group, is known as a *descriptive norm* in social influence research. In one past scientific intervention, signs placed throughout Arizona's Petrified Forest National Park asked visitors not to steal petrified wood, accompanied with a message stating that either very few or very many visitors had stolen wood in the past. Participants who read that very few people had stolen wood were less likely to do so themselves (Cialdini, 2003). Similarly, when hotel guests are asked to recycle towels along with the majority of past guests from their particular room (i.e., an injunctive norm of environmental protection), they are more likely to abide by this request (Goldstein, Cialdini, & Griskevicius, 2008). Future research or policymakers might likewise alter perceived peer prevalence with the intention of reducing not only hacktivist and hacker intentions, but perhaps even illegal downloading and trolling behaviors.

Ethnic and cultural differences. Future research would also do well to examine ethnic differences in risk, payoff, and attack likelihood estimations as they relate to hacktivist behaviors, particularly as we observed a series of significant differences across ethnic groups in our findings. These differences may be culturally related, but may also stem from the nature of the target used in our call to action: Because the focus was on a controversial topic whose future resides in the hands of the United States government, international students—which form a considerable number of students at the university sampled within this study—may have responded differently from students whose citizenship resides solely in the United States. The government's decision on Net Neutrality might not, for example, directly impact international students if it takes effect after they return to their home country; alternatively, some international students might have hailed from countries whose Internet access laws involve censorship and limited use (e.g., Burma, Saudi Arabia, China, Iran). Because we did not ask country of origin in our demographics section, we cannot examine whether responses differed by international student status within the present study.

Attack targets. Similarly, it must be noted that the targets of our hypothetical hacktivist attack comprised government branches (i.e., the Federal Communications Commission, Department of Justice), which not only pose a greater risk on potential hackers, but which also would make the chance of successful attack—let alone prevention of Net Neutrality—very slim. By presenting participants with a hypothetical hacktivist scenario against the government, we may have primed greater perceived risks, lower perceived payoffs, and (as per game theory) lower attack likelihood overall. In the future, research might examine how our observed trends differ when the targets of hypothetical or primed hacktivist attacks comprise individuals (as has been the case with Anonymous and Hal Turner; see Appendix A.1) or small, non-governmental groups. One might expect risks to seem smaller by comparison, while chances of success—whether it be conveying a message or publicly punishing the group—might seem greater. These hypothesized trends may, perhaps, explain the shift of Anonymous activity from large-scale, often governmental issues to matters of individual and organizational concern following the arrests of LulzSec members in 2012 (Shakarian, Shakarian, & Ruef, 2013). If so, more risk averse members of Anonymous and similar groups might be expected to fix their sights on targets of the latter classification rather than on government branches.

Primes. As was mentioned previously, future research would benefit from further exploring our key phenomena of interest within this study—that is, deindividuation, power motives, and justice motives—through different means. Beyond examining replicability across various forms of deindividuation and motive primes, such investigations would be of key importance to determine the conditions that would most elicit hacktivist attacks. It is our hope that by constructing such a situational profile, we could better inform policymakers, law enforcement, and government officials of the times and events most conducive toward mobilizing individuals to engage in hacktivist attacks.

Our findings suggest that by priming a dispersed sense of identity through *we* language, participants perceive less risk underlying inherently risky activities. If the mechanism of this finding is indeed deindividuation, it should be the case that participants are most likely to take a chance by carrying out a hacktivist attack when they feel least identifiable. Past research on the psychological effects of darkness suggest that participants feel unidentifiable and, in a sense, anonymous when placed in a dark environment, exhibiting increased aggression, antisocial behavior, and self-interest (Johnson & Downing, 1979; Zhong, Bohns, & Gino, 2010). As such, we may expect to see a surge in hacktivist activity at nighttime or in dark milieus, which would necessitate increased needs for cybersecurity after business hours. A study using in-person participants might experimentally investigate this possibility by placing participants in completely dark, computer-equipped testing rooms to observe whether perceived risk and attack likelihood substantially increase. Those asked to participate in the experiment in a bright room, however, should feel identifiable and individuated, thereby heightening their perceptions of risk. Similarly, we might expect hacktivist or illegal online behaviors to occur predominately while in private situations, rather than in settings where passersby might be watching. Although, for example, school or public library computers boast unidentifiability in both MAC and IP addresses, they are also placed in highly individuating settings, which should predict a decrease in hacktivist activity.

It is also important to determine when, and with what severity, individuals will respond to a trigger event (e.g., government talks related to Net Neutrality) through illegal or particularly risky means. Following our findings, we might expect that when justice motives are high and the issue at hand holds particular consequence over an individual's lifestyle or values, men and women will be most likely to pursue a chance at retaliation or conveying a disdainful message. Governments and organizations might do well to expect heightened risk of cyber-attack in the midst of, or immediately following, a controversial topic or event, particularly among subpopulations that might feel most threatened by, or angered over, the issue (e.g., lower and middle-class socioeconomic groups in response to Occupy Wall Street). In an experimental setting, justice motives could be primed through false news stories after gauging participants' views on a series of politically or societally charged topics (e.g., abortion, climate change, immigration, privacy). News stories would be manipulated to suggest a new bill in a nearby state that, if passed, would ensure the opposite of the participants' viewpoint on a given issue, with the intent of deceptively inducing a sense of social injustice or unfairness. They would then be presented with a call to action proposing a hacktivist attack against the passing of the bill, thereby supporting the participant's viewpoint and possibly serving as a key motive and perceived payoff underlying the attack.

Taken together, we believe that although the current study has offered answers to several questions with which we started, future research stands to expand these findings in realms of broader ethnic groups, of perceived descriptive norms, and of varying definitions and actualizations of our deindividuation and motive primes.

Contributions and Implications

The present findings contribute to our existing knowledge of game theoretic approaches toward decision-making strategies, such that when facing a risky, anti-government but pro-justice situation, college-aged participants weigh payoffs—rather than the ratio of payoffs to risks—before making an attack decision. Should this effect be replicated among a wider range of demographic groups and attack targets (e.g., individuals, non-governmental groups), this finding may suggest that we are not, in fact, rational decision-makers in domains where hacking and online activism are involved.

Furthermore, our findings underscore the importance of three key elements in the prediction of attack likelihood: (1) anonymity, group identification, and deindividuation, (2) power motives, and (3) perceived social injustice. The strong relationships between our manipulation check variables (i.e., assessing these elements) and attack likelihood speak to the importance of these factors in gauging the likelihood of carrying out a hacktivist attack, as was originally suggested through the strong themes of *we*-focused, group-priming language, power, and social injustice in our two preliminary analyses of past Anonymous activity (see Pilot Studies 1 and 2). It is our hope that future research will further explore these particular factors as they relate to illegal online activities and decision-making strategies.

The present study also contributes to the construction of a demographic profile of hacktivists in a time when hacking and online activism are increasingly on the rise, despite recent government sanctions and executive orders (e.g., President Obama's "War on Hackers"; Graham, 2015). In past research on social activism, it has been found that protests regarding a given ethnic group or gender-, socioeconomic-, or career-focused topic often comprise activists who define themselves through those distinct groups (e.g., farmers protesting a detrimental agricultural bill); however, in cases of broader controversies that affect a sizable proportion of the population, social activists tend to be young, politically active, White, highly educated, and male (Walgrave, Rucht, & Van Aelst, 2007)—qualities that have also described many well-known Anonymous members in the past (Coleman, 2014; Yar, 2013).[16]

A key difference distinguishes the former from the latter, however: While social activists seek public recognition and outward awareness through their protests, tactics, and movements, hacktivists seek the unidentifiable, the secretive, and the anonymous. Illustrating this important distinction, social activists have been shown to share information about important causes (e.g., Save Darfur) through Facebook and similar mediums not to recruit or even inform their friends and followers, but rather for peer influence and outward appearance (Lewis, Gray, & Meierhenrich, 2014). Such a desire is all but impossible in an online realm where most Anonymous members are just that—anonymous from one another. With the exception of in-person protests against the Church of Scientology and Occupy Wall Street, in which masked members could first

[16] It must be noted, however, that the degree of demographic overlap between social activists and hacktivists may change depending on the nature of the trigger event at hand. When faced with events that (unlike Net Neutrality) are not politically charged, hacktivist respondents may diverge from the demographic norms of typical social activists: They might become less politically active, less educated, and perhaps higher in impulsive sensation-seeking if targeting an individual or group for a cause that they do not strongly support or believe in.

come face-to-face (Coleman, 2014), Anonymous is renowned in part due to the mystery and lack of hierarchy that defines it.

As such, we know little about the motives underlying hacktivism as a newer, riskier, and still emerging form of activism. If public recognition is both difficult and dangerous, other motives must be at play to mobilize hundreds, if not thousands, of individuals in pursuit of a single cause. The present study highlights three such factors, and it is our hope that additional elements of importance continue to emerge in the coming years. Our end aim is to inform policymakers, lawmakers, and Internet citizens of the key predictors of, and possible protections against, hacktivist behaviors.

On this note, we end with a quote from the formerly prominent hacker Loyd Blankenship, who after his 1986 arrest published *The Conscience of a Hacker*, now regarded as a cornerstone and ethical foundation for hacker groups:

> We exist without skin color, without nationality, without religious bias, and you call us criminals. […] My crime is that of judging people by what they say and think, not what they look like. My crime is that of outsmarting you, something that you will never forgive me for. I am a hacker, and this is my manifesto. You may stop this individual, but you can't stop us all…after all, we're all alike (Blankenship, 1986).

And so, whether tied together by empowerment, anonymity, or the pursuit of social justice, it appears that hacktivism will continue to represent those most willing to dodge risk, most compelled to seek chance, and most drawn to donning the mask of a computer screen.

REFERENCES

Albanesius, C. (2011, September). Cyber crime costs $114b per year, mobile attacks on the rise. *PC Magazine*. Retrieved from http://www.pcmag.com/article2/0,2817,2392570,00.asp.

Alter, A. L., & Kwan, V. S. (2009). Cultural sharing in a global village: Evidence for extracultural cognition in European Americans. *Journal of Personality and Social Psychology*, *96*(4), 742.

Ames, D. R., & Bianchi, E. C. (2008). The agreeableness asymmetry in first impressions: Perceivers' impulse to (mis)judge agreeableness and how it is moderated by power. *Personality and Social Psychology Bulletin*, *34*, 1719-1736.

Anderson, C. & Galinsky, A. D. (2006). Power, optimism, and risk-taking. *European Journal of Social Psychology*, *36*, 511-536.

AnonymousOfficial24 (2014, May 5). *Anonymous - The Monopoly [Comcast, Net Neutrality]* [Video file]. Retrieved from https://www.youtube.com/watch?v=9CMFabT7Tyk.

Antón, P. S., Anderson, R. H., Mesic, R., & Scheiern, M. (2013). *The Vulnerability assessment & mitigation methodology: Finding and fixing vulnerabilities in information systems* (Report No. OMB 0704-0188). RAND National Defense Research Institute.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596.

Ayres, J., & Hopf, T. (1992). Visualization: Reducing speech anxiety and enhancing performance. *Communication Reports*, *5*(1), 1-10.

Ayres, J., & Hopf, T. S. (1990). The long ‐ term effect of visualization in the classroom: A brief research report. *Communication Education*, *39*(1), 75-78.

Ayres, J., & Hopf, T. S. (1985). Visualization: A means of reducing speech anxiety. *Communication Education*, *34*(4), 318-323.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*(7), 462.

Beauregard, K. S., & Dunning, D. (1998). Turning up the contrast: self-enhancement motives prompt egocentric contrast effects in social judgments. *Journal of Personality and Social Psychology*, *74*(3), 606.

Berger, J. & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, *49*(2), 192-205.

Blankenship, L. (1986). The conscience of a hacker. *Phrack Magazine*, *18*(3).

Bochner, S., & Hesketh, B. (1994). Power distance, individualism/collectivism, and job-related attitudes in a culturally diverse work group. *Journal of Cross-Cultural Psychology*, *25*(2), 233-257.

Brown, J. (2014). Notes to the underground: Responsibility claims as agenda-setting and standard-setting by activist and terrorist groups.

Carney, D. R., Cuddy, A. J., & Yap, A. J. (2010). Power posing brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science*, *21*(10), 1363-1368.

Chan, M. (2010). The impact of email on collective action: A field application of the SIDE model. *New Media & Society*, *12*(8), 1313-1330.

Chinchani, R., Iyer, A., Ngo, H., & Upadhyaya, S. (2004). *A Target-centric formal model for insider threat and more* (Technical Report No. 2004-16). Buffalo, NY: University of Buffalo.

Churchill, S. D. (1991). Reasons, causes, and motives: Psychology's illusive explanations of behavior. *Theoretical & Philosophical Psychology*, *11*(1), 24-34.

Cialdini, R. B. (2003). Crafting normative messages to protect the environment. *Current Directions in Psychological Science*, *12*(4), 105-109.

Coleman, G. (2014). *Hacker, hoaxer, whistleblower, spy: The Many faces of Anonymous*. Brooklyn, NY: Verso Books.

Colesky, M. R. & Van Niekerk, J. (2012, September). *Hacktivism: controlling the effects*. Paper presented at The 2012 Annual Conference on WWW Applications, Durban, South Africa.

Colvin, C. R., & Griffo, R. (2007). The psychological costs of self-enhancement. In E. C. Chang (Ed.), *Self-criticism and self-enhancement: Theory, research, and clinical implications* (pp.123.140). Washington, DC: American Psychological Association.

Costa, P.T., Jr. & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual*. Odessa, FL: Psychological Assessment Resources.

Crocker, J., Thompson, L. L., McGraw, K. M., & Ingerman, C. (1987). Downward comparison, prejudice, and evaluations of others: effects of self-esteem and threat. *Journal of Personality and Social Psychology*, *52*(5), 907.

Crosby, F. (1976). A Model of egoistical relative deprivation. *Psychological Review*, *83*, 85-113.

Dahan, M. (2013). *Hacking for the homeland: Patriotic hackers versus hacktivists*. Proceedings of the 8th International Conference on Information Warfare and Security. Denver, CO: ICIW.

Diener, E. (1979). Deindividuation, self-awareness, and disinhibition. *Journal of Personality and Social Psychology*, *37*(7), 1160-1171.

Dixit, A. K. & Nalebuff, B. J. (2008). *The Art of strategy: A Game theorist's guide to success in business & life*. New York, NY: W. W. Norton & Company, Inc.

Espenschied, J. (2012, June). A discussion of threat behavior: Attackers & patterns. Paper presented at The *Microsoft Corporation and NATO CyCon* (*Cooperative Cyber Defence Centre of Excellence*). Seattle, WA: Microsoft Trustworthy Computing.

Festinger, L., Pepitone, A., & Newcomb, T. (1952). Some consequences of de-individuation in a group. *The Journal of Abnormal and Social Psychology*, *47*(2S), 382-389.

Galinsky, A. D., Ku, G., & Wang, C. S. (2005). Perspective-taking and self-other overlap: Fostering social bonds and facilitating social coordination. *Group Processes & Intergroup Relations*, *8*(2), 109-124.

Galinsky, A. D., Gruenfeld, D. H., & Magee, J. C. (2003). From power to action. *Journal Of Personality and Social Psychology*, *85*(3), 453.

Gardner, W. L., Gabriel, S., & Lee, A. Y. (1999). "I" value freedom, but "we" value relationships: Self-construal priming mirrors cultural differences in judgment. *Psychological Science*, *10*(4), 321-326.

Gellman, B. (2012, April 18). The World's 100 Most Influential People: 2012: Anonymous. *TIME*. Retrieved from http://content.time.com/time/specials/packages/article/0,28804,2111975_2111976_2112122,00.html.

Gergen, K. J., Gergen, M. M., & Barton, W. H. (1973). Deviance in the dark. *Psychology Today*, *7*(5), 129-130.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, *35*(3), 472-482.

Gosling, S. D., Rentfrow, P. J., & Swann, Jr, W. B. (2003). A Very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504-528.

Graham, R. (2015, January 15). President Obama Is Waging a War on Hackers. *WIRED*. Retrieved from http://www.wired.com/2015/01/president-obama-waging-war-hackers.

Greenberg, J. (1987). A Taxonomy of organizational justice theories. *The Academy of Management Review*, *12*(1), 9-22.

Greitzer, F. L., Kangas, L. J., Noonan, C. F., Dalton, A. C., & Hohimer, R. E. (2012). Identifying at-risk employees: Modeling psychosocial precursors of potential insider threats. In *Proceedings of the 45th Hawaii International Conference on System Science* (pp. 2392-2401). Maui, HI: IEEE.

Greitzer, F. L., Kangas, L. J., Noonan, C. F., & Dalton, A. C. (2010). *Identifying at-risk employees: A behavioral model for predicting potential insider threats* (Report No. PNNL-19665). U.S. Department of Energy.

Griskevicius, V., Tybur, J. M., Gangestad, S. W., Perea, E. F., Shapiro, J. R., & Kenrick, D. T. (2009). Aggress to impress: Hostility as an evolved context-dependent strategy. *Journal of Personality and Social Psychology*, *96*, 980-994.

Griskevicius, V. & Kenrick, D. T. (2013). Fundamental motives for why we buy: How evolutionary needs influence consumer behavior. *Journal of Consumer Psychology*, *23*(3), 372-386.

Hackenbrack, K. (1992). Implications of seemingly irrelevant evidence in audit judgment. *Journal of Accounting Research*, 126-136.

Hai-Jew, S. (2013). Action potentials: extrapolating an ideology from the Anonymous hacker socio-political movement (a qualitative meta-analysis). In C. M. Akrivopoulou & N. Garipidis (Eds.), *Digital Democracy and the Impact of Technology on Governance and Politics: New Globalized Practices*. Hershey, PA: IGI Global.

Helweg-Larsen, M., & Shepperd, J. A. (2001). Do moderators of the optimistic bias affect personal or target risk estimates? A review of the literature. *Personality and Social Psychology Review*, *5*, 74–95.

Higgins, E. T. (1989). Self-discrepancy theory: What patterns of self-beliefs cause people to suffer. *Advances In Experimental Social Psychology*, *22*, 93-136.

Highhouse, S., & Gallo, A. (1997). Order effects in personnel decision making. *Human Performance*, *10*(1), 31-46.

Hofer, J. & Chasiotis, A. (2011). Implicit motives across cultures. *Online Readings in Psychology and Culture*, *4*(1), 2-16

Hofstede, G. (1980). Motivation, leadership, and organization: do American theories apply abroad?. *Organizational Dynamics*, *9*(1), 42-63.

Jennings, M. K. (1991). Thinking about social injustice. *Political Psychology*, *12*(2), 187-204.

Johnson, R. D., & Downing, L. L. (1979). Deindividuation and valence of cues: effects on prosocial and antisocial behavior. *Journal of Personality and Social Psychology*, *37*(9), 1532.

Jordan, T. & Taylor, P. A. (2004). *Hacktivism and cyberwars: Rebels with a cause*? New York, NY: Routledge.

Kenrick, D. T., Griskevicius, V., Neuberg, S. L., & Schaller, M. (2010). Renovating the pyramid of needs contemporary extensions built upon ancient foundations. *Perspectives on Psychological Science*, *5*(3), 292-314.

Knappenberger, B. (Producer & Director). (2012). *We are legion: The Rise of the hacktivists* [Motion picture]. United States: Luminant Media.

Krawczyk, K. (2014, June). Cyber crime costs the world more money than some natural disasters do. *Digital Trends*. Retrieved from http://www.digitaltrends.com/computing/new-study-says-cyber-crime-costs-hundreds-of-billions-per-year.

Kwan, V. S. Y., Wojcik, S. P., Miron-shatz, T., Votruba, A. M., & Olivola, C. Y. (2012). Effects of symptom presentation order on perceived disease risk. *Psychological Science*, *23*(4), 381-385.

Kwan, V. S. Y., Kuang, L. L., & Zhao, B. (2008). In search of the optimal ego: When self-enhancement bias helps and hurts adjustment. In H. Wayment & J. Bauer (Eds.) *Quieting the ego: Psychological benefits of transcending ego* (pp. 43-52). Washington, DC: American Psychological Association.

Lammers, J., Stapel, D. A., & Galinsky, A. D. (2010). Power increases hypocrisy moralizing in reasoning, immorality in behavior. *Psychological Science*, *21*(5), 737-744.

Lee, E. (2007). Deindividuation effects on group polarization in computer-mediated communication: The role of group identification, public-self-awareness, and perceived argument quality. *Journal of Communication*, *57*(2), 385-403.

Lerner, M. J. (2003). The Justice motive: Where social psychologists found it, how they lost it, and why they may not find it again. *Personality and Social Psychology Review*, *7*(4), 388-399.

Lerner, M. J. (1977). The Justice motive: Some hypotheses as to its origins and forms. *Journal of Personality*, *45*(1), 1-52.

Lerner, M. J. (1975). The Justice motive in social behavior: Introduction. *Journal of Social Issues*, *31*(3), 1-19.

Lewis, K., Gray, K., & Meierhenrich, J. (2014). The Structure of online activism. *Sociological Science*, *1*, 1-9.

Liu, P. (2005). *A game theoretic approach to cyber-attack prediction* (Report No. DOE/ER/25527). U.S. Department of Energy: Office of Science.

Manion, M., & Goodrum, A. (2000). Terrorism or civil disobedience: toward a hacktivist ethic. *ACM SIGCAS Computers and Society*, *30*(2), 14-19.

Mansfield-Devine, S. (2011). Hacktivism: assessing the damage. *Network Security*, *8*, 5-13.

Martin, J., Brickman, P., & Murray, A. (1984). Moral outrage and pragmatism: Explanations for collective action. *Journal of Experimental Social Psychology*, *20*(5), 484-496.

Maslow, A.H. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370-396.

Massa, F. G. (2011, January). Out of bounds: Anonymous' transition to collective action. In *Academy of Management Proceedings*, *1*, 1-6. San Antonio, TX: Academy of Management.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology: General*, *90*, 227-234.

Newcomb, T. M. (1950). Acquiring motives and attitudes. In T. M. Newcomb (Ed.) *Social Psychology* (pp. 107-146). Fort Worth, TX: Dryden Press.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *LIWC: Linguistic Inquiry and Word Count* [Computer software], Version LIWC2007.

Perrin, F. A. C. (1923). The psychology of motivation. *Psychological Review*, *30*(3), 176-191.

Pfleeger, S. L., & Caputo, D. D. (2012). Leveraging behavioral science to mitigate cyber security risk. *Computers & Security*, *31*(4), 597-611.

Phillips, W. (2013). The house that Fox built: Anonymous, spectacle, and cycles of amplification. *Television & New Media*, *14*(6), 494-509.

Pilkington, E. (2013, November). Jailed Anonymous hacker Jeremy Hammond: 'My days of hacking are done'. *The Guardian*. Retrieved from http://www.theguardian.com/technology/2013/nov/15/jeremy-hammond-anonymous-hacker-sentenced.

Ponemon Institute (2012). Cyber security on the offense: A study of IT security experts. Traverse City, MI: *Ponemon Institute LLC*. Retrieved from http://security.radware.com/uploadedFiles/Resources_and_Content/Attack_Tools/CyberSecurityontheOffense.pdf.

Postmes, T., Spears, R., & Lea, M. (1998). Breaching or building social boundaries? SIDE-effects of computer-mediated communication. *Communication Research*, *25*(6), 689-715.

Postmes, T., Spears, R., Sakhel, K., & de Groot, D. (2001). Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin*, *27*(10), 1243-1254.

Rogers, A. (2014, August). What Anonymous is doing in Ferguson. *TIME*: Crime. Retrieved from http://time.com/3148925/ferguson-michael-brown-anonymous.

Rogers, M. (1999). *Modern-day Robin Hood or moral disengagement: Understanding the justification for criminal computer activity*. The Center for Education and Research in Information Assurance and Security at Purdue University. Retrieved from http://homes.cerias.purdue.edu/~mkr/moral.doc.

Schröder, T., & Thagard, P. (2013). The affective meanings of automatic social behaviors: Three mechanisms that explain priming. *Psychological Review*, *120*(1), 255-280.

Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Social Psychology*, *3*(2), 102-116.

Shakarian, P., Shakarian, J., & Ruef, A. (2013). Cyber attacks by nonstate hacking groups: The Case of Anonymous and its affiliates. In P. Shakarian, J. Shakarian, & A. Ruef (Eds.) *Introduction to Cyber-Warfare: A Multidisciplinary Approach*. Sebastopol, CA: Syngress.

Shepperd, J. A., Carroll, P., Grace, J., & Terry, M. (2002). Exploring the causes of comparative optimism. *Psychologica Belgica*, *42*(1/2), 65-98.

Shim, W., Allodi, L. & Massacci, F. (2012). *Crime pays if you are just an average hacker*. Paper presented at the ASE International Conference on Social Informatics. CyberSecurity 2012.

Sinclair, S., & Smith, S. W. (2008). Preventative directions for insider threat mitigation via access control. In S. J. Stolfo, S. M. Bellovin, S. Hershkop, A. D. Keromytis, S. Sinclair, and S. Smith (Eds.) *Insider attack and cyber security: beyond the hacker* (pp. 165-194). New York, NY: Springer.

Spears, R., Postmes, T., Lea, M., & Wolbert, A. (2002). When are net effects gross products? the power of influence and the influence of power in computer-mediated communication. *Journal of Social Issues*, *58*(1), 91-107.

Stryker, C. (2012). *Hacking the future: Privacy, identity, and anonymity on the Web*. New York, NY: Overlook Press.

Sumner, M., & Samuel, A. G. (2007). Lexical inhibition and sublexical facilitation are surprisingly long lasting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 769-790.

Swann, W. B. (1983). Self-verification: Bringing social reality into harmony with the self. In J. M. Suls & A. G. Greenwald (Eds.), *Social psychological perspectives on the self* (Vol. 2, pp. 33-66). Hillsdale, NJ: Erlbaum.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks/Cole Publishing.

Timeline of events associated with Anonymous (n.d.). In *Wikipedia*. Retrieved from http://en.wikipedia.org/wiki/Timeline_of_events_associated_with_Anonymous.

Trope, Y (1983). Self-assessment in achievement behavior. In J. M. Suls & A. G. Greenwald (Eds.), *Social psychological perspectives on the self* (Vol. 2, pp. 93-121). Hillsdale, NJ: Erlbaum.

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *8*(4), 336-342.

Turgeman-Goldschmidt, O. (2005). Hackers' accounts hacking as a social entertainment. *Social Science Computer Review*, *23*(1), 8-23.

Turner, J. H. (1987). Toward a sociological theory of motivation. *American Sociological Review*, 15-27.

Turner, J. & Oakes, P. (1986). The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence. *British Journal of Social Psychology*, *25*(3), 237–252.

Tyler, T. R. (1994). Psychological models of the justice motive: Antecedents of distributive and procedural justice. *Journal of Personality and Social Psychology*, *67*(5), 850-863.

United Nations (2007). *Universal declaration of human rights*. Retrieved from http://www. un.org/en/documents/udhr/index.shtml.

Vegh, S. (2003). Classifying forms of online activism. In M. McCaughey & M. D. Ayers (Eds.) *Cyberactivism: Online Activism in theory and Practice* (pp. 71-95). New York, NY: Routledge.

Vincent, J. (2013, December). $183,000 fine for man who joined Anonymous attack for 'one minute'. *The Independent: Technology*. Retrieved from http://www.independent.co.uk/ life-style/gadgets-and-tech/183000-fine-for-man-who-joined-anonymous-attack-for-one-minute-8995609.html.

Voiskounsky, A. E., & Smyslova, O. V. (2003). Flow in computer hacking: A model. In *Web and Communication Technologies and Internet-Related Social Issues: International Conference on Human Society @ Internet* (pp. 176-186). Berlin, Germany: Springer Berlin-Heidelberg.

von Neumann, J. (1928). Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, *100*(1), 295-320.

Walgrave, S., Rucht, D., & Van Aelst, P. (2007). Socio-demographics: Typical new social movement activists, old leftists or normalized protesters? In S. Walgrave & D. Rucht (Eds.) *Protest politics: Antiwar mobilization in advanced industrial democracies* (in preparation).

White, A. E., Johnson, K. A., & Kwan, V. S. Y. (2014). Four Ways to Infect Me: Spatial, Temporal, Social, and Probability Distance Influence Evaluations of Disease Threat. *Social Cognition*, *32*(3), 239-255.

Wines, M. & Fitzsimmons, E. G. (2014, August 21). What started as a local protest in Missouri grows into a center of national activism. *The New York Times*, p. A14.

Wyer, R. S., & Srull, T. K. (Eds.). (1989). *Social intelligence and cognitive assessments of personality* (Vol. 2). Psychology Press.

Yar, M. (2013). *Cybercrime and society*. New York, NY: SAGE Publications Limited.

Zhong, C. B., Bohns, V. K., & Gino, F. (2010). Good Lamps Are the Best Police Darkness Increases Dishonesty and Self-Interested Behavior. *Psychological Science*, *21*(3), 311-314.

Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order vs. deindividuation, impulse and chaos. In W. J. Arnold & D. Levine (Eds.), *Nebraska symposium on motivation* (Vol. 17, pp. 237–307). Lincoln, NE: University of Nebraska Press.

Zimbardo, P. G. (1971). The Power and pathology of imprisonment. *Congressional Record* (Serial No. 15, 1971-10-25). Hearings before Subcommittee No. 3, of the Committee on the Judiciary, House of Representatives, Ninety-Second Congress, First Session on Corrections, Part II, Prisons, Prison Reform and Prisoner's Rights: California. Washington, DC: U.S. Government Printing Office.

Zimbardo, P. G. (2004). A Situationist perspective on the psychology of evil: Understanding how good people are transformed into perpetrators. In A. G. Miller (Ed.) *The Social psychology of good and evil*. New York, NY: The Guilford Press.

Zuckerman, M., Eysenck, S. B., & Eysenck, H. J. (1978). Sensation seeking in England and America: Cross-cultural, age, and sex comparisons. *Journal of Consulting and Clinical Psychology*, *46*(1), 139-149.

Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., & Kraft, M. (1993). A comparison of three structural models for personality: The Big Three, the Big Five, and the Alternative Five. *Journal of Personality and Social Psychology*, *65*(4), 757.

APPENDIX A

PILOT STUDY MATERIALS

A.1. *Pilot study 1: 25 Anonymous case studies*

Dog Poop Girl. In 2005, passengers on a subway in South Korea watched as a woman's small dog pooped on the floor of the subway car. When the woman made no move to clean it up, several passengers asked her to do so, one offering her a tissue. The woman used the tissue to wipe the dog, but did not touch the floor. A passenger took photos of the woman and incident and posted them online, which members of Anonymous spread under the label of "Dog S*** Girl"; within days, her name and personal information were released on various websites. As the harassment continued, including mentions of her family, the woman was forced to leave university and issue a public apology, during which she threatened suicide if the harassment did not stop.

Habbo Hotel. Habbo Hotel was an early social networking site that allowed users to explore the animated world of Habbo while meeting other users under the guise of personally designed avatars. In mid-2005, Habbo's site moderators banned African-American avatars, stating that the decision ensured better gameplay for its users. In protest, members of Anonymous gained access into Habbo's site, created dozens of African-American avatars characterized by business suits and afros, which gathered at the entrance of the site's pool area. "Pool's Closed" became a catch-phrase for these raids.

Hal Turner. Hal Turner is an American white nationalist most notable for broadcasting The Hal Turner Show, a radio program that occasionally lobbied for "pro-White" rallies, denied the existence of the Holocaust, and supported the shooting of illegal immigrants as punishment for their illegal status. In 2006 and 2007, members of Anonymous took Turner's website offline, costing him thousands of dollars in bandwidth bills. Turner later sued 4chan and other Anonymous-related websites; however, the lawsuit was later dismissed.

Dusty the Cat. In 2009, 14-year-old Kenny Glenn uploaded two videos to YouTube in which he beats and abuses a cat named Dusty. Members of Anonymous were able to trace his YouTube account to a MySpace username, where they then located his address, phone number, and additional personal information. Glenn was arrested and the cat placed in a safe home; however, Anonymous continued to encourage other members to shame and humiliate Glenn's family and younger brother, who filmed both videos.

BRB Church. In late 2007, 53-year-old Chris Forcand revealed in online message boards that he was attracted to, and interested in pursuing sexual relations with, underage girls. Members of Anonymous disguised themselves as 13-year-old girls and lured Forcand into exposing himself and making clear moves toward these supposed girls, at one point asking if they could engage in cybersex once he got back from church (his words "brb church" became a catch-phrase for his later sentencing). These members then notified the Toronto police, who arrested Forcand and charged him with two counts of luring children under the age of 14.

#OpDarknet. In 2011, members of Anonymous discovered that Freedom Hosting, among other web hosting companies, was associated with child pornography sites. When Freedom Hosting refused to remove this content, Anonymous shut down approximately 40 offending sites and published a statement that read, "Remove all child pornography content from your servers. Refuse to provide hosting services to any website dealing with child pornography. This statement is not just aimed at Freedom Hosting, but everyone on the Internet. It does not matter who you are, if we find you to be hosting, promoting, or supporting child pornography, you will become a target." They subsequently released the log-in information for more than 1,500 members of the offending sites. The operator of Freedom Hosting was arrested soon thereafter, as were more than 140 frequent visitors to child pornography sites.

Project Chanology. In early 2008, the Church of Scientology produced a video featuring an interview with Tom Cruise, who revealed aspects of the religion that the Church wished to keep confidential. When the video was leaked to YouTube, the Church issued a copyright violation claim against YouTube, requesting its removal. Members of Anonymous, who perceived this action as a form of Internet censorship, prevented access to Scientology websites and inundated Churches of Scientology with prank calls (including to SWAT teams), black faxes, and thousands of pizza orders.

Zhang Ya. After the 2008 Siuchan earthquake, a young woman named Zhang Ya released a YouTube video in which she complains about the earthquake and its victims, stating that she would rather more attention be paid to her rather than the earthquake; some victims deserved to die, and preferably sooner; many survivors were too unattractive to be featured on television; and the news coverage was preventing her favorite shows from airing. Members of Anonymous discovered and spread this video along with Ya's personal information, going so far as to include her blood type. Ya was arrested and held in custody for three days for breaking laws of defamation and endangering public stability.

Operation Titstorm. In 2010, the Australian government attempted to censor and outlaw pornography that featured small-breasted women and female ejaculation. In protest under claims of sexism and censorship, members of Anonymous took down a series of Australian government websites including the Australian Parliament House, which was unavailable for three days following the attacks.

WikiLeaks: Operation Payback. In late 2010, the document archive website WikiLeaks came under intense pressure to remove published U.S. diplomatic cables (i.e., communications). In response, members of Anonymous took down websites that declared themselves anti-WikiLeaks, including Amazon, PayPal, MasterCard, and Visa.

Arab Spring. In 2010 and 2011, members of Anonymous led citizens of Tunisia, Egypt, and Libya into an upsurge against their respective governments during the Tunisian Revolution, Egyptian Revolution, and Libyan Civil War. Groups that sided with dictators or pro-censorship laws were particularly targeted for website attacks and defacements. Anonymous also released the e-mail accounts and passwords of major North African and Middle Eastern government officials, including those from Bahrain, Egypt, Jordan, and Morocco.

HBGary Federal. In early 2011, the chief executive of security firm HBGary Federal announced that his firm had successfully infiltrated the Anonymous group, the findings from which he would present at a later conference. In retaliation, members of Anonymous hacked the HBGary Federal website, replacing the homepage with a statement that Anonymous should not be messed with. Anonymous took control of the company's e-mail and phone systems, sending 68,000 spam messages, erasing files, and limiting access to company phone lines. They also took control of the chief executive's Twitter account, where they posted his current address and social security number.

Operation Sony. George Hotz, best known for releasing the first software to jailbreak iPhones, breached the system security of Sony's PlayStation 3. This breach, which Hotz released on his blog in early 2011, allowed any software to be run on a PS3 system. Sony filed a lawsuit against Hotz, soon thereafter gaining access to the IP addresses of all visitors to his blog. Members of Anonymous responded to this claimed obstruction of freedom by causing a major outage of the PlayStation Network and Sony's website.

Operation Orlando. In 2011, members of Food Not Bombs—an organization that feeds surplus food from grocery stores and restaurants to those in need—were arrested for feeding the homeless in a park in Orlando, Florida. Members of Anonymous responded by gaining access to and defacing a different Orlando-related website every day, including that of Orlando International Airport and the Orlando mayor's re-election site.

Chinga la Migra. In 2011, the Arizona Department of Public Safety sought the passage of Arizona SB 1070, the broadest and most strict anti-illegal immigration measure in recent U.S. history. Members of Anonymous and LulzSec, which viewed this law as unjust racial profiling, released hundreds of highly sensitive documents taken from the Arizona DPS. They released the names, addresses, social security numbers, Internet passwords, e-mails, and voicemails of dozens of Arizona border patrol officers, as well as officer chat logs containing racist remarks and documents evincing at least one officer of sex offender status. They stated that they wanted the officers to "experience just a taste of the same kind of violence and terror they dish out on an everyday basis."

Occupy Wall Street. In support of the Occupy Wall Street movement, members of Anonymous released thousands of names, ranks, addresses, phone numbers, passwords, and social security numbers of police officers in various cities around the country directly related to the Occupy movement. More than 40 related websites were taken down or otherwise defaced.

Los Zetas. In late 2011, Anonymous discovered that Los Zetas—considered the most dangerous and technologically advanced drug cartel in Mexico—was holding an Anonymous member hostage. Anonymous threatened Los Zetas, stating they would release identifying information about cartel members and collaborators, which would likely lead to their prosecution, execution, or targeting by rival cartels. Los Zetas responded that for every piece of personally identifiable information released, they would kill ten innocent people. Los Zetas later released their hostage within the timeframe Anonymous had set.

Stratfor. On Christmas Eve, 2011 members of Anonymous stole thousands of e-mail addresses and credit card numbers from wealthy and corporate clients of Stratfor, an international security firm that failed to aptly encrypt (i.e., protect) their clients' data. They donated all obtained funds to charities such as American Red Cross, CARE, and Save the Children. Early Christmas morning, Anonymous Tweeted, "Not so private and secret anymore?" while hactivist group Lulzsec wrote, "Y u no bother encrypting?"

#SOPAblackout. The Stop Online Piracy Act (SOPA) and Protect Intellectual Property Act (PIPA) bills were introduced in early 2012 as an attempt to shut down illegal download websites, as well as video- and movie-streaming websites such as Megaupload. Furthermore, Internet service providers (e.g., Cox, Time Warner Cable) would, if passed, have the ability to restrict Internet speeds and access depending on customer's data plans, similar to cable or phone plans. Members of Anonymous viewed such movements as direct attacks on a free and uncensored Internet, and engaged in the attack and defacement of government (e.g., U.S. Department of Justice, FBI) and copyright or recording industry websites (U.S. Copyright Office, MPAA, Warner Bros. Music, RIAA).

Opération Québec. In 2012, Quebec Bill 78 was passed, restricting the freedom of association—that is, the freedom to gather or take collective action in a group's interests—following weeks of student protests. Members of Anonymous released a video that urged Quebec to let their citizens protest in line with freedom of speech and opinion; when the government did not adhere to this demand, Anonymous gained access to several government websites and released a two-hour-long video of a government party, at which former U.S. presidents and Canadian politicians were present.

Operation Japan. In 2012, members of Anonymous took down the Japanese Business Federation website after Japanese copyright laws were amended to fine anyone in possession of pirated material (e.g., pirated music or movies) up to $25,000 and two years in prison.

Uganda LGBT Rights. In 2012, Anonymous gained access to two major Ugandan government websites, including that of the president, to protest the country's strict anti-gay laws. These hacks directly followed Uganda's first Pride Parade, in spite of Uganda's policy that homosexuality is punishable by up to 14 years in prison.

Hong Kong National Education. In 2012, the Hong Kong government organization known as the National Education Centre revised its education curriculum for children between the ages of 6 and 18, not grading based on learned factual information, but rather on emotional attachment to—and approval of—the Communist Party of China. In response, Anonymous threatened the Hong Kong government, later leaking classified government documents and defacing its websites.

Steubenville Rape. In 2012, a high school girl in Steubenville, Ohio was gang raped while unconscious from alcohol use. In early 2013, only a subset of the men responsible were charged in court. Members of Anonymous gained access to these men's e-mails and phone records, revealing information about additional men involved in the gang rape. They released incriminating videos, photos, and tweets of all involved participants, leading to their later sentencing.

Operation Free North Korea. In response to the Kim Jong-un administration, members of Anonymous engaged in a series of attacks on the North Korean government, demanding that it adopt a free and democratic government, abandon its nuclear ambitions, and provide uncensored Internet access, among other impositions. They have since waged "Cyber War" against the North Korean government, releasing more than 15,000 usernames and passwords associated with the regime; hacking into main government websites, Twitter, and Flickr accounts; and uploading an image of Jong-un's face with a pig-like snout and Mickey Mouse tattoo on his chest over the text, "Threatening world peace with ICBMs and Nuclear weapons/Wasting money while his people starve to death."

A.2. *Pilot study 1: Definitions of six identified motives*

Self-protection. Desire to protect oneself from harm or negative consequences. This is often reactionary against something perceived as a personal threat. Note that "self" here can be either individual or collective.

Status, power. Desire for power (i.e., the ability to make an impact on others' beliefs or behaviors) or status (prestige, reputation). This may be seeking power for oneself (to enhance one's own self worth), or seeking prestige among a group of others (e.g., other hackers, the public), either implicitly or explicitly.

Sensation-seeking. Desire for experiences and feelings that are varied, novel, complex, and intense. A readiness to take physical, social, legal, and financial risks for the sake of these experiences. Doing things "for fun" or out of curiosity.

Third-party punishment. Desire of an outside observer (i.e., third party) to punish a person or group of people for violating social norms, even though the third party is not directly affected by the violation. Also considered "altruistic punishment", described as "playing God". This has a strict morality/sinning/criminality component.

Social injustice. Desire to correct unfairness or injustice of a society, possibly based on inequalities, suppression, or burdens placed on certain subgroups. Possibly a desire to maintain personal rights, including freedom of speech and expression. This has a strict legal/unjust component.

Economic redistribution. Desire to redistribute wealth to bring unequal groups to equal standing in terms of wealth or other resources (e.g., Robin Hood complex).

A.3. *Pilot study 2: 11 Anonymous slogans*

1. We are the 99%.
2. The Internet as we know it will end. FIGHT BACK.
3. We see. We judge.
4. The people should not be afraid of their government. The government should be afraid of its people.
5. Y u no bother encrypting?
6. We are legion. We do not forgive. We do not forget.
7. Authority can't break down a movement if there isn't a leader to corrupt.
8. We stand for freedom of speech, the power of the people.
9. We have no leadership.
10. We're speaking as one, and it's as a collective.
11. You want to see Anonymous rise up? Try to shut down its message. Then you'll see what Anonymous can do.

APPENDIX B

STUDY MATERIALS

B.1. *Venues for identifying and recruiting participants*

- School of Computing, Informatics, Decision Systems Engineering (CIDSE)
- Residential Life (University Housing) for Ira A. Fulton Schools of Engineering
- ASU Computer Science Facebook group (801 members)
- ASU Computer Science Club
- ASU Women in Computer Science (WCS)
- ASU Computer Systems Engineering Facebook group (176 members)
- ASU Advanced Technology Innovation Center
- Students enrolled in CST 100 and 200; CSE 110, 200, and 205 (Polytechnic Campus)
- Students enrolled in CIS 235, 340, 345, 365, 425, 430, 440 (50+ students)
- CIDSE Mentoring Program
- ASU Association for Computer Systems Security
- Software Developers Association of ASU (SoDA)
- ASU eSports Association
- ASU Linux User's Group (ASULUG; Tempe and West Campuses)
- ASU Department of Information Systems Club (DISC)

Note: Participant recruitment took place over winter break, during which few students responded to requests for participation either from professors, organization leaders, or experimenters. Due to slow recruitment, an IRB modification was filed requesting to post flyers (comprising the same IRB consent message given to all other recruited participants) inside all buildings housing the above majors.

B.2. *Guided visualization third-person scenario*

What is your sex? *Male, Female*

How old are you? [*Drop-down from 17 or younger* [*exit*] *to 60 or higher*]

[*Redirect to gender-specific scenario*]

Guided Visualization Scenario

Jordan is in **his/her** junior year in ASU's Ira A. Fulton Schools of Engineering. Last summer, **s/he** completed a computer security internship at Cisco. For a class assignment, **s/he** was asked to write a paper on current and ongoing legislation related to Net Neutrality. Net Neutrality advocates an equal Internet that does not discriminate by user, content, or platform; however, after reading through recent news articles on this controversy, **s/he** finds **himself/herself** feeling increasingly opposed to Net Neutrality due to the ease with which government branches and Internet providers could exercise control over clients' and citizens' access.

After searching extensively through Google results on Net Neutrality, **s/he** finds **himself/herself** in a barebones forum with posts dated this morning. The forum's name is

"Operation Net Neutrality," and the post at the top of the archive was posted by a user whose handle, like the rest in the forum, is simply "anonymous."

Please continue to the next page to read this exact post. **You will be allowed to continue once you have finished reading.**

B.3. *Manipulation: Anonymous call to action*

Operation Net Neutrality

Note: Will collect hidden data pertaining to (1) whether participant attempts to click any of the false links below, and (2) how long they remain on the page before continuing.

The Internet as **we/you** know it is on the brink of falling into the hands of corrupt corporations. Net Neutrality at its core brings a just, transparent, and equal Internet to all, but those lurking behind the shadow of Net Neutrality are trying to destroy this equality.

The Federal Communications Commission is only a puppet controlled by multi-million-dollar Internet providers who would provide faster Internet services to high-paying customers, and who would be able to censor any Internet content they wish to hide.

This will violate and suppress **our/your** freedom of communication and expression.

(*Power motive*) **The people should not be afraid of their government. The government should be afraid of its people.**

(*Justice motive*) **EXPECT JUSTICE. DEMAND CHANGE.**

(*Justice + Power motive*) **The people should not be afraid of their government. The government should be afraid of its people. EXPECT JUSTICE. DEMAND CHANGE.**

(*Control*) [Nothing]

Whether [**you are**] rich or poor, young or old, the Internet should allow all people to seek information and communicate globally. **We/You** must not turn over **our/your** rights to the highest bidder. A free, open, and equal Internet is essential to a just world.

What We/You Must do to Stop Them

During the week of January 23rd, run this bot.[17] It auto-clicks the ads of dozens of Internet providers. Their ad account will be suspended for suspected click-fraud, losing a primary source of their income.

On January 23rd at 9:14 AM EST, DDoS the FCC and Department of Justice with an LOIC flood, which will take their services offline. Download LOIC here. IP addresses are listed here.

At 9:36 AM EST, DDoS these primary Internet providers controlling the FCC.

At 10:18 AM EST, follow these attacks with thousands of spam e-mails, black faxes, and prank phone calls to these same targets.

(*Power motive*) **The Internet as we/you know it will end. FIGHT BACK. WE/YOU CAN WIN THIS.**

(*Justice motive*) **[We/You] must not forgive censorship. [We/You] must not forget the denial of our human rights.**

(*Justice + Power motive*) **The Internet as we/you know it will end. FIGHT BACK. WE/YOU CAN WIN THIS. [We/You] must not forgive censorship. [We/You] must not forget the denial of our human rights.**

(*Control*) [Nothing]

B.4. *Perceived risks, payoffs, and attack likelihood*

Jordan finds **himself/herself** contemplating whether **s/he** might join in this call to action, particularly weighing potential *payoffs* (successfully protesting Net Neutrality) against potential *risks* (being caught).

The following questions ask about the *risks* Jordan may or may not face. By *risks*, we mean the riskiness of getting caught while taking part in this call to action.

- How **risky do you think it would be for Jordan** to take part in this attack?
- If Jordan takes part in this attack, what do you think is the **likelihood that s/he will get caught?**
- If Jordan is caught, **how severe do you think his/her punishment will be?**

[17] The four hyperlinks that appear to be embedded in the call manipulation text are false; participants will see cobalt blue, underlined text that indicates a hyperlink, but that does not redirect to another location.

The following questions ask about the *payoffs* Jordan may or may not gain. By *payoffs*, we mean either psychological or tangible benefits that could stem from taking part in this call to action.

- To what extent could Jordan **benefit in terms of his/her pride** by taking part in this attack?
- To what extent could Jordan **benefit others** by taking part in this attack?
- To what extent could Jordan **challenge himself/herself** by taking part in this attack?
- To what extent could Jordan **boost his/her status** among his/her peers by taking part in this attack?
- To what extent could Jordan **help fight Net Neutrality** by taking part in this attack?
- To what extent could Jordan **raise public awareness** by taking part in this attack?
- If Jordan takes part in this attack, would you classify his/her benefits as **more personally focused or more societally focused?**
- If Jordan takes part in this attack, what do you think is the **likelihood that s/he will successfully benefit** from his/her participation in the end?
- If Jordan is successful, **how sizable do you think HIS/HER overall payoffs will be?**
- How likely do you think Jordan is to participate in this call to action?

Manipulation Checks

- To what extent do you think Jordan might **feel a sense of empowerment?**
- To what extent do you think Jordan views **this situation as an example of social injustice?**
- How **anonymous (i.e., unidentifiable) do you think Jordan feels** while reading this call to action?
- To what extent do you think Jordan **identifies with (i.e., feels a part of) the group of people** who will take part in this attack?
- Outside of class assignments, how common do you think it is among your peers to carry out attacks such as this one?

B.5. *Sensation-seeking*

Zuckerman Kuhlman Personality Questionnaire (ZKPQ)—Impulsive Sensation-Seeking (ImpSS)

{0% – 100% me, in 10% increments}

1. I tend to start a new task or project, without much advance planning on how I will do it
2. I usually think about what I am going to do before I do it
3. I tend to do things on impulse
4. I very seldom spend much time on the details of planning ahead
5. I like to have new and exciting experiences and sensations even if they might be a little scary to me

6. Before I begin a complicated job or project, I tend to make careful plans
7. I would like to take off on a trip with no preplanned or definite routes or timetable
8. I enjoy getting into new situations where I can't predict how things will turn out
9. I like to do certain things just for the thrill of it
10. **If you are reading this question, please select 100**
11. I tend to change interests frequently
12. I sometimes like to do things that are a little frightening
13. I will try anything once
14. I would like the kind of life where I am on the move and traveling a lot, with lots of change and excitement
15. I sometimes do crazy things just for fun
16. I like to explore a strange city or section of town by myself, even if it means getting lost
17. I prefer friends who are excitingly unpredictable
18. I often get so carried away by new and exciting things and ideas that I never stop to consider possible complications
19. I am generally an impulsive person
20. I tend to enjoy "wild" uninhibited parties

B.6. *Ten-Item Personality Inventory*

O = *Openness to experience*; C = *Conscientiousness*; E = *Extraversion*; A = *Agreeableness*; *Emotional Stability* = ES; (Rev) *indicates a reverse-scored item*.

Here are a number of personality traits that may or may not apply to you. Please indicate the extent to which you agree or disagree with each statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

{0% – 100% me, in 10% increments}

I see myself as:

1. Extraverted, enthusiastic. (E)
2. Critical, quarrelsome. (Rev; A)
3. Dependable, self-disciplined. (C)
4. Anxious, easily upset. (Rev; ES)
5. Open to new experiences, complex. (O)
6. Reserved, quiet. (Rev; E)
7. Sympathetic, warm. (A)
8. Disorganized, careless. (Rev; C)
9. Calm, emotionally stable. (ES)
10. Conventional, uncreative. (Rev; O)

B.7. *Demographic Items*

What ethnicity do you identify with?

- Black/African-American
- Native American
- Asian/Asian-American
- White/Caucasian
- Hispanic/Latino
- Middle Eastern
- Other ____________________

Please indicate your parents' or guardians' combined household income.

Under $10,000	*$10,000- $19,999*	*$20,000- $29,999*	*$30,000- $39,999*	*$40,000- $49,999*	*$50,000- $74,999*	*$75,000- $99,999*	*$100,000- $150,000*	*Over $150,000*

Is English your first language?

Yes *No*

[*If No*]

How long have you spoken English?

0-1 years *1-2 years* *2-3 years* *3-4 years* *4-6 years* *6-8 years* *8+ years*

What is your stance on Net Neutrality, from 0 (strongly opposed) to 100 (strongly for)?

{0% – 100%, in 10% increments}

Redirect for Raffle Entry

What is your name? ____________________
What is your e-mail? ____________________
What is your phone number? ____________________
If your name is drawn for a Roku 3, how would you prefer that we contact you?
Phone *E-mail*

What is your primary campus?

Tempe	*Polytechnic*	*West*	*Downtown*	*Online*	*Lake Havasu*

APPENDIX C

ADDITIONAL TABLES AND FIGURES

Figure 5. *Boxplot of response distributions: Risk*

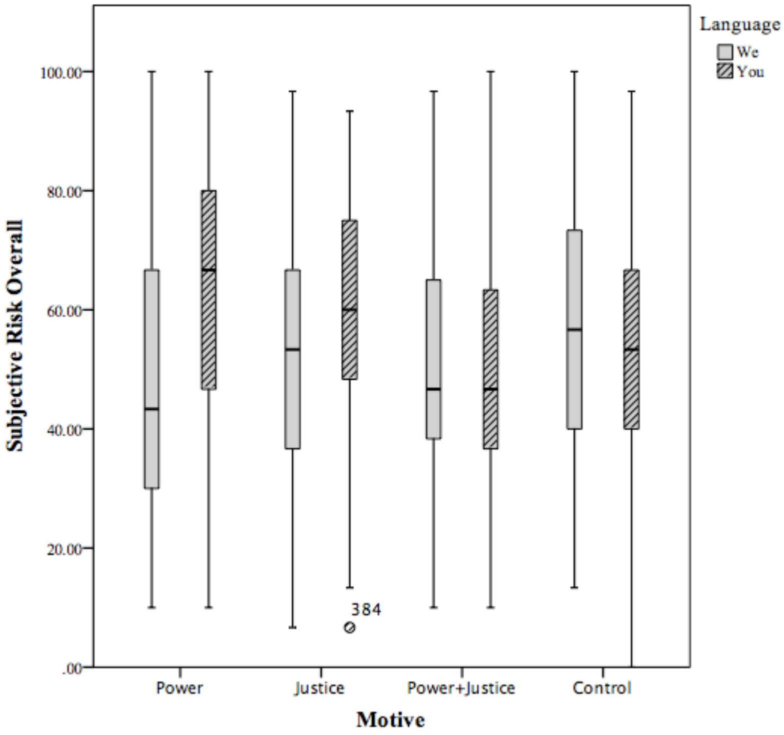


Figure 6. *Boxplot of response distributions: Payoff*

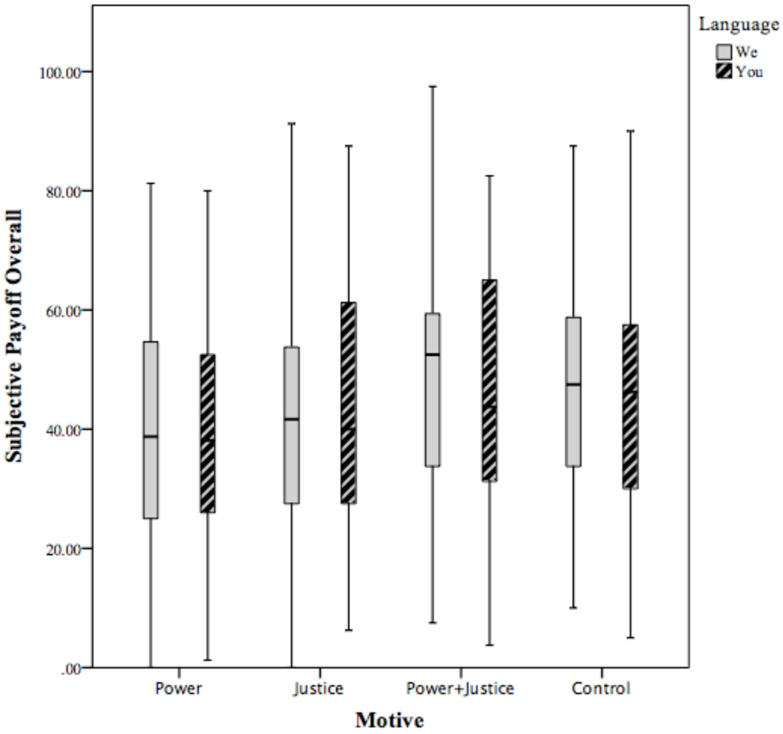
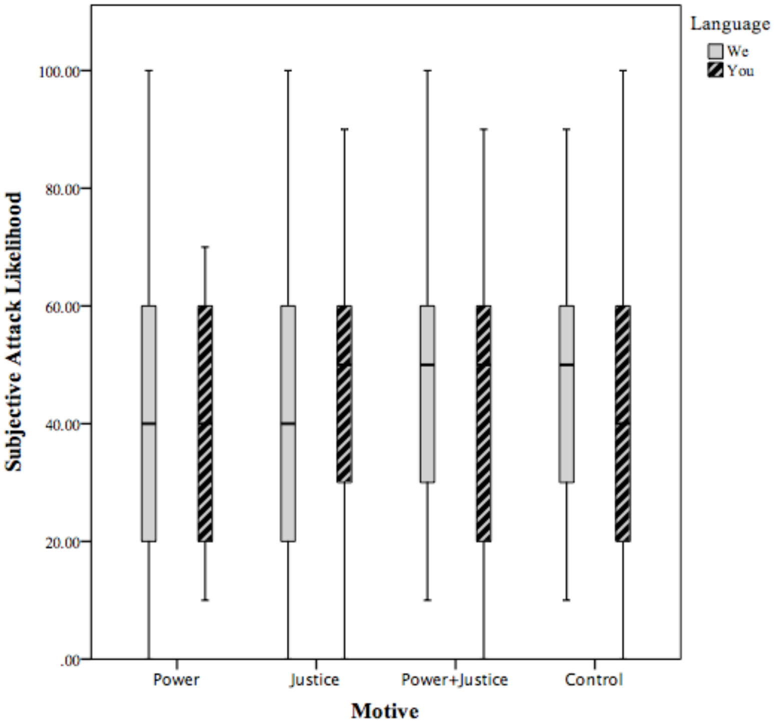


Figure 7. *Boxplot of response distributions: Attack likelihood*



Outliers are depicted as a circle beyond the scope of the “whiskers” (i.e., two standard deviations beyond the mean); extreme cases are depicted with an asterisk

Table 9. *Intercorrelation matrix: All variables of interest*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
2	.031	—	—	—	—	—	—	—	—	—	—	—	—	—
3	-.022	.551***	—	—	—	—	—	—	—	—	—	—	—	—
4	.020	.410***	.368***	—	—	—	—	—	—	—	—	—	—	—
5	-.032	.141**	.144**	.202***	—	—	—	—	—	—	—	—	—	—
6	.005	.017	.079†	.016	.408***	—	—	—	—	—	—	—	—	—
7	.026	-.068	-.037	-.018	-.397***	.050	—	—	—	—	—	—	—	—
8	.065	.053	.100*	.075	.332***	.289***	.001	—	—	—	—	—	—	—
9	.013	.023	.038	-.031	-.074	.156***	.125**	-.009	—	—	—	—	—	—
10	.032	-.009	.018	-.037	-.066	.190***	.031**	.238***	.202***	—	—	—	—	—
11	-.046	-.066	-.027	-.051	.023	.046	-.053	-.058	-.121**	-.062	—	—	—	—
12	.021	.414***	.404***	.126**	.076	.078†	-.011	.032	.023	-.036	.021	—	—	—
13	-.010	.453***	.391***	.148**	.123***	.188***	.018	.062	.078†	-.023	-.007	.586***	—	—
14	-.085†	.151***	.234***	.047	.060	.073	.022	.027	-.011	.050	.032	.263***	.216***	—
15	-.008	.340***	.379***	.225***	.034	.065	-.015	.114*	-.026	.020	-.003	.496***	.463***	.340***

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$
1: RiskOverall, 2: PayoffOverall, 3: AttackLikely, 4: PeerPrevalence, 5: ImpSS, 6: Openness, 7: Conscientiousness, 8: Extraversion, 9: Agreeableness, 10: EmotionalStability, 11: Stance on Net Neutrality, 12: MC Empowered, 13: MC Social Injustice, 14: MC Anonymity, 15: MC GroupIdentification

Table 10. *Means and significance: Risk*

Condition	Power	Justice	Joint	Control
We	46.512 (23.838)$_{a}$	52.449 (22.595)$_{a,c}$	50.000 (20.763)$_{a}$	56.256 (21.723)$_{a,c}$
You	61.897 (24.781)$_{b,c}$	59.104 (20.838)$_{b,c}$	50.151 (19.282)$_{a}$	52.867 (21.191)$_{a,c}$

Means with unmatched subscripts are significantly different, *p* < .10

Table 11. *Means and significance: Payoff*

Condition	Power	Justice	Joint	Control
We	40.415 (19.912)	42.640 (21.140)	47.155 (21.166)	46.043 (18.258)
You	40.259 (19.182)	44.522 (20.910)	45.735 (20.153)	44.192 (18.792)

No means were significantly different from one another, *p* < .10

Table 12. *Means and significance: Attack likelihood*

Condition	Power	Justice	Joint	Control
We	42.619 (26.232)	43.061 (24.934)	48.036 (20.839)	45.753 (19.644)
You	39.483 (18.583)	46.364 (23.247)	42.830 (23.808)	42.432 (23.514)

No means were significantly different from one another, *p* < .10