2016-12-03

Testing the Assumptions of the Provocation Doctrine

Carlton J. Patrick
University of Miami, carlton.patrick@gmail.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_theses

UNIVERSITY OF MIAMI

TESTING THE ASSUMPTIONS OF THE PROVOCATION DOCTRINE

By

Carlton J. Patrick

A THESIS

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Master of Science

Coral Gables, Florida

December 2016

UNIVERSITY OF MIAMI

A thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science

TESTING THE ASSUMPTIONS OF THE PROVOCATION DOCTRINE

Carlton J. Patrick

Approved:

_____
Debra Charles (Lieberman), Ph.D.
Professor of Psychology

_____
Michael McCullough, Ph.D.
Professor of Psychology

_____
Tamara Rice Lave, J.D., Ph.D.
Professor of Law

_____
Guillermo Prado, Ph.D.
Dean of the Graduate School

PATRICK, CARLTON J.   (M.S., Psychology)
                                                       (December 2016)

Testing the Assumptions of the Provocation Doctrine

Abstract of a thesis at the University of Miami.

Thesis supervised by Professor Debra Lieberman.
No. of pages in text. (37)

The provocation doctrine, which mitigates the punishment for killings made "in the heat of passion," is a longstanding and often criticized area of United States law. At the crux of these criticisms is the doctrine's imprecision: courts and lawmakers continue to struggle to both delineate specific standards for its application and to offer a satisfying rationale for its continued existence. Here, we use an evolutionary-computational model of anger and intimate partner violence to inform these problems. Using the archetypal case of provocation—a male discovering his partner's infidelity—we surveyed 1,939 males to address and test two of the doctrine's core assumptions: that a partner's infidelity is an especially potent catalyst for male intimate partner violence, and that anger motivates people to act in ways that they do not perceive as justified while in a calmer state. Then, we test several predictions about the specific cognitive mechanisms underlying the actor's behavior in the archetypal heat of passion case.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

In most jurisdictions in the United States, homicides committed in the "heat of passion" are treated as manslaughter and are punished less severely than premeditated or intentional killings. In some jurisdictions, the punishment for battery can also be mitigated (Volokh, 2014). The doctrine of heat of passion—sometimes referred to as the doctrine of provocation—is premised on the assumption that one who acts in a highly emotional state is less culpable than one who acts when calm and collected.

The provocation doctrine is one of the oldest legal doctrines of the United States and can be traced as far back as twelfth-century English legal decisions (Horder, 1992). It is also one of the most frequently criticized areas of law. This state of affairs is probably not a coincidence: because of the ancient roots of the doctrine, it remains reasoned on a folk-intuitional conception of human psychology that predates Charles Darwin, William James, the cognitive revolution, neuroscience, and virtually all modern theories of behavior. As a result, heat of passion jurisprudence is marked, and hampered, by imprecision. Jurors, for example, must apply vague standards that ask them to evaluate whether a provoked killer acts "in the heat of blood" or from a "transport of passion," and courts, when attempting to explain the phenomenon, often rely on nebulous descriptions such as "a dethronement of the reasoning faculty" and "a concession to the frailty of human nature" (Rozelle, 2005; State v. Lee, 1982).

The goal of this current study to reduce some of this imprecision and bring the heat of passion framework into the twenty-first century by reframing the phenomenon in terms of the modern psychological sciences. Relying heavily on a computational model of cognition and on insights from evolutionary psychology, I aim to both inform the

underlying assumptions on which the heat of passion doctrine is based, as well as add some descriptive clarity to the behavioral model that the law deploys.

A Brief History of the Provocation Doctrine

The provocation doctrine was adopted by the United States from English common law. Though the roots of the policy can be traced to the 1100's, a recognizable form of the rule emerges in England in the 1600's, requiring the killing to have occurred while the killer was in the "heat of blood" and that the heat of blood must have been brought about by provocation that was "sufficiently grave" (Horder, 1992). By the late eighteen-hundreds, we see an annunciation of the doctrine by the United States Supreme Court in Andersen v. United States (1898) that closely resembles the framework in place today:

> The law, in recognition of the frailty of human nature, regards a homicide committed under the influence of sudden passion, or in hot blood produced by adequate cause, and before a reasonable time has elapsed for the blood to cool, as an offen[s]e of a less heinous character than murder.

The traditional common law standard is typically structured so that a defendant must show that (a) a reasonable person in the defendant's situation would have been adequately provoked, (b) the defendant was in-fact provoked and became emotionally charged to an extent that he lost self-control, (c) a reasonable person in the defendant's situation would not have had sufficient time to "cool off" between provocation and killing, and (d) the defendant did not, in-fact, cool off before killing his victim (Fontaine, 2009).

Though this standard differs between jurisdictions and has undergone subtle degrees of periodic modification, the most significant change came when the American Law Institute promulgated the Model Penal Code (MPC) in 1962 and broadened the common law definition of heat of passion in two significant ways. First, it expanded the umbrella of killings covered by the defense to include homicides "committed under the influence of extreme mental or emotional disturbance for which there is a reasonable explanation or excuse" and, second, it removed the mandatory "cooling off" period imposed by the traditional standard (MPC § 210.3(1)(b), 1980; Robinson & Dubber, 2007; Clavel, 2011).

The MPC was influential: thirty-four states have either directly adopted the "extreme mental or emotional disturbance" language or revised their statutes to varying degrees to comport with the expanded view of the MPC (Robinson, 2011). Today, the vast majority of jurisdictions in the United States—whether they track the language of the traditional common law standard or of the MPC—recognize some form of heat of passion homicide. And in all jurisdictions that recognize the heat of passion "defense," the standard remains remarkably similar to the standard first incorporated by the Supreme Court over a century ago.

Criticisms of the Doctrine

Normative criticisms. The normative discourse surrounding the provocation doctrine is typically concerned with whether heat of passion should serve as a mitigating condition in the first place and, if it should, under what rationale. Generally, these debates revolve around the accuracy and applicability of the behavioral assumptions on which the doctrine is based. For example, the heat of passion defense is fundamentally premised on

the assumption that an extreme state of anger causes you to act in a way that you otherwise would not. Specifically, it assumes that anger can cause a person to act in a way that is more culpable, or morally compromised, than they would in a calm and rationale state. Though this assumption is built into the heat of passion doctrine, the rationale for it varies widely depending on which jurisdiction's precedent you examine, or which scholar you ask. The traditional common law analysis, for instance, typically justifies heat of passion on retributive grounds, reasoning that the defendant's heightened emotional state renders him less morally responsible for his actions, and—as a result— less *deserving* of punishment. The defendant acts not "from a bad or corrupt heart, but rather from the infirmity of passion to which even good men are subject" (Paz v. State, 2000).

Conversely, several prominent provocation scholars suggest a more utilitarian justification, implying that heat of passion homicides are less-deterrable than intentional killings. Fontaine (2010) notes that the heat of passion defense reduces murder to manslaughter "via the rational that the killer acted, due to emotional disturbance, out of a largely impaired capacity to act otherwise" (p. 84). Dressler (1982) similarly remarks that being in the heat of passion "affects choice-capabilities, not mere opportunities" and that it "makes us less able to respond in a legally and morally appropriate fashion" (p.464).

The commentaries to the MPC combine both rationales in recognizing that the provocation doctrine is both a concession to "human weakness" and to "non-deterrability" and noting that provocation killings "may be as much attributable to the extraordinary nature of the situation as to the moral depravity of the actor" (MPC § 210

Commentaries, 1980). In sum, the heat of passion defense is—as Dressler's (1982) famous article phrased it—a defense in search of a rationale.

A second behavioral assumption underlying the doctrine's normative rationale is that there is something psychologically distinct about infidelity that makes it a particularly potent catalyst for intense anger and violence. Historically, the paradigmatic heat of passion case was known as *in flagrante delicto*: an (almost always) male defendant, upon discovering his romantic partner in the act of adultery, kills his partner, her paramour, or both (Coker, 1992). One of the first articulated examples of infidelity-based mitigation comes from the seventeenth century English decision in *Manning's Case* (1671), where the defendant was sentenced to a burning of the hand, but the executioner was directed to "burn him gently, because there could be no greater provocation than this."

As the standards for heat of passion have evolved and shifted, various other circumstances have come and gone as to what constitutes adequate provocation—false arrest, violent assault, and mutual combat, for example—but adultery has steadfastly remained at the core of the doctrine. Today most scholars agree that even though the language of the modern standard is broad enough to encapsulate an array of emotional disturbances, the quintessential case remains that of a jealous husband killing his spouse after witnessing or learning of her infidelity (Clavel, 2011; Miller, 2010; Rozelle, 2005; Broussard, 2012; Coker 1992). Scholars have repeatedly pointed out the consequences of this: that heat of passion can be used as an "abuse excuse" and that it is disproportionately beneficial to males while leaving other categories of defendants—such

as females who kill their husbands after years of physical abuse, but not during 'the heat of passion'—without a similar safety net (Miller, 2001; Ramsey, 2010).

Descriptive criticisms. Like the normative criticisms of the heat of passion doctrine, the descriptive reservations likewise stem from a lack of institutional rigor. Despite centuries of jurisprudence in England, and a minimum of 150 years in the United States, courts and legal scholars have made little progress in clarifying the psychological mechanisms underlying violence committed in a highly emotional state. As a result, jurors are given ambiguous standards and very little direction in applying them. Jurors, for instance, are offered almost no guidance as to what constitutes an *extreme emotional disturbance* or what it means to act *in the heat of passion* and so must rely on their own common-sense intuitions in evaluating the defendant's mental state (e.g. People v. Beltran, 2013). This descriptive gap not only results in non-uniform application of the heat of passion defense, but also leaves lawmakers and policymakers at a disadvantage for predicting and preventing future episodes. Without a more instructive cognitive model, those responsible for preventing (as opposed to simply punishing) acts committed in a highly emotional state remain mired in vague and uninstructive folk psychology.

An Evolutionary Account of Emotions

To help reframe the issue with a more instructive model, our analysis draws heavily from an evolutionary perspective of emotions. As the language surrounding the provocation doctrine indicates, part of the persistent ambiguity of the doctrine may be due, at least in part, to the origins of the violent impulses. As social and behavioral scientists have recognized for some time, a great amount of the brain's processing— including much of the activity that initiates and regulates emotions—takes place outside

of conscious awareness (Bargh & Morsella, 2008; Domasio, 1999; Ohman et al., 2000). In fact, considerable research by cognitive scientists suggests that the experience or *feeling* of an emotion is just the conscious back-end output of a more complex subconscious computational process (see Winkielman & Berridge, 2004, for a review). Because humans do not have ready introspective access to their subconscious processes and are thus unable to articulate them, generating concrete, testable theories of the computational parameters of emotions can sometimes prove difficult (Nisbett & Wilson, 1977).

Evolutionary theory helps assuage this problem by producing a priori assumptions about the function and structure of the various subconscious processing systems of the brain. By reconstructing the conditions of our hunter-gatherer ancestors, researchers can reverse engineer subconscious processes (such as emotions) and—by identifying the historical problems that the brain evolved to solve—formulate testable hypotheses about the information processing architecture that was naturally-selected to accomplish these tasks.

Evolutionary psychologists often conceptualize emotions as modes of operation, or cognitive "software programs" that evolved to configure and arrange an individual's mind and body to effectively solve certain ancestrally recurrent situations (Nesse, 1990; Pinker, 1997; Cosmides & Tooby, 2000; Tooby & Cosmides, 2008). In this view, emotions function by recognizing ancestrally recurrent situations and orienting the individual towards effectively navigating them by up-regulating and tweaking certain cognitive, behavioral, and physiological systems, and down regulating and altering others into the configuration best suited to handle the task at hand.

Investigating an emotion from an evolutionary perspective thus entails both ultimate and proximate levels of analysis. At the *ultimate* level researchers attempt to understand what ancestral situation, or class of situations, a given emotion evolved to navigate. The ultimate explanation of behavior then informs a *proximate* analysis of the neurocomputational architecture used to perform this task. That is, a proximate analysis of emotions involves investigating (i) the specific information—or cues—that the brain relies on to identify the ancestrally-recurrent situation, (ii) the computational infrastructure and algorithmic rules that process the incoming information and initiate and regulate the emotional response, and (iii) the cascade of cognitive and physiological adjustments orchestrated by the emotion to solve the problem (Tooby & Cosmides, 2008; Lieberman & Patrick, 2014). Though a heat of passion defense can be premised on many different emotions, the most common emotion—and the emotion most salient to the archetypal case—is anger (Dressler, 1982). Thus, to understand the psychological and behavioral mechanics of being "the heat of passion" I depend on an ultimate and proximate analysis of intense anger.

The Cognitive Architecture of Anger

The highly-social environment of our hunter gatherer ancestors was rife with opportunity for individuals to exploit one another. Competition over scarce, zero-sum resources—be they tangible resources such as food – or intangible resources such as reputational value, positioning within a hierarchy, or the favor of a potential mate—meant that, ceteris paribus, individuals had an incentive to obtain the greatest possible benefit for themselves, even (and sometimes especially) if it meant exploiting others. To navigate this incentive structure, selection should have organized cognitive and

behavioral systems to deter exploitation and negotiate for the best possible resolution to potential conflicts. This, in crux, is what an evolutionary model of anger posits. In this view, anger is a neurocognitive program designed to orchestrate aggression and punishment (via the imposition of costs or withholding of benefits) or the *threat* of punishment, in order to increase the amount of value placed the angry individual's welfare so as to resolve conflict in favor of the angry individual and prevent exploitation. Usually termed the recalibrational theory of anger, it is, as McCullough and colleagues have succinctly phrased it, "the logic of deterrence" (McCullough, Kurzban & Tabak, 2013).

If the ultimate goal of anger is to increase the amount of value placed on the angry individual's welfare a proximate explanation of anger thus entails identifying the information-processing machinery in the brain that executes this task. One prominent theory suggests that the fulcrum on which this process depends is an internal regulatory variable termed the welfare tradeoff ratio (WTR) (Sell, 2009; Tooby et al., 2008). A WTR is a perpetually-updated cognitive parameter that represents the amount of value one individual places on the welfare of another person, relative to their own. A high WTR towards someone means that you value their welfare highly, relative to your own. A low WTR indicates the opposite. Likewise, if you expect a high WTR from another person, you expect them to value your welfare highly relative to their own, and vice versa.

What elements might go into the brain's calculation of how high to value another's welfare, or how high to expect them to value yours? Generally, at least in the context of potential conflict, WTR calculations are computed using factors that indicate

the relative "bargaining positions" of the individuals in question, including the relative benefits that individuals can bestow on one another (now and in the long term), as well as their abilities to impose costs in a potential conflict (Sell et al., 2009; Petersen et al., 2012). These factors can include, among other things: relative fighting ability, the number and strength of the allies someone has, their social skills and competences, their access to resources, their degree of relatedness, and their standing within a hierarchy (Petersen et al., 2010). To illustrate an example, if an individual has a higher relative status than another person within a hierarchy, (and presumably along with it the ability to recruit the influence of others, the ability to inflict costs and the ability to bestow benefits as a result of the higher status), that individual should, holding other factors constant, lower his or her WTR towards the lower-status person and raise the minimally-acceptable WTR value that he or she expects.

Within this view, the particular WTR weightings of a given relationship (toward another and expected from another), wherever they lay, will determine the proximate parameters of the anger response. The expected WTR value is thought to serve as the algorithmic threshold: when the actor detects that another individual has exploited the actor by expressing a lower WTR than the actor expects (by "expects," I mean unconsciously calculated as the minimally acceptable value), the anger program is initiated. Thus, measuring anger from a computational perspective requires two variables: the expected WTR from the other person, and the degree of deviation from that expected WTR value. The intensity of the anger response is predicted to be dictated by the size of the deviation: the larger the exploitation (the further the demonstrated WTR from what the actor expects), the more anger.

The activation of anger then automatically and unconsciously triggers a cascade of cognitive and physiological adjustments that orients the individual into the "mode of operation" selected to carry out its ultimate function of recalibration: aggression may be motivated; goal choice oriented towards punishment or threat of punishment; the actor may now perceive ambiguous stimuli as hostile; what was once a prohibitively-costly gamble of physical conflict may now become a viable option; the actors face, voice and posture might assume a threatening expression; and the actors physiology altered such that their heart rate changes, blood is pumped to the limbs, and breaths become shorter as their body readies to yell, fight or otherwise resolve the situation. Once the costs have been imposed, recalibration has been achieved, or the information indicating the need for an anger response otherwise removed, the program is de-initiated and the individual returned to a more neutral resting state.

Importantly, while this model of anger implies a very strong relationship between anger and behavior, I do not propose that they should be construed as having a perfect one-to-one relationship. A number of factors might influence the link between the two. An individual may, for instance, experience an intense state of anger but might be unable to punish because the exploiting individual is stronger, or has many allies, or because the exploited individual must rely on the exploiting individual for future benefits. Moreover, if the ultimate goal of anger is recalibration, it—in addition to being achieved through the infliction of costs or withholding of benefits—might also be achieved by just *threatening* to inflict costs or withhold benefits, or even by simply displaying anger in order to induce guilt in the exploiting party.

Anger in Romantic Relationships.

Though an inferior demonstrated WTR value may be a universal mechanism for anger, the precise criteria that go into weighting WTR values or that are used for identifying exploitation will depend specifically on the context of any given relationship. In the context of romantic relationships, WTR's might be determined by a range of informational cues indicating the value each individual can offer to the relationship, including criteria such as physical attractiveness or intelligence (indicating good genes for potential offspring) as well as material capital or social influence (indicating potential resources to be contributed to the relationship and offspring). Furthermore, romantic relationships present their own unique subset of ways in which partners can exploit and anger one another, along with a corresponding subset of counter-exploitative strategies that might be employed as part of the anger response. For example, partners can exploit one another via deception, misappropriation of pooled resources, disproportionate parental or relationship investment, and infidelity. By the same token, partners can respond with counter-exploitative strategies such as yelling, making demeaning comments to the partner or about the partner to others, emotional or psychological abuse, inducement of guilt, withdrawal of resources, threats of—and actual—desertion, physical harm and murder (Buss, 1992; Frieze, 2005).

From an evolutionary point of view, sexual infidelity—the act at issue in the paradigmatic heat of passion case—presents a particularly costly form of exploitation to a male partner. Because females are fertilized internally, in the ancestral environment (a setting without birth control or paternity tests) men could never be truly sure that their partner's child was their own and so faced the potential risk of investing resources in (i.e.,

raising and caring for) a rival's offspring. Though women also incur costs as a result of a partner's sexual infidelity—especially to the extent that the infidelity results in the male's reallocation of personal, material and parental resources to the other partner and offspring—they do not face the same peril of paternity uncertainty that males do. As a result of this asymmetric distribution of risk, men evolved a greater proclivity for sexual jealousy, mate-guarding, and intimate partner violence than women, presumably because male sexual proprietariness limits female access to potential reproductive rivals, thereby increasing paternity probability (Buss & Duntley, 2011; Wilson & Daly, 1996).

Research has demonstrated that detection or suspicion of infidelity is a key predictor of intimate partner violence (Daly et al., 1982, Daly and Wilson, 1988). In one particularly salient example, researchers found that women who were battered by a spouse while pregnant were more likely to be carrying the child of man other than her current partner (Martin et al., 2004; Taillieu & Brownridge, 2010). Nonetheless, even if sexual infidelity represents a potentially severe exploitation and a potent catalyst for intense anger, extreme violence and homicide are, from an evolutionary point of view, apparently paradoxical counter-exploitative strategies. The reason for this is straightforward: if the object of the anger response in this context is to recalibrate the behavior of the romantic partner so that a successful romantic relationship can be maintained, then extreme or fatal costs defeat the purpose inasmuch as they leave little or no room for reconciliation. Likewise, if the infidelity renders the relationship unsalvageable, there seems to be little incentive for the wronged partner to not adopt the ostensibly less-costly alternative of just abandoning the relationship.

Though the vast majority of cases of spousal infidelity do follow one of these paths, there are still significant numbers that result in cases of extreme violence or homicide, and that appear to follow the predictions generated by an evolutionary model. For example, suspected or discovered infidelity is the modal cause of uxoricides committed in the United States (Shackelford et al., 2003) and jealousy is the leading cause of spousal homicide worldwide (Buss, 2000). Also, as predicted by the asymmetric risks of parental confidence, this phenomenon is decidedly unbalanced between the sexes: across a range of cultures and political systems, intimate partner homicides account for nearly half of all female deaths, while just 6% of male deaths each year (Fox & Zawitz, 2007). In addition, because intimate partner violence is tied to securing a reproductive monopoly, we also observe a relationship between intimate partner violence and fertility—the probability of a woman being stalked, battered by her spouse, and murdered by a jealous husband are all highest as the woman approaches peak lifetime fertility (Buss, 2000; Shackelford et al., 2003). Thus, even if cases of extreme physical violence and homicide represent a relatively small subset of the total reactions to infidelity, they nevertheless represent a large subset of the total female homicides and, moreover, seem to reliably follow patterns hypothesized by an evolutionary model.

Chapter 2: The Current Study

In many respects, the legal framework for heat of passion overlaps with the predictions generated by an evolutionary model. Both, for example, posit sexual infidelity as a uniquely powerful source of anger for males, and both conceive of anger as a potential source of behavioral inclinations that exceed moral judgments made in a calm and collected state. However, where the heat of passion doctrine remains reasoned on an early, makeshift psychological framework, an evolutionary model offers a more coherent underlying model of behavior that enables more rigorous investigation. Thus, my aim with the current study is twofold: first, to add a deeper layer of understanding to the timeworn behavioral assumptions on which the normative conventions of the traditional doctrine depend, and second, to use the computational model outlined above to generate and test novel predictions about the cognitive processes that underlie such behavior. To accomplish this, I recruited a sample of heterosexual male participants who had been wronged in a previous or current relationship and explored the following research questions:

Question 1: Is there something specific to a female partner's infidelity (as opposed to other forms of harm) that makes it a particularly potent catalyst for male aggression? To assess the assumption that infidelity is a special case, I examined whether there were differences between those participants who had been harmed via infidelity, and those who had been harmed in a different manner by their partner, in both anger and the degree to which they wanted to retaliate. Because, across evolutionary time, infidelity has posed such a singular threat to male biological fitness, I expected to find that those

15

who had been cheated on would be angrier and more willing to engage in retaliatory measures, even when controlling for the degree to which they were hurt by their partner's actions.

Question 2: Does anger motivate people to act in ways that they may not perceive as being justified in a calmer state? To assess whether anger motivates people to behave outside of what they feel is justified, I examined whether anger could be used to predict the difference between participants' desire (upon discovering their partner's infidelity) to engage in certain behaviors (e.g. physically harm) and their belief in how justified they would have been in engaging in those behaviors. Because behavioral changes might be necessary to accomplish anger's ultimate goal of recalibration, and adjustments to moral determinations not necessarily crucial to achieve these ends, I predicted that higher levels of anger would lead to larger differences between how participants acted, and what they reported as being justified.

Question 3: Does the amount of value that a male expects his partner to place on his own welfare depend on the mate-value of his partner? To assess this, I examined whether evolutionarily-salient predictors indicating mate value correlate with the participants' expected WTR from their partner. Because mate value is one component of the "bargaining positions" hypothesized to inform WTR calculations, I expected to find that the predictors would significantly correlate with expected WTR.

Question 4: In the context of a romantic relationship, is the anger response triggered by an individual placing less value on their partner's welfare than the partner expects? To assess this, I examined whether the size of the deviation in revealed versus expected WTR would correlate with the amount of anger reported, as well as the various

behavioral responses. Based on the hypothesis that WTR is a universal fulcrum for anger, both inside and outside of romantic relationships, I expected to find that the deviation in WTR would positively correlate with all items.

Participants

Participants were adult heterosexual males residing in the United States recruited via Amazon's Mechanical Turk to complete an online survey administered though Qualitrics. Participants received $2.50 as compensation for completing the survey, which was advertised as a "relationship survey" in which participants would be asked "a series of questions about a past relationship and about how you felt in that relationship." Twenty-seven participants were excluded after admitting to not being heterosexual males (after completing the survey and being assured of compensation), and one an additional participant was excluded based on obviously suspicious data indicating unreliable responses, yielding a total of 1,939 participants.

Procedure and Materials

After providing informed consent, participants were asked whether they had "ever been in a relationship in which the woman you were involved with cheated on you?" and split into two separate survey paths based on the results. Participants who answered "Yes" (n = 1,209, 62%) then completed one version of the survey in which subsequent questions asked them about their partner's infidelity. Participants who answered "No" (n = 730, 38%) completed another version of the survey in which subsequent questions asked them about a harmful incident in which their partner engaged in a specific behavior that harmed the individual (e.g. "lying to you, stealing from you, yelling at you, insulting you or otherwise making you feel bad about yourself"). The two conditions were

identical except that questions in the second condition replaced the words "the infidelity" with "the harmful incident." In both conditions, participants were instructed to focus on the one relationship and single incident and subsequent questions were randomized within sections. The sections were completed by participants in the same order, but the questions were randomized within each section.

Pre-incident measures. In this section participants completed items that related to the pre-incident conditions of the relationship, each on a Likert-type scale. This included items such as relative welfare trade-off ratio measures (*I did not value her well-being at all* to *I valued her well-being so much that I did not consider my own at all*), relative physical attractiveness (*she was much less attractive* to *she was much more attractive*), relative mate value (*she was much more of a catch* to *I was much more of a catch*), and relative socio-economic status (*I came from a much lower SES than her* to *I came from a much higher SES than her*).

Incident-related measures. In this section participants completed items that related to circumstances surrounding the incident. Each participant answered two open-ended questions which required 250-character responses. The first question asked participants to: *Please take 1-3 minutes and describe in a short paragraph the circumstances around [the harmful incent/the infidelity], what your partner did, and how you discovered it*. The second question asked participants to: *Please take a moment to recollect how the harmful incident made you feel. Take 2-3 minutes and describe what you felt like in the moment you discovered your partner had taken the harmful action*.

Participants also completed a series of recall measures, each on a Likert type scale, that included items such as: how surprised the participant was by the incident (*not*

surprised at all to *extremely surprised*), how many people knew about the incident (*no one else knew* to *very many people knew*), how much the participants reputation was harmed by the incident (*not at all* to *completely ruined*), how angry the incident made the participant (*not angry at all* to *extremely angry*), after the incident, how likely the participant thought it was that the partner would do it again (*not likely at all* to *extremely likely*), how hurt the incident made the participant (*not hurt at all* to *extremely hurt*), and post-incident relative welfare measures (*I did not value her well-being at all* to *I valued her well-being so much that I did not consider my own at all*).

Behavioral and moral measures. Participants completed two sections designed Ft. Lauderdale, FL - FLLto measure both behavioral and moral responses to the incident. The behavioral section presented participants with certain behaviors and asked them to recall and rate how much they wanted to engage in such behaviors on a Likert-type scale ranging from *1-not at all* to *5-extremely badly*. The behaviors consisted of: (1) *yell at my partner,* (2) *push my partner,* (3) *gossip about my partner to ruin her reputation*, (4) *physically hurt my partner*, (5) *hit or slap my partner*, and (6) *walk away from the relationship and never see her again*. The moral section presented participants with the same behaviors and asked them to rate how justified they would have been to have engaged in such behaviors on a Likert-type scale ranging from *1-not justified at all* to *5-completely justified*.

Analysis and Results

Infidelity as a Special Case.

To examine the differences between the type of hurtful action (infidelity vs. non-infidelity) in the amount of anger elicited and the desire to engage in the 6 mate-directed

behaviors (push, hit/slap, physically hurt, yell, gossip about, and walk away), I conducted

7 independent samples *t*-tests. In addition, because the results between the infidelity and

other-harm conditions could be due to average differences in the gravity of the harm—i.e.

infidelity being on average a more hurtful action that the other forms of harm, regardless

of type—I also conducted 7 independent ANCOVA analyses in which the degree to

which the participant reported being hurt by the harmful action was included as the

covariate. This was used to control for the *degree* of harm caused by the action to

ascertain whether there was a significant difference based on the type (i.e. infidelity or

non-infidelity) of harm across the 6 behaviors. Table 1 reports both the unadjusted results

from the *t*-tests (not assuming equal variances where appropriate) as well as the adjusted

mean values from the ANCOVA analyses. Because the two standardized regression

slopes of the two groups differed by less than .4, no corrections or adjustments were

applied to the ANCOVA data (Wu, 1984). Table 1 reports significance for the *t*-tests; all

ANCOVA analyses resulted in functionally similar *p* values for the *F*-tests (all significant

at a *p* < .001 level).

Table 1. Unadjusted mean responses from t-tests with adjusted mean responses from ANCOVA in parentheses, standard deviations, *p* values, and cohen's *d* values for the different behavioral responses.

Behavior	Infidelity	SD	Non-Infidelity	SD	Difference	*p*	*d*
Yell	3.82(3.72)	1.29	3.21(3.38)	1.34	0.61	<.001	0.47
Push	1.86(1.83)	1.27	1.44(1.49)	0.93	0.42	<.001	0.36
Physically hurt	1.7(1.67)	1.17	1.35(1.39)	0.86	0.35	<.001	0.33
Hit or slap	1.83(1.80)	1.27	1.41(1.49)	0.90	0.42	<.001	0.37
Gossip about	2.45(2.40)	1.45	1.9(1.97)	1.21	0.55	<.001	0.40
Walk away	4.05(3.99†)	1.20	3.22(3.32†)	1.42	0.83	<.001	0.65

Note: Each question asked, at the time of discovering the partner's harmful action, how much the participant wanted to engage in each behavior and was presented on a scale that stretched between 1 ("not at all") and 5 ("extremely badly").

† Includes a significant interaction between the groups and the covariate degree of harm in the ANCOVA analysis $F(1,1935) = 18.42$, $p < .001$, $\eta p^2 = .009$.

As expected, there was a significant difference in anger between the infidelity (M = 4.11, SD = 1.02) and non-infidelity groups (M = 3.60, SD = 1.02), $t(1531.61) = -10.712$, $p < .001$, cohen's $d = .5$ (with equal variances not assumed). All six questions were significantly different in the predicted direction: those who had been cheated on reported an increased desire to engage in the retaliatory behaviors. The ANCOVA means suggest that these effects may also be present when controlling for the degree of initial harm. To further test this, I conducted a follow-up MANCOVA analysis using the more conservative and robust Pillai's Trace test (Tabachnick & Fidell, 2001; Scheiner, 2001) on the six results across the two groups (infidelity, non-infidelity) with the level of harm included as the covariate. Results supported the ANCOVA findings by demonstrating a significant multivariate effect of infidelity on the different behaviors after controlling for the effect of level of harm, Pillai's Trace = .073, F (6, 1931) = 25.19, p < .001, ηp^2 = .073.

**Motivational Properties of Anger**

To explore the triangular relationship between motivation, justification, and anger (the relationship at issue in heat of passion cases), I next plotted the means of motivation and moral acceptability for each of the behaviors in question, each at the different levels of reported anger (n at level 1 = 27, level 2 = 76, level 3 = 172, level 4 = 391, level 5 = 543) for those individuals who had reported discovering their partner's infidelity (N = 1,209). *Figure 1* (next page) illustrates the results. The data plots seem to imply different patterns for the physical violence measures (push, physically hurt, hit or slap) and the non-physical violence measures (yell, gossip, walk away). Most saliently, the motivation and moral acceptability measurements for three violence measures suggest the pattern

predicted by the heat of passion doctrine: at low levels of anger, motivation and judgments of moral acceptability are the same, but they begin to separate at high levels of anger, with motivation exceeding judgments of moral acceptability.

To test whether the patterns suggested by Figure 1 were meaningful, a series of Spearman correlations (Siegel & Castellan, 1988) were conducted among anger, motivation, and judgments of moral acceptability. First, I obtained zero-order correlations between anger and the six motivational items, and between anger and the six items of moral judgment. The results are illustrated in *Figure 2* and *Figure 3* (following page); anger was significantly correlated with all 12 items. To determine whether, as assumed by the heat of passion doctrine, anger corresponds with motivation in excess of judgments of moral acceptability, a corresponding set of partial correlations were obtained. First, between anger and judgments of moral acceptability while controlling for motivation, and second between anger and motivation while controlling for judgments of moral acceptability. As predicted and shown in *Figure 3*, when controlling for judgments of moral acceptability, there are still significant correlations between anger and the six motivational items. However, as shown in *Figure 2* when controlling for motivation, the significant positive correlations between anger and judgments for the three physical violence measures disappear, but there are still significant correlations between anger and yell, gossip, and walk away.

As a follow-up analysis, I tested whether these effects were also present in the non-infidelity group. A second series of Spearman correlations (Siegel & Castellan, 1988) was obtained among the three measures for all six behaviors. Anger (M = 3.60, SD = 1.02) was significantly positively correlated (zero-order) with all six motivational

Figure 1. Mean levels of reported motivation and moral acceptability.
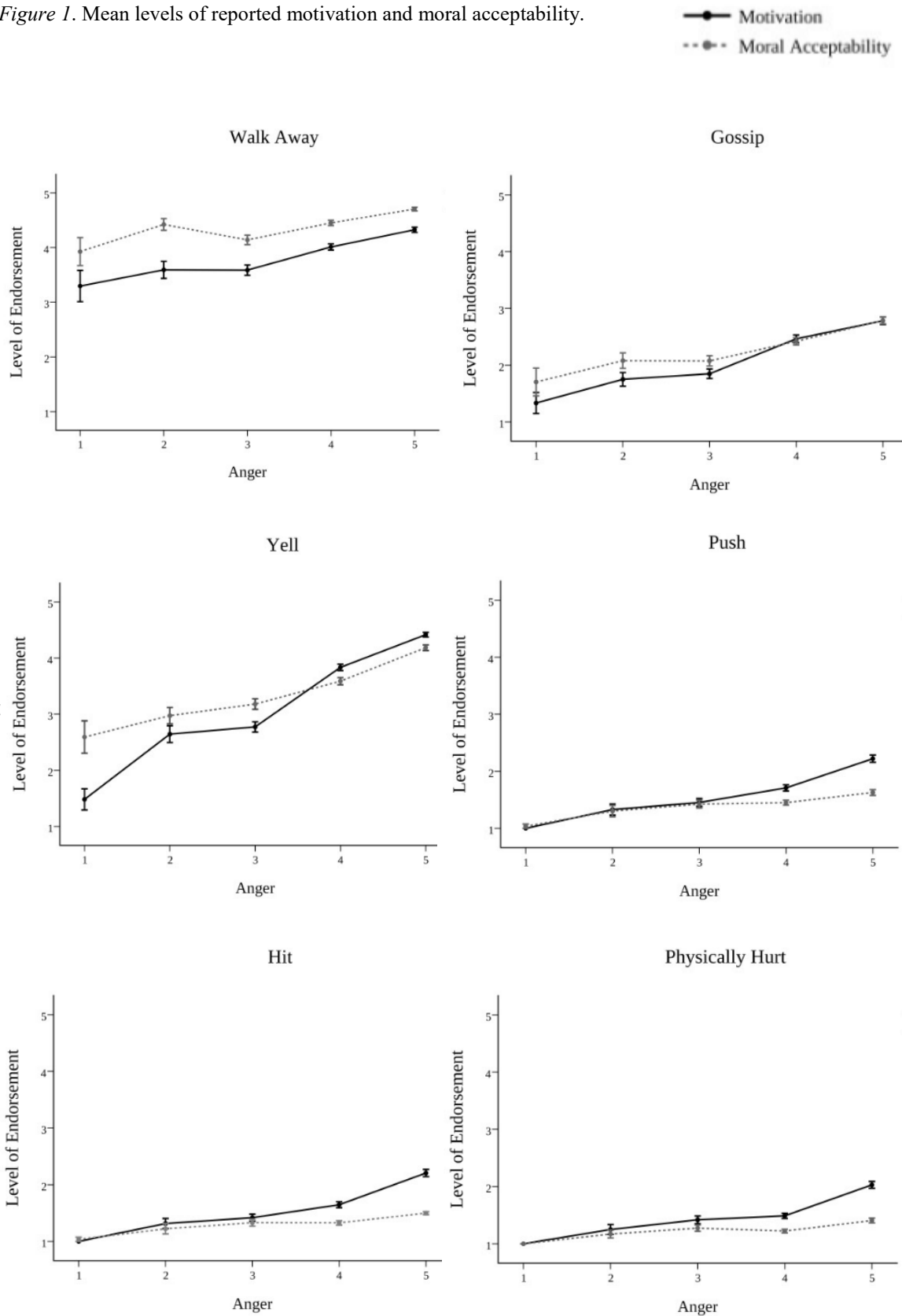
items: yell $r_s(730) = .49$, $p < .001$, gossip $r_s(730) = .20$, $p < .001$, walk away $r_s(730) =$ .35, $p < .001$, push $r_s(730) = .18$, $p < .001$, physically hurt $r_s(730) = .21$, $p < .001$, hit or slap $r_s(730) = .21$, $p < .001$, but *not* with all six moral acceptability items. There were significant correlations between anger and yell $r_s(730) = .40$, $p < .001$, gossip $r_s(730) = .18$, $p < .001$, walk away $r_s(730) = .35$, $p < .001$, push $r_s(730) = .09$, $p = .020$, and physically hurt $r_s(730) = .07$, $p = .045$, but not between anger and hit or slap $r_s(730) = .07$, $p = .063$.

Next, to test whether anger predicted behavioral motivations beyond what could be accounted for by judgments of moral acceptability, we obtained partial correlations between anger and each of the motivation and moral acceptability measures while controlling for the other. When controlling for moral acceptability, anger remained significantly positively correlated with motivation across all six behaviors: yell $r_s(730) = .32$, $p < .001$, gossip $r_s(730) = .12$, $p = .001$, walk away $r_s(730) = .16$, $p < .001$, push $r_s(730) = .17$, $p < .001$, physically hurt $r_s(730) = .18$, $p < .001$, hit or slap $r_s(730) = .21$, $p < .001$. Conversely, while controlling for motivation, anger remained significantly positively correlated with judgments of moral acceptability for gossip $r_s(730) = .11$, $p = .005$, walk away $r_s(730) = .17$, $p < .001$, yell $r_s(730) = .12$, $p = .001$, but not push $r_s(730) = -.02$, $p = .544$, physically hurt $r_s(730) = -.04$, $p = .347$, or hit or slap $r_s(730) = -.06$, $p = .086$. Thus, just as in the infidelity group, when controlling for judgments of moral acceptability, there are still significant correlations between anger and the six motivational items. However, when controlling for motivation, the significant positive correlations between anger and judgments for the three physical violence measures disappear, but there are still significant correlations between anger and yell, gossip, and
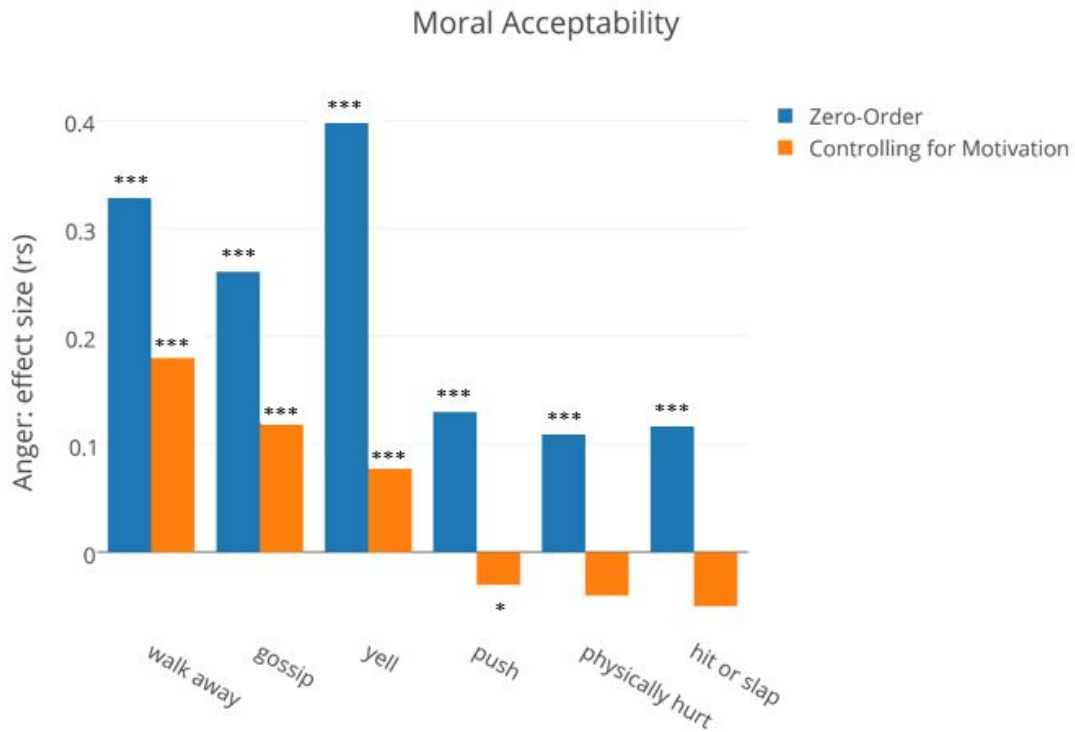
walk away.

Moral Acceptability

Figure 2. Spearman's correlations between anger and moral acceptability. *Note*: * *p* < .05, *** *p* < .001.

Motivation

Figure 3. Spearman's correlations between anger and motivation. *Note*: all correlations significant at *p* < .001.

The Computation of Provocation

To test the evolutionary-computational model outlined above, I conducted a series of Spearman correlations corresponding to the three hypothesis listed in the prior section. First, as displayed in *Figure 4*, I found very small but significant correlations between the participants' expected WTR from their partner and their partner's relative physical attractiveness, socio-economic status (SES), and overall value in the dating market (measured by asking who would be the "better catch"). The more physically attractive the partner, the higher her SES, and the higher her relative value in the dating market, the *lower* the expected WTR reported by the participant. Conversely, the higher the partner's WTR relative to the participant, the greater the expected WTR.
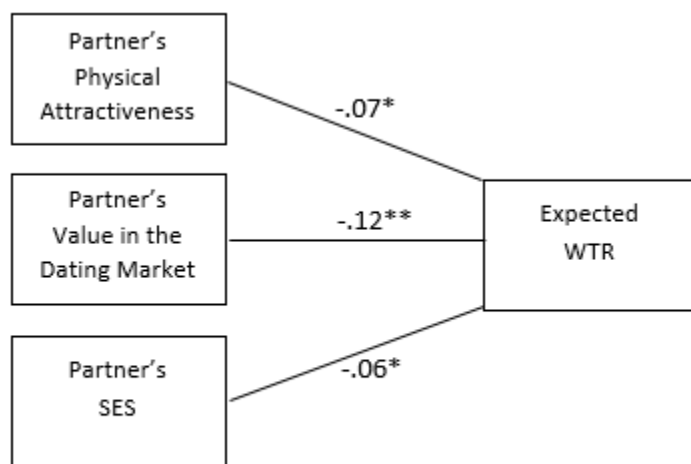

Figure 4. Spearman's correlations between predictors and expected WTR from partner. *Note: *$p \leq .01$, **$p \leq .001$.

We then examined the relationship between the decrement in expected WTR and anger. I used two methods to calculate the decrement in expected WTR. The first was to subtract post-incident welfare tradeoff measures from pre-incident measures. However, because I was wary of participants' abilities to accurately differentiate pre and post-incident welfare tradeoff measures (that is, able to recall how much they believed their

partner valued the participant's well-being prior to the incident *and* subsequent to the incident) without the post-incident measure biasing the pre-incident, I also estimated WTR decrements by asking the participants to recall how surprised they were when they discovered the incident. This is because surprise may be thought to index the change in WTR: the greater deviation from what was expected, the higher the surprise. Though both values were significantly correlated with anger, the WTR estimate using the participant's reported level of surprise $r_s(1939) = .37$, $p < .001$ showed a larger effect than the WTR estimate obtained through subtracting the pre and post incident measures $r_s(1939) = .10$, $p < .001$.

To then test the remaining relationships, I obtained Spearman correlations between the decrement in expected WTR and the six motivational items. There were small significant correlations between change in WTR and yell $r_s(1939) = .10$, $p < .001$, gossip $r_s(1939) = .09$, $p < .001$, and walk away $r_s(1939) = .09$, $p < .001$, with non-significant correlations with push $r_s(1939) = .03$, $p = .234$, physically hurt $r_s(1939) = .02$, $p = .328$, and hit or slap $r_s(1939) = .04$, $p = .090$. However, when using reported level of surprise as the measure for deviation in expected WTR, there were significant correlations for all six items: yell $r_s(1939) = .25$, $p < .001$, push $r_s(1939) = .08$, $p = .001$, physically hurt $r_s(1939) = .08$, $p = .001$, hit or slap $r_s(1939) = .07$, $p = .001$, gossip $r_s(1939) = .07$, $p = .001$, walk away $r_s(1939) = .16$, $p = .000$. Nevertheless, all correlations were small and significant likely because of the large number of participants.

CHAPTER 3: DISCUSSION

I began my analysis by asking whether and to what degree two of the core assumptions of the provocation doctrine aligned with the predictions generated by an evolutionary model. Consistent with our hypotheses, I found (a) that infidelity, when compared to other types of harm, had an especially potent effect on both anger and inclinations to retaliate and (b) that anger accounted for a significant amount of the variance of motivation (in a positive direction) for the physical violence measures in excess of that explained by levels of moral acceptability. In other words, there was a significant relationship between anger and the motivation to engage in physical violence above and beyond what was considered justified.

We then sought to augment these results by testing an evolutionary-computational model that answers what type of informational cues initiate and dictate the intensity of the anger response and its behavioral outputs. The hypotheses generated by our computational model were confirmed: I found significant relationships between the different mate value measurements and expected WTR, between change in WTR and Anger, and between change in WTR and the various behavioral responses. However, these results should be taken with the limitation that one attempted measure of change in WTR (surprise) showed significant results for all behaviors, while another (subtracting pre and post measures), only for three. Moreover, all such correlations were small.

This model is intended as proof of concept for the computational mechanics underlying conflict in romantic relationships. Of course a much larger set of informational cues is integrated into the cognitive calculations outlined here. Relative physical attractiveness, value on the dating market, and relative SES, for example,

28

represent only a small subset of the cues that might be used to compute expected WTR from a romantic partner—hence the small size of the otherwise significant correlations. Any variable indicating potential value to the actor could conceivably be included in such calculations. Likewise, there are likely many other components (besides our measures of change in WTR) that dictate the anger response. Our model is meant to be illustrative: it is an analytical tool that, like any other, requires sharpening. None the less, its power is demonstrated here in its ability to generate coherent, non-obvious predictions about the kinds of information and algorithmic relationships that ultimately comprise the anger response.

These insights are especially useful when contrasted against the rhetoric typically used in the heat of passion discourse. Rather than relying on folk-intuitional phrases such as *a dethronement of the reasoning faculty* or a *transport of passion*, the cognitive model outlined here permits a more rigorous examination of the mental processes of 'provoked' defendants. By using evolutionary theory to reverse-engineer the anger response, researchers and legal scholars can make empirical headway into descriptive quagmire of the doctrine that is so frequently criticized.

Two important caveats for legal analysis should be noted. First, the model here is meant to be illustrative, but it should not be over-construed as exhaustive *or* under-construed as merely speculative. Instead, our model should be thought of as a useful starting point for removing the aura of mysticism surrounding heat of passion and for constructing a more accurate probabilistic model of behavior. Our model, for example, does not posit that infidelity is necessary *or* sufficient to engender intense anger and physical violence. After all, the majority of men who discover that their spouse has been

unfaithful do not engage in extreme physical violence, nor are all cases of intense anger attributable to infidelity. Rather the conclusion to be drawn is that, ceteris paribus, infidelity is one *type* of evolutionarily-recurrent pressures that could the response.

The second caveat of the model is that it is confined to descriptive explanation. It can inform normative questions, as I will argue shortly, but the biologically-based explanations for behavior proffered here cannot in and of themselves mandate any normative conclusions. Explaining the natural foundations of behavior is not tantamount to justifying the behavior, and to contend otherwise would mean committing the "naturalistic fallacy" of assuming that natural equals good and thus concluding ought from is (Jones & Goldsmith, 2005). Instead, as I explain in greater detail below, legal scholars wishing to draw normative conclusions from these results should instead focus on the type of "if-then" inferences enabled by such explanations.

Normative Implications and Conclusion

The heat of passion jurisprudence, as outlined in this article's introduction, troublingly lacks a coherent normative framework and is instead supported by a patchwork assemblage of divergent rationales. Because of this, our findings here can only sharpen and clarify the questions at stake; answering them still requires selecting a specific normative objective or ideology. For example, our results here reinforce the assumption that infidelity can be an especially potent anger catalyst for males, given the asymmetric risks of parental confidence. But this does not in and of itself mandate a normative conclusion—a normative objective must still be chosen. If society wishes to privilege *fairness*, for instance, then our model serves to explain the decidedly unbalanced male tendency towards anger and violence and could be marshalled as

evidence to justify the existing discrepancy produced by sex differences in cognition and behavior. However, if society choses to adopt *equality* as their normative end, then lawmakers should reconsider the traditional privilege granted to cases of infidelity-based heat of passion defenses—despite the explanations for this difference produced by our model—in favor of a more gender-neutral rule.

Consider, as an additional example, the assumption that anger causes individuals to act in ways they do not feel are justified when in a calm and collected state. Our findings here fall in line with literature demonstrating that being in an emotional state affects how information is processed and translated into behavior during the duration of the emotion, including in ways that temporarily motivate morally-suspect behavior (Arielly & Lowenstein, 2006). Again, though, what these findings dictate depends crucially on the normative goals being pursued, including whether those goals are based on retributive or utilitarian framework. Retributivist justifications, for instance, based on the premise that the provoked individual is incapable of forming intent or acts "without reason" are contradicted by our findings. Though the actor may not be acting with consciously formed intent, the subconscious systems regulating the anger response are, in fact, reasoning quite accurately as is demonstrated by our consistently-observed criteria underlying the response. As the legal scholar Kyron Huigens (paraphrasing the English legal philosopher John Gardner) notes (Huigens, 2009):

> The insane actor acts *against* reason, but the provoked actor does not. Very much to the contrary: the provoked actor acts *for* reasons. The paradigmatic cuckolded husband acts as he does for the reason that he has been

cuckolded. He might not act with cool deliberation, but he does act for a

reason.

Thus, if a retributivist justification for the heat of passion defense is to be maintained, it

should be based on arguments emanating from accurate functional distinctions, such as

the distinction between consciously and subconsciously-formed intent, not on timeworn

intuitions of emotions as ethereal, non-quantifiable processes.

Similarly, utilitarian justifications based on the premise that a heightened state of

anger renders an individual "less-deterable" should likewise be revisited. Though the

computational procedures governing anger are taking place subconsciously, that does not

mean that the threat of punishment is not being incorporated into the emotional and

behavioral response. Indeed, further research is needed to know if, how, and to what

extent the threat of punishment is incorporated into the computational processing of

emotion and anger in this context. After all, for every angry man who kills his wife after

catching her cheating, there are orders of magnitude of men who might be angry enough

to kill their wife, but do not. This sentiment is vividly demonstrated by participants'

answers to our prompt to *describe what you felt like in the moment you discovered your*

partner's infidelity, which included, for example: "I felt like killing someone. I swear I

could've lifted a car…" "I was ready to kill both of them…" "I still do not know how I

managed to not kill her…" "It was probably the one time in my life that I could have

committed double homicide and not had any remorse…" "I wanted her and him dead. I

just wanted both of them removed from life…" "I felt like shooting her and her lover,

thank God, she had hid the gun…" "I was so angry that had they both or either one of

them been around, I may well have ended up in prison for murder…" "I wanted her dead. Extreme anger and rage. I wanted to strangle the life out of her. She was a heartless bitch…" and "I felt like beating her to death. I wanted her to die, and I wanted the last thought to go through her head to be that she brought this on herself. If I could have done it without getting caught, I believe I would have killed her and not felt any remorse."

The computational template I have outlined here should prove useful for both legal and empirical scholars examining intimate partner violence and the heat of passion defense. Our preliminary findings offer some concrete descriptive explanations, but more importantly suggest proof of concept for a powerful framework for future investigation. By reverse engineering our cognitive processes as solutions to the recurrent problems of our hunter-gatherer ancestors, investigators are provided with a blueprint for generating testable hypotheses about the nuerocomputational architecture designed to achieve these solutions. As I have demonstrated here in the context of the heat of passion defense, this process allows scholars to not only sharpen descriptive models of behavior (thus enhancing our ability to predict and influence conduct) but also to add a deeper layer of understanding to that behavior and inform the normative analysis surrounding it.

References

Andersen v. United States, 170 U.S. 481, 510 (1898).

Bargh, J. A., & Morsella, E. (2008). The unconscious mind. *Perspectives on Psychological Science*, *3(1)*, 73-79.

Broussard, D. B. (2012). Principles for passion killing: An evolutionary solution to manslaughter mitigation. *Emory LJ, 62*, 179.

Buss, D. M. (1992). Manipulation in close relationships: The five factor model of personality in interactional context. *Journal of Personality, 60*, 477–499.

Buss, D. M. (2000). *The dangerous passion*. New York: The Free Press

Buss, D. M., & Duntley, J. D. (2011). The evolution of intimate partner violence. *Aggression and Violent Behavior*, *16*(5), 411-419.

Clavel, N. Y. (2011). Righting the wrong and seeing red: Heat of passion, the Model Penal Code, and domestic violence. *New Eng. L. Rev., 46*, 329.

Coker, D. K. (1992). Heat of passion and wife killing: Men who batter/men who kill. *S. Cal. Rev. L. & Women's Stud., 2*, 71.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. *Handbook of emotions, 2*, 91-115.

Daly, M., Wilson, M. I., & Weghorst, S. J. (1982). Male sexual jealousy. *Ethology & Sociobiology*, 3, 11–27.

Damasio, A.R. (1999). The feeling of what happens: Body and emotion in the making of consciousness. New York: Harcourt Brace.

Dressler, J. (1982). Rethinking heat of passion: A defense in search of a rationale. *Journal of Criminal Law and Criminology*, 421-470.

Fontaine, R. G. (2009). Adequate (non) provocation and heat of passion as excuse not justification. *University of Michigan Journal of Law Reform*, *47*, 207.

Fontaine, R. G. (2010). An attack on self-defense. *American Criminal Law Review*, Vol. 47, 57-89.

Fox, J. A., & Zawitz, M. W. (2000). *Homicide trends in the United States: 1998 update*. US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

Frieze, I. H. (2005). *Hurting the one you love: Violence in relationships*. Pacific Grove, CA: Thompson/Wadsworth.

Greenhouse, S. W., & Geisser, S. (1958). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95-112.

Horder, J. (1992). *Provocation and responsibility*: Clarendon Press Oxford.

Huigens, K. (2009). Symposium: The nature, structure, and function of heat of passion/provocation as a criminal defence. A critical introduction to the symposium. *U. Mich. JL Reform, 43*, 1.

Jones, O. D., & Goldsmith, T. H. (2005). Law and behavioral biology. *Columbia Law Review*, *105*, 405-502.

Lieberman, D., & Patrick, C. (2014). Are the behavioral immune system and pathogen disgust identical?. *Evolutionary Behavioral Sciences*, *8*(4), 244.

Manning's Case, 83 Eng. Rep. 112 (1671).

Martin, S. L., Harris-Brit, A., Li, Y., Moracco, K. E., Kupper, L. L., & Campbell, J. C. (2004). Changes in intimate partner violence during pregnancy. *Journal of Family Violence, 19*, 201–210.

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2013). Cognitive systems for revenge and forgiveness. *Behavioral and Brain Sciences, 1*(1), 1-15.

Miller, A. E. (2010). Inherent (gender) unreasonableness of the concept of reasonableness in the context of manslaughter committed in the heat of passion. *Wm. & Mary J. Women & L., 17*, 249.

Miller, E. L. (2001). (Wo)manslaughter: Voluntary manslaughter, gender, and the Model Penal Code. *Emory LJ*, *50*, 665.

Model Penal Code § 210.3(1)(b) (Official draft and revised comments 1980).

Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human nature*, *1*(3), 261-289.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological review*, *84*(3), 231.

Ohman, A., Flykt, A., & Lundqvist, D. (2000). Unconscious emotion: Evolutionary perspectives, psychophysiological data and neuropsychological mechanisms. In R.D.

Lane, L. Nadel, & G. Ahern (Eds.), *Cognitive neuroscience of emotion.* New York: Oxford University Press.

People v. Beltran, 301 P. 3d 1120 (Cal. 2013).

Petersen, M. B., Sell, A., Tooby, J., & Cosmides, L. (2010). Evolutionary psychology and criminal justice: A recalibrational theory of punishment and reconciliation. From *Human morality & Sociality: Evolutioanry & Comparitive Perspectives*. Henrik Hogh-Olesen (Ed.), Palgrave MacMillian, New York, pp. 72-131.

Petersen, M. B., Sell, A., Tooby, J., & Cosmides, L. (2012). To punish or repair? Evolutionary psychology and lay intuitions about modern criminal justice. *Evolution and Human Behavior, 33*(6), 682-695. doi: DOI 10.1016/j.evolhumbehav.2012.05.003

Pinker, S. (1997). *How the mind works*. NY: Norton.

Ramsey, C. B. (2010). Provoking change: Comparative insights on feminist homicide law reform. *The Journal of Criminal Law and Criminology*, 33-108.

Robinson, P. H. (2011). Abnormal mental state mitigations of murder: The U.S. perspective. In M. A. B. Reed, M. (Ed.), *Loss of Control and Diminished Responsibility: Domestic, Comparitive and International Perspective*: Ashgate Publishing, Ltd.

Robinson, P. H., & Dubber, M. D. (2007). The American model penal code: A brief overview. *In International and Interdisciplinary Journal*, *10*(3), 319-341.

Rozelle, S. D. (2005). Controlling Passion: Adultery and the Provocation Defense. *Rutgers LJ, 37*, 197.

Shackelford, T. K., Buss, D. M., & Weeks-Shackelford, V. (2003). Wife-killings committed in the context of a lovers triangle. *Journal of Basic and Applied Social Psychology*, 25, 137–143.

Scheiner, S. M. 2001. Multiple response variables and multispecies interaction. In SM Scheiner and J Gurevitch (eds.), *Design and analyses of ecological experiments*, 2nd ed. Oxford University Press, Oxford, England.

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences of the United States of America, 106*(35), 15073-15078. doi: DOI 10.1073/pnas.0904312106

Siegel, S., & Castellan Jr, N. J. (1988). *Nonparametrics statistics*. NY: McGraw-Hill.

State v. Lee, 321 Wis. 2d 1, 6 (1982) quoting *Johnson v. State,* 129 Wis. 146, 160 (1906).

Tabachnick, B. G., & Fidell, L. S., (2001). *Using multivariate statistics*. Pearson

Taillieu, T. L., & Brownridge, D. A. (2010). Violence against pregnant women: Prevalence, patterns, risk factors, theories, and directions for future research. *Aggression and Violent Behavior*, *15*, 14–35.

Tooby, J., & Cosmides, L. (2008). The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of Emotions, 3rd Ed.* (pp. 114-137.) NY: Guilford.

Tooby, J., Cosmides, L., Sell, A., Lieberman, D., & Sznycer, D. (2008). Internal Regulatory Variables and the Design of Human Motivation: A Computational and Evolutionary Approach. *Handbook of approach and avoidance motivation*, 251.

Volokh, E. (2014). The 'heat of passion' voluntary manslaughter theory. *The Washington Post*, 20 November. Available at: http://www.washingtonpost.com/news/volokh-conspiracy/wp/2014/11/20/the-heat-of-passion-voluntary-manslaughter-theory-responsibility-and-punishment/ (accessed 29 January 2015).

Wilson, M. I., & Daly, M. (1996). Male sexual proprietariness and violence against wives. *Current Directions in Psychological Science, 5*, 2–7.

Winkielman, P., & Berridge, K. C. (2004). Unconscious emotion. *Current Directions in Psychological Science*, *13*(3), 120-123.

Wu, Y. W. B. (1984). The effects of heterogeneous regression slopes on the robustness of two test statistics in the analysis of covariance. *Educational and Psychological Measurement*, *44*(3), 647-663.