Western University Scholarship@Western

Electronic Thesis and Dissertation Repository

8-22-2022 2:00 PM

A Demonstration of the Utility of Formalizing Theory: Mapping, Evaluating, and Testing the Theory of Devaluation of Alternatives

Rebecca Koessler, The University of Western Ontario

Supervisor: Lorne Campbell, *The University of Western Ontario* A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Psychology © Rebecca Koessler 2022

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Social Psychology Commons

Recommended Citation

Koessler, Rebecca, "A Demonstration of the Utility of Formalizing Theory: Mapping, Evaluating, and Testing the Theory of Devaluation of Alternatives" (2022). *Electronic Thesis and Dissertation Repository*. 8743.

https://ir.lib.uwo.ca/etd/8743

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Abstract

Devaluation of alternatives is often identified as a key relationship maintenance strategy that is used to combat the threat of attractive alternatives. However, since the theory was proposed by Johnson and Rusbult (1989), few efforts have been dedicated to evaluating the progression of the theory and assessing the quantity and quality of the existing evidence that claims to corroborate it. The present research fulfilled this need by formalizing the theory of devaluation of alternatives using a metatheoretical approach that involved categorization, deconstruction, reconstruction via theory mapping, and evaluation (Chapter 2). P-curve meta-analyses were used to determine whether *p*-hacking or selective reporting could be detected in a selection of the literature (Chapter 3). These processes revealed many of the theoretical assumptions were understudied and lacked sufficient corroboration, and evidence of selective reporting was not detected. Guided by these findings, the assumption that attractive alternatives are threatening was selected and tested using a novel experimental manipulation (Chapter 4). Following the completion of a compatibility questionnaire, participants were shown fabricated compatibility results with their current partners and a real alternative partner of their choosing. Compatibility with the alternative was manipulated to be high, average, or low, with higher compatibility representing stronger threat. Quantitative and qualitative results indicated that participants across conditions did not significantly differ in their experiences of negative affect or threat. While participants reported experiencing negative emotions, disappointment was experienced to a greater extent than anxiety or insecurity. These results indicate a failure to corroborate the primary assumption of devaluation of alternatives. In addition, across conditions participants expressed devaluation by criticizing the questionnaire/algorithm in a free response, while quantitative measures demonstrated devaluation in all conditions except the low-threat condition. These results indicate conflicting evidence for when devaluation is elicited. This research demonstrated that theories that are thought to be well-founded may not be. Social psychology is a theoryrich discipline; however, our empirical pursuits would be much improved by dedicating effort towards formalizing and evaluating the current state and verisimilitude of our existing theories to best determine how much has been accomplished and where to focus future research efforts.

ii

Keywords

Derogation of alternatives, devaluation, attractive alternatives, relationship threat, relationship maintenance, theory formalization, theory evaluation, *p*-curve analysis.

Summary for Lay Audience

Devaluation of alternatives is a theory that suggests people in romantic relationships are threatened by potential attractive alternative partners and are motivated to act to protect their relationships by identifying and emphasizing negative aspects of the alternative to make them seem less tempting. However, since the theory was proposed by Johnson and Rusbult (1989), the development of the theory and quality of the evidence that claims to support it has not been evaluated. I formalized the theory of devaluation of alternatives by describing the original version, defining key terms, identifying underlying assumptions, visualizing the theory, and evaluating the existing evidence (Chapter 2). In addition, because the published literature is biased towards including studies that have statistically significant results, p-curve meta-analyses were used to determine whether this bias was present in a selection of the literature (Chapter 3). These processes revealed the theoretical assumptions were understudied and lacked sufficient support. Guided by these findings, an important assumption of the theory - that attractive alternatives are threatening - was selected and tested using a new experimental manipulation (Chapter 4). Following the completion of a compatibility questionnaire, participants were shown fake compatibility results with their current partners and a real alternative partner of their choosing. Compatibility with the alternative was manipulated to be high, average, or low, with higher compatibility representing stronger threat. Participants across conditions did not differ in their experiences of negative emotion or feelings of threat. Participants reported experiencing negative emotions, but disappointment was reported more than anxiety or insecurity. These results do not support the primary assumption of devaluation of alternatives. In addition, across conditions participants comparably expressed devaluation by criticizing the questionnaire/algorithm in written responses, while quantitative measures demonstrated devaluation in all conditions except the low-threat condition, which indicates conflicting evidence for when devaluation occurs. This research demonstrated that seemingly wellfounded theories may not be. Social psychology is a theory-rich discipline; however, our research would be much improved by dedicating efforts towards formalizing and evaluating the current state of our theories to determine how much has been accomplished and where to focus future research efforts.

iv

Acknowledgments

I would like to express the utmost gratitude to my advisor, Dr. Lorne Campbell, for your support and guidance over the years. Thank you for giving me the time and space to pursue other interests and activities outside of the research domain. Your flexibility allowed me to develop new skills, discover more about myself, and ultimately experience more happiness and joy throughout this program. Though research is not my true passion, I'm thankful that the research I've done has been in collaboration with you.

I would also like to extend my thanks to, Dr. Erin Heerey, Dr. Alex Benson, and Dr. Paul Tremblay for providing feedback and suggestions throughout this project, and especially for your patience and understanding towards the end. Thank you all for your time and thoughtfulness in my departmental and senate exams.

Thank you to Emma Labovitz, my research assistant, who did impressive work getting the compatibility study up and running and assisting with the qualitative coding.

To my friends – words cannot express how grateful I am to feel so deeply cared for, loved, and supported by you. Dr. Nicolyn Charlot, your friendship has been indispensable throughout the past six years. I can always depend on you to validate to my complaints, come up with solutions to my problems, and call me out on my bullshit – as only true friends do. I have always admired your creativity, determination, and fearless originality. You inspire me to be better, and I am so thankful that we were able to go through this journey together. Marina Gross, I have learned so much from you. You continually remind me that my emotions are valid, that rest is important, and that accomplishments deserve to be celebrated. Thank you for writing with me the past couple months – while most of the time I wasn't looking forward to writing, I always looked forward to seeing you.

Finally, thank you to my family for supporting me throughout this entire process. Knowing that I made you proud is one of the best feelings in the world.

Honorable mention goes to my therapist and my cat – the two secret ingredients which, in my opinion, all but ensure the successful completion of a doctoral program.

Table of Contents

A]	BSTRACT		II
SI	UMMARY FO	R LAY AUDIENCE	IV
٨	CKNOWI FD	GMENTS	V
T.	ABLE OF CO	NTENTS	VI
L]	IST OF TABL	ES	IX
L	IST OF FIGUI	RES	X
L	IST OF APPE	NDICES	XII
C	HAPTER 1		1
1	INTRODU	CTION	1
	1.1 Rese	ARCH OVERVIEW	2
	1.1 KESE	ARCH OVERVIEW	2
C	HAPTER 2		4
2	FORMAL	IZING THE THEORY OF DEVALUATION OF ALTERNATIVES	4
	2.1 Bene	FITS AND CHALLENGES OF FORMALIZING THEORY	4
	2.2 Form	IALIZING DEVALUATION OF ALTERNATIVES	7
	2.2.1 Ca	itegorizing Devaluation of Alternatives	8
	2.2.2 De	econstructing Devaluation of Alternatives	10
	2.2.2.1	Core Constructs	
	2.2.2.2	Theoretical Assumptions	
	2.2.3 Re	constructing Devaluation of Alternatives with Theory Mapping	20
	2.3 Asse	SSING DEVALUATION OF ALTERNATIVES	23
	2.3.1 Pi	ogressive vs. Degenerative Research Programs	23
	2.3.1.1	Evolution of the Theory of Devaluation of Alternatives	
	2.3.2 Ve	risimilitude and the Lakatosian Defense	
	2.3.2.1	CA1: External Stressors Are Threatening (Or Perceived as Threatening)	
	2.3.2.2	PA1: More Intense Stressors Are More Threatening	
	2.3.2.3	PA2: External Stressors Threaten Commitment	
	2.3.2.4	CA2: When Threatened the Relationship Maintenance System Is Activated	
	2.3.2.5	CA3: When the Relationship Maintenance System Is Activated, Relationship Mainte	nance
	Mechani	sms May Occur	
	2.3.2.	-	
	2.3.2.		
	2.3.2.	5.3 Evaluation	
	2.3.2.6	CA4: Relationship Maintenance Mechanisms Are Enacted in Attempt to Neutralize I	External Threats
		51	

	2.3.2	.7 PA3: Commitment Must Be Sufficiently High for Relationship Maintenance Mechanism	is to Be
	Activ	vated 55	
	2.4 D	ISCUSSION	
CI	HAPTER 3	3	65
3	EVAL	UATING DEVALUATION OF ALTERNATIVES WITH P-CURVE META-ANA	LYSES.65
	3.1 In	TRODUCTION TO <i>P</i> -CURVE ANALYSIS	65
	3.2 N	IETHODS	67
	3.3 R	ESULTS	75
	3.3.1	P-Curve A	76
	3.3.2	P-Curve B	79
	3.4 D	ISCUSSION	
CF	TAPTER 4	l	85
4		IPIRICAL TEST OF CORE ASSUMPTION 1: ARE ATTRACTIVE ALTERNAT	
PA	RINERS	PERCEIVED AS THREATENING?	85
	4.1 In	VTRODUCTION	
	4.1.1	Review of Relevant Research	86
	4.1.2	Threat Perception and Response	
	4.1.3	The Current Study	
	4.2 N	Iethods	90
	4.2.1	Participants	91
	4.2.2	Procedure	
	4.2.3	Measures	
	4.2.3	.1 Demographics	96
	4.2.3	.2 Commitment	97
	4.2.3	.3 Positive and Negative Affect	97
	4.2.3	.4 Compatibility Questionnaire	97
	4.2.3		
	4.2.3	1 5	
	4.2.3		
		ESULTS	
	4.3.1	Descriptive Statistics	
	4.3.2	Efficacy of Random Assignment	102
	4.3.3	Confirmatory Hypotheses and Moderation Explorations	
	4.3.3	1	
	4.3.3		
	4.3.3		
	4.3.3		
	4.3.3	.5 Perceived Accuracy of Alternative Compatibility Result	

4.3.3.6	Conclusions of Confirmatory Analyses	
4.3.4 Qu	ualitative Analyses	
4.3.5 Ac	dditional Exploratory Analyses	
4.3.5.1	Differences in Measures Between Part I and Part II	
4.3.5.2	Estimated Partner and Alternative Compatibility in Part II	
4.3.6 Re	esults Summary	
4.4 DISC	USSION	
CHAPTER 5		
5 GENERAI	L DISCUSSION	
REFERENCES		
APPENDICES.		172

List of Tables

TABLE 1. METHODOLOGICAL CHARACTERISTICS OF STUDIES AND EXPERIMENTS INVOLVING DEVALUATION OF		
ALTERNATIVES		
TABLE 2. P-CURVE DISCLOSURE TABLE 71		
TABLE 3. DESCRIPTIVE STATISTICS FOR PARTICIPANT DEMOGRAPHIC AND RELATIONSHIP CHARACTERISTICS 91		
TABLE 4. DESCRIPTIVE STATISTICS OF PARTICIPANTS' SELECTED ALTERNATIVES		
TABLE 5. DESCRIPTIVE STATISTICS OF STUDY MEASURES		
TABLE 6. MODEL COMPARISONS OF MODERATED MULTIPLE REGRESSION ANALYSES WITH EVALUATION OF THE		
COMPATIBILITY QUESTIONNAIRE/ALGORITHM AS THE OUTCOME VARIABLE		
TABLE 7. MODEL COMPARISONS OF MODERATED MULTIPLE REGRESSION ANALYSES WITH PERCEIVED		
ACCURACY OF THE ALTERNATIVE COMPATIBILITY RESULT AS THE OUTCOME VARIABLE		
TABLE 8. CONCLUSIONS OF CONFIRMATORY ANALYSES 122		
TABLE 9. FREQUENCY COUNTS OF CODES REPORTED WITHIN EACH GROUP AND WITHIN THE FULL SAMPLE124		
TABLE 10. EXEMPLARS EXTRACTED FROM THE CONTENT ANALYSIS DIVIDED BY CONDITION 125		

FIGURE 1. THEORY MAP LEGEND
FIGURE 2. THEORY MAP OF JOHNSON AND RUSBULT'S (1989) ORIGINAL PROPOSAL OF DEVALUATION OF
ALTERNATIVES
FIGURE 3. THEORY MAP OF EXTERNAL THREAT INDUCED MAINTENANCE PROCESSES
FIGURE 4. P-CURVE A
FIGURE 5. CUMULATIVE META-ANALYSIS OF P-CURVE A
FIGURE 6. P-CURVE B
FIGURE 7. CUMULATIVE META-ANALYSIS OF P-CURVE B
FIGURE 8. CORRELATION MATRIX OF STUDY MEASURES
FIGURE 9. DIFFERENCES IN NEGATIVE AFFECT ACROSS THREAT CONDITION
FIGURE 10. DIFFERENCES IN FEELINGS OF THREAT ACROSS THREAT CONDITION
FIGURE 11. DIFFERENCES IN PERCEPTIONS OF THE QUESTIONNAIRE/ALGORITHM ACROSS THREAT CONDITION 109
FIGURE 12. PARTNER COMPATIBILITY (PI) AS A MODERATOR OF THE INFLUENCE OF THREAT CONDITION ON
EVALUATIONS OF THE COMPATIBILITY QUESTIONNIARE/ALGORITHM
FIGURE 13. ALTERNATIVE COMPATIBILITY (PI) AS A MODERATOR OF THE INFLUENCE OF THREAT CONDITION ON
EVALUATIONS OF THE COMPATIBILITY QUESTIONNIARE/ALGORITHM
FIGURE 14. DIFFERENCES IN PERCEIVED ACCURACY OF THE ALTERNATIVE COMPATIBILITY RESULT ACROSS
Threat Condition
FIGURE 15. COMMITMENT AS A MODERATOR OF THE INFLUENCE OF THREAT CONDITION ON PERCEPTIONS OF
ACCURACY OF THE ALTERNATIVE COMPATIBILITY RESULT117
FIGURE 16. ATTRACTION TO ALTERNATIVE AS A MODERATOR OF THE INFLUENCE OF THREAT CONDITION ON
PERCEPTIONS OF ACCURACY OF THE ALTERNATIVE COMPATIBILITY RESULT
FIGURE 17. WILLINGNESS TO DATE ALTERNATIVE AS A MODERATOR OF THE INFLUENCE OF THREAT CONDITION
ON PERCEPTIONS OF ACCURACY OF THE ALTERNATIVE COMPATIBILITY RESULT
FIGURE 18. PARTNER COMPATIBILITY (P1) AS A MODERATOR OF THE INFLUENCE OF THREAT CONDITION ON
PERCEPTIONS OF ACCURACY OF THE ALTERNATIVE COMPATIBILITY RESULT
FIGURE 19. ALTERNATIVE COMPATIBILITY (P1) AS A MODERATOR OF THE INFLUENCE OF THREAT CONDITION ON
PERCEPTIONS OF ACCURACY OF THE ALTERNATIVE COMPATIBILITY RESULT
FIGURE 20. DIFFERENCES IN ESTIMATES OF PARTNER COMPATIBILITY BETWEEN PART I AND PART II WITHIN
CONDITION
FIGURE 21. DIFFERENCES IN ESTIMATES OF ALTERNATIVE COMPATIBILITY BETWEEN PART I AND PART II
WITHIN CONDITION
FIGURE 22. DIFFERENCES IN COMMITMENT BETWEEN PART I AND PART II WITHIN CONDITION
FIGURE 23. DIFFERENCES IN EVALUATIONS OF THE QUESTIONNAIRE/ALGORITHM BETWEEN PART I AND PART II
WITHIN CONDITION
FIGURE 24. DIFFERENCES IN PARTNER COMPATIBILITY AT PART II ACROSS THREAT CONDITION
FIGURE 25. DIFFERENCES IN ESTIMATES OF PARTNER COMPATIBILITY ACROSS THREAT CONDITION

List of Figures

APPENDIX A. RESEARCH ETHICS APPROVAL

Chapter 1

1 Introduction

The search for a romantic partner often requires a substantial amount of effort. Relationshipseeking individuals must fearlessly traipse into environments that contain fellow singles, be willing to display and promote their attractive qualities, and make evaluative judgements about which individuals would be fitting partners. But even when a relationship has successfully been initiated the work does not stop. Partners in romantic relationships often must protect their relationship from a variety of potential threats, for example, partner transgressions (McNulty & Dugas, 2019), incompatible preferences (Rusbult et al., 2001), life stressors or changes (Randall & Messerschmitt-Coen, 2020), and attractive others (Johnson & Rusbult, 1989; Kelley, 1983; Rusbult, 1983). Consequently, coupled individuals must mitigate threatening circumstances to maintain their relationships.

Recognition that relationships could be destabilized by threatening circumstances led to the development of a theory to explain how coupled persons combat the specific threat of attractive alternative partners. In their seminal article, Johnson and Rusbult (1989) proposed the theory of devaluation of alternatives, which posits that individuals who are highly committed to their relationship will devalue potential alternative partners to protect their relationship with their current partner. Johnson and Rusbult identified the constructs of interest and proposed possible underlying mechanisms that could explain the phenomenon of devaluation of alternatives, providing a preliminary foundation on which to test the theory and for others to build upon. Their theory gained a great deal of traction and immensely influenced how relationship with devaluation often identified as a prominent and well-founded relationship maintenance strategy.

When new theories are proposed, uncertainties due to gaps in existing knowledge are acceptable and expected. Subsequent research efforts contribute to the refinement of theories through attempts at falsification, and such efforts increase our comprehension and (ideally) move us closer to the truth. However, findings derived from research informed by oversimplified theories

may be vague and ambiguous. Consequently, only limited and cautious conclusions are allowed, hindering the pace at which we chase the truth. The theory of devaluation of alternatives has followed this latter path, with the theory being referred to and relied upon in the research literature in its most simplified form. Specifically, much of the devaluation of alternatives literature is characterized by slight methodological variants and directional evidence in alignment with the theory, representing only weak corroboration of the theory (Meehl, 1990). Despite research in this area spanning 30 years, very few novel facts (Lakatos, 1970) about this phenomenon have been discovered and what we do know is only inconsistently and feebly supported by existing evidence.

1.1 Research Overview

The existing literature on devaluation of alternatives suffers from theoretical and methodological limitations which have effectively impeded the ability to increase our understanding of the phenomenon, including its causes and consequences. Due to overreliance on vague conceptualizations of the theory, past research is lacking in breadth and depth, and we are without an explicit account of what this theory proposes. As a result, we are ill prepared to determine where and how to focus future research efforts to refine and test the theory. Knowing romantic relationships can be destabilized by external forces, it is imperative the impediments in this research area are addressed so we may be better positioned to explain and predict how devaluation may mitigate the threat of attractive alternatives.

I argue that the theory of devaluation of alternatives must be formalized, evaluated, and subjected to rigorous tests to best assess its current state and utility. I argue further that these processes are informative, beneficial, and necessary to appropriately determine where to focus future research efforts that are intentionally designed to *expand* our understanding of devaluation of alternatives and increase its practical utility.

In this dissertation, I:

- a) formalized the theory of devaluation of alternatives (Chapter 2),
- b) assessed the extent to which the existing literature has advanced the theory since its conception and addressed its theoretical assumptions (Chapter 2),

- c) assessed the strength of evidentiary support for the phenomenon and determined whether indicators of selective reporting or publication bias were present through *p*-curve meta-analyses (Chapter 3),
- d) conducted an experiment that tests a specific core theoretical assumption identified through the process of theory formalization (Chapter 4), and
- e) asserted that the combination of theory formalization and strong methodology is essential to advance our understanding of devaluation of alternatives (Chapter 5).

Research on the theory of devaluation of alternatives is currently stagnant. Through this work, I aspire to revitalize and refocus research in this area and motivate fellow scholars to work towards transforming this degenerative research program into a more progressive one.

Chapter 2

2 Formalizing the Theory of Devaluation of Alternatives

Johnson and Rusbult's (1989) introduction of the theory of devaluation of alternatives was notable in that the core constructs of the theory were identified, two potential underlying causal mechanisms were proposed, and three initial studies were conducted using diverse methods that showed preliminary support for their ideas. However, subsequent scholars investigating this phenomenon have often relied on a simplified model of the theory when designing their empirical studies, and little effort has been directed at explicitly identifying the underlying assumptions that combine to construct the broader theory.

In this chapter, I discuss why social psychologists may find theory refinement challenging and justify why the process of theory formalization offers great benefit. Then, building on existing metatheory methodology (Quistberg & Sakaluk, 2021), I will formalize the theory of devaluation of alternatives through the processes of categorization (2.2.1), deconstruction (2.2.2), reconstruction (2.2.3), and evaluation (2.3). Categorization (Wallis, 2010) involves providing basic descriptive information about the theory. Deconstruction (Meehl, 2004, Smaldino, 2017) involves identifying, defining, and where appropriate, operationalizing the core constructs of the theory and identifying and describing the theoretical assumptions that offer explanations for how the core constructs relate to one another. Reconstruction (Gray, 2017) involves visualizing the theory, and evaluation (McGovern, 2018; Meehl, 1990) involves subjecting the formalized theory to critical scrutiny.

2.1 Benefits and Challenges of Formalizing Theory

Theories are broad structures of knowledge that facilitate the integration of past empirical observations and allow for predictions of future observations (Gawronski & Bodenhausen, 2015; Gray, 2017). Theories consist of statements that identify and define specific constructs of interest and describe how the constructs are related to each other

(Cronbach & Meehl, 1955; Meehl, 1990). Theories are valuable tools because they provide the opportunity to make inferences about specific outcomes under certain conditions without having to collect data under those conditions (Borsboom, 2013). Some have even argued that collecting data should only be employed as a "last resort" to determine which of competing sets of theoretical predictions are more accurate (Muthukrishna & Henrich, 2019). However, the utility of theory is determined in part by the extent to which it is formalized. The lack of formalized theory in social psychology is at least partially responsible for our inability to integrate collections of evidence, idling our efforts to increase our understanding of human behavior. Amplified attention and improvement in this domain may stimulate the development of more progressive research programs and lead to more meaningful advances in our knowledge, even if these advances are slow-moving or marginal in size.

Turning attention towards theory may be a challenging endeavor in part because the past decade in the field's history has been characterized by the "replicability crisis" or "credibility revolution" (Vazire, 2018). This movement spurred a widespread overhaul of how psychologists conduct research, including the introduction of open science practices such as preregistration, registered reports, power analyses, advanced statistical analyses, and large multi-lab replication studies. Limitations in methodological and statistical approaches were (and still are) the main targets of improvement, while negligible attention has been directed towards the importance and utility of theory in psychological research (Borsboom, 2013; Gray, 2017; Muthukrishna & Henrich, 2019; Wallis, 2010).

The discernible preference that psychological researchers have for collecting data over developing theory is not wholly a consequence of the credibility revolution (although arguably it is an influential culprit today), but rather a seemingly longstanding tendency. Sixty years ago, Forscher (1963) compared theorists and scientists to builders and brickmakers, whose purposes were to make and arrange bricks (facts) guided by blueprints (theories) in such a way that they created useful edifices (explanations for human behavior). Forscher suggested that a significant shift had occurred in the field, where much of researchers' time, resources, and efforts were focused on brickmaking while the goal of creating edifices fell by the wayside, resulting in the field of psychology

being "covered with loose bricks" (p. 339). It seems that not much has changed, as researchers have more recently observed that "we are facing a general neglect or agnostic state regarding theoretical issues" (Fiedler, 2004, p. 123), and some have even suggested we are suffering from theoretical amnesia, a "philosophical disorder...[where] researchers no longer know what a theory is, which means that they can neither recognize its presence nor its absence" (Borsboom, 2013, para.7).

A possible explanation for the preference of collecting data over developing theory is that theorizing in social psychology is extremely challenging. When it comes to theory, social psychology is incomparable to the physical sciences, where theories can be described using precise mathematical equations and hypotheses can be derived that predict numeric values down to the decimal point (Meehl, 1967). Rather than objective properties that can be directly observed or measured such as mass, distance, or temperature, social psychologists are "concerned with the study of actual, imagined, or anticipated person-toperson relationships in a social context" (Allport, 1954, as cited in Deutsch & Krauss, 1965, p.3). Confronted with the challenge of measuring unobservable constructs, attempting to control innumerable extraneous forces that may influence certain variables of interest, and accepting that, to at least some degree, all observations will be situated in the unique social and historical context in which they occur, developing social psychologists have resorted to rely on empiricism and statistics (Borsboom, 2013) to compensate for the "primitive state" (Meehl, 1990, p. 108) of psychological theory.

While social psychological theories may never be as precise or specific as those found in the physical sciences, there are actions we can take to improve how we theorize and how we use and evaluate theories in psychological research. The field of social psychology is not lacking psychological theory, but most theories in the field are vague and informal. These are known as verbal models (Smaldino, 2017). Verbal models are problematic

¹ Not to mention the issue of all theories in psychology being weak or defective by way of being incomplete (Meehl, 1990).

because their inherent ambiguity makes them easy to defend and protect, rather than inviting challenges and opportunities for refinement (Smaldino, 2017). Lakatos (1970) described the evolution of scientific progress as being brought about by "bold speculations...some of which are later eliminated by hard, conclusive refutations and then replaced by still bolder, new and, at least at the start, unrefuted speculations" (p. 173). While verbal models are appropriate when used to introduce and describe new theories, the underlying expectation is that as more empirical investigations are undertaken, as more data is collected and as our knowledge increases, theories will evolve to be more explicit, more well-defined, and less ambiguous (Deutsch & Krauss, 1965). However, with much of our focus on methodology and statistical analysis, we may be losing sight of the overall goal of scientific research – the development and refinement of theories that explain and predict human behavior. Without a broader structure connecting pieces of evidence, the result of our efforts will simply be a collection of facts. And as Poincare (1905) observed, "a collection of facts is no more a science than a heap of stones is a house" (p. 141).

2.2 Formalizing Devaluation of Alternatives

The verbal model of devaluation of alternatives has become the most prominent conceptualization of the theory upon which the existing literature is founded. Despite a clear need, thus far no comprehensive efforts have been taken to formalize the theory. Formalizing a theory involves identifying, defining, and/or operationalizing relevant constructs, describing the relationships between constructs (including potential mechanisms responsible for these relationships), and identifying the assumptions inherent in the underlying logic of the theory (Smaldino, 2017). Formal models attempt to organize complexity, and in so doing facilitate more structured and specific evaluations of the theory by increasing its falsifiability. Formalizing theories can inform and guide the development of studies that are specifically designed to challenge the theory, for example, by testing predictions that place an effect within a certain range rather than in a simple direction (Fiedler, 2004; Meehl, 1990). Subjecting the theory and its postulates to "risky" tests provides the opportunity for strong corroboration, as compared to weak corroborations where an observed effect is found to differ significantly from a null, or

"chance" value (Meehl, 1967, 1990).² Making a conscious effort towards formalizing our theories may substantially improve the way we design, evaluate, and interpret our research. Additionally, it encourages intention and reflection about how our present research builds on prior work. The following sections draw on existing literature to categorize, deconstruct, reconstruct, and evaluate the theory of devaluation of alternatives.

2.2.1 Categorizing Devaluation of Alternatives

Categorizing the theory of devaluation of alternatives involves identifying the central theorists, the year the theory was proposed, the description of the theory in the theorists' original words, and the application or purpose of the theory (Quistberg & Sakaluk, 2021). The basic descriptive information presented below was sourced entirely from Johnson and Rusbult (1989) – the foundation on which subsequent literature investigating and refining this theory rests.

The theory of devaluation of alternatives was originally proposed by Dennis J. Johnson and Caryl E. Rusbult in 1989, who sought to answer the questions: "When highly committed persons are confronted with attractive alternative partners, what are their options? By what processes do individuals manage to maintain commitment?" (p. 967). Johnson and Rusbult proposed that "greater commitment is associated with tendencies to devalue alternative partners, and that this process is most marked when the alternative poses the greatest threat to the current relationship: when the alternative is exceptionally attractive, and when the individual is faced with an actual opportunity to become

² Meehl (1967) demonstrated that the null hypothesis is "[quasi-]always false" (p. 112), and asserted that, assuming high levels of power and precision, the probability of finding corroborating evidence in support of a theory, inferred from the rejection of a directional null hypothesis significance test, approaches 50%. In other words, one has a ~50% chance of finding support for a theory by rejecting a directional null hypothesis, regardless of the quality or verisimilitude of the theory that informed the development of the hypothesis. Meehl also observed this issue was compounded by psychologists' tendency to commit the logical fallacy of affirming the consequent - interpreting the observation of the alternative hypothesis as substantiating evidence "confirming" the theory, as opposed to utilizing the logically valid deductive criteria of *modus tollens* – interpreting the failure to observe the alternative hypothesis as a refutation of the theory (i.e., falsification).

involved with that person" (p. 968). Devaluation of alternatives is a theory designed to explain *one* strategy³ individuals may employ to avoid temptation and maintain their commitment to their current partners and relationships.

In their formative article, Johnson and Rusbult (1989) proposed two potential explanations for devaluation of alternatives. The motivational explanation, informed by dissonance theory, suggests that when a committed individual feels attracted to someone who is not their partner, the dissonance between feelings of attraction for an alternative and feelings of loyalty and commitment to their partner creates a sense of discomfort. This discomfort stimulates the individual with the wandering eye to devalue personal qualities of the potential alternative or the hypothetical relationship the individual would have with them. For example, upon seeing an attractive alternative, a partnered person might think "That person is much too short for me," or "They don't have a great sense of humor, they probably wouldn't make me laugh as much as my current partner," in an effort to alleviate the discomfort they experienced due to feeling attracted to the alternative. According to this explanation, devaluation is elicited by the desire to reduce dissonance.

The perceptual explanation, informed by the concept of comparison level in interdependence theory (Thibaut & Kelley, 1959), suggests that individuals who are more

³ Johnson and Rusbult (1989) recognized that other strategies may be employed to mitigate the threat of attractive alternatives. Incidentally, devaluation of alternatives is more appropriately categorized as a subtheory of a broader theory describing the process of responding to external threats. Attractive alternatives are an exemplar of an external threat, and devaluation is one possible strategy, but not the only strategy, that may be used to neutralize such threats. Their research may be considered an extension of Kelley's (1983) identification of various processes of commitment, of which "reducing the availability and attractiveness of alternative relationships" (p. 305) was a part. Such processes "served to maintain promembership causal systems and to link them, over time, to the interaction" (p. 298). While Kelley's work may be considered the origin point of theoretical groundwork for this area of research, Johnson and Rusbult's paper may have been one of the first to formally test a proposed relationship maintenance strategy. Previous work may have recognized the phenomenon of relationship or commitment maintenance but had not yet embarked on formalizing it as a theory or exploring it empirically (Kanter [1968] could be considered an exception, but her sociological research involved the population of utopian communities in the US between 1780 and 1860 rather than those in romantic relationships). This research has since expanded, resulting in the creation and recognition of a broader theory of relationship maintenance, of which devaluation, among other strategies is a part. This will be addressed thoroughly in the reconstruction section (2.2.3).

highly committed to and satisfied with their relationships would have a higher comparison level. One's comparison level is indicative of the relationship outcomes a person believes they deserve and can realistically attain and is determined by both past relationships and the outcomes achieved in one's current relationship. An alternative partner would need to meet or exceed one's comparison level to be considered a viable alternative to the current partner and to put the individual "at risk" of leaving their current relationship. Logically, individuals are unlikely to leave their current relationship unless the alternative is of higher quality, or the relationship is likely to be more satisfying. Those with high comparison levels would devalue potential alternative partners by considering them to be less attractive and appealing than they "objectively" are. Individuals with higher comparison levels, therefore, are less likely to be drawn away from their relationship by an alternative partner. According to this explanation, devaluation is elicited by high levels of satisfaction and commitment in one's current relationship.

Before any complex analytic or evaluative endeavors are begun, the theory must first be described. The categorization process is beneficial because it necessitates tracing a certain theory back to its origins. Knowledge of the early form of a theory aids the understanding of its evolution and allows for subsequent revisions, refutations, deviations, and improvements to be more easily identified.

2.2.2 Deconstructing Devaluation of Alternatives

The first step in deconstructing the theory of devaluation of alternatives involves identifying and defining the core constructs of the theory and describing how the core constructs relate to each other (Wallis, 2014).

2.2.2.1 Core Constructs

Expanding the descriptive information elicited from the process of categorizing the theory of devaluation of alternatives, the following core constructs were identified and described using information from Johnson and Rusbult (1989):

Devaluation (or derogation) – Johnson and Rusbult (1989) cited Thibaut and Kelley's (1959) definition of devaluation: the act of "diminishing the value of the unattainable [person]...by taking a 'sour grapes' attitude toward the rewarding aspects of the interaction or by emphasizing the negative, cost-increasing aspects of it" (p. 175). Devaluation is the behavior that, when effective, is hypothesized to neutralize the threat posed by attractive alternatives.

Alternative – Johnson and Rusbult (1989) considered "dating another person or other persons or being without a romantic involvement" (p. 969) as potential alternatives to one's current romantic partner. Alternatives are considered a source of threat to relationship stability.

Attractive – "Attractive" is often included as a descriptor of "alternatives." The measures Johnson and Rusbult (1989) used in their study were adapted from Rusbult's (1980; 1983) previous work. In 1980, Rusbult conceptualized attractiveness as intelligence, personality, physical attributes, and wit (p. 177). In 1983, Rusbult included a variety of items intended to measure the rewards and costs of participants' current partners, including: personality, attitudinal similarity, intelligence, physical appearance, physical proximity, similarity of values, need complementarity, sense of humor, shared interests, similarity of habits or pastimes, quality of the sexual relationship, embarrassing habits, unattractive personal qualities, unattractive attitudes about relationships, and reliability. In addition, the participant's loss of personal freedom, the relationship's monetary cost, and the relationship's time costs were taken into consideration (p. 104).

Threat – The core construct of threat is least explicitly defined. Johnson and Rusbult (1989) describe commitment as having the tendency to fall to "'threatenable' low levels" (p. 967), however, more commonly the presence (Kelley, 1983; Leik & Leik, 1977; Rusbult, 1983; Thibaut & Kelley, 1959), availability (Johnson & Rusbult, 1989), and attractiveness (Brehm, 1956) of potential alternative *partners* is regarded as the "threat." Threatening characteristics of alternatives other than a single alternative partner are not identified or discussed. Alternatives must be perceived as a threat to elicit devaluation.

Relationship maintenance mechanisms – Johnson and Rusbult (1989) cited Kelley (1983) in reference to relationship maintenance mechanisms, of which they propose devaluation of alternatives to be an exemplar. Kelley defines relationship maintenance mechanisms (which he refers to as "processes of commitment") as "processes that promote the person's being in a state in which the causal conditions favoring continued membership stably outweigh those acting against it" (p. 296). In their article, Johnson and Rusbult provided examples of such processes other than devaluation of alternatives, such as "invest[ing] numerous irretrievable resources in the relationship…pledg[ing] their mutual fidelity…rely[ing] on virtue...[and] adopt[ing] a broad time perspective" (p. 967).

Commitment – While commitment is not explicitly defined conceptually, it is likely appropriate to assume Johnson and Rusbult (1989) endorsed Kelley's (1983) conceptual definition of commitment, as it is inherent in the definition of relationship maintenance mechanisms identified above. For clarity, Kelley defined commitment as existing "when the total set of relevant causal conditions stably generates a resultant that is supportive of continued membership in the relationship" (p. 293). Johnson and Rusbult (1989) operationalized commitment in terms of likelihood of ending one's relationship in the near future, how long one would like their relationship to last, how attractive an alternative would need to be in order to leave one's relationship, degree of attachment to one's partner, and the extent to which one is committed to their relationship (p. 969). In short, commitment is a measure of durability and stability within a relationship. In relation to the theory of devaluation of alternatives, Johnson and Rusbult suggested that commitment must be sufficiently high to elicit a threat response (i.e., devaluation) in response to an attractive alternative. If commitment is not high, the presence of an attractive alternative may be regarded as an opportunity rather than a threat. Consequently, approach behaviors may be elicited rather than devaluation.

Identifying the explicit conceptual definitions of the key constructs are helpful when evaluating a theory because they provide a benchmark that can be used to assess the scope in which these concepts have been operationalized in the literature. These comparisons aid in identifying what aspects of the theory have been tested robustly, and

those that have been tested narrowly, revealing specific areas to focus further research efforts.

2.2.2.2 Theoretical Assumptions

The second part of the process of theory deconstruction involves identifying underlying theoretical assumptions that, when taken together, constitute the foundation upon which the theory is based (Wallis, 2014).

Theoretical assumptions can be categorized as either part of the "hard core" or the "protective belt" of a research program (Lakatos, 1978). Popper and Lakatos differed in their opinions of which assumptions should be subject to criticism, revision, and replacement, with Popper arguing for those in the hard core while Lakatos argued for those in the protective belt. In reality, researchers often target both (Dienes, 2008). The more relevant distinction between the hard core and the protective belt is that the assumptions that constitute the hard core of the theory are considered critical to observe the phenomenon of interest, in this case, devaluation of alternatives. In contrast, the assumptions that constitute the protective belt provide extra layers of complexity and nuance to the core assumptions and may be removed or revised without affecting the hard core (Cooper & Shallice, 2000). Assumptions may also be added to the protective belt to propose explanations for evidence found to contradict what the existing theory predicted. This is only appropriate if the new assumption is testable and falsifiable. If these conditions are satisfied, such an addition is representative of a progressive research program.

The assumptions of the theory of devaluation of alternatives are described below. Assumptions that make up the hard core of the theory are labeled "CA", while assumptions that make up the protective belt are labeled "PA". Three core assumptions and four peripheral assumptions were identified.

CA1: External stressors are threatening (or perceived as threatening)

Factors external to the relationship or independent of the relationship partners' actions can be stressful and threatening. Two common external stressors or threats are potential alternatives for oneself, the focus of most of the research guided by this theory, and for one's partner. While romantic relationships often convey a variety of positive benefits to those who enter them, there may always be alternatives that offer more than those currently realized. Generally, we strive to attain the best possible outcomes. In Kelley's (1983) words, "If they are offered to him, he will always prefer outcomes better than those he has, no matter how favorable the level of outcomes he has reached, for if they are offered they are instigated, and if they are better it is true by definition that he will prefer them" (p. 80). If an attractive alternative is judged to potentially offer better outcomes than one's current partner, then that person may more closely evaluate the outcomes of their current relationship. This aligns with how many scholars have conceptualized "threats," that being as influential forces that may "change the decisional context in which their intended targets make decisions" (p. 127, Milburn, 1977). Driven by our motivation to achieve the best possible outcomes, recognition that an alternative may offer greater benefits adds a complicating factor to the decisional context and may stimulate evaluative processes that could destabilize existing relationships.

PA1: More intense stressors are more threatening

This peripheral assumption extends CA1 by suggesting that the severity of external stressors influences the extent to which they are threatening. Using attractive alternatives as an example, to be threatening or perceived as threatening an alternative must meet some threshold of attractiveness that likely varies from person to person. Once that threshold is reached, greater levels of a specific attractive feature or attractiveness across multiple domains may result in stronger perceptions of threat. For example, consider a couple P and O, with O displaying average physical attractiveness, as a 6 out of 10. If P starts working with a new colleague Q, who is slightly more attractive than O (7 out of 10), Q may be perceived as threatening. But if another colleague R is extremely attractive (9 out of 10), R may be perceived as more threatening than Q. Despite the simplicity of this example, it illustrates the hypothesized positive relationship between attractiveness

and perceptions of threat. Greater quantities of attractive traits and/or specific attractive traits that are more salient would constitute greater threats.

PA2: External stressors threaten commitment

While CA1 posits that external stressors are threatening, this peripheral assumption suggests that external stressors specifically threaten commitment. This is a peripheral assumption because it is plausible that something other than commitment may be threatened (e.g., attachment, self-construal), though commitment is most often claimed to be influenced by external threats. With attractive alternatives in mind as a specific external threat, the logic underlying this assumption is founded on interdependence theory (Thibaut & Kelley, 1959) and investment theory (Rusbult, 1980; 1983). Romantic relationships are characterized by interdependence, meaning that to some extent one's well-being or happiness is dependent on their partner and how well their partner meets their needs (Kelley & Thibaut, 1978). Individuals have a comparison level that encompasses what they believe they can attain in their relationships, and they also have a comparison level for alternatives – a standard used to evaluate whether one's current relationship is satisfying enough or whether available alternatives have the potential to be more satisfying. If one's satisfaction with their relationship falls below their comparison level for alternatives, it is possible that they would end their relationship⁴. Similarly, investment theory dictates that dependence⁵ increases as satisfaction with the relationship increases, investment size increases, and quality of alternatives decreases (Rusbult, 1983; Rusbult et al., 1994). The presence of attractive alternatives in one's environment then,

⁴ This is not to say that the moment an alternative seems to be more beneficial that a relationship will be abandoned. Fluctuations in the forces that push and pull individuals towards and away from relationships are normal and to be expected. But when a consistent or extremely salient shift occurs away from membership in a relationship it is logical to assume that the relationship would likely end (Kelley, 1983).

⁵ Dependence and commitment are positively related, such that as dependence increases, commitment becomes stronger (Rusbult, 1983; Rusbult et al., 1994), however, these constructs are distinct. Commitment is determined by one's intent to persist in their relationship, one's attachment to their partner, and thinking and acting as if their relationship will survive long-term (Rusbult et al., 2001). In other words, commitment is the "sense of allegiance that is established with regard to the source of one's dependence" (p.95), whereas dependence is a characteristic property of a relationship. However, for the purpose of the current discussion distinguishing between dependence and commitment is not of high importance.

may increase one's comparison level of alternatives and quality of alternatives, both of which negatively affect commitment.

CA2: When threatened, the relationship maintenance system is activated

Once individuals have begun a romantic relationship, they preserve, enhance, and defend it. This dynamic process is called relationship maintenance and these mechanisms are galvanized in part as protective responses to perceived threats to one's relationship (Agnew & VanderDrift, 2015; Ogolsky et al., 2017). Stated simply, if a relationship is threatened it's a problem and relationship maintenance mechanisms are the solution (Rusbult et al., 2001). In their review of relationship maintenance, Agnew and VanderDrift (2015) classified threats to relationships into three types: noncorrespondence of outcomes (when a decision or action would benefit oneself but not one's partner), internal threats (destructive behavior enacted by oneself or one's partner), and external threats (hazards outside of the dyad, such as attractive alternatives).

Relationship maintenance mechanisms have been theorized to exist for a variety of reasons (Ogolsky et al., 2017). From an evolutionary standpoint, relationship maintenance mechanisms may have evolved because they effectively kept partners together despite the existence of threats (e.g., extradyadic affairs). Being coupled may have increased resource acquisition, the likelihood of sexual reproduction, the successful rearing of offspring, and the likelihood of survival in general (Buss, 1998; Buss & Schmitt, 1993). From an attachment perspective (Hazan & Shaver, 1987), the nature of one's attachment to their romantic partner, and the pursuit of achieving a secure bond, likely stimulates relationship maintenance processes to support those bonds (Etcheverry et al., 2013). And from an interdependence perspective, coupled persons may feel more motivated to behave in relationship sustaining or enhancing ways as their dependence on each other and their relationship increases. Incurring costs to enact relationship maintenance behaviors becomes acceptable and preferable because one's relationship with their partner is rewarding and worthy of protection (Kelley & Thibaut, 1978). Each of these perspectives offer logical explanations for the existence of a system that encourages membership in a dyad.

CA3: When the relationship maintenance system is activated, relationship maintenance mechanisms may occur

CA2 advanced that the relationship maintenance system is activated in threatening circumstances, with the presence of attractive alternatives identified as an example of an external threat. When the system is activated, this third core assumption suggests that individuals are primed and prepared to engage in behaviors (i.e., relationship maintenance mechanisms or strategies). Devaluation is *one* such strategy that may emerge, but other strategies may be just as appropriate and likely to occur. Devaluation and a related strategy used to address attractive alternatives are described below.

Devaluation of alternatives involves perceiving the qualities of alternatives as more costly or negative than they "objectively" are. Devaluation of alternatives can be verbal and deliberate. Kelley (1983) described a hypothetical situation in which one or both members of a couple may "break off interaction with persons who might constitute alternative partners. The renunciation of these outsiders may also involve downgrading their attractiveness, with elaborations on their negative qualities and unfavorable comparisons with the current partner" (p. 306). Doing so consciously, may provide demonstrable evidence that one is committed to their relationship (Kelley, 1983). Devaluation has also been suggested as a process that operates unconsciously and automatically for the purpose of engaging in "mental restructuring toward the goal of enhancing couple well-being" (Rusbult et al., 2001, p. 103). Rusbult provided a similar example of devaluation, describing the process as "subtly minimiz[ing] alternative partners' abilities or attributes ('I bet he has no sense of humor')" (p. 105).

Inattention is also a strategy that has been suggested to mitigate the threat of attractive alternatives, in addition to being effective as a more preventative mechanism that serves to maintain relationship stability even in the absence of threats (Agnew & VanderDrift, 2015; Miller, 1997). When attractive alternatives may be perceived as potential threats, coupled persons may divert their attention away from the alternative. Doing so may decrease the likelihood of interaction or reduce thoughts of the alternative (i.e., out of sight, out of mind). Miller (1997) summarized this strategy as "'what one doesn't know'

(about alluring alternative attractions) 'can't hurt one' (or, at least one's relationship)" (p. 764).

Devaluation and inattention are the most common strategies covered in the literature that are affiliated with responding to the specific external threat of attractive alternatives. However, other relationship maintenance strategies could also be effective responses to this specific threat, and other yet-to-be identified relationship maintenance strategies could be effective as well.

CA4: Relationship maintenance mechanisms are enacted in attempt to neutralize external threats

Implementing a certain strategy in response to a threat is an *attempt* to reduce or defeat the threat, but success is not always guaranteed. However, the existence of relationship maintenance mechanisms arguably serves as evidence of their adaptiveness to neutralize external threats, otherwise they would not exist.

PA3: Commitment must be sufficiently high for relationship maintenance mechanisms to be activated

Committed individuals are likely to be more heavily invested in their relationships and are typically more likely to be satisfied with their partners than those who are less committed (Rusbult et al., 2001). Committed individuals, therefore, feel their relationships are worthy of maintaining and are likely motivated to think, feel, and act in ways to protect their relationships from potential threats, like attractive alternatives. It may be argued then, that a sufficient degree of commitment is necessary for relationship maintenance mechanisms to be enacted.

Johnson and Rusbult (1989) suggested that the elicitation of devaluation of alternatives necessitates a high level of commitment. From the dissonance perspective, feelings of attraction toward an alternative would not likely cause discomfort or distress if one was not committed to their partner. In the absence of significant investment or a sense of obligation or loyalty to their current partner and relationship, feelings of attraction toward another would not conflict with the weak feelings one has toward their current partner.

Without this conflict, mentally restructuring one's thoughts towards the alternative would be unnecessary and would preclude devaluation from occurring.

The perceptual perspective rests on the assumption that those who are highly committed are likely also highly satisfied. Their satisfaction with their partner and their relationship inflates their comparison level, consequently resulting in a higher bar that potential alternatives must meet to constitute a threat to one's current relationship. With such high expectations established from their satisfying relationships, alternatives may pale in comparison. In contrast, individuals who are less satisfied or committed may have lower comparison levels and alternatives may be evaluated more positively (i.e., not perceived as threats but rather as opportunities) without the influence of elevated levels of satisfaction with their current relationships.

This assumption is considered peripheral because CA2 suggests that threat alone can activate the relationship maintenance system – that threat is the necessary precursor to the employment of maintenance mechanisms. While action in response to threat is a phenomenon that is well established, the role commitment plays in the activation of relationship maintenance strategies within this process is more uncertain. As the theory currently stands, relationship maintenance processes are expected to occur when a threat is perceived, regardless of the level of commitment.

Parsing apart a theory into a series of specific assumptions is beneficial in a variety of ways. From an assessment standpoint, identifying assumptions facilitates highly focused and detailed evaluations of the literature and aids in detecting specific assumptions that may not be strongly corroborated. These elements can then be the target of future research to test the strength of the foundation of the theory. With research design in mind, identifying assumptions allows for research to be tailored specifically to test a certain assumption, or at least encourages transparency in stating which assumptions are considered as true (without testing) when interpreting results. And, with collaboration and integration in mind, identification of a theory with fellow researchers in the field. Since different researchers may endorse and operate under different assumptions, being explicit

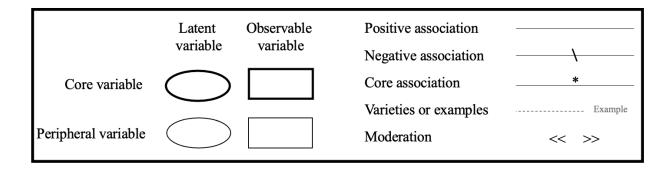
in one's understanding of the theory clearly establishes the background knowledge upon which a certain study or theoretical argument is based.

2.2.3 Reconstructing Devaluation of Alternatives with Theory Mapping

Once a theory has been through the categorization and deconstruction processes where key concepts and assumptions are identified and defined, the next step in formalizing theory is reconstruction. One method of reconstruction involves creating a theory map (Gray, 2017). In tandem with the information extracted from deconstruction, a reconstructed theory map aids researchers in visualizing and communicating their understanding of a certain theory. This can be especially helpful in detecting and eliminating redundant theoretical elements and identifying potential disagreements or inconsistencies between researchers' theoretical views (Gray, 2017).

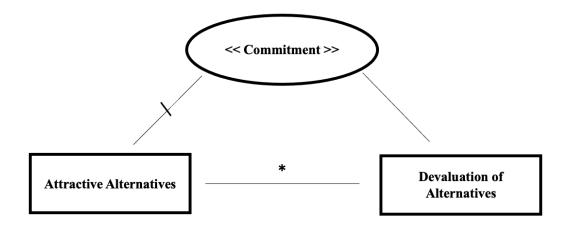
Gray's (2017) theory mapping technique involves identifying and characterizing relevant constructs (e.g., core or peripheral, latent or observable) and visualizing how they are related (e.g., positive relationship, negative relationship, moderation). Notably, Gray emphasized theory maps are not intended to replace computational or statistical models and that this technique can be adapted to fulfill researchers' specific purposes. For this endeavor, Gray's original notation was used to construct the theory maps presented here (see Figure 1); no significant modifications were made.

Figure 1. Theory Map Legend



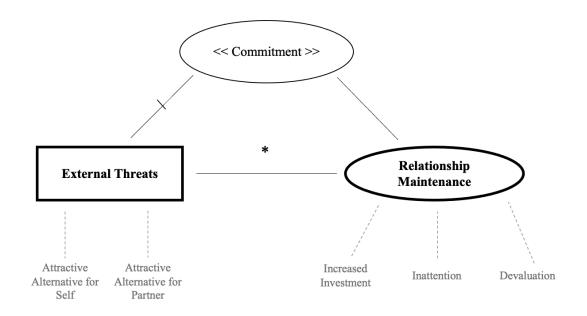
This first theory map represents how devaluation of alternatives may be visualized based on Johnson and Rusbult's (1989) original proposal (see Figure 2). The map consists of three core concepts: attractive alternatives, devaluation of alternatives, and commitment. Attractive alternatives and devaluation of alternatives are observable entities, while commitment is a latent psychological construct. The presence of attractive alternatives is hypothesized to negatively affect commitment, signifying that commitment is what is being threatened (PA2). When attractive alternatives are present and perceived to be threatening, devaluation is hypothesized to occur, which represents the core processes of the theory (CA1, CA2, CA3). Commitment is suggested to moderate this process, such that devaluation is only expected to occur when commitment is sufficiently high, while devaluation would be less likely to occur when commitment is low (PA3). Lastly, when devaluation occurs, commitment is hypothesized to stabilize or increase (CA4).

Figure 2. Theory Map of Johnson and Rusbult's (1989) Original Proposal of Devaluation of Alternatives



As alluded to earlier, devaluation is only one possible strategy that can be employed to counter external threats. Similarly, attractive alternatives are only one possible type of external threat. Consequently, this original conception and test of devaluation of alternatives may be more accurately regarded as a sub-theory of the larger theoretical framework of relationship maintenance rather than a stand-alone theory on its own. Figure 3 represents a theory map of threat-induced maintenance processes (Agnew & VanderDrift, 2015), which more broadly encompasses the process of responding to external threats.

Figure 3. Theory Map of External Threat Induced Maintenance Processes



The theory map of external threat-induced maintenance processes displays similarities with the theory map of devaluation of alternatives. This should be unsurprising as use of devaluation in response to attractive alternatives can be considered a specific example derived from this parent map. However, some notable changes exist. Attractive alternatives were transferred from a core feature to an example of an external threat, which constitutes a core, observable feature of the map. Another example of an external threat was added, such as one's partner having an attractive alternative (i.e., a potential mate poaching situation; Buss, 1998; Lee & O'Sullivan, 2018). Similarly, devaluation of alternatives was relegated to the status of an example, rather than a core feature, and was replaced with the broader core latent construct of relationship maintenance. Additional examples of relationship maintenance mechanisms that can be used to combat external threats were added to the map (e.g., inattention, increased investment). The examples of the core constructs provided in the theory map are in no way comprehensive.

The last notable change is the demotion of commitment from a core construct to a peripheral construct. While it is plausible that external threats do affect commitment specifically, they may influence other relationship or psychological characteristics. For example, external threats like attractive alternatives might be threatening to one's partner, who may react negatively. In this scenario one's commitment is not at risk, but perhaps having a satisfied and secure partner is. Therefore, *what* is threatened is less essential to

specify. What is more important, and what is necessary to occur for relationship maintenance mechanisms to be activated is the detection and recognition of a specific stimulus as a threat. Relegating commitment to a peripheral element aligns the theory map with the third peripheral assumption identified in the previous section, that suggests commitment must be sufficiently high for relationship maintenance mechanisms to be activated.

Thus far, the conception of devaluation of alternatives has been described, core constructs defined, and assumptions identified and categorized as core or peripheral. Through the processes of deconstruction and reconstruction, it was observed that both attractive alternatives and devaluation may more appropriately be considered as one of many types of external threats and mechanisms of relationship maintenance, respectively. Rather than being thought of as its own independent theory, devaluation can more appropriately be considered a sub-theory that exists within the broader theory of relationship maintenance in response to external threats.

2.3 Assessing Devaluation of Alternatives

The final step in theory formalization is evaluation. While there are multiple theoretical and methodological approaches to evaluating theories (e.g., Meehl, 1990; Wallis, 2010), I structured this evaluative approach around Lakatos's (1970) distinction between progressive and degenerative research programs, the concept of verisimilitude and two principles that Meehl (1990) considered necessary to warrant the Lakatosian defense of an existing theory.

2.3.1 Progressive vs. Degenerative Research Programs

When seemingly falsifying evidence is observed, theories can be protected by constructing auxiliary theories or hypotheses that offer suggested explanations for the evidence. However, these auxiliaries can vary in quality and purpose. Popper asserted a need to distinguish between auxiliaries that contribute to scientific progress and those that do not, which he termed *ad hoc* hypotheses (Lakatos, 1970). Lakatos proposed that auxiliary theories are theoretically progressive "if each new theory has some excess

empirical content over its predecessor; that is, if it predicts some novel, hitherto unexpected fact" (p. 183), and is also empirically progressive "if some of this excess empirical content is also corroborated, that is, if each new theory leads us to the actual discovery of some *new fact*" (p. 183). In contrast, auxiliary theories are considered degenerative if the theory offers an explanation for counterevidence in a way that is content decreasing, meaning it does not lead to novel predictions and facts.

To assess the extent to which devaluation of alternatives as a research program is progressive or degenerative, I will review how this sub-theory evolved over time. Examining the life history of the theory is one way to look at the "big picture" – the broader research program. Zooming out to take this perspective is important because "The more you focus on the details within individual studies, the more you lose sight of the connections between these studies, missing the forest for the trees" (Trope & Liberman, 2011, as cited in Gray, 2017, p.3). While we are often quick to critique methodological details of specific studies, we should also scrutinize the state of the body of literature the studies exist in. Doing so may reveal large gaps, areas of heavy activity, debates, disagreements, and other potentially useful observations that may be relevant when evaluating a theory.

2.3.1.1 Evolution of the Theory of Devaluation of Alternatives

Following Johnson and Rusbult's (1989) article introducing devaluation of alternatives, Simpson et al. (1990) was the first to build on the phenomenon. Simpson et al.'s oft cited contribution to this area of research was their operationalization of attractive alternatives as photos of physically attractive magazine models. From their research they concluded that potential alternatives did not have to be realistic to elicit derogation (in contrast to Johnson and Rusbult's assertion that relationship maintenance processes only operate when faced with realistic threats). This study influenced much of the subsequent research involving devaluation in which authors often used photographs (in many cases facial photographs) as the experimental stimuli to represent potential alternatives⁶.

After these studies, research pivoted to competing theories and explanatory mechanisms. Miller (1997) suggested that an important precursor to devaluing attractive alternatives was being *aware* of such alternatives. Miller proposed inattention to alternatives as a relationship maintenance strategy which stimulated further studies of his own and research by other scholars as well (DeWall et al., 2011; Koranyi & Rothermund, 2012; Ma et al., 2019a, 2019b; Maner et al., 2008, 2009; Zhang et al., 2017). Subsequent research, however, has sometimes blurred the distinction between devaluation and inattention (see Ogolsky et al., 2017).

Bazzini and Shaffer's (1999) work suggested an alternate explanation for differences in evaluations of potential alternative partners between highly and less committed individuals. Instead of evaluations of potential alternatives being devalued by highly committed individuals, less committed individuals may be enhancing their evaluations, making highly committed individuals' evaluations seem like underestimations in comparison. Bazzini and Shaffer's work represented the first challenge, or competing theory, to devaluation of alternatives (the devaluation vs. enhancement debate). Additionally, they recognized an interesting issue in this area of research, which is identifying who evaluates the attractiveness of individuals most objectively. For example, if single and coupled persons differ in their evaluations of an alternative, whose evaluations are most objective? Are evaluations more objective when coming from samesex evaluators or opposite-sex evaluators? Identifying the "true neutral" evaluators would create comparisons that would allow more confident assertions for when devaluation or enhancement occur. The observation of enhancement behaviors represented a progressive addition to the theory of devaluation of alternatives. Enhancement offered a competing explanation for the observed effect, representing the proposal and discovery of a "new

⁶ Often (but not exclusively) photographs were supplemented with information in cover stories that informed the participants the photographs were real people.

fact." Importantly, it was testable and elicited corroborating evidence. The proposal of a tendency to enhance evaluations of alternatives created a need for an auxiliary theory to make sense of the evidence that seemed to contradict devaluation of alternatives.

In response to this development, Lydon et al. (1999, 2003) proposed the commitment calibration hypothesis, which speculated that devaluation would only be observed when the level of threat was calibrated with one's level of commitment. If threat exceeds one's level of commitment, relationship maintenance mechanisms may not be elicited or may be ineffective due to the threat being too strong. When commitment exceeds the level of the threat, the potential threat may not be perceived as such, precluding activation of the relationship maintenance system. However, when the level of threat is comparable to the level of one's commitment (i.e., threat and commitment are calibrated), relationship maintenance mechanisms like devaluation will occur in response. This represents another progressive shift in the theory's life history. The calibration hypothesis offered a testable explanation for why enhancement may be observed while specifying new facts that should be discovered if the hypothesis was correct.

The next prominent phase of research in this area employed a more cognitive approach and introduced another set of competing hypotheses about devaluation: whether it is explicit or implicit. Much of the past research at this point measured devaluation by ratings of attractiveness of potential alternatives, a measure considered by Plant et al. (2010) to be conscious and explicit. Plant et al. utilized an interesting paradigm to identify differences in threat detection by outfitting images of opposite-sex alternatives with guns or neutral objects and measuring response times and the correct hit rate for responses of "SHOOT" or "DON'T SHOOT" (the correct response being to choose to shoot images containing a gun). Other approaches to measuring implicit devaluation involved asking men to evaluate women actors at different times throughout their menstrual cycle (Miller & Maner, 2010), assessing differences in the attractiveness of participants' memories of alternative faces (Cole et al., 2016; Karremans et al., 2011), and exploring the neural correlates of devaluation (Meyer et al., 2011; Miller et al., 2012). Other research explored additional variables that were hypothesized to influence

the ability to derogate, such as self-regulation and regulatory focus (Lydon & Karremans, 2015; Rodrigues et al., 2017) and executive control (Pronk et al., 2011).

Devaluation research that took place roughly between 2008 and 2017 is characterized by its focus on methodological techniques rather than designing studies or using findings to refine the theory. Instead, researchers referred to the "well documented" nature of the effect and focused on explaining the rationale behind their methodological designs. At this stage, it is difficult to determine according to Lakatos's criteria, whether this is progressive or degenerative, as the focus of this phase of research was largely operational – varying how alternatives are presented or how they are evaluated, rather than conceptual – suggesting alternative hypotheses or theories derived from past literature and connected to the theory proper. Notably, this phase coincides with the initiation of the credibility revolution, providing some important historical context and a likely explanation for why the focus turned towards methods and away from theory, as previous scholars have observed (Borsboom, 2013; Fiedler, 2004).

The early life of the research program of devaluation of alternatives was quite progressive, with researchers interacting deeply with each other's previous findings – connecting, challenging, and integrating knowledge – collectively refining the theory with each publication. However, the later (more recent) research in this area appears to be quite the opposite. There is little connection, little integration, and scarce (perhaps even nonexistent) evidence of suggested theoretical refinements. Lakatos asserted that "It is a succession of theories and not one given theory which is appraised as scientific or pseudo-scientific. But the members of such series of theories are usually connected by a remarkable *continuity* which welds them into *research programmes*. This *continuity*...plays a vital role in the history of science; the main problems of the logic of discovery cannot be satisfactorily discussed except in the framework of a *methodology of research programmes*" (1970, p. 191). Instead of attaching a label of progressive or degenerative to the research program as it exists today, a more appropriate conclusion may be that a research program no longer exists.

2.3.2 Verisimilitude and the Lakatosian Defense

The second approach I will use to evaluate the sub-theory of devaluation of alternatives is to assess the degree of verisimilitude this sub-theory demonstrates. Verisimilitude is a concept that qualifies the "truth-likeness" of a theory and is suggested to be correlated with "the degree of evidentiary support, the number, variety, and stringency of empirical tests that the theory has passed or failed" (Meehl, 1990, p. 113). Meehl suggested that the extent of a theory's verisimilitude can be appraised by two principles – the Lakatos principle ("money in the bank") and the Salmon principle ("the damn strange coincidence"). These principles can be used to make determinations about when theories should continue to be refined or when they should be abandoned.

According to Lakatos's principle, theories have high verisimilitude and should continue to be revised if they have a successful record of surviving numerous strict and diverse tests, in other words, they have ample "money in the bank." In addition, theories with core assumptions found to have strong support can be interpreted as having higher verisimilitude than those who have a strong periphery but a weak core (Meehl, 1990).

Salmon's principle asserts that "the main way a theory gets money in the bank is by predicting facts that, absent the theory, would be antecedently improbable" (Meehl, 1990, p. 115). One way to consider Salmon's principle satisfied is by finding evidence for theoretically derived point or range predictions. Similarly, corroborated predictions that rank order group observations can be considered more indicative of strong theory. Meehl (1990) recognized that point predictions were difficult to make and rare in soft sciences such as psychology. He identified another method of establishing Salmon's principle that may be considered more realistic and achievable by psychologists. While a theory may not be strong enough to posit a point prediction, a theory may specify that certain observations (i.e., within an *a priori* determined range) will be recorded by two or more "qualitatively diverse observational avenues" (p. 120).

To determine whether devaluation of alternatives fulfills the Lakatos and Salmon principles I assessed the strength, quantity, and robustness of results for each assumption inherent in the sub-theory by examining the methodological variety and the precision of the predictions in the published literature.

2.3.2.1 CA1: External Stressors Are Threatening (Or Perceived as Threatening)

The assumption that attractive alternative partners are threatening is at the very heart of the sub-theory of devaluation of alternatives, however, very few studies have attempted to measure threat perception. Threat is commonly treated as a construct to manipulate and is rarely measured as an outcome variable or as a manipulation check. Johnson and Rusbult (1989) cited four publications (Kelley, 1983; Leik & Leik, 1977; Rusbult, 1983; Thibaut & Kelley, 1959) following their assertion that "one of the major threats to the stability of a relationship is the presence of an attractive alternative" (p. 967). Each of these publications was examined to determine the extent to which their content supports Johnson and Rusbult's claim.

Thibaut and Kelley (1959), Leik and Leik (1977) and Kelley (1983) were all theoretical works that either involved proposing a theory or describing existing research in a particular theoretical domain. None of these publications focused specifically on external threats nor explicitly referenced the especially threatening nature of attractive alternatives. Rather, this was assumed from logical inferences. For example, from an interdependence theoretical perspective, it is logical to assert that the presence of an attractive alternative partner (assuming recoupling is likely or at least possible) would increase one's CL_{alt}, which *could* have a destabilizing effect on one's current relationship. However, so might the idea of being alone (i.e., this is another potential alternative to being in a relationship). The extent of destabilization would also depend on one's comparison level and current relationship outcomes (Thibaut & Kelley, 1959). Similarly, Leik and Leik (1977) suggested that when the exchanges (i.e., benefits) of a committed relationship could be gained in a different involvement, it would constitute a "challenge" to commitment. Leik and Leik took a more muted stance than what one would assume based on Johnson and Rusbult's (1989) citation. The closest Leik and Leik came to suggesting that alternatives are a major threat is their observation that "a confrontation

(even if only covertly recognized) between alternative partners is most likely resolved by reduction in involvement with the 'loser'" (p. 317), with the loser being either the partner or the alternative. Leik and Leik stated that a challenge of an attractive alternative could result in "a variety of jealousies" (p. 317), and if successful, may result in "the relationship which was committed...return[ing] to a lesser involvement" (p. 317). Kelley (1983) identified and described the causal conditions of commitment and processes of commitment. He listed some examples of "conditions that act to push or draw the person out of the relationship, such as the psychological costs (effort, anxiety) experienced with the partner, the attractiveness of alternative relationships that are precluded by the present one, internal and external pressures to experiment with such alternatives, and so on" (p. 289). However, as in the previous examples, the threatening nature of attractive alternative partners is not distinguished from other potential relationship destabilizers. It is unclear then, what specific aspects of these works support the claim that alternative partners are a *major* threat to relationship stability. While some may dismiss this as a minor matter of semantics, this assertion is a core postulate upon which the theory of devaluation is based, and therefore warrants this level of scrutiny.

The remaining study Johnson and Rusbult (1989) cited in their foundational article on devaluation of alternatives was Rusbult's (1983) work that tested the Investment Model. The study took place over a 7-month period and involved 34 participants. Results showed a negative relation between commitment and quality of alternatives (operationalized by two items: "In general, how appealing are your alternatives [dating another person or other persons, or being without a romantic involvement]?", and "All things considered, how do your alternatives compare to your current relationship?"). Those whose relationships ended (N = 10) reported a slight increase in alternative quality, with the effect driven by participants who ended their relationships compared to those who were left by their partners. Despite these results emerging in the expected direction, a small effect coupled with the miniscule sample size represents only weak support that attractive alternatives are threatening. In addition, the items used to measure alternatives is muddied by the inclusion of potential alternatives aside from other romantic partners. Consequently, "alternative quality" in this study is not specific to attractive alternative partners. Despite the empirical nature of this work, Rusbult's (1983) cross-sectional

results do not directly support the claim that attractive alternative partners constitute a "major threat" to relationship commitment or stability. Taken together, the studies Johnson and Rusbult (1989) cited did not offer convincing, strong, or specific evidence of the salient and threatening nature of attractive alternatives, though the citations appear repeatedly throughout the literature as references for evidence of this claim.

To my knowledge, investigations of whether attractive alternatives were perceived as threats were not undertaken until 2008, almost 20 years after the sub-theory was proposed. Lydon et al. (2008) was interested in gender differences and the extent to which the implicit threat of an attractive alternative was effective in eliciting feelings of threat. In Study 1, coupled participants (N = 300) were asked to imagine themselves in a scenario that involved an interaction with a same-sex person, or with an attractive opposite-sex alternative who either made a reference to their dating partner or asked for the participant's phone number. Participants then completed a word-fragment completion task with 16 word fragments, six of which had either commitment or neutral completion options, three had threat or neutral completion options (i.e., thr_t [threat/throat], be_a_e [beware/became], t_m_ing [tempting/tumbling]), and seven were filler neutral words. In Study 2, virtual reality technology was used to investigate the extent to which men and women would spatially distance themselves from an attractive alternative. Coupled participants (N = 115) were simply tasked with arranging four images (three were neutral objects, one was an attractive opposite-sex, age-appropriate person) in their virtual reality space in a manner pleasing to them. Half of the participants filled out a relationship attitudes survey meant to prime them with thoughts of their relationships, while the other half only filled out demographic questions. In both studies, women in the threat condition differed significantly from all other conditions, by reporting more threat words (Study 1) and by placing the alternative image farther away from them (Study 2). Notably, in Study 1, the main effects of experimental condition and gender on number of threat word completions were not reported, and pairwise comparisons were reported despite the interaction failing to reach the threshold for significance. Interpretation of these results should be cautious as they heavily depend on whether participants considered a hypothetical (Study 1) and static virtual reality image (Study 2) of an opposite-sex person as a potential alternative and a realistic threat. No measures involving perceptions of the

"alternatives" were solicited that could shed light on this issue. Consequently, despite utilizing implicit measures and a clever operationalization of feeling threatened (distance between oneself and the source of the threat), sizeable hesitancy accompanies using the results as evidence for or against the claim that attractive alternatives are particularly threatening.

Plant et al. (2010) explicitly recognized that whether attractive alternatives elicit an automatic threat response had yet to be directly tested (aside from Lydon et al., 2008). Drawing on a popular task used in the stereotyping literature, the authors adapted a computer simulation task wherein participants identified targets as dangerous ("SHOOT") or not dangerous ("DON'T SHOOT") based on whether the target had a gun. The authors varied the facial attractiveness of the targets (attractive vs. average), the gender of the target (man vs. woman), and the object the target was holding (gun vs. neutral object), and also included participant gender (man vs. woman) and relationship status (committed vs. single) in their analysis. The hypothesized five-way interaction emerged as significant, with follow-up tests revealing effects only within committed participants. Specifically, committed men were more likely to make errors when faced with unarmed attractive woman targets compared to armed attractive woman targets, indicating that attractive targets were perceived as threats. No differences in error rates were found when the woman targets were of average attractiveness, suggesting these targets were not perceived as threats. No significant differences were found with women or single participants.

Plant et al. (2010) interpreted these results as evidence in support of the claim that attractive alternatives stimulate automatic threat responses, especially for committed men (in contrast to Study 2 in Lydon et al., 2008). Both studies, however, are limited in that the alternative targets were static photographs, with Plant et al.'s being a photograph of only alternatives faces. As mentioned above, the external validity of such stimuli must be considered when making conclusions from these studies and generalizing them to the presence of or interactions with attractive alternatives in reality. In addition, while Plant et al.'s adaptation of a paradigm used in the stereotyping literature for identifying threat was crafty, such paradigms (i.e., using SHOOT vs DON'T SHOOT) are often used in the

context of implicit racial bias, where there exists a clear and well-documented stereotype that Black individuals are considered *physically* dangerous and threatening. This paradigm, especially when used in research with police officers who carry weapons (Plant & Peruche, 2005), may be argued to be a more realistic and an externally valid methodological design for measuring threat perception and response. However, applying such a paradigm to the context of romantic relationships is inappropriate because attractive alternatives are not claimed to be physically threatening and do not require nor considered capable of stimulating impulse reactions involving violence. Operationalizing a threat response in terms of a decision to shoot the target or not is inapt and extreme in this context.

The assumption that attractive alternative partners are major threats to the stability of a relationship was originally based on theoretical reasoning and logic. Strong theoretical rationale, while acceptable at early stages of theory development, should be accompanied by equally strong and robust empirical support. However, in the 30 years since devaluation of alternatives has been proposed, only three studies have measured threat perception in response to attractive alternative partners and have only done so using implicit measures. Returning to Lakatos's principles, it is clear this assumption has little to no "money in the bank." The quantity of evidence is severely lacking and the strength of the effect is questionable due to the inconsistent gender differences found between Lydon et al.'s (2008) and Plant et al.'s (2010) studies. Regarding robustness, while diverse methods were used to both manipulate and measure threat, each study has weak external validity. With the Lakatos principle unsatisfied, the Salmon principle follows suit. The dearth of evidence coupled with the conflicting gender difference findings indicates little convergence in results between studies using different methods. Consequently, this core assumption presently lacks verisimilitude.

2.3.2.2 PA1: More Intense Stressors Are More Threatening

While few researchers have attempted to measure threat perception in the context of devaluation of alternatives, many have manipulated the characteristics of target alternative partners, such as level of attractiveness (Johnson & Rusbult, 1989; Karremans

et al., 2011; Ma-Kellams et al., 2017; Meyer et al., 2011; Plant et al., 2010; Ritter et al., 2010; Rodrigues & Lopes, 2013), age-appropriateness (Simpson et al., 1990), availability (Karremans & Verwijmeren, 2008; Lydon et al., 2003), and expressed or reciprocal interest (Bazzini & Shaffer, 1999; Lydon et al., 1999; Lydon et al., 2008). But the degree of threat posed by the targets was always implied, never measured. Consequently, evidentiary support for this assumption exists exclusively in the theoretical realm. Therefore, an attempt to assess the degree to which Lakatos and Salmon's principles are satisfied is moot. The theoretical logic that supports this assumption is described below.

Researchers who have operationalized and manipulated threat in their investigations of devaluation of alternatives seem to endorse the validity of this assumption. For example, Lydon et al. (2008) made such a distinction when they stated, "now imagine a real-world interaction with an attractive alternative, rather than the casual perusal of a picture and a file folder. The attractive person is showing a great interest in you, maintaining eye contact, and maybe even being a bit flirtatious. The threat is strong" (p. 51). Here, Lydon et al. critiqued a common method of operationalizing threat in the literature – exposing participants to static photographs of attractive individuals meant to represent potential alternatives to their current partners. Often participants were exposed to a single photograph, sometimes only a photograph of the target's face. Methodologically, a stronger manipulation of threat would involve interacting with a realistic, available, and interested potential alternative. In parallel, such an alternative in the real world would logically constitute a greater threat to one's current partner and relationship compared to an alternative who was not realistic (e.g., a celebrity), unavailable (e.g., already coupled), and uninterested (e.g., affections or attraction was not reciprocated).

This assumption is defendable by multiple theoretical perspectives. From an interdependence perspective, more attractive alternatives (in terms of their salience or number) would increase CL_{alt} to a greater extent than less attractive alternatives, lowering the threshold for which one may be at risk of leaving their current relationship. Since CL_{alt} could be thought of as a specific place on a continuous scale, the sensitivity of CL_{alt} to fluctuations in the quality of alternatives is inherent in the nature of the construct. From an evolutionary theoretical perspective, the physical attractiveness specifically of a

potential sexual partner is more highly regarded by both men and women, as it acts as a signal of good health and good genes (Gangestad & Thornhill, 1997). Following this logic, more physically attractive alternatives may pose stronger threats to coupled individuals compared to less physically attractive alternatives, as they signify greater temptation and could potentially offer greater rewards (Maner et al., 2008).

This assumption remains in the periphery of the theory of devaluation of alternatives because the extent or degree of the threat posed by an alternative is somewhat tangential to understanding relationship maintenance mechanisms, such as devaluation of alternatives. If the alternative reaches some arbitrary and independent threshold for being perceived as a threat, the relationship maintenance system should be activated. Perhaps more salient threats may elicit stronger, more frequent, or a diverse variety of maintenance mechanisms, or perhaps similar behaviors may be used in response to threats of varying degree. Regardless, the perception of threat, rather than the degree of threat, is the necessary precursor of relationship maintenance.

2.3.2.3 PA2: External Stressors Threaten Commitment

Interdependence theory and investment theory are often used to explain why and how attractive alternatives affect commitment, and as such, many empirical investigations of derogation of alternatives consider commitment a key component of the theory and include a measure of commitment in their research. The association between exposure to or quality of attractive alternatives and commitment has been explored using diverse methods including longitudinal studies, cross-sectional self-report studies, and experiments.

Johnson and Rusbult (1989) found that participants whose relationships ended during a longitudinal study had increased positive evaluations of their actual potential alternative partners compared to participants whose relationships endured throughout the study, and those with higher commitment had lower evaluations of alternative partners. In contrast, exploring exposure to attractive alternative partners through participants' own unique social media behaviors and networks revealed a nonsignificant (and positive) relationship between exposure to alternatives and commitment (de Lenne et al., 2019). This evidence

contains conflicting support for the influential relationship between commitment and attractive alternatives, and its cross-sectional nature prevents inferences about directionality and causality.

Experiments have shown similarly weak and conflicting evidence for the association between commitment and attractive alternatives. Hadden (2016) measured commitment before and after exposing participants to target photographs (10 attractive and 10 unattractive opposite-sex faces), but neither measure of commitment was associated with physical attractiveness or sex appeal ratings of the attractive or unattractive photographs and commitment did not predict whether the participants considered the targets potential partners. When differences between pre- and post-measures of commitment were analyzed, ratings of attractive alternatives and the extent to which participants considered targets to be potential partners did not significantly predict decreases in commitment, though the effect emerged in the expected negative direction. When commitment was measured implicitly in Lydon et al.'s (2008) study after participants imagined an interaction with a same-sex or opposite-sex person, results revealed a significant effect of gender and a significant effect of experimental condition, likely⁷ such that participants who were women or in the threat condition completed more word fragments with commitment words rather than neutral words. In both these experimental contexts, exposure to facial photographs of attractive individuals or simply imagining an interaction with an opposite-sex person may have marginal negative effects on commitment.

While much of the literature has focused on commitment as the target of external threats such as attractive alternatives, other constructs could be affected as well. For example, Johnson and Rusbult (1989) found that satisfaction was also related to evaluations of alternative partners, though commitment was found to be the stronger predictor. Similarly, Linardatos and Lydon (2011) recognized that attractive alternatives were

⁷ Marginal means and standard deviations were not reported, therefore the nature of the group differences (i.e., which elicited higher commitment related word completions) cannot be confidently stated.

thought to decrease commitment, satisfaction, and dependence. Researchers have also explored activation of the attachment system in response to relationship threats such as impeding separation, insecurity with one's partner, and the possibility of their partner's infidelity (Birnbaum et al., 2008; Birnbaum et al., 2011; Mikulincer et al. 2002), as well as in response to internal threats such as hurtful behavior by a partner (Birnbaum et al., 2019). Curiously, the activation of the attachment system in response to the threat of attractive alternatives has yet to be explored. A third possible target of the threat of attractive alternatives could be one's own self-concept or self-esteem, as implied by the motivational explanation of devaluation proposed by Johnson and Rusbult (1989). Attractive alternatives may create distressing and dissonant feelings within a person, threatening not their commitment but their view of themselves as a person or as a partner.

While these potential processes are less researched in the context of attractive alternatives, they do constitute plausible alternative or additional components of the theory of devaluation of alternatives. Therefore, the assumption that attractive alternative partners threaten commitment specifically remains in the protective belt of the theory rather than the hard core, as other constructs or aspects of oneself or one's relationship could be threatened and comparably elicit relationship maintenance mechanisms such as devaluation.

Though the cross-sectional evidence that tested the association between attractive alternatives and commitment was beneficial in that participants had actual alternative partners in mind, the results were inconsistent. Similarly, experimental evidence did not show strong or consistent evidence for significant effects of exposure to attractive alternatives on commitment. These studies do involve a diverse use of methodological designs, manipulations and measures, however, stable results across these studies were not found. These findings taken together with the presence of highly plausible personal and relationship-oriented characteristics that may be vulnerable to threat aside from commitment that have not yet been ruled out as alternative explanations leads to the conclusion that this assumption also lacks verisimilitude.

2.3.2.4 CA2: When Threatened the Relationship Maintenance System Is Activated

The activation of the relationship maintenance system can only be inferred from direct observations of proposed relationship maintenance mechanisms and the precipitating factors that precede them. If physiological, behavioral, or cognitive maintenance mechanisms are observed in response to a threatening situation or individual, the relationship maintenance system can be said to have been activated⁸. This assumption is identical to that of other theories that suppose the activation of a latent system prior to enacting a certain behavior or set of behaviors. For example, the activation of the attachment system in response to a threat motivates an individual to seek comfort from their attachment figure (Birnbaum et al. 2008, Birnbaum et al., 2011; Bowlby, 1969/1982;) and deciding what action to take in response to environmental stimuli or specific action goals depends on the activation of latent schema networks (Cooper & Shallice, 2000; Norman & Shallice, 1986). The possible behavioral responses to the activation of the relationship maintenance system are identified and evaluated in-depth in the following section.

2.3.2.5 CA3: When the Relationship Maintenance System Is Activated, Relationship Maintenance Mechanisms May Occur

A variety of strategies have been identified and observed following the activation of the relationship maintenance system. Devaluation is one possible maintenance mechanism that could be elicited in response to threatening attractive alternatives. Due to devaluation

⁸ As mentioned previously, the relationship maintenance system can also be activated in the absence of threat, for example, when individuals want to maintain or increase their interdependence and stability with their partner (Agnew & VanderDrift, 2015). However, different types of relationship maintenance mechanisms are activated and utilized to serve different purposes. Devaluation of alternatives is considered a maintenance mechanism that exists specifically to combat the external threat of attractive alternatives. This assumption will therefore be evaluated in the context of threatening circumstances, specifically of attractive alternatives, rather than in preventative circumstances (e.g., instances where interdependence or commitment is intended to be enhanced) or in response to other external threats (e.g., one's partner having attractive alternatives).

being the mechanism of interest in this dissertation, this assumption will mainly be evaluated by examining this specific maintenance mechanism rather than the broad spectrum of mechanisms that might be employed to combat this specific external threat.

2.3.2.5.1 Methodological Review

Table 1 shows a selection of studies and experiments since 1989 that measured devaluation of alternatives by *manipulating* exposure to alternative targets. While efforts were made to ensure the inclusion of studies in this table was comprehensive, it is possible that relevant and eligible studies were missed. However, the present selection of 36 studies across 20 publications likely constitutes a strong and representative sample of the literature. The evaluation of this selection of literature will first involve a description and critique of the various methodological approaches used to stimulate and measure devaluation. Then, the results of these studies will be summarized. In addition, research that was not included in Table 1 will be discussed as well.

		Methods																		
		Pa	mitment	Manipulation(s)										Dependent Variable(s)						
Authors	Study	Ν	Sample	Scale (e.g., hi vs. lo)	Rel. Status (e.g., dating vs. single)	Target Attractiveness Target Availability Target Interest		Target Gender	Target Gender Evaluation Perspective		Multiple Targets	Interaction w/Target Imagine Scenario		Target Physical or Sexual Attractiveness Other Target		Desire to Meet Target	Pote		Behavioral Response	
Johnson & Rusbult (1989)	2	273	Ugrad	Х		Х				Х	\mathbf{X}^{F}				X	Х	Х	Х		
	3	76	Ugrad	Х										Х	X	Х	Х	Х		
Simpson et al. (1990)	1	204	Ugrad		X	Х						Х			X					
	2	197	Ugrad		X	Х			Х			Х			Х					
Bazzini & Shaffer (1999)	1	98	Ugrad		X		37	X		Х	37		ala.	Х	X			X		
	2	104	Ugrad		Х		Х	Х			Х	Х	*		X			Х		
Lydon et al. (1999)		134	Grad & Community	Х	Х			Х			Х		*		Х	Х		Х		
Lydon et al. (2003)		174	Ugrad	Х			Х				Х		*		X	Χ		Х		
Karremans & Verwijmeren	1	20 (M)	Ugrad		Х						Х		Х							Х
(2008)	2	53	Ugrad		Х						Х		Х							Х
	3	76 (M)	Ugrad		Х		Х				Х		Х		Х					X
Miller & Maner (2010)		38 (M)	Ugrad		Х						Х		Х		Х	Х				
Ritter et al. (2010)	1	102	Ugrad	Х	Х	Х						Х						Х		
	2	93	Ugrad	Х	Х	Х						Х						Х		
VanderDrift et al. (2011)	1	62	Ugrad	Х							\mathbf{X}^{NP}		Х			Х	Х	Х		
	2	145	Ugrad	Х								Х	*				Х			
Karremans et al. (2011)		82	Ugrad?		Х	Х					\mathbf{X}^{F}				Х					
Meyer et al. (2011)		14	Unspecified	Х		Х						Х						Х	Х	Х
Pronk et al. (2011)	2	21 (M)	Ugrad								Х		Х		Х					Х
	3	65	Ugrad		Х						Х		Х		Х		Х			
Rodrigues & Lopes (2013)	2	136	Ugrad?	Х	Х	Х					\mathbf{X}^{F}				Х			Х		
Visserman & Karremans	1	67 (W)	Ugrad	Х	Х						Х		*			Х				
(2014)	2	87 (W)	Ugrad	Х	Х						Х					Х				Х

Table 1. Methodological Characteristics of Studies and Experiments Involving Devaluation of Alternatives

	3	78 (W)	Ugrad	Χ				Х			*		Х				Х
Petit & Ford (2015)		72	Ugrad		Х					Х		Х				Х	
Hadden (2016)	1	562	Ugrad	Х		Х				Х		Х			Х		
	2	185	Ugrad	Х				Х		Х		Х		Х			
	3	209	MTurk	Х				Х				Х		Х			
Wang et al. (2016)	1	350 (W)	Online	Х	Х					\mathbf{X}^{F}							Х
	2	1000	Online	Х	Х					\mathbf{X}^{F}							Х
		(W)															
Rodrigues et al. (2017)	1	230	Online	Х	Х				\mathbf{X}^{F}					Х			
	2	55	Ugrad	Х					\mathbf{X}^{F}					Х			
	3	132	Ugrad	Х					\mathbf{X}^{F}					Х			
Ma-Kellams et al. (2017)	3	134	Ugrad &						Х			Х					
			Community						Λ			Λ					
	4	156	MTurk							Х		Х					
McNulty et al. (2018)	1	224 (C)	Community	Х	\mathbf{X}^1	Х		Х		\mathbf{X}^{F}		Х					

Note. Interpretation of notation within table: In the *N* column, (M) signifies the sample consisted of only men (M), (W) signifies only women, and (C) indicates the sample consisted of both members of couples. In Sample column, Ugrad = undergraduate, Grad = graduate, MTurk = Amazon's Mechanical Turk. Sample descriptions with a "?" indicate the type of sample was assumed based on the average age reported in the study. The "X"s indicate the presence of the column's feature in a study. In the Single Target and Multiple Targets columns, "X^{NP}" indicates no photograph of or visual contact with the target was provided and "X^F" indicates the target was presented as a photo of the target's face only. Asterisks (*) in the Interaction w/Target column indicate the participants were told they would be interacting with a target in the future as part of the cover story, but the interaction never occurred, while an "X" in this column indicates the participants did interact with a target. "X¹" in the McNulty et al. (2018) row indicates that an independent sample of singles (*N* = 66) was used to compare to the sample of couples.

Of the included studies, approximately half were published pre-credibility revolution (1989-2011) while the other half were published after awareness of the replication crisis in psychology was prevalent (2013-2018). The median sample size of the included studies was 103 participants. Seventy five percent of studies (n = 27) used undergraduate samples, 13.89% recruited participants online (n = 5), 5.56% used a combination of student and community samples (n = 2), 2.78% sampled community members (n = 1), and 2.78% did not specify the population the sample was recruited from (n = 1).

Only three studies (8.33%) omitted a measure of commitment from their methods, which may suggest researchers generally endorse the assumption that attractive alternatives threaten commitment rather than other relationship-relevant characteristics or constructs (e.g., attachment, satisfaction, self-construal). Commitment was either measured using a multi- or single-item self-report scale (e.g., Rusbult Investment Model – Commitment Subscale; Rusbult et al., 1998), based on relationship status (e.g., married vs. exclusively dating vs. casually dating vs. single), or both. The number of studies that included either one type of commitment measure and those that included both types of measures were relatively equal.

In this literature, commitment was often treated as a quasi-independent variable, with relationship status determining different groups or continuous commitment scale scores being divided into distinct groups (e.g., high, average, low). These two methods of operationalizing commitment are generally known to represent different types of commitment. Commitment measured on a continuous scale represents attitudinal or moral commitment, while commitment measured by relationship status represents structural commitment (Lydon et al., 1999). Measures of attitudinal commitment are beneficial because minor fluctuations in one's state level of commitment are detectable, which is especially useful and appropriate when exposing individuals to potentially threatening stimuli and observing their response. However, when in a relationship, attitudinal commitment is typically high, potentially resulting in ceiling effects, minimal variance, and/or the absence of a group with "low" commitment if scores are dichotomized. While in theory continuous scores allow for more nuanced observations, in practice, the naturally elevated levels of commitment within relationships may impede the detection of practically significant differences. Structural commitment, on the other hand, has been used to compare responses to attractive alternatives between coupled and single individuals,

which was useful for investigations designed to determine whether coupled individuals were devaluing potential alternatives or whether single individuals were enhancing their evaluations of potential partners (i.e., the devaluation vs. enhancement debate; Bazzini & Shaffer, 1999). Comparing coupled individuals to singles also allow for comparisons between objectively distinct groups, compared to dichotomizing continuous measures of commitment. However, operationalizing commitment by relationship status is imprecise (Simpson et al., 1990), as commitment can fluctuate naturally and fall to critically low levels (Johnson & Rusbult, 1989) in circumstances such as impending dissolution. While continuous measures of commitment can detect such detail, structural measures cannot.

Exposure to attractive alternatives has most commonly been operationalized by showing participants a single photograph of an opposite-sex target (50%, n = 18), or multiple photographs of opposite-sex targets (38.89%, n = 14). Out of these manipulations, six authors (16.67%) specified the use of facial photographs, though it may be reasonable to assume that many stimuli consisted of faces only, with possible exceptions being studies that used photos from magazine advertisements (Hadden, 2016; Petit & Ford, 2015; Simpson et al., 1990). Approximately 17% (n = 6) of studies that used photographs of alternatives also supplemented the manipulation of threat by facilitating an interaction between the participants and a potential alternative either online or in-person, and the same number of studies included anticipation of a future meeting with a potential alternative as part of the cover story. Only one study that involved an interaction between the participant and alternative target did not allow the participant to see what the alternative looked like. Finally, the least common method used to manipulate attractive alternative was present or interacted with the participant (2.78%, n = 2).

A notable limitation of the use of photographs to operationalize an attractive alternative is that such stimuli are static and only provide a limited amount of information from which to make evaluative inferences aside from those assessing physical attractiveness. This issue is compounded when only a single stimulus photograph is used because different individuals may have different preferences and opinions on what features they consider to be attractive. Generally, the prevalence of the use of photographic stimuli demonstrates researchers' reliance on using physical attractiveness as both an operationalization of threat (i.e., more attractive

alternatives are more threatening) and as a target of cognitive devaluation (i.e., attractive alternatives will elicit negative evaluations of physical attractiveness). Despite these limitations, use of physical attractiveness in this context is justifiably appropriate since this dimension is easily displayed and noticed compared to other interpersonal attributes (Berscheid & Walster, 1974; Simpson et al., 1990), and is a realistic and relatively easy characteristic to manipulate in experimental settings.

Each of the operationalizations of attractive alternatives in the selected set of studies suffer from one major limitation, which Lee and O'Sullivan (2018, 2019) identified as "fleeting extradyadic attraction." Each of the paradigms involve short-term exposure to potential alternatives and may not be comparable, or only marginally so, to persistent attraction to an extradyadic individual in reality – both in terms of the level of threat that each situation exudes as well as what responses follow. However, attempts at creating circumstances in which fleeting extradyadic attraction exists allows for real-time and immediate evaluations of the targets to be assessed. Observations of devaluation behaviors in the lab may be more feasible compared to observing or measuring such behaviors in reality. Lee and O'Sullivan acknowledge that the ability to capture such evaluations and behaviors during an active period of extradyadic attraction to a real potential alternative partner is difficult, and perhaps due to the difficulty has not yet been attempted. The only study that assessed real-time evaluations of potential alternatives was Overall and Sibley (2008), who measured attraction to opposite-sex alternatives using a longitudinal daily diarystudy design. However, their measures did not focus on a specific or persistent attractive extradyadic target of the participants' attention. Instead, they measured attraction to any nonfamilial opposite-sex others with whom the participants interacted.

The characteristics of the target attractive alternatives were often manipulated to provide varying levels of threat. The attractiveness of the alternatives was manipulated in 27.78% (n = 10) of the selected studies, with the remaining studies using only highly attractive targets (72.22%, n = 26) in their methods. The availability of the target (i.e., whether the target was in a relationship or single) was manipulated in 8.33% of studies (n = 3), as was whether the target expressed interest in the participant. The evaluation perspective of the participant (i.e., whether the participant was evaluating the target from their own perspective or from a distanced perspective) was

manipulated in 2.78% of studies (n = 2). Lastly, the gender of the target (i.e., same- or oppositesex as participant) was manipulated in 13.89% (n = 5) studies.

Johnson and Rusbult (1989) proposed that alternatives that were more highly attractive and who were realistic alternatives (i.e., available) constituted greater threats. However, later research showed availability was not a necessary condition for devaluation to occur (Simpson et al., 1990). Scholars have also asserted that reciprocal interest compounds the level of threat (Bazzini & Shaffer, 1999; Lee & O'Sullivan, 2019; Lydon et al., 1999). As mentioned in PA1, these assumptions have rarely been properly tested (with operationalizations of threat perception as the outcome of interest). Akin to relying on photographs as the operationalization of attractive alternatives (indicating the prominence of the target's physical attractiveness), physical attractiveness was most often manipulated in this selection of studies compared to other relevant characteristics like alternative availability and reciprocal interest. As stated above, physical attractiveness is a feature that individuals heavily rely on to make initial evaluative judgments. While physical attraction is an important element of interpersonal attraction and partner selection, other features like warmth and kindness, status and dominance, wealth, or compatibility may exert comparable or even greater influences on threat perception, attraction, and subsequent devaluation (Lydon et al., 2008, manipulated the warmth/intimacy and power/dominance of an alternative but measured coupled participants' willingness to accommodate rather than devaluation). In context, it seems unlikely that brief exposure to a pretty picture of a stranger would cause intense enough feelings of threat for the relationship maintenance system to activate and respond. Currently, the devaluation literature is characterized by its focus on physical attractiveness which suggests the domain in which devaluation has been observed to occur is particularly limited.

Similarly, devaluation is quite often operationalized by ratings of physical or sexual attractiveness of the target or attraction towards the target alternative (58.33%, n = 21). Other qualities of the target such as intelligence or kindness were employed less than half as often (25%, n = 9). Participants were asked to report whether they would consider the target as a potential partner in 33.33% of the studies (n = 12) and whether they desire to meet the target in 27.78% of the studies (n = 10). Twenty-five percent of studies (n = 9) included a behavioral

measure of devaluation (e.g., mimicry, flirting behavior, memory of targets) and only 5.56% of studies (n = 2) measured physiological responses (brain activation and pupil dilation).

When rating a target's personal qualities, especially physical attractiveness or sex appeal, variations in the wording of items prompting participants to make their evaluations may conflate ratings of attractiveness with attraction *toward* the alternative. For example, "how physically attractive is the target?" may elicit different ratings than "how attracted are you to the target?" (Ma-Kellams et al., 2017). Perhaps it is possible to appreciate another's beauty without feeling personally attracted towards them or feeling like one's perceptions of a current partner or relationship are negatively affected by such thoughts. In addition, soliciting ratings of characteristics, interest in meeting, and consideration of attractive alternatives as potential partners are explicit and conscious cognitive evaluations, however, such items may be influenced by other factors other than perceptions of threat or a desire to protect one's relationship. For example, many authors have noted that rating attractive others highly in a lab setting when relationship status has been reported or is known by the experimenter may motivate responding in a socially desirable way to appear as a loyal romantic partner or an individual of noble character (Bazzini & Shaffer, 1999; Gonzaga et al., 2008). The issue of enhancement versus devaluation emerges here as well, as "objective" ratings of attractiveness or other qualities may be difficult if not impossible to determine.

In terms of the remaining measures, devaluation is often considered to occur both consciously and unconsciously. Physiological measures are employed to detect implicit and automatic responses to attractive alternatives. However, inferences must be made about the meaning of physiological responses, as such responses may have multiple causes. For example, pupil dilation may signal arousal (Laeng & Falkenberg, 2007) or perception of a threat (Bitsios et al., 1996; Henderson et al., 2017), but it may also signal cognitive effort (Kahneman, 2012). Physiological measures offer insight as to early autonomic responses to attractive alternatives but how well these measures reflect the conceptual definition of devaluation may be questionable. Behavioral measures of devaluation suffer from similar limitations of both direct questions and physiological measures. Behaviors such as flirting or smiling in an interaction with an alternative may be indicative of social desirability, politeness, or extroversion rather than interpersonal attraction. However, relatively automatic behaviors such as mimicry are less susceptible to this

limitation. While devaluation has been operationalized in cognitive, behavioral and physiological ways, how well these operational choices accurately reflect devaluation and the broad domain in which devaluation could occur should be carefully considered when making conclusions about their derived results.

The research identified in Table 1 involved manipulations of characteristics of alternative targets, categorizing participants into groups based on commitment level or relationship status, or both. A few studies not included in Table 1 have been conducted in this area that involve self-reporting information from participants' actual experiences. For example, Overall and Sibley (2008) conducted a daily-diary study and asked participants to rate how attracted they were to appropriate individuals who were not their partners. Cross-sectional studies have asked participants to report on their exposure to attractive alternatives on social media (de Lenne et al., 2019) and report on actual experiences of extradyadic attraction and subsequent use of a variety of relationship maintenance mechanisms, including devaluation (Lee & O'Sullivan, 2018, 2019). Interestingly, it seems as if research has only recently turned toward investigating actual use and frequency of use of devaluation behaviors in reality. While it is possible that behaviors used in lab studies may also be used outside of the lab, the opposite is also possible – that exposure to or presence of attractive alternatives in the real world may require different, more intensive, or a variety of maintenance behaviors to counter the supposed threat. Additional self-report studies that solicit information about actual alternatives and participants' responses is a necessary complement to the gratuitous amount of existing experimental and quasi-experimental research in this area.

2.3.2.5.2 Results Review

The majority of research testing devaluation of alternatives (and that is included in Table 1) involves measuring differences in ratings (mostly of physical attractiveness) between different quasi-independent groups or between experimental conditions involving the manipulation of the characteristics of the target alternative(s). Ten studies found participants in relationships reported less attraction to or lower ratings of attractiveness of an alternative target or targets compared to those who were single (Study 1, Bazzini & Shaffer, 1999; Study 2, Johnson & Rusbult, 1989; Karremans & Verwijmeren, 2008; Karremans et al., 2011; McNulty et al., 2018; Miller &

Maner, 2010; Petit & Ford, 2015; Study 2, Rodrigues & Lopes, 2013; Rodrigues et al., 2017; Study 1 and 2, Simpson et al., 1990). Differences between coupled and uncoupled participants were also observed in regard to rating negative behaviors of alternatives (Visserman & Karremans, 2014) and memory of attractive faces (Study 1 and 2, women only, Wang et al., 2016). However, when interest in dating the target alternative(s) or viewing the alternative as a potential partner was measured as a dependent variable, three studies failed to find a difference between coupled and noncoupled participants (Study 1 and 2, Bazzini & Shaffer, 1999; Ritter et al., 2010).

When differences between participants of varying levels of commitment were investigated, those with higher commitment were found to report significantly lower attraction to target alternatives compared to those with lower commitment (Study 3, Johnson & Rusbult, 1989; Lydon, 2003). Similar results were only found in one of three of Hadden's (2016) studies. Breaking down the relationship between commitment and threat further, Lydon et al. observed lower ratings of the alternative when the level of threat matched the participants' level of commitment for moderate threat/commitment (1999, 2003) and high threat/commitment (1999).

When the availability of the alternative was manipulated (as opposed to attractiveness), Johnson and Rusbult (1989, Study 2) found inconsistent evidence for differences in perceptions of the alternative between coupled and noncoupled participants. Cole et al. (2016) found that coupled participants rated alternatives as less attractive when they were revealed to be available (Study 1) and interested in dating (Study 2) compared to alternatives that were not available or interested in dating, while no differences were found for single participants. In contrast, Karremans and Verwijmeren (Study 3, 2008) found that the availability of the alternative did not significantly affect ratings of attraction to the alternative.

Aside from self-report measures, across three studies Karremans and Verwijmeren (2008) found that coupled participants mimicked an attractive alternative less than single participants. However, when using the physiological measure of pupil dilation, no differences were observed after exposure to attractive alternatives (Petit & Ford, 2015).

Some research explored the influence of measures of self-regulation and executive control on evaluation of alternatives. When involved participants had available cognitive resources, they

tended to judge potential alternatives less often as potential partners (Ritter et al., 2010) and reduced their desire to meet an attractive other (Pronk et al., 2011). Meyer et al. (2011) found results in a similar pattern, with coupled participants selecting more targets as potential partners under time pressure compared to no time pressure, but the effect did not reach statistical significance.

Of the research not included in Table 1 that involved reports from participants' real-life experiences, Lee and O'Sullivan found that relationship maintenance strategies involving devaluation of alternatives were reportedly used by the majority of two samples of participants who reported experiencing extradyadic attraction (2018, 2019). However, only two devaluation strategies were included in a larger 7-item factor that involved self-regulation, so the frequency with which devaluation alone was used cannot be specifically reported.

2.3.2.5.3 Evaluation

Returning to the evaluation criteria, Lakatos's principle suggests a theory has high verisimilitude if it has passed numerous strict and diverse tests. To assess methodological diversity, we can compare the conceptual definitions identified during the deconstruction process to the operationalizations used in this literature. Alternatives were conceptually defined as involvement with another person or persons or being single. However, all studies that have tested devaluation have operationalized alternatives as being an alternative partner. Of the operationalizations of alternative partners in this literature, the majority used photographs. As mentioned above, such operationalizations, even when accompanied with a cover story or promise of a future interaction, may only be weak operationalizations of alternatives and of threat. While a handful of studies did involve an interaction with an alternative, the alternative targets were always strangers and always simulated a situation of fleeting extradyadic attraction (if attraction was experienced at all). In other words, exposure to alternatives rarely included either an actual, known alternative to the participants, or an alternative to whom a participant was attracted to long-term (Lee and O'Sullivan, 2018 and 2019 being the exceptions). Regarding operationalizations of attractiveness, a clear preference exists for physical attractiveness and/or sex appeal despite the original conceptual definition including multiple and more diverse characteristics, such as personality, intelligence, shared interests, and similarity of values to

name a few. This represents another example of scholars employing only a narrow test of a much broader concept and indicates a clear lack of operational diversity. Relatedly, devaluation has often been measured using ratings of physical attractiveness or attraction to the target alternative(s). While this measure is appropriate when physical attractiveness is either what is being manipulated or what is considered the threatening aspect of a target alternative, the conceptual definition of devaluation is again, much broader. Devaluation was defined as emphasizing the negative cost increasing aspects of a possible interaction or engagement with the target, or by diminishing the value of the target person. Ratings of attraction fall under this latter part of the devaluation definition; however, no efforts have been made to measure whether individuals emphasize negative aspects of an interaction or relationship with the target. As with the other concepts, opportunity exists to expand measures of devaluation.

In terms of the strength and robustness of the results, when comparisons were made between coupled and single individuals, differences in ratings of attractiveness were quite clear. However, no such differences were found with more approach-oriented dependent variables such as interest in dating the target or viewing the target as a potential partner. When participants were divided based on high versus moderate/low commitment, less consistent evidence was found for differences in ratings of attraction toward the targets. When alternative characteristics other than attractiveness were manipulated (e.g., availability), inconsistent evidence was found for differences between coupled and single individuals. Similarly mixed results were found when behavioral and physiological measures were used.

The literature testing this assumption is lacking diversity because alternatives were most often operationalized using only photographs and attractiveness and devaluation were narrowly operationalized in terms of physical attractiveness. In terms of strictness, while no point predictions were made in the entire body of reviewed literature, researchers did often specify which groups were expected to differ significantly from one another, especially when 2 x 2 designs were utilized. In addition, while quantitatively more predictions were supported rather than refuted, refutations carry more weight than do numerous accepted predictions that are only directional in nature (Meehl, 1990). It is therefore difficult to argue that this literature has passed strict tests. The Lakatos principle for this assumption is not satisfied. Consequently, the lack of sufficient diversity suggests that the Salmon principle is also not satisfied.

2.3.2.6 CA4: Relationship Maintenance Mechanisms Are Enacted in Attempt to Neutralize External Threats

Though the devaluation of alternatives literature is relatively small, the focus of most research in this area involves measuring evaluations of alternatives to detect devaluation. However, there is a surprising dearth of research designed to assess the efficacy of devaluation. The purpose of relationship maintenance processes activated in response to threat is to reduce or eliminate the source of the threat, but there is very limited evidence that supports the assumption that devaluation is effective in neutralizing the threat of attractive alternatives and protecting one's relationship.

When interdependence and investment theory were increasing in popularity, cross-sectional and longitudinal studies were conducted that showed one's quality of alternatives (Rusbult, 1980, 1983) was associated with commitment and relationship longevity. Around the same time, theories of relationship maintenance were in early stages, with the research focus revolving prominently around the identification of various relationship maintenance processes and conditions under which the system is activated, rather than the efficacy of specific maintenance mechanisms. To my knowledge, the first attempt at investigating the effectiveness of devaluation of alternatives did not occur until relatively recently, in 2018.

McNulty et al. (2018) conducted a longitudinal study over three and a half years with 224 newlyweds (113 couples). Participants completed baseline measures of commitment and infidelity (among others) before a laboratory session in which they completed a devaluation of alternatives task where they rated the attractiveness of same- and opposite-sex photographs of highly attractive and average-looking targets. At each six-month follow up, participants reported their commitment, infidelity, and whether their relationship had dissolved. Preliminary analyses revealed that compared to a sample of single individuals (N = 66), the newlyweds rated the highly attractive targets as less attractive than the single participants. Interestingly, McNulty et al. (2018) included both same-sex and opposite-sex targets in the dependent measure⁹ and only

⁹ McNulty et al. (2018) reported exposing participants to four photos from each of the following categories: highly attractive men, highly attractive women, average-looking women, and average-looking men. Then they used "eight

reported inferential statistics divided by gender. Devaluation was not significantly associated with relationship dissolution, martial satisfaction, changes in commitment over time and commitment at the end of the study. Devaluation did emerge as significantly associated with infidelity, such that participants who devalued alternatives less at the initial lab session were more likely to engage in infidelity at some point during the study. A robustness test indicated this significant association disappeared when the presence of infidelity agreed upon by both partners (i.e., Partner A and Partner B both reported Partner B's unfaithfulness) was used as a dependent variable. In comparison to the single participants, faithful partners were found to evaluate attractive alternatives as significantly less attractive than singles, while unfaithful partners had comparable ratings as single participants.

The inclusion of same- and opposite- targets in the analysis in this research, however, is atypical in the devaluation literature where measurement of devaluation towards exclusively opposite-sex targets is the norm. Newlyweds may be inclined to devalue same-sex targets because they are also viewed as threats, specifically as potential mate poachers. However, this explanation conflicts with past research which has shown that same-sex targets are not often subject to devaluation (Hadden, 2016; Simpson et al., 1990). Instead, other maintenance mechanisms are often employed in potential mate-poaching situations, such as increasing investment in the poachable partner, increasing vigilance, and increasing signals of possessiveness (Lee & O'Sullivan, 2018).

The unusual inclusion of same- and opposite-sex targets in the devaluation measure coupled with the ambiguity around whether same- and opposite-sex targets or only opposite-sex targets were used in the analyses with relationship outcomes necessitates interpreting these results with caution. However, if the reported results are representative of reality, it would seem that devaluation is not strongly associated with commitment or relationship longevity. Devaluation may be more strongly associated with the likelihood of infidelity, which influences the

attractive photos to create an index of devaluation" (p. 80), suggesting they combined the attractiveness ratings of highly attractive men and women targets in their measure of devaluation. However, when reporting the results of the analyses involving longitudinal data, the authors report "participants' ratings of the attractive, opposite sex targets" (p. 82). It is unclear why the authors would report analyses with different operationalizations of devaluation (first with highly attractive same- and opposite-sex targets, and second with highly attractive opposite-sex targets only).

probability of relationship dissolution in a stronger, more consistent manner (see also Betzig, 1989; Fincham & May, 2017). In addition, as with most of the manipulations of exposure to an attractive alternative in the literature, McNulty et al. used photos of attractive individuals rather than an actual person (e.g., actor or fellow participant) or a real person known to the participant. Photos alone, especially when unaccompanied by a cover story providing details about the existence and characteristics of the target (e.g., their relationship status, interest in the participant, etc.) may not be seen as realistic threats.

Attempting to better assess the association of realistic threats with devaluation tendencies, Lee and O'Sullivan (2018) investigated which relationship maintenance mechanisms were reportedly used during an actual period of extradyadic attraction in reality. When faced with attraction to a target outside one's relationship, participants used three types of strategies: Proactive Avoidance, Relationship Enhancement, and Low Self-Monitoring and Derogation¹⁰. The researchers included seven strategies in the Low Self-Monitoring and Derogation factor, with two items describing devaluative behaviors explicitly ("Told myself that this other person was bad for me", "Looked for unflattering things in this other person"). Explorations of the use of maintenance strategies between participants who experienced adverse outcomes such as extradyadic contact or relationship dissolution and those who did not only revealed a significant difference in use of Low Self-Monitoring and Derogation strategies, such that those who experienced adverse outcomes reported lower self-monitoring and derogation strategies than those who did not. Specifically, this pattern held for participants who flirted with their extradyadic target than those who did not flirt, but no differences were found for other specific extradyadic behaviors (romantic infidelity, sexual infidelity, relationship termination; replicated in Study 2). Direct logistic regression indicated the three-factor model could correctly classify whether participants engaged in adverse outcomes 67% of the time, with Low Self-Monitoring and Derogation emerging as the strongest predictor of adverse outcomes, such that those who demonstrated lower self-monitoring and derogation were 1.5 times more likely to engage in extradyadic behavior (replicated in Study 2).

¹⁰ This factor, unlike the others, had negative loadings. This factor indicates a lack of effort to emotionally or cognitively manage attraction to an extradyadic other (Lee & O'Sullivan, 2018).

Akin to McNulty's et al. (2018) findings, Lee and O'Sullivan (2018) found that a lack of devaluation was associated with flirtatious behavior, a potential precursor to infidelity. While McNulty et al. found that less devaluation was associated with future infidelity, Lee and O'Sullivan did not find such an association for romantic or sexual infidelity, nor relationship dissolution. In fact, none of the three factors of relationship maintenance mechanisms were significantly associated with infidelity or relationship dissolution, which may indicate contradictory evidence to the assumption that such mechanisms are effective in neutralizing threats to romantic relationships (Lee & O'Sullivan, 2018). Regarding the effectiveness of devaluation specifically, the Low Self-Monitoring and Derogation factor only contained two items that could confidently be said to describe devaluation behaviors, while the remaining five items were more representative of self-monitoring. Caution should be taken when discussing these findings in relation to the efficacy of devaluation alone, as the independent influence of devaluation in the absence of other self-monitoring behaviors cannot be established based on the reported results. Nonetheless, Lee and O'Sullivan's initial foray into describing what maintenance strategies are commonly enacted when faced with a threatening attractive alternative in reality, as well as their associations with adverse relationship outcomes, represent a promising avenue for future research.

Lee and O'Sullivan (2019) further investigated the efficacy of relationship maintenance mechanisms in a later study. Participants recruited from MTurk reported their commitment, engagement in infidelity, and relationship maintenance mechanism use during periods of extradyadic attraction in a preliminary survey (N = 287) and 2-month follow-up survey (N = 131). Approximately 60% and 50% of participants reported experiencing extradyadic attraction, with 80% and 85% reporting use of Self-Monitoring and Derogation maintenance strategies in the first survey and follow-up survey, respectively. Structural equation modeling revealed that use of relationship maintenance mechanisms reported in the initial survey did not significantly predict romantic or sexual infidelity at the 2-month follow-up. Lastly, explorations of the influence of reciprocal attraction revealed a significant difference such that participants who felt reciprocal attraction to their alternative reported higher use of Self-Monitoring and Derogation strategies than did participants who did not reciprocate attraction to their alternative.

Although a similar pattern of results was found in both Lee and Sullivan's (2018, 2019) studies it is notable that use of maintenance mechanisms in both studies were reported in a dichotomous fashion – participants either used a strategy or not. This measurement method is limited in that the frequency or consistency of use of maintenance strategies remains unknown (Lee & O'Sullivan, 2019). The association between relationship maintenance mechanisms and adverse relationship outcomes may emerge if use is measured in a more specific and detailed way, or if use is measured during the period of extradyadic attraction. This method did reveal that those experiencing reciprocated attraction utilized more maintenance strategies, suggesting greater efforts to protect one's relationship from an extradyadic threat. However, the effectiveness of the quantity of maintenance strategies was negligible.

The efficacy of devaluation strategy in neutralizing external threats is therefore not wellestablished. The limited research that has explored the efficacy of this maintenance mechanism seems to contain more convincing evidence that devaluation is *not* effective in neutralizing the threat of attractive alternatives. But there exists some notable methodological weaknesses that limit the ability to confidently and comprehensively determine the role devaluation plays in protecting relationships. For instance, given that there are a limited number of studies testing this assumption, and both of Lee and O'Sullivan's (2018, 2019) studies are more descriptive than predictive in nature, it is premature to make determinations about the extent to which the Lakatos and Salmon principles are satisfied. More research using diverse measurement methods and study designs will be essential going forward to clarify the validity of this core assumption.

2.3.2.7 PA3: Commitment Must Be Sufficiently High for Relationship Maintenance Mechanisms to Be Activated

The final peripheral assumption of devaluation of alternatives is that devaluation is not elicited unless commitment is sufficiently high. In other words, an individual must feel their relationship is worthy of protection for protective relationship maintenance mechanisms to be activated (Johnson & Rusbult, 1989).

Of the studies in the literature that have only used attractiveness as their operationalization of threat, weak evidence exists for the association between commitment and alternative

devaluation. Across two studies exploring the influence of self-regulatory resources on evaluations of alternatives, Ritter et al. (2010) found no significant associations between commitment and consideration of attractive others as potential partners (though in Study 1 the association was negative) in either the depletion or control conditions. Similarly, Rodrigues et al. (2017) explored associations between commitment, regulatory focus and attraction to a single alternative target. Across two studies, commitment was not found to significantly influence ratings of attraction to the target. Hadden (2016) explored commitment and ratings of physical and sexual attractiveness and desire to meet attractive alternatives (Study 1 used facial photographs, Study 2 and 3 used magazine ads). Commitment was only found to be significantly negatively associated with ratings of attractiveness in Study 3, which involved an MTurk sample compared to undergraduate samples used in the other studies. Lydon et al. (2003) found commitment was significantly negatively associated with attraction to an alternative only in a moderate-threat condition (where the alternative was described as single and a future interaction with the participant was expected) compared to a low-threat (the alternative was described as single and a future interaction was not expected) and an unavailable condition (the alternative was described as in a relationship and a future interaction was expected). Using a physiological measure, Miller et al. (2012) explored the association of commitment with brain activation in regions associated with approach or avoidance motivations after exposure to photos of three attractive alternatives. Participants with moderate levels of commitment expressed marginally significantly greater activation in brain areas associated with avoidance than participants with higher or lower commitment. Of these studies, few statistically significant associations exist between commitment and evaluation of alternatives, casting doubt over the claim that commitment must be necessarily high for this behavior to occur.

In the only study that involved participants reflecting on their own experiences of extradyadic attraction, Lee and O'Sullivan (2019) asked participants to identify which actions were taken to maintain their relationships with their partners. A significant positive correlation was found between commitment and use of self-monitoring and derogation strategies (though only two of seven items of the factor are representative of derogation). Interestingly, two other factors that embodied different types of relationship maintenance mechanisms were either negatively related with commitment (relationship enhancement strategies) or not related to commitment (proactive avoidance). These findings, like the others, represent weak evidence in support of the assumption

that commitment must be sufficiently high for relationship maintenance mechanisms to be activated.

Lydon et al. (1999) modified this proposition further and dubbed it the calibration hypothesis suggesting that the relationship maintenance system becomes activated only when the caliber of threat is equal to the caliber of commitment (i.e., commitment does not necessarily need to be high, but it needs to be on par with the potential threat). When a potential threat is lower than one's level of commitment, it will not be perceived as a threat and therefore no protective or corrective action is needed. If the potential threat exceeds one's level of commitment, relationship maintenance mechanisms may be used less frequently or not at all due to the individual being insufficiently motivated to act to protect their relationship.

To test the calibration hypothesis, Lydon et al. (1999, 2003) used two operationalizations of different levels of commitment and threat to determine whether devaluation occurred when these characteristics were comparatively equal. In the first study, commitment was categorized as high or low based on the combination of responses to both an attitudinal measure of commitment and the participants' relationship statuses. Participants considered to be low on commitment had low attitudinal scores and were unmarried, those considered to be moderately committed had a high affiliation on only one commitment measure, and those considered to be highly committed had high scores on attitudinal commitment and were married. In the second study, a measure of attitudinal commitment appears to have been dichotomized, but the cut-off value used for distinguishing between moderate and low commitment was not explicitly stated. The target alternative's interest in the participant was manipulated in the first study and the likelihood of a future interaction with an alternative who was indicated to be single was manipulated in the second study. Devaluation was observed with highly committed participants under high threat conditions only in the first study, and with moderately committed participants under moderate threat conditions in both studies, compared to conditions with discordant levels of commitment and threat exposure indicating support for the calibration hypothesis.

Rodrigues and Lopes (2013) also claimed to test the calibration hypothesis, however, the authors considered high threat to be exposure to an attractive alternative (in contrast to Lydon et al., 1999, who would consider this operationalization to be only moderately threatening) and did not

have a second manipulation for moderate or low threat. Scores on a commitment measure were median split, with scores less than or equal to 6.5 considered low, and those above 6.5 considered high (on a scale of 1 - 7). Results indicated highly committed participants reported less attraction to an attractive target compared to those with lower commitment. The discrepancy between the level of threat posed by the stimuli and the ceiling effects of the commitment measure limit the strength of support these results provide for the calibration hypothesis.

Despite the respectable amount of research that has examined this assumption, there is not consistent evidence for the assertion that commitment must be sufficiently high for relationship maintenance mechanisms, specifically devaluation, to be engaged. In general, Lakatos's principle of passing strict and diverse tests is not satisfied. However, Lydon et al.'s (1999, 2003) proposition of the calibration hypothesis and their initial studies that showed support for the hypothesis represent a solid demonstration of Salmon's principle of a "damn strange coincidence." The predictions made by the calibration hypothesis are specific and were supported in preliminary tests. However, more research in general, and research using more diverse operationalizations of commitment and threat, are needed to demonstrate more convincing support for this hypothesis.

In part due to inconsistent evidence, this assumption remains in the periphery of the theory. As argued when discussing CA1, the essential element needed to stimulate relationship maintenance mechanisms is the perception of an external threat. The level of commitment of an individual is less relevant than the detection of threat itself. Threat perception and commitment may be related, as this assumption suggests. While the calibration hypothesis represents an additional layer of complexity, adding nuance to our understanding of the theory, the essential phenomenon of devaluation and its underlying mechanisms would remain in the absence of any calibration between threat and commitment. That said, with more research, development of the theory, and conversations between interested scholars this assumption may transition to the hard core in the future.

2.4 Discussion

Devaluation of alternatives is a phenomenon touted as well-documented and established (Meyer et al., 2011). However, many empirical tests and discussions of devaluation are based on a verbal model of the theory. Reliance on a verbal model instead of a formal model makes it difficult to identify limitations in our understanding and gaps in the literature, as well as design future research to challenge and refine the theory. In this chapter, adapting the metatheory methodology of Quistberg and Sakaluk (2021), the theory of devaluation of alternatives was formalized through the processes of categorization, deconstruction, and reconstruction through theory mapping (Gray, 2017). The information gained from the formalization process was compared to and complemented by the identification, description, and evaluation of the assumptions of the theory according to Lakatos's and Salmon's principles (Meehl, 1990).

First, the theory of devaluation of alternatives was categorized and described as it was first presented in Johnson and Rusbult (1989). This process identified the original state of the theory, providing a comparison point for subsequent iterations. The examination of evidence of the first core assumption, that external stressors (such as attractive alternatives) are threatening revealed that little efforts have been made to find support for this claim. Consequently, whether more attractive or available alternatives cause more intense feelings of threat are only supported by theoretically based logic. Second, Johnson and Rusbult suggested that commitment must be sufficiently high for relationship maintenance mechanisms such as devaluation to be elicited. Again, targeted examination of this specific assumption revealed inconsistent evidence for requisite high levels of commitment (operationalized by correlations between commitment and observations of devaluative behavior). Both of these original assertions, that more attractive and available alternatives are more threatening, and that commitment must be high for devaluation to be observed, were both relegated to the protective belt of the theory. They also mentioned, quite explicitly in their introduction, that devaluation was not the only strategy that could be used to counter the threat of attractive alternatives. This represented the first indication that devaluation was not its own theory, but rather a component of a larger theory of relationship maintenance which developed over time.

During the deconstruction process, original conceptual definitions of the key constructs were extracted from Johnson and Rusbult's (1989) seminal article, revealing some interesting gaps in the existing literature. First, devaluation was defined by having a "sour grapes" attitude toward potentially rewarding aspects of an alternative relationship and by emphasizing negative, cost-increasing aspects of a potential relationship with an extradyadic partner. While evaluating a potential alternative less positively or attractively compared to how they "objectively" may be falls under the "sour grapes" conceptual umbrella, the examination of the literature revealed no efforts to measure devaluation by measuring the extent to which individuals emphasize the cost-increasing aspects of an alternative relationship. In a similar vein, relationship "alternatives" were originally defined as being with another person, persons, or being single. Again, a review of existing research found that alternatives were operationalized, almost exclusively, as a single alternative partner. The only exceptions involved studies with measures of alternative quality, which often only explored quality of alternatives in relation to relationship commitment or longevity rather than devaluation behaviors. The third example of a preference for a certain aspect of a conceptual definition of a key construct involved the qualifier of "attractive" to describe alternatives. While early research such as Johnson and Rusbult (1989), Simpson et al., (1990) and Lydon (1999) involved evaluations of characteristics other than physical attractiveness, most of the devaluation literature conceptually and operationally considered "attractive alternatives" to be those which displayed high levels of physical attractiveness *only*. Very few empirical investigations involved participants evaluating characteristics of a target alternative that may be especially relevant when engaging in cost-benefit analyses or debating whether to leave one's current partner. The process of simply comparing the original definitions to definitions used in the literature exposed major limitations in how research involving this phenomenon has been approached. This highlighted the narrow scope in which devaluation has been studied in the past 30 years and revealed a variety of areas for future research.

One of the most prominent discoveries that emerged during the formalization process, especially deconstruction and reconstruction, was that devaluation of alternatives can be situated within the larger theory of relationship maintenance. While Johnson and Rusbult (1989) acknowledged that devaluation was only one of many potential strategies that could be used to counter the threat of attractive alternatives, so too are attractive alternatives only one possible source of external threat. Consequently, devaluation of alternatives can be conceptualized as a specific sub-theory of a theory of relationship maintenance in response to external threat (with Agnew and VanderDrift, 2015, proposing two other sources that elicit relationship maintenance behaviors). Notably, the theory of relationship maintenance has also yet to be formalized. What started out as an effort to formalize the theory of devaluation of alternatives turned into the formalization of a broader theory which should prove useful for investigations of other combinations of external threats and maintenance strategies, for example, an attractive alternative for one's partner (i.e., a potential rival) and the use of devaluation or increased investment. Interestingly, commitment was proposed to play a key role in devaluation of alternatives, but the role of commitment was not as explicitly defined in the relationship maintenance literature more broadly. By mapping the specific sub-theory of devaluation of alternatives alongside the more general map of relationship maintenance in response to external threats, a plausible auxiliary assumption was added to the larger model.

Finally, the last process in the formalization process was evaluation, which involved a description of the evolution of the theory and an evaluation of each of the identified assumptions. Tracing the evolution of the theory revealed debates in the literature (devaluation vs. enhancement, conscious vs. unconscious), points in time when auxiliary assumptions were proposed (e.g., the calibration hypothesis), and forays into related areas (e.g., inattention towards alternatives). The assumptions were evaluated using two criteria, the Lakatos and Salmon principles (Meehl, 1990). The first core assumption, that external stressors are threatening, failed to satisfy both principles when research specifically investigating the threatening nature of attractive alternatives was reviewed. Consequently, the first peripheral assumption, that more intense stressors are more threatening was lacking empirical evidence and thus failed both principles as well. Similarly, the second peripheral assumption posited that commitment is specifically targeted by external threats. While theoretically based logic supports the idea, empirical research is insufficient to support this claim and many plausible targets of threat exist that

have not yet been ruled out. As such, this peripheral assumption was also concluded to be lacking verisimilitude.

The second core assumption, that an underlying relationship maintenance system is activated in response to threat is unable to be directly tested, and thus can only be assessed through the third core assumption, which asserts that relationship maintenance mechanisms can be observed following the activation of such a system. With the focus purely on devaluation as a maintenance strategy, there was a large amount of research reporting observations of devaluation. However, most of the literature used undergraduate samples and most relied on the brief exposure to photographs of attractive targets as an operationalization of potential attractive alternatives. Such manipulations, however, represent only fleeting extradyadic attractions (Lee & O'Sullivan, 2018, 2019) and the extent to which such experiments indicate reality is questionable at best. In line with the observations that emerged from deconstruction in terms of how key constructs are operationalized, much has yet to be discovered when it comes to the use of devaluation once relationship maintenance mechanisms are stimulated. Thus, this core assumption was also found to be lacking in verisimilitude.

The fourth core assumption, that devaluation is effective in neutralizing the threat of attractive alternatives, also was shown to be lacking substantial evidence. Due to the small number of studies investigating this assertion, conclusions of the verisimilitude of this assumption are premature. Lastly, the peripheral assumption that commitment must be sufficiently high was an interesting case. While an admirable amount of research examined whether commitment was associated with devaluation, inconsistent results were found. However, Lydon et al.'s (1999, 2003) calibration hypothesis came closest to satisfying Salmon's principle of a "damn strange coincidence," since the derived hypotheses were specific and supported across multiple studies. Overall, examination of the assumptions independently proved most informative indeed.

The adopted approach to theory formalization used in the present research is notably only one of many possible approaches. While the processes of categorization and deconstruction are more straightforward, in that different researchers embarking on these

tasks would likely emerge with highly similar outcomes, the process of reconstruction is more subjective and variable in employment. For example, theory mapping is only one possible technique that could be used to reconstruct a theory, and this approach has limitations. Theory maps, when not supplemented with clearly identified assumptions or descriptions of the theory may be vague and appear overly simplified. Further, theory maps may be confused or conflated with statistical models when their purpose is purely to visualize a theory "at a glance." But, this technique is helpful in comparing different conceptualizations of a theory, or, in the present case, illustrating how devaluation of alternatives was a sub-theory derived from a similar, but more general theory of relationship maintenance elicited by external threats. Relatedly, there is also not a single agreed upon method that is most appropriate to use to evaluate the state of a theory (Quistberg & Sakaluk, 2021). While I elected to use the Lakatos and Salmon principles, it is possible that different conclusions about the current quality of devaluation of alternatives may be made if different evaluative criteria were applied.

That being said, according to Lakatos's principle theories should only continue to be refined if they have a successful record of surviving numerous strict and diverse tests. In addition, Salmon's principle suggests that theories must be capable of predicting facts that would be improbable without the theory. From the current evaluation of the theoretical assumptions, it seems that the theory of devaluation of alternatives fails to sufficiently satisfy either of these principles and may be appropriately qualified as a degenerative or at the very least, stagnant, research program. However, the evidence does suggest that devaluation is likely a behavior or tendency that exists (CA 3). What is needed then, is a transformation of how scholars approach researching this phenomenon. Specifically, methodological approaches need to be diversified, operational definitions of key constructs need to be expanded, efforts need to be invested in corroborating the underlying assumptions, and scholars need to be mindful of existing research to ensure their new empirical pursuits appropriately and effectively build off what has already been done. These suggestions represent what research programs might look like when solid methodology is married with strong theory, and when the motivator and overall goal of psychological research is the pursuit of the truth.

This chapter involved the formalization and conceptual evaluation of the theory of devaluation of alternatives. The next chapter describes a systematic approach to quantifying and interpreting the extent to which evidentiary value is demonstrated in the existing published research on devaluation of alternatives and whether the literature contains evidence of selective reporting and/or questionable research practices.

Chapter 3

3 Evaluating Devaluation of Alternatives with *P*-Curve Meta-Analyses

In this chapter, a meta-analytic method that uses *p*-values as data points was used to evaluate the strength of the evidence of devaluation of alternatives and investigate whether any indications of selective reporting or questionable research practices were present in the literature.

3.1 Introduction to *P*-Curve Analysis

P-curve meta-analyses are beneficial to conduct because the publishing process is vulnerable to and afflicted by biases, the most prominent being a preference for statistically significant results. Consequently, published literature that is included in meta-analyses, systematic reviews, and in introductions to empirical articles may be similarly biased. With this knowledge, researchers must be cautious when interpreting the strength of a body of evidence for a hypothesized effect, and certainly cannot base such judgements on the quantity of evidence (e.g., number of published studies), because statistically significant results in the published literature may be due to Type I errors or questionable research practices (QRPs), rather than representations of true effects in the population. Selective reporting and publication bias describe the tendencies for researchers, reviewers, and journal editors to prefer submitting and publishing statistically significant results (*p*-values < .05) while storing null results in the "file drawer" (Rosenthal, 1979; Sterling, 1959).

The first tangible evidence of publication bias emerged in 1959, when Sterling used simple counts of articles within four of the top psychology journals and found that 95-99% involved a rejection of the null hypothesis and none involved replication studies. From these data Sterling concluded that statistically significant results were more likely to be published compared to null results, that replications are uncommon (perhaps because a single published statistically significant result acted as sufficient evidence for a true effect), and due to the clear pervasiveness of positive results, that more experiments are performed than are published (Sterling, 1959). More recently, a review of over 2,000 published articled across 20 academic disciplines (spanning the hard sciences and social sciences) revealed 84% of articles included results that

demonstrated statistically significant support for the alternative hypothesis over the null (Fanelli, 2010). Of the disciplines included in the review, psychology and psychiatry articles reported the highest frequency of statistically significant results (91.5%), and social science disciplines were found to be 2.3 times as likely to report statistically significant results compared to disciplines in the physical sciences (Fanelli, 2010). A follow up review that almost doubled the sample of articles found similar results, such that social science disciplines showed more prominent evidence of a preference for publishing statistically significant results (Fanelli, 2012). Consequently, when using meta-analytic approaches, it is important to use methods that identify and estimate the prevalence of publication bias. These methods offer more realistic approximations of strength and quality of evidence because larger systemic practices that influence what is "allowed" in the published literature are accounted for rather than ignored.

Selective reporting and publication bias are two sides of the same coin, with selective reporting describing researchers' own biases for submitting statistically significant results for publication (though arguably this is both cause and consequence of publication bias). Evidence of selective reporting also emerged early in the field's history. In a survey of 36 authors and 39 reviewers of the Journal of Personality and Social Psychology in 1973, results revealed only a 6% probability of submitting null results for publication. Perhaps even more concerning, there was a 28% probability that researchers would not conduct further studies regarding potential effects that did not emerge as statistically significant (Greenwald, 1975). In a recent sample of 221 grant funded studies, 80% had been written up with intent to publish or present at a conference. Of the 48% of studies that were published, only 10% reported null results. Of the 20% that were not written up, 65% included null results. When the grant awardees were asked why some results were not written up, they mentioned anticipating that null results would not be published, loss of interest in the null finding, and a result being published after statistical significance was reached in a different sample (Franco et al., 2014). Relatedly, O'Boyle et al. (2014) compared the number of supported, unsupported, added, and dropped hypotheses from original dissertations on topics relevant to management to their published article counterparts. Of ~2,000 dissertation hypotheses, 45% were statistically significant, while 66% of the ~1,000 published hypotheses were statistically significant. Dissertation hypotheses that were dropped from the published articles were 1.5 times as likely to be nonsignificant than significant, and hypotheses that were added to journal articles (i.e., hypotheses not present in the dissertations) were twice as likely to

be statistically significant than hypotheses that were dropped. O'Boyle et al. suggested such discrepancies likely resulted from QRP's such as selective reporting, HARKing (hypothesizing after results are known; Kerr, 1998), and *p*-hacking (e.g., dropping or combining experimental conditions, excluding outliers, analyzing five dependent variables and only reporting the two that "worked," etc.; Simmons et al., 2011). These findings offer direct evidence to support Sterling's (1959) conclusion that more experiments are performed than are published, indicating that selective reporting and publication bias are influential factors that need to be considered to properly estimate the strength of the published evidence for a certain effect when using meta-analytic methods.

Recently, Simonsohn et al. (2014) developed an approach called *p*-curve analysis to evaluate the strength of evidentiary support for specific effects and assess the extent to which results within a body of literature are indicative of selective reporting and/or publication bias. Even if the amount of evidence that purports the existence of a true effect is sizeable, the strength of the evidence may be weak because of the prevalence of a QRP called *p*-hacking, where researchers make numerous choices during data analysis that eventually result in a *p*-value below .05. Much of the empirical literature investigating devaluation of alternatives was conducted prior to the credibility revolution or replication crisis, which began around 2011. At this time, fewer practices and safeguards existed within academia to discourage or prevent QRPs. Therefore, the devaluation of alternatives literature may be at risk of being affected by *p*-hacking, selective reporting, and publication bias. To complement the theoretical evaluation of derogation of alternatives in Chapter 2, I used *p*-curve analyses to determine whether there was evidence of selective reporting in the published literature that involved empirical tests of the theory.

3.2 Methods

Following Simonsohn's et al. (2014) instructions, I created a set of rules to select studies to include in the *p*-curve analyses. Using Western's academic institution search engine (Omni), I searched for articles in peer-reviewed journals published between 1980 and 2020 with Psychology as the subject. With these criteria selected, I conducted two keyword searches: first, a search with "Derogation of Alternatives" in any field, and "Attractive" in any field, then "Attractive Alternatives" in any field and "Commitment" in any field. Of these two searches, I only selected articles from the first 100 results.

Using psychology as a selection criterion was required because the theory of devaluation of alternatives was proposed in the context of close, romantic relationships, and the focus of this dissertation is on devaluation of alternatives in this context. The date range was selected because devaluation of alternatives was proposed in 1989 and the *p*-curve analyses were conducted in early 2021. Including a cut off for the searches (i.e., only selecting from the first 100 studies that emerge from the search) was intended to make study selection manageable and less time-consuming for researchers who may attempt to reproduce these searches. Lastly, the search terms "derogation of alternatives", "attractive", "attractive alternatives", and "commitment" are keywords and essential components to the theory of devaluation of alternatives and would be very likely to appear in the publications of studies that include a discussion or investigation of devaluation of alternatives. Based on these criteria, 38 relevant articles were identified across both searches.

Once relevant articles were selected, I followed Simonsohn's et al. (2014) instructions to select eligible *p*-values from the articles. To be included in a *p*-curve analysis, the selected *p*-values must test the hypothesis of interest, have a uniform distribution under the null, and be independent of the other *p*-values included in the *p*-curve analysis. Further, reported *p*-values must be less than 0.05. Since many articles consist of multiple studies or multiple analyses that test a central hypothesis, I decided in advance to select the *p*-values that corresponded to the first hypothesis that seemed relevant to devaluation of alternatives (e.g., pilot studies or initial hypotheses that were manipulation checks were not included). If the first relevant hypothesis was not accompanied by reported results that met the eligibility criteria for *p*-value inclusion, the article was not included in the *p*-curve.¹¹

Of the 38 articles, one did not include specific, identifiable hypotheses (Rusbult & Buunk, 1993), two included non-parametric tests or descriptive statistics only (Hui & Molden, 2014; Lee &

¹¹ Prior to selecting appropriate p-values, I had not considered which p-value to include should a hypothesis be tested using more than one dependent variable, and more specifically, whether I should include an article that reported a p-value that was not significant for the first dependent variable but significant for the second reported dependent variable. Consequently, I treated these as ambiguous cases and ran the p-curve analyses with and without their respective p-values. Notable differences in the main findings of the p-curve analyses are identified where necessary.

O'Sullivan, 2018), five did not include a test statistic for determining the significance of a regression coefficient (Birnbaum et al., 2019; Johnson & Rusbult, 1989; Miller & Maner, 2010; Miller et al., 2012; VanderDrift et al., 2011), five reported *p*-values greater than 0.05 (Lydon et al., 2008; Ma et al., 2019a; Ma et al., 2019b; Visserman & Karremans, 2014; Wang et al., 2016), two included significant *p*-values that were revealed to be nonsignificant upon recalculation (Linardatos & Lydon, 2011; Meyer et al., 2011), and upon closer inspection five involved inattention to alternatives rather than devaluation of alternatives (Maner et al., 2008, 2009; McNulty et al., 2018, Zhang et al., 2017, DeWall et al., 2011). In addition, one article was removed because the study context involved high school seniors' devaluation of colleges, not romantic relationships (Lyubomirsky & Ross, 1999). Consequently, 17 eligible articles were included in the initial *p*-curve analysis (*p*-curve A). See Table 2 for the *p*-curve disclosure table (an expanded version of the *p*-curve disclosure table is available on the OSF; https://osf.io/huz4v).

The articles included in the initial *p*-curve analysis fit the originally determined inclusion criteria, but upon closer inspection, many incorporated variables that are not essential components of the derogation of alternatives theory. I conducted a second *p*-curve analysis (*p*-curve B) that retained only the studies that used simple experimental designs with threat as a manipulated variable, relationship status or commitment as a quasi-independent variable, and measures of attractiveness or liking towards an alternative as the dependent variable. These studies represent straightforward and comparable approaches to test devaluation of alternatives and exclude covariates and experimental manipulations other than threat, removing potential muddying influences on the *p*-curve analysis. Nine studies were removed from the *p*-curve A for this analysis. Studies were removed for focusing on variables or manipulations unrelated to the essential elements of devaluation of alternatives (Ritter et al., 2010; Rodrigues & Lopes, 2017; Ziegler-Hill et al., 2020), the investment model (Miller, 1997), and not having a measure to evaluate the alternative in terms of attractiveness as a dependent variable (de Lenne et al., 2019; Plant et al., 2010; Pronk et al., 2011; Ma-Kellams et al., 2017; Lee & O'Sullivan, 2019). Consequently, eight articles were included in the second *p*-curve analysis (*p*-curve B).

P-curve analyses were conducted with version 4.06 of the online *p*-curve app (<u>http://www.p-</u> <u>curve.com</u>). The supplementary materials on the OSF include detailed information about article selection and inclusion (<u>https://osf.io/5c7pk</u>) and an extended *p*-curve disclosure table that identifies dependent variables, robustness *p*-values and identification of ambiguous cases (<u>https://osf.io/huz4v</u>). In addition, lists of the *p*-values included in each analysis are for reproducibility (<u>https://osf.io/p2rx6</u>), and original output from the *p*-curve app for each analysis is also provided (<u>https://osf.io/4j93q/</u>).

Original paper	<i>p</i> -curve	(1) Quoted text indicating hypothesis of interest	(2) Study design	(3) Key statistical result	(4) Quoted text from original paper with statistical results	(5) Results	Re- computed <i>p</i> -value
Bazzini & Shaffer (1999)	A, B	But even so, we presume that differences in ratings as a function of dating status will be apparent only in the high threat/opportunity condition, with participants in the low-threat/opportunity condition being a closer approximation to an unbiased baseline for judging the attractiveness of the SP [stimulus persons].	2 (gender: male vs female) x 2 (threat/opportunity: high vs low) x 2 (dating status: exclusive daters vs. nonexclusive daters)	difference of means, <i>F</i>	The ANOVA produced a significant main effect for dating status, $F(1, 94) = 7.12$, $p < .001$, that was qualified by the two-way interaction between dating status and threat/opportunity, $F(1, 94) = 4.50$, $p < .04$, that is shown in Figure 1.2.	<i>F</i> (1, 94) = 4.50, <i>p</i> < .04	0.03653
Cole et al. (2016)	A, B	We expected perceptual downgrading to occur when participants were in a relationship and when the target presented a potential threat to their relationship (i.e., was single). That is, we expected that participants in relationships compared with single participants would perceive a single, attractive target as less attractive. When the target was described as already in a relationship, we expected no differences between participants who were single and those in relationships.	2 (participant status: single vs. in a relationship) x 2 (target threat: low vs. high)	difference of means, F		<i>F</i> (1, 127) = 6.82, <i>p</i> = .01	0.0101
Karremans et al. (2011)	A, B	According to the derogation hypothesis, the classification image of the attractive mate generated by romantically involved participants should be less attractive than the image generated by uninvolved participantswe also examined romantically involved and uninvolved participants' memory of <i>un</i> attractive mates. According to the reasoning underlying the derogation effect, romantically involved individuals would not necessarily devaluate unattractive mates, as they do not pose a potential threat to the current relationship.	2 (relationship status: involved vs. uninvolved) x 2 (attractiveness: attractive vs. unattractive)	difference of means, F	An interaction effect between relationship status and attractiveness condition was obtained, $F(1, 61) =$ 14.06 , $p < .001$, $\eta^2 = .19$.	<i>F</i> (1, 61) = 14.06, <i>p</i> < .001	.00040
Karremans & Verwijmeren (2008)		In Study 1 we wanted to provide initial evidence for the basic effect, namely, that people involved in a romantic relationship would mimic an attractive alternative to a lesser extent than people not involved in a relationship.	status: involved vs. not involved) x 2		This analysis revealed a strong main effect of relationship status on mimicry, $F(1, 19) = 11.35$, $p < .01$, $\eta_p^2 = .42$.	F(1, 19) = 11.35, $p < .01$	0.00322

Table 2. P-Curve Disclosure Table

Lydon et al. (2003)		Devaluation/calibration hypotheses 1. Moderately committed participants will rate the target as more appealing in the low-available/low- threat condition than in the high-available/moderate- threat condition.		difference of means, <i>F</i>	However, these main effects were qualified by the predicted interaction between relationship status and threat level, $F(2, 103) = 3.89, p < .05$.	<i>F</i> (2, 103) = 3.89, <i>p</i> < .05	0.02352
Lydon et al. (1999)		Our overall prediction was a three-way interaction between attitudinal commitment, structural commitment and adversity. Under moderate adversity or threat, those high on one type of commitment would rate the alternative as less attractive compared to those low on both types of commitment and those high on both types of commitment. Under high adversity, we predicted that those high on one or the other type of commitment would no longer devalue and instead appear similar to those low on both measures of commitment. However, we expected under high adversity, those high on both measures of commitment would begin to devalue the alternative.	2 (threat/adversity: moderate vs. high) x 2 (structural commitment: married vs. dating) x 2 (attitudinal commitment: high vs. low)	test statistic for regression coefficient, t	status by attitudinal commitment) was significant $B = 41$ sr [sominortial r]	<i>t</i> = 2.71, <i>p</i> < .01 <i>N</i> = 134	0.00764
Petit & Ford (2015)	A, B	According to the motivational hypothesis, coupled and noncoupled participants should show the same level of pupil dilation in response to the opposite-sex models but express different degrees of attractiveness on a self- report measure. Coupled participants should censor or override their initial, automatic perceptions and report less perceived attractiveness.		difference of means, <i>F</i>		<i>F</i> (1, 70) = 4.74, <i>p</i> < .03	.03284
Simpson et al. (1990)	Α, Β	Accordingly, we hypothesize that individuals involved in ongoing dating relationships (daters), relative to those not involved in relationships (nondaters), should perceive highly desirable opposite-sex persons as less physically and sexually attractive.		means, F	As is evident in Table 1, individuals involved in ongoing dating relationships found the opposite-sex persons to be significantly less physically and sexually attractive than did individuals not involved in dating relationships, $F(1, 197) = 11.28, p < .001$.	<i>F</i> (1, 197) = 11.28, <i>p</i> < .001	0.00094
Lee & O'Sullivan (2019)		Relationship commitment would be positively associated with the range of monogamy maintenance efforts used.	longitudinal cross- sectional	correlation, r		r = .17, n = 287, p < .05	0.00387

					Derogation use ($r = .17$, $n = 287$, $p < .05$), and was unrelated to Proactive Avoidance, $r =03$, $n = 287$, $p = .76$.		
Plant et al. (2010)	A	We hypothesized that participants in a committed, heterosexual relationship would automatically perceive attractive opposite-sex targets as a threat and, therefore, would be biased toward identifying those targets as having a gun.	(object: gun ys	difference of means, <i>F</i>		<i>F</i> (1, 109) = 4.21, <i>p</i> < .05	0.04258
Ritter et al. (2010)	A	Based on the reasoning outlined above, we expected that among romantically involved participants relatively high (as compared to low) self-regulation capacity would result in less interest in the attractive opposite-sex others. Such findings are not expected with regard to unattractive others, and are not expected to occur for single participants.	2 (relationship status: romantically involved vs. single) x 2 (attractiveness: attractive vs. unattractive) x 2 (condition: no depletion vs. depletion)	difference of	Most importantly, the analysis yielded the predicted three-way interaction of attractiveness, relationship status, and condition, $F(1, 96) = 5.06$, $p < .03$, $\eta_p^2 = .05$.	<i>F</i> (1, 96) = 5.06, <i>p</i> < .03	0.02677
Rodrigues & Lopes (2017)	A	Regulatory focus interacts with relationship status on initial attraction judgments (Study 1). Specifically, Hypothesis 1A: Single individuals with a predominant focus on promotion will indicate greater initial attraction to an attractive other, compared to those with a predominant focus on prevention; Hypothesis 1B: This association between predominant focus on promotion and initial attraction should be weaker, or even disappear, for romantically involved individuals	cross-sectional	test statistic for regression coefficient, <i>t</i>	land relationship status $h = 0.08$ VF -	t(226) = - 2.70, p = .008	0.00746
de Lenne et al. (2019)	A	Exposure to alternative partners on SNS [social networking sites] is positively related to pursuing of alternative partners on SNS. In turn, this pursuing behavior is negatively related to relationship commitment.	cross-sectional	cni-square,	The main model showed a good fit with the data (Figure 1): Chi-square $\chi^2(101) = 284.29, p < .001$	$\chi^2 (101) =$ 284.29, <i>p</i> < .001	<.00001

Ma-Kellams et al. (2017)	А	We predicted that more physically attractive men would experience briefer marriages and be more likely to divorce. Here, we focused on actual PA [physical attractiveness].	archival analysis	correlation, <i>r</i>	married, Pearson's $r =23$, $p < .001$; those rated as more physically	r =23, p < .001 N = 238	0.00035
Miller (1997)		I hypothesized that higher satisfaction, investments, and commitment would all be associated with reduced attention to potential alternatives.	correlational	correlation, <i>r</i>	relationships.	(from Table 1) r =67, p < .01 N = 187	<.00001
Pronk et al. (2011)		We reasoned that if executive control is associated with people's ability to stay faithful, on the basis of their own experiences, people low in executive control should report relatively more difficulty in staying faithful		correlation, <i>r</i>	Most important and in line with our predictions, there was a positive significant relation between reaction times on the EAST and participants' self-reported level of difficulty in staying faithful, $r(72) = .30$, $p = .01$. Thus, slower response latencies on the executive control task – indicating a lower level of executive control – were related to a higher level of difficulty in staying faithful.	r(72) = .30, p	0.00941
Zeigler-Hill et al. (2020)	A	We predicted that narcissistic rivalry would have an indirect association with commitment through attitudes toward alternative romantic partners. That is, we expected that attitudes toward alternative romantic partners would provide at least a partial explanation for the association that narcissistic rivalry had with commitment.	cross-sectional	for mediation	Tests of mediation showed thatnarcissistic rivalry ($a_2b = -0.10$, SE = .05, $z = -2.07$, $p = .04$, 95% CI [- 0.20, -0.01]) had negative indirect associations with commitment through attention to alternatives.	.04	0.03845

Note. The inclusion of Bazzini and Shaffer (1999) may be considered an ambiguous case, as the analysis identified in this table tested the first relevant hypothesis but used the second reported dependent variable, as the analysis that used the first reported dependent variable was not statistically significant. *P*-curve A was run with and without this result. Robustness tests were identified for three studies (Karremans & Vewijmeren, 2008; Petit & Ford, 2015; Miller, 1997). Details of robustness tests are available in the extended *p*-curve disclosure table on the OSF. *P*-curves A and B were run with the results identified in this table, and again with the robustness results replacing the original results. These additional *p*-curves are available on the OSF (https://osf.io/4j93q/).

3.3 Results

P-curve analysis involves the interpretation of the shape of a distribution of independently sourced *p*-values to identify studies that represent true effects from those that may have been obtained through *p*-hacking or selective reporting. When the null hypothesis is true (i.e., an effect is nonexistent), the distribution of *p*-values is uniform, meaning each *p*-value that could emerge from a significance test is equally likely to occur. However, when the alternative hypothesis is true (i.e., a true effect exists), the distribution of *p*-values is right-skewed, meaning smaller *p*-values (those closer to zero) are more likely to be observed compared to higher *p*-values (Lakens, 2015; Simonsohn et al., 2014). Therefore, a collection of *p*-values, whereas a collection of *p*-values with significant left skew would be representative of *p*-hacking and/or selective reporting (Simonsohn et al., 2014).

Statistical inferences can be made from *p*-curves about whether a collection of studies involving a specific effect contains evidential value by determining whether there is statistically significant right skew. If significant right skew is obtained, the set of studies included in the *p*-curve *as a whole* can be said to contain evidential value. Significant right skew is determined to exist if the distribution of *p*-values below .025 (termed the half *p*-curve) is significant at the p < .05 level, or if the full *p*-curve and the half *p*-curve are both significant at the p < .05 level. Examination of the half *p*-curve in addition to the full *p*-curve provides a more robust and more accurate estimate of evidential value, as it addresses the possibility of plausible levels of ambitious *p*-hacking (i.e., *p*-hacking beyond p < .035), while still maintaining an acceptable level of power (Simonsohn, et al., 2015).

A *p*-curve that does not have statistically significant right skew can be explained by the inclusion of small or nonexistent effects, estimates that contain considerable amounts of noise, or a sample of *p*-values that is too small to reach statistical significance (Simonsohn et al., 2014). To investigate these potential issues further, a follow up test that compares the *p*-curve of interest to a *p*-curve that would be obtained if the collection of studies was powered at 33% (an arbitrary cutoff; Simonsohn et al., 2014) can be conducted. Doing so adjusts the significance test to

determine whether the *p*-curve of interest is significantly different from a small effect rather than testing to determine whether it is significantly different from zero. If the *p*-curve of interest is significantly flatter than the 33% power *p*-curve, the collection of studies can be interpreted as lacking evidential value. If the test does not reach significance, the *p*-curve analysis is considered inconclusive, and the quantity of studies included in the *p*-curve, as well as their level of power, should be increased (note this is an appropriate interpretation of this test only if the right skew test was not statistically significant).

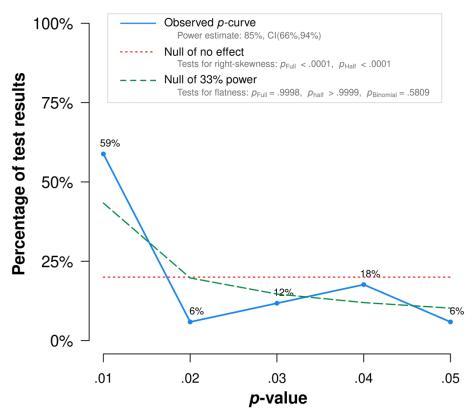
P-curve analyses also provide information about how the *p*-curve is affected when the highest and lowest *p*-values are removed from the analysis to detect whether the overall results of the *p*curve analysis are significantly influenced by extreme values. Specifically, the *p*-values of the right skew test for the full and half *p*-curve and the 33% power test of the full *p*-curve are plotted excluding extreme values in ascending order (for the lowest values) and descending order (for the highest values). Changes in the overall significance of these tests may indicate the sensitivity of the *p*-curve analysis to the presence of extreme values included in the selected set of studies.

Lastly, *P*-curve analysis calculates an estimate of statistical power for the set of studies by comparing the *p*-curve to the expected *p*-curves for each value of power between 5% and 99%. The expected *p*-curve that best fits the observed *p*-curve represents the estimate of statistical power for the set of studies.

3.3.1 *P*-Curve A

The initial *p*-curve (*p*-curve A; Figure 4) that included 17 studies emerged as significantly rightskewed for both the half (z = -6.83, p < .001) and full *p*-curves (z = -6.40, p < .001), suggesting the presence of evidential value for this set of effects. When compared to a *p*-value distribution with 33% power, no significant difference emerged, suggesting this set of effects represented a detectable and non-negligible effect (half *p*-curve: z = 7.33, p > .9999; full *p*-curve: z = 3.48, p =.9998). The estimated power for this set of studies was 85% (90% CI: 66%, 94%). When the ambiguous case (Bazzini & Shaffer, 1999) is removed, the key findings remain unchanged and the estimated power of the set of studies increased to 88% (90% CI: 72%, 96%).

Figure 4. P-Curve A



Note: The observed *p*-curve includes 17 statistically significant (p < .05) results, of which 12 are p < .025. There were no non-significant results entered.

Note. The ambiguous case, Bazzini and Shaffer (1999) is included.

When *p*-curve A is run with the three identified robustness results substituted in for their respective original values, the main findings remain unchanged. There is significant right skew for both the half (z = -4.92, p < .001) and full *p*-curves (z = -4.72, p < .001), and no significant difference emerged in the 33% power test, suggesting the evidential value of this set of effects was not inadequate or absent (half *p*-curve: z = 5.89, p > .9999; full *p*-curve: z = 2.17, p = .985). However, the estimated power for the set of studies decreased to 69% (90% CI: 42%, 86%). The robustness *p*-curve can be found on the OSF (https://osf.io/ybdk8).

Figure 5 shows the results of the cumulative meta-analysis. When the five lowest *p*-values are dropped from the full and half *p*-curve tests for right skew, the *p*-values of the significance tests exceed .05, with the *p*-value of the half *p*-curve appearing to equal or exceed .1. As mentioned above, significant right-skew is determined to exist if the half *p*-curve is significant at the p < .05 level, or if the full *p*-curve and the half *p*-curve are both significant at the p < .1 level. When the

five lowest *p*-values are excluded, this set of studies can no longer be said to demonstrate significant right skew (the same pattern emerges for the robustness *p*-curve). When examining the effect of excluding the lowest *p*-values on the 33% power test, the test remains nonsignificant, suggesting that the set of studies elicits a *p*-curve that is not significantly flatter than what would be observed if the studies were powered at 33%. With the five lowest studies are excluded, the *p*-curve is inconclusive. When the highest *p*-values are sequentially eliminated, there are no changes in significance for the full *p*-curve tests of right skew and 33% power.

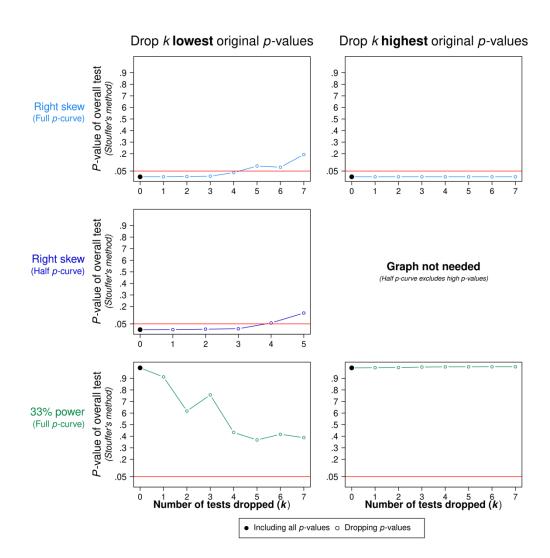


Figure 5. Cumulative Meta-Analysis of *P*-Curve A

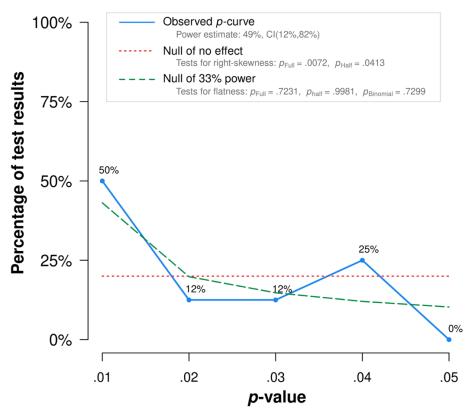
Note. The ambiguous case, Bazzini and Shaffer (1999) is included.

P-curve A was well-powered. According to simulations (R code created and shared by Simmons & Simonsohn, 2017), when all studies in the selected set were assumed to have on average 33% power, the *p*-curve analysis would have an 78.29% chance of detecting evidential value. If the studies were powered to 50% on average, the *p*-curve analysis would have a 95.04% chance of detecting evidential value. Powered at 80% on average would result in a *p*-curve analysis powered at over 99.99% (the adapted R code is available on the OSF: <u>https://osf.io/cgzvf</u>). Exploring the power of *p*-curve with varying numbers of real and null effects, *P*-curve A would have greater power than this set of studies had on average even if 8 studies examined real effects and 9 examined null effects.

3.3.2 *P*-Curve B

The second *p*-curve (*p*-curve B; Figure 6) that included 8 studies emerged as significantly rightskewed for the full *p*-curve (z = -2.45, p = .0072) and the half *p*-curve (z = -1.74, p = .0413), which suggests the presence of evidential value for this set of effects. When compared to a *p*value distribution with 33% power, no significant difference emerged, suggesting the size of this set of effects was detectable and non-negligible (full *p*-curve: z = 0.59, p = .7231; half *p*-curve: z = 2.89, p = .9981). The estimated power for this set of studies was 49% (90% CI: 12%, 82%). When the ambiguous case (Bazzini & Shaffer, 1999) is removed, the key findings remain unchanged and the estimated power of the set of studies increases to 60% (90% CI: 19%, 88%).

Figure 6. P-Curve B



Note: The observed *p*-curve includes 8 statistically significant (p < .05) results, of which 6 are p < .025. There were no non-significant results entered.

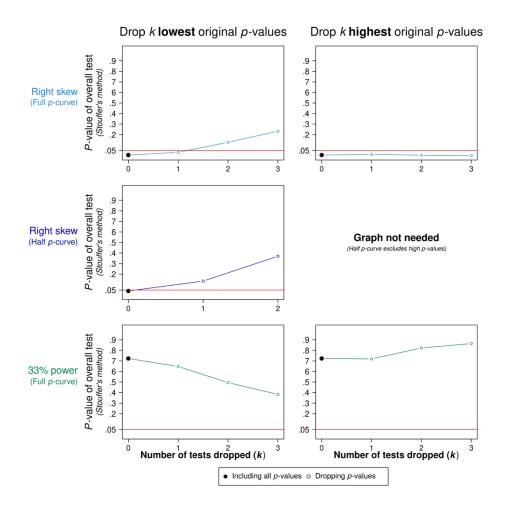
Note. The ambiguous case, Bazzini and Shaffer (1999) is included.

When *p*-curve B is run with the two identified robustness results substituted in for their respective original values, the main findings remain unchanged. There is significant right skew for both the half (z = -1.87, p = .0304) and full *p*-curves (z = -2.54, p = .0055), and no significant difference emerged in the 33% power test, suggesting the evidential value of this set of effects was not inadequate or absent (half *p*-curve: z = 2.98, p = .9986; full *p*-curve: z = 0.68, p = .7532). The estimated power for the set of studies was 51% (90% CI: 13%, 83%). The robustness *p*-curve can be found on the OSF (https://osf.io/4j93q/).

Figure 7 shows the results of the cumulative meta-analysis. When the lowest *p*-value is dropped from the full and half *p*-curve tests for right skew, the *p*-values increase. Specifically, the full *p*-curve *p*-value appears to closely approach .05 and the half *p*-curve *p*-value appears to approach or equal .1. Consequently, with the most extreme *p*-value excluded from the analysis, we should be cautious to interpret this set of studies as demonstrating the presence of evidential value.

When the next lowest *p*-value is removed, both the full and half *p*-curve right skew *p*-values exceed .1, demonstrating a clear inability to detect evidential value. When examining the effect of excluding the lowest *p*-values on the 33% power test, the test remains nonsignificant, suggesting that the set of studies elicits a *p*-curve that is not significantly flatter than what would be observed if the studies were powered at 33%. However, as the number of removed extreme small *p*-values increases, the *p*-value for the 33% power test does decrease substantially. This combination of results suggests that if the one or two most extreme *p*-values are excluded from the analysis, the *p*-curve becomes inconclusive (dropping the two lowest *p*-values in the robustness *p*-curve leads to the same result). When the highest *p*-values are sequentially eliminated, there are no changes in significance for the full *p*-curve tests of right skew and 33% power.





Note. The ambiguous case, Bazzini and Shaffer (1999) is included.

P-curve B was adequately powered. According to simulations (R code created and shared by Simmons & Simonsohn, 2017), when all studies in the selected set were assumed to have on average 33% power, *P*-curve B would have a 51.49% chance of detecting evidential value. If the studies were powered to 50% on average, the *p*-curve analysis would have a 74.65% chance of detecting evidential value. Powered at 80% on average would result in a *p*-curve analysis powered at 97.97% (the adapted R code is available on the OSF: <u>https://osf.io/cgzvf</u>).

Exploring the power of *p*-curve with varying numbers of real and null effects, *p*-curve B would have greater power than the set of studies had on average if 6 studies examined real effects and 2 examined null effects.

3.4 Discussion

P-curve analyses were conducted using two samples of studies from the devaluation of alternatives literature. P-curve A which included a broader selection of the literature detected evidential value, presence of a non-negligible effect, and showed the included studies were adequately powered. P-curve B which included a more specific subset of the literature detected evidential value, presence of a non-negligible effect, but indicated the included studies were poorly powered. Both *p*-curves were rendered inconclusive when the lowest *p*-values were dropped, which suggests we should be cautious in concluding that selective reporting can be ruled out as an explanation for each set of statistically significant findings, as we cannot say that each set of studies robustly demonstrates the presence of evidential value. The *p*-curve analyses were adequately powered, as simulations demonstrated. However, *p*-curve analyses are often biased towards detecting evidential value in a set of effects (Simonsohn et al., 2014). In other words, it is difficult for p-curve analyses to fail to detect evidential value (fail to reject the null of right-skew) unless there is intense *p*-hacking or publication bias at play. This reinforces the caution with which we should regard the preliminary detection of evidential value in the present p-curves. Simonsohn et al. (2014) suggest including more studies and higher-powered studies in *p*-curve analyses to increase the likelihood of detecting definitive results.

The *p*-curve analyses also indicated that the selection of studies involving devaluation of alternatives often contains less than the typically recommended level of power (80%) and can even fall to an unacceptable average level of 50% power. Studies with inadequate levels of power decrease the likelihood of detecting a true effect when one exists, consequently wasting valuable resources such as time and money in addition to stymying the progression of research programs. Future researchers aiming to test devaluation of alternatives should take great care and effort to ensure their studies are highly powered, at or above 80% for the effect size of interest. Doing so would increase the likelihood of detecting true effects in general and would increase the quality of *p*-curve analyses that include such high-powered results.

Conducting *p*-curve analyses provided the opportunity to assess the strength of the existing literature and determine whether the selected samples displayed evidence of selective reporting or *p*-hacking. A selection rule for the articles and their respective eligible *p*-values included in the *p*-curves were determined *a priori*. However, different selection rules would likely result in the inclusion of different studies and *p*-values, which could potentially lead to different conclusions. The inferences made from *p*-curve analyses, therefore, should not be confidently expanded beyond the scope of the specified set of studies from which they emerge to the entirety of the devaluation of alternatives literature. Though the selected set of articles represent a sample of the devaluation of alternatives literature, it cannot be said that the selected articles and pvalues are representative of the literature as a whole -a determination that may be difficult or impossible to make, or at the very least, challenging for a community of scholars to agree on. However, the information elicited from the present *p*-curve analyses can be used to make cautious recommendations for future *p*-curve analyses and future studies that aim to test or incorporate the theory of devaluation of alternatives. P-curvers interested in devaluation of alternatives could consider conducting other *p*-curve analyses with different selection criteria, perhaps broader criteria than those used here.

Overall, these analyses indicate that conclusive evidence of selective reporting and/or *p*-hacking was *not* detected, and that this literature likely involves a true effect and contains evidentiary value. However, these findings are less robust than we might prefer and some of the past research was shown to be insufficiently powered. These analyses provided a quantitative and

meta-analytical evaluation of a selection of the literature and identified areas for future researchers to improve.

Chapter 4

4 An Empirical Test of Core Assumption 1: Are Attractive Alternative Partners Perceived as Threatening?

Through the process of formalizing the theory of devaluation of alternatives, specific theoretical assumptions were identified and categorized as part of the hard core or protective belt of the theory. Examining the evidence for each of these assumptions in the context of devaluation of attractive alternatives literature revealed that many of the assumptions, as their name implies, have simply been assumed to be true when further testing and corroboration is both possible and necessary. This chapter describes an experiment that was specifically designed as a test of the first core assumption of the theory of devaluation of alternatives, that attractive alternatives (external stressors) are threatening.

4.1 Introduction

Recent reviews of the relationship maintenance literature have identified two broad motivations that underly relationship maintenance efforts: relationship enhancement and responding to relationship threats (Agnew & VanderDrift, 2015; Ogolsky et al., 2017). In Chapter 2, devaluation of alternatives was classified as a sub-theory of the broader theory of relationship maintenance that involves a specific relationship threat (attractive alternatives) and a specific maintenance strategy (devaluation of alternatives). Inherent in this simple summary of the sub-theory is the characterization of attractive alternatives as threatening. However, as demonstrated in Chapter 2, little research has attempted to measure outcomes that signal perceptions of this proposed threat, and existing research has relied on narrow operationalizations of attractive alternatives such as brief exposures to photographs, which provided only weak and inconsistent support for the claim. This represents a concerning lack of verisimilitude of the primary core assumption of the sub-theory and the essential precedent of devaluation. The main purpose of the current study was to measure threat responses quantitatively and qualitatively after exposure to a novel experimental manipulation of an external relationship threat.

4.1.1 Review of Relevant Research

Only three studies have operationalized and measured threat perception in the context of devaluation of alternatives (refer to Chapter 2 for a more detailed overview and evaluation of this literature). Notably, each operationalized attractive alternatives and measured threat in different ways. Attractive alternatives were manipulated by asking participants to imagine a scenario where an attractive classmate asked for their phone number (Study 1; Lydon et al., 2008), by priming participants with thoughts of their relationship then exposing them to an image of an attractive other in virtual reality (Study 2; Lydon et al., 2008), and by exposing participants to repeated images of attractive others who were armed with a gun or neutral object in a computer simulation task (Plant et al., 2010). Feelings of threat were operationalized implicitly through word-completion tasks (Study 1; Lydon et al., 2008), the measured distance between the participant and the alternative (Study 2; Lydon et al., 2008), and the error rates of decisions to shoot an armed target or not shoot an unarmed target (Plant et al., 2010). Results from both of Lydon et al.'s studies showed women experienced implicit perceptions of threat, by completing more word fragments with threat words and by placing the image of the attractive alternative farther away in virtual space. In contrast, Plant et al. found that men behaviorally expressed feelings of threat, such that they more often mistakenly chose to "shoot" attractive, unarmed women targets.

These results represent preliminary evidence in support of the claim that attractive alternatives stimulate feelings of threat. However, the research teams found that feelings of threat were observed for participants of different genders, indicating that evidence for this claim is inconsistent. In addition, the operationalizations of attractive alternatives were limited in that the static stimuli (i.e., photographs) were not identified or described as potential dating partners (Study 2, Lydon et al., 2008; Plant et al., 2010). Thus, whether participants regarded the stimuli as such is unknown. In addition, "attractive" was operationalized solely through physical attractiveness, and in Plant et al. (2010) specifically facial attractiveness. Similarly, in the imagined scenario operationalization, the participants were asked to visualize the scenario with a classmate who asked for their number to study later. While this could be argued to be a threatening scenario, it is also possible for opposite-sex persons to be platonic classmates or friends, especially when the purpose of swapping phone numbers was explicitly said to be for

studying rather than a date or other romantic or sexual interaction. And, while participants were instructed to imagine an actual attractive person known to them, the question of whether such an imagined scenario is threatening is questionable. All three operationalizations represent potential situations of "fleeting extradyadic attraction" (Lee & O'Sullivan, 2018, 2019), but whether the participants experienced attraction to the target alternative(s) was not measured in any of the studies and such attraction should not be assumed to have occurred. Consequently, the chosen manipulations of attractive alternatives and/or threatening circumstances leave much room for improvement, as the severity of the presented "threat" was likely quite low. Regarding the operational choices for measuring perceived threat, each of these studies measured threat at implicit or automatic levels. These approaches were beneficial because potential threats are often perceived quickly and automatically, and subsequently influence affect, cognitions, and behaviors. Implicit and impulsive behavioral responses are useful because they allow for perceptions of threat to be observed close to when they occur. Threat perception can also be observed using other methods, such as through physiological responses or verbal responses. To increase our understanding of how attractive alternatives are perceived, the methods used to measure threat perception must be diversified to demonstrate the robustness (or lack thereof) of the effect.

4.1.2 Threat Perception and Response

Threats are persons or circumstances that convey the potential for loss and are perceived to be outside an individual's means of coping (Tomaka et al., 1993). Threats are also considered influential forces that introduce "what ifs" to decisional contexts (Milburn, 1961, 1969; Tedeschi, 1970). Threats vary in the extent to which they are perceived as sources of real danger, and this is partially determined by characteristics of the threat. For example, ambiguous threats may create feelings of uncertainty in the perceiver, which may stimulate greater negative affect or impulsive actions, while threats considered to be uncredible or of low esteem may only be successful in eliciting hostility, if any response at all (Millburn, 1977). Characteristics of the perceiver also influence threat perception. For example, perceivers may consider their knowledge of and past experiences with the source of the threat when making evaluations about its severity. Personal characteristics of the perceiver, especially those relating to coping resources and abilities, may also influence threat perception, such as attachment (Hart et al., 2005) and

defensive self-esteem (Jordan et al. 2003). Threats add complicating information to decisional contexts that carry the weight of potentially incurring losses, but the extent to which threats are seriously regarded depends on both the characteristics of the threat and the characteristics of the threat perceiver.

If attractive alternative partners are threats, they should embody the characteristics of threats outlined above. Attractive alternatives, as past researchers have noted (Kelley, 1983; Rusbult, 1980, 1983) represent temptation, stimulating coupled persons to think "what if" and evaluate their current circumstances and hypothetical benefits of switching out their partner for another. In so doing, alternative partners may act as forces that draw individuals away from their partners and relationships. The potential loss then, would be the ending of the current relationship, which, despite the possibility of quickly entering a new engagement, may be distressing and considered to be a strain on one's resources and abilities to cope. The observation that threats of low esteem or meaning often fail to elicit responses corresponds with the peripheral assumption of the subtheory that more attractive alternatives are more threatening. Alternatives who fail to reach this threshold would not stimulate a response or subsequent mitigating action. Lastly, attractive alternatives may be evaluated differently based on the qualities and characteristics of the perceiver. For example, persons who have been in relationships longer or who are extremely committed (either attitudinally or structurally) may not find alternatives tempting even if they are attractive and available. Or, if such individuals do perceive attractive alternatives as threats, they may feel equipped enough to combat the threat and thus may not feel intense feelings of stress or discomfort.

Once a threat is noticed it is evaluated. Lazarus and Folkman (1984) introduced the concept of cognitive threat appraisal – a mechanism for threat perception. Cognitive appraisal has two parts: an assessment of the degree of risk posed by the potential threat (primary appraisal) and an assessment of the extent of available resources to manage the potential threat (secondary appraisal). This suggests the potential to observe two waves of responses to threatening stimuli – an initial immediate response, and a more calculated response once the circumstance has been situated within the context of an individuals perceived coping abilities (Lazarus & Folkman, 1984). Responses to threats can take a variety of forms and may include multiple response systems (Hodgins et al., 2010). For example, threats may elicit negative affect (Tomaka et al.,

1993; Millburn; 1977; Hodgins et al., 2010) and stimulate physiological processes such as heart rate (Hodgins et al., 2010) and pupil dilation (Bitsios et al., 1996; Henderson et al., 2017). Threat responses may be controllable (e.g., verbal content) or more automatic (e.g., cardiovascular responses), and are also more likely to be expressed (and consequently, observed) behaviorally as the severity of the threat increases (Hodgins et al., 2010). Thus, threats imply potential loss which individuals respond to in affective, cognitive, behavioral, and physiological ways.

Much research on devaluation presents attractive alternatives as quite powerful, such that lowered ratings of attractiveness following brief exposures to physically attractive targets are interpreted as evidence of protective measures (i.e., devaluation). In the past, observing devaluation has led researchers to conclude that the behavior was caused by feeling threatened by the attractive alternative. However, this represents the logical fallacy of affirming the consequent (If p then q, q therefore p). While devaluation is hypothesized to occur in response to threats, it is illogical to assert that if devaluation is observed that feeling threatened is the cause. This is especially inappropriate when so little research has been dedicated to measuring threat perception following exposure to or interaction with attractive alternatives. The incorporation of past research on threat perception into the study of devaluation of alternatives is beneficial because it provides guidance on what outcomes may be expected. Knowledge of the existence of primary and secondary threat appraisals necessitate the use of multiple measures and a design that provides participants the opportunity to reflect on their available resources to cope and manage the threat. Physiological responses may occur such as elevated heart rate or dilated pupils, and feelings of negative affect could occur along with negative evaluations of the threat. Use of multiple and diverse measures to detect threat perception would be a strong contribution to existing literature involving the threatening nature of attractive alternatives.

4.1.3 The Current Study

Despite repeated claims that attractive alternatives represent "threats" in the relationship maintenance literature, little empirical evidence exists to support this claim. Research testing this claim is limited in both quantity and scope and has not considered relevant background knowledge about threat perception and response into their methodological decisions. At present, this primary core assumption of the sub-theory of devaluation of alternatives lacks adequate

support. The current study aimed to test this assumption by measuring threat perception and devaluation following exposure to a threatening stimulus. Participants in exclusive, monogamous relationships completed a compatibility questionnaire and were told their scores would be evaluated by an algorithm that would calculate their compatibility with their partners. Under the guise of collecting more data to further refine the algorithm, the participants also completed the questionnaire with a single and attractive person in mind who was not their partner. Participants received experimentally manipulated and randomly assigned compatibility results for both their partner and alternative, except in the control condition where participants were only shown their compatibility was either "high" (88%), "average" (73%), or "low" (36%). Participants then were asked to provide open-ended responses immediately after viewing their compatibility results and completed measures of positive and negative affect, a single item asking about feelings of threat, an evaluation of the compatibility questionnaire/algorithm, and the perceived accuracy of their results, among other measures.

Based on prior research involving threat perception and response and devaluation of alternatives, I predicted that participants in the high threat condition would report greater negative affect (H1), greater feelings of threat (H2), stronger negative perceptions of the questionnaire/algorithm (H3), and lower perceived accuracy of the alternative compatibility result (H4) than those in the low threat condition. Additionally, each of these outcomes were be subjected to exploratory moderation analyses, to determine whether the responses are moderated by relevant personal and relationship characteristics and beliefs. Specifically, relationship length, commitment, perceived partner compatibility, feelings toward the alternative, and initial perceptions of the questionnaire/algorithm were explored as moderators.

4.2 Methods

This study was preregistered on the Open Science Framework (OSF; <u>https://osf.io/bkvy3</u>), and all adopted and created materials are publicly available on the project page (<u>https://osf.io/w2h3k/</u>). This study was approved by the University of Western Ontario Non-Medical Research Ethics Board in March 2022.

4.2.1 Participants

To determine the sample size needed to conduct appropriately powered confirmatory analyses, I ran an *a priori* power analysis using the G*Power application (Version 3.1.9.6; Faul et al., 2007). For a one-way ANOVA with alpha set at 0.05 and 80% power, I would need 492 individuals (123 participants per condition) to detect an effect with a small effect size (Cohen's f = 0.15, $\eta^2 = 0.02$), and this was the target minimum sample size of participants who successfully met the eligibility criteria, missing data, and data quality and attention checks.

Of the 760 participants who provided consent in Part I, 367 (48.29%) consented to complete Part II. After the preregistered eligibility criteria, missing data criteria, and data quality/attention checks were implemented, 207 participants were included in the final sample. The obtained sample size did not meet the target sample size dictated by the *a priori* power analysis. A sensitivity analysis was conducted using G*Power to determine the effect size the sample is powered to detect. Alpha was set at 0.05, power at 80%, the number of groups at four and the total sample size at 207. Results revealed that with the obtained sample size it was possible to reliably detect effect sizes of medium size with 80% power (Cohen's f = 0.29, $\eta^2 = 0.078$). While higher levels of power are preferred because they allow for small effect sizes to reach statistical significance, the obtained sample may be limited in that it is only adequately powered to detect effects of medium size. While more limited than anticipated, implementing such extensive data cleaning criteria should better ensure that only high-quality data from attentive, thorough participants who believed (or at least provided no indication of skepticism) of the deception was used for the analyses. For specific details regarding how many participants were excluded and for what reason, please refer to the supplemental materials on the OSF (https://osf.io/rsdbc/). Demographic and relationship characteristics of the full sample and for each experimental condition are shown in Table 3. No significant differences emerged between conditions regarding the distribution of these characteristics.

Table 3. Descriptive Statistics for Participant Demographic and RelationshipCharacteristics

Demographics			Condition			Condition Differences
	ALT_High $(N = 48)$	ALT_Avg ($N = 48$)	ALT_Low $(N = 54)$	P_Only $(N = 57)$	All (<i>N</i> = 207)	

Age, M(SD)†	23.96 (4.86)	24.92 (6.26)	24.02 (5.68)	23.04 (4.14)	23.94 (5.27)	F(3, 203) = 1.12 p = .34
Relationship Length in years†	2.96 (2.75)	3.61 (4.3)	2.76 (2.77)	3.28 (3.28)	3.15 (3.31)	F(3, 203) = 0.65 p = .59
Gender†	()	(112)	()	(0.20)	(0.01)	$\chi^2(3) = 0.20,$
	7	8	9	9	33	<i>p</i> = .90
Man	(3.38%) 40	(3.86%) 39	(4.35%) 41	(4.35%) 48	(15.94%) 168	
Woman	(19.32%)	(18.84%)	(19.81%)	(23.19%)	(81.16%)	
Non-binary, gender fluid or not listed*	1 (0.48%)	1 (0.48%)	4 (1.93%)	0 (0.00%)	6 (2.90%)	
Sexual Orientation†						$\chi^2(3) = 5.34,$ p = .14
Heterosexual	41 (19.81%)	41 (19.81%)	37 (17.87%)	45 (21.74%)	164 (79.23%)	
Bisexual	4 (1.93%)	5 (2.42%)	11 (5.31%)	5 (2.42%)	25 (12.08%)	
Lesbian/Gay, Pansexual, Queer, Prefer not to answer or not listed*	3 (1.45%)	2 (.97%)	6 (2.90%)	7 (3.38%)	18 (8.70%)	
Relationship Status†						$\chi^2(3) = 1.05,$ p = .79
Exclusively dating	43 (20.77%)	40 (19.32%)	48 (23.19%)	50 (24.15%)	181 (87.44%)	
Married	5 (2.42%)	8 (3.86%)	6 (2.90%)	7 (3.38%)	26 (12.56%)	
Partner Location [†]						$\chi^2(6) = 7.56,$ p = .27
Co-habitating	12 (5.80%)	19 (9.18%)	18 (8.70%)	17 (8.21%)	66 (31.88%)	
Same city/town	14 (6.76%)	16 (7.73%)	21 (10.14 %)	15 (7.25%)	66 (31.88%)	
Different city/town	22 (10.63%)	13 (6.28%)	15 (7.25%)	25 (12.08%)	75 (36.23%)	
Frequency of In- Person Contact [†]	(10.05%)	(0.28%)	(1.2370)	(12.0070)	(30.2370)	$\chi^2(6) = 7.94,$ p = .24
Every day	16 (7.73%)	20 (9.66%)	24 (11.59%)	22 (10.63%)	82 (39.61%)	1
Few days/week	11 (5.31%)	17 (8.21%)	16 (7.73%)	16 (7.73%)	60 (28.99%)	
Few days/month	16 (7.73%)	10 (4.83%)	6 (2.90%)	15 (7.25%)	47 (22.71%)	
Few times/year*	5 (2.42%)	1 (0.48%)	8 (3.86%)	4 (1.93%)	18 (8.70%)	

 $\underbrace{(2.4270)}_{(0.40\%)} \underbrace{(3.80\%)}_{(1.93\%)} \underbrace{(8.70\%)}_{(8.70\%)}$ *Note.* Demographic categories marked with an asterisk (*) indicate that it was omitted from the chi-square analysis. Variables marked with a cross (†) indicate it was included in the random assignment assessment.

Participants also reported information about a real potential alternative partner. Participants reported how their relationship with the alternative was best described (e.g., acquaintance, friend, colleague/peer from work or school, ex-partner), how long they have known the alternative, the alternative's relationship status (if known), the location of the alternative relative to the participant, how often the participant interacts with the alternative in-person, how well they know the alternative, the degree to which they are attracted to the alternative, and if they could see themselves dating the alternative if they were not with their current partner. Descriptive statistics for these items are presented in Table 4. There were no significant differences in these characteristics across conditions.

Alternative Descriptives			Condition			Condition Differences
	ALT_High $(N = 48)$	ALT_Avg ($N = 48$)	ALT_Low ($N = 54$)	P_Only $(N = 57)$	All (<i>N</i> = 207)	
Relationship Length in years†	6.64 (5.11)	5.43 (4.13)	5.35 (4.05)	4.76 (4.56)	5.52 (4.48)	F(3, 162) = 1.29, p = .28
Relation [†]						$\chi^2(3) = 2.81,$ p = .42
Acquaintance	6 (2.90%)	5 (2.42%)	2 (0.97%)	8 (3.86%)	21 (10.14%)	<i>p</i> – .+2
Friend	36 (17.39%)	27 (13.04%)	36 (17.39%)	40 (19.32%)	139 (67.15%)	
Colleague/Peer from work or school*	2 (0.97%)	7 (3.38%)	7 (3.38%)	1 (0.48%)	17 (8.21%)	
Ex-partner*	3 (1.45%)	5 (2.42%)	7 (3.38%)	5 (2.42%)	20 (9.66%)	
Not listed*	1 (0.48%)	4 (1.93%)	2 (0.97%)	3 (1.45%)	10 (4.83%)	
Relationship Status†	× /			~ /	~ /	$\chi^2(9) = 8.06,$ p = .53
Single	28 (13.53%)	23 (11.11%)	37 (17.87%)	29 (14.01%)	117 (56.52%)	
Casually dating	5 (2.33%)	8 (3.86%)	5 (2.42%)	9 (4.35%)	27 (13.04%)	
Exclusively dating	7 (3.38%)	6 (2.90%)	7 (3.38%)	10 (4.83%)	30 (14.49%)	
Married*	1 (0.48%)	1 (0.48%)	1 (0.48%)	2 (0.97%)	5 (2.42%)	
Unknown	7 (3.38%)	10 (4.83%)	4 (1.93%)	7 (3.38%)	28 (13.53%)	
Relationship Location†	× /	` '		~ /	、 /	$\chi^2(3) = 2.07,$ p = .56
Co-habitating*	1 (0.48%)	1 (0.48%)	0 (0.00%)	0 (0.00%)	2 (0.97%)	

Table 4. Descriptive Statistics of Participants' Selected Alternatives

Same city/town	23	20	20	22	85	
Buille elly/town	(11.11%)	(9.66%)	(9.66%)	(10.63%)	(41.06%)	
Different city/town	24	27	34	35	120	
2	(11.59%)	(13.04%)	(16.43%)	(16.91%)	(57.97%)	2
Frequency of In-						$\chi^2(6) = 5.03,$
Person Contact [†]						p = .54
Every day*	1	1	2	1	5	
Every duy	(0.48%)	(0.48%)	(0.97%)	(0.48%)	(2.42%)	
Few days/week	4	6	5	9	24	
I'W Uays/ WEEK	(1.93%)	(2.90%)	(2.42%)	(4.35%)	(11.59%)	
Few days/month	8	4	11	10	33	
rew days/monul	(3.86%)	(1.93%)	(5.31%)	(4.83%)	(15.94%)	
Few times/year	35	37	35	36	143	
rew times/year	(16.91%)	(17.87%)	(16.91%)	(17.39%)	(69.08%)	
Attraction to						
Alternative Items						
I lim are the are really	3.40	3.33	3.80	3.63	3.55	F(3, 203) = 2.46,
I know them well†	(0.96)	(1.06)	(0.90)	(.99)	(.99)	p = .06
I am attracted to	3.62	3.88	3.81	3.51	3.70	F(3, 203) = 1.88,
them†	(0.84)	(0.84)	(0.91)	(0.97)	(0.90)	p = .13
I could see myself						•
dating them if I	3.17	3.25	3.19	3.21	3.20	F(3, 203) = .06,
was not with my	(1.08)	(1.02)	(1.07)	(1.11)	(1.07)	p = .98
partner†						•

Note. Demographic categories marked with an asterisk (*) indicate that category was omitted from the chi-square analysis. The Attraction to Alternative items were rated on a scale from 1 to 5, with higher ratings indicating higher levels of agreement with the item. Variables marked with a cross (†) indicate it was included in the random assignment assessment.

The participants in the final sample were in their current relationships for an average of around three years. Most of the sample consisted of women, those who identified as heterosexual, and those exclusively dating one partner. Participants reported expectedly high levels of commitment at Part I and estimated their compatibility with their partners highly as well. Estimated compatibility with partners exceeded that of compatibility ratings with alternatives. The alternatives identified by the participants were mostly friends and were known to the participant for over five years on average. As instructed, almost 70% of participants identified an alternative that was not in an exclusive relationship. Most alternatives were reported to live in a different city and participants indicated in-person contact with the alternatives occurred only a few times a year. Responses to this question, however, should be interpreted cautiously due to the impact COVID-19 has had on limiting in-person gatherings. Participants rated knowing their alternative well, being attracted to them, and seeing themselves dating their alternative all above the midpoint on the provided scale.

4.2.2 Procedure

This study was approved by the University of Ontario Research Ethics Board in Spring 2022. All ethics materials are available on the OSF (<u>https://osf.io/m4u5b/</u>). Participants were recruited via mass email at the University of Western Ontario. All registered students from every degree program and affiliate college who had not opted out of receiving mass emails were invited to participate in Part I of the study. Interested participants had to be at least 18 years of age, fluent in English, and in an exclusive romantic relationship for at least three months with one partner who was also at least 18 years of age. To participate in Part I, participants must have completed Part I and had to input their personalized ID code created during Part I. Participants were given the opportunity to enter a raffle for one of three \$100 CAD Amazon.ca gift cards during each part of the survey. Completion of Part I earned one entry into the raffle while completion of Part II earned two entries into the raffle.

Prior to providing consent for Part I, participants must have passed a screening questionnaire that assessed whether they met the eligibility criteria set out in the initial recruitment advertisement. After participants consented to participate in Part I, they answered demographic questions about themselves, their current romantic partners, their relationship, and their commitment to their partner. Participants were told that the researchers had developed a questionnaire and algorithm that in previous tests had been shown to be highly accurate in determining romantic compatibility. Participants filled out the compatibility questionnaire about their partner. Next, participants were told that it would be helpful to have as much data as possible to refine the algorithm, therefore, the researchers would appreciate if they could fill out the compatibility questionnaire with someone else in mind who is not their partner. This person should be someone they know, who they may be attracted to, and who is single (ideally). The participants then filled out the compatibility questionnaire for that individual, the "alternative". Lastly, the participants were asked to report their initial perceptions of the compatibility questionnaire and algorithm. The participants created a unique ID code under the guise of it being necessary to access their "personalized results" during Part II and could provide their email to be entered into the raffle draw.

Approximately two weeks after the initial launch of Part I, participants were emailed the link to Part II. Prior to providing consent for Part II, participants must have passed a screening questionnaire that asked whether they participated in Part I and that asked them to provide the ID code they created in Part I to retrieve their "personalized" compatibility results. After providing consent and filling out some demographic questions, participants were randomly assigned to see one of four possible compatibility results that were displayed using colorful infographics (https://osf.io/xrjek/). Each infographic included a descriptive keyword at the top indicating the overall level of compatibility (i.e., highly, moderately, mildly), four subsections with respective percentages and brief general descriptions (communication & trust, values & beliefs, interests & personality, attraction), and a single percentage representing overall compatibility. In the "control" condition, participants were only shown their "algorithm-generated" compatibility result with their partner (76%). In the experimental conditions, participants were shown their compatibility result with their partner (76%) and with their alternative, with the alternative compatibility result being either high (88%), average (73%), or low (36%). After participants viewed their results, they were asked to provide qualitative and quantitative feedback about their initial feelings upon seeing the results, the perceived accuracy of the results, their perceptions of the compatibility questionnaire and algorithm, and their commitment to their partner. Participants were then debriefed, clearly informed that their compatibility results were fabricated, asked to not discuss the study for at least three weeks (when data collection would be complete), and thanked for their time. Participants were given the opportunity to enter the raffle again.

4.2.3 Measures

All measures used in the study are available on the OSF (<u>https://osf.io/zxswq/</u>).

4.2.3.1 Demographics

Participants reported their age, gender identity, sexual orientation, current relationship status, relationship length, their location relative to their partner, and how frequently they see their partners in person. Participants also provided demographic information about their current partners, including their age, gender identity, and sexual orientation.

4.2.3.2 Commitment

Participants completed the 7-item commitment subscale of the Investment Model of Commitment (Rusbult, Martz & Agnew, 1998) in Part I ($\alpha = 0.81$) and Part II (postmanipulation; $\alpha = 0.81$). Seven items that were rated on a scale of 0 (do not agree at all) to 8 (agree completely) assessed the extent to which participants want and anticipate their relationship to be long lasting, how upset they would feel if the relationship was over, their level of attachment, their belief they will be with another partner within the year, and their hopes of always being with their current partner. Two items were reversed scored ("I would not feel very upset if our relationship were to end in the near future"; "It is likely that I will date someone other than my partner within the next year"), then averaged to create a single index of commitment, with higher scores indicating greater commitment.

4.2.3.3 Positive and Negative Affect

Participants completed the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) in Part II. Participants reported on a scale of 1 (very slightly or not at all) to 5 (very much) the extent to which they felt positively (e.g., excited, enthusiastic; $\alpha = 0.75$) and negatively (e.g., distressed, angry; $\alpha = 0.72$) immediately after viewing their compatibility results. Positive and negative affect items (10 of each) were averaged independently, with higher scores indicating greater feelings of the respective type of emotion. An item was added to the scale ("threatened") to directly assess how threatened participants were feeling. This item was analyzed independently. Prior to being shown the PANAS, participants were also invited to describe how they felt immediately after viewing their compatibility results in an open-ended format ("Take a moment to reflect on how you feel about your results. What are your initial thoughts, feelings, and/or physiological reactions?").

4.2.3.4 Compatibility Questionnaire

A compatibility questionnaire was created consisting of approximately 33 items that participants completed for themselves and their current partner as well as themselves and a real potential alternative partner of their choice. To increase believability, we aimed to include questions that tapped into various aspects of romantic relationships, including trust and communication,

97

problem solving, similarity of interests and personality, alignment of values and beliefs, and physical and sexual attraction. In addition, we included some items that embraced common poppsychology concepts, such as astrological sign and love languages. We designed questions to have a variety of response formats, including multiple choice, rating scales, sliding scales, ranking preferences, select-all-that-apply and open-ended responses to make it difficult to determine how compatibility scores may be "calculated" by the algorithm. The compatibility questionnaires set up the deception of the experimental manipulation. The questionnaires were not validated, scored in any way, or used in any analyses.

4.2.3.5 Perceptions of the Compatibility Questionnaire/Algorithm

Participants completed a short 6-item scale created by the researchers in Part I ($\alpha = 0.75$) and Part II ($\alpha = 0.82$) that assessed their perceptions of the legitimacy of the compatibility questionnaire, algorithm, and the researchers. Items were rated on a scale from 1 (strongly disagree) to 5 (strongly agree), with an additional option labeled "unable to determine" that was not included in the analyses. Two items were reversed scored ("It is difficult to believe that the algorithm has shown to be 'highly accurate' in preliminary tests"; "The algorithm needs to be significantly revised before being used in more formal contexts"), then averaged to create a single index of perceptions of the questionnaire and algorithm, with higher scores indicating greater perceptions of legitimacy.

4.2.3.6 Perceived Compatibility Estimates

Participants reported their perceived levels of compatibility with their partner and their alternative using a sliding scale that ranged from 0% to 100%, with higher percentages indicating greater compatibility. All participants reported this information in Part I. In Part II, the control group only reported their perceived compatibility with their partner, while the participants in the experimental conditions reported their perceived compatibility with their partner and their alternative.

4.2.3.7 Perceptions of Accuracy of Compatibility Results

Participants reported how accurate they perceived their compatibility results to be with their partner and their alternative using a sliding scale that ranged from 0% to 100%, with higher percentages indicating greater accuracy. Participants in the control group only reported perceived accuracy of compatibility results with their partner, while the participants in the experimental conditions reported the perceived accuracy of the compatibility results with their partner and their alternative.

4.3 Results

The analytic plan was preregistered before data collection began (<u>https://osf.io/bkvy3</u>). Assumptions of the employed statistical tests were assessed prior to conducting the confirmatory analyses and planned exploratory moderations. Violations of assumptions are described where necessary. Deviations from the preregistration are transparently noted, justified, and detailed supplementary materials are available for in-depth review of the assumption assessments.

4.3.1 Descriptive Statistics

Descriptive statistics of each measure by experimental condition and across the full sample are provided in Table 5. Commitment measured at both Part I and Part II was approaching the maximum value of the scale. On average, participants rated their perceived compatibility with their partners as higher than their perceived compatibility with their alternatives at Part I and Part II. Perceptions of the questionnaire/algorithm at Part I and II were similar, with perceptions emerging around the midpoint of the scale. Participants perceived the accuracy of their compatibility results with their partners to be higher (~66%) than that of their alternative results (~56%). Finally, feelings of positive affect, negative affect, and threat approached the minimum of the scale.

Measure			Condition			Condition Differences
	ALT_High (N = 48)	ALT_Avg (N = 48)	$\begin{array}{l} \text{ALT_Low} \\ (N = 54) \end{array}$	P_Only (N = 57)	All $(N = 207)$	

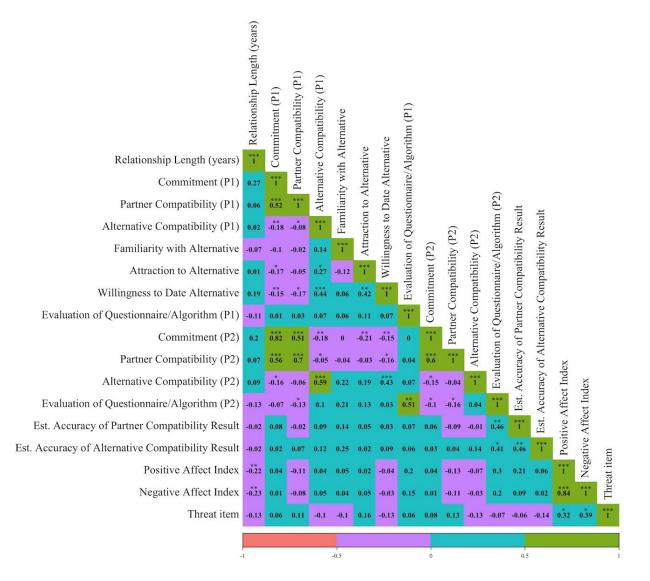
Table 5. Descriptive Statistics of Study Measures

Commitment†	7.22 (.87)	7.28 (.93)	7.16 (1.36)	7.23 (1.13)	7.22 (1.09)	F(3, 203) = 0.09, p = .97
Partner Compatibility†	82.12 (15.36)	84.52 (9.71)	82.48 (14.20)	86.07 (12.76)	83.86 (13.20)	F(3, 203) = 1.05, $p = .37$
Alternative Compatibility†	60.21 (18.73)	56.83 (17.24)	55.87 (16.77)	57.70 (19.55)	57.61 (18.07)	F(3, 201) = 0.52, p = .67
Perception of Questionnaire & Algorithm†	3.62 (.52)	3.43 (.73)	3.44 (.66)	3.50 (.66)	3.49 (.65)	F(3, 199) = 0.83, $p = .48$
Part II						
Commitment	7.16 (.91)	7.26 (.81)	6.88 (1.55)	7.08 (1.21)	7.09 (1.17)	
Partner Compatibility	83.98 (9.67)	85.67 (11.17)	83.54 (12.72)	86.11 (9.47)	84.84 (10.83)	
Alternative Compatibility	58.48 (22.33)	53.52 (19.85)	46.35 (16.53)		52.53 (20.11)	
Perception of Questionnaire & Algorithm	2.97 (.81)	2.98 (.88)	3.33 (.70)	3.29 (.75)	3.15 (.80)	
Est. Accuracy Partner Compatibility	59.81 (22.79)	61.90 (21.08)	71.87 (18.35)	69.53 (19.94)	66.12 (20.97)	
Est. Accuracy Alternative Compatibility	45.79 (28.11)	46.73 (26.53)	74.65 (24.34)		56.48 (29.48)	
Positive Affect	1.73 (.67)	1.79 (.64)	1.82 (.48)	1.78 (.60)	1.78 (.60)	
Negative Affect	1.86 (.69)	1.75 (.52)	1.80 (.40)	1.69 (.49)	1.77 (.53)	
Threat Item	1.62 (1.03)	1.42 (.71)	1.26 (.52)	1.33 (.55)	1.40 (.72)	

Note. Commitment scores ranged from 0 to 8, with higher scores indicating greater commitment. Compatibility estimates and perceived accuracy estimates ranged from 0% to 100%, with higher scores indicating greater compatibility and accuracy, respectively. Perception of Questionnaire/Algorithm scores ranged from 1 to 5, with higher scores indicating greater positive perceptions. Positive and negative affect and the threat item scores ranged from 1 to 5, with higher scores indicating greater indicating greater endorsement of the respective type of affect. Variables marked with a cross (†) indicate it was included in the random assignment assessment.

To complement the descriptive information, the correlations between the main measures are presented in Figure 8. Commitment, estimated compatibility with partner and alternative, and evaluation of the questionnaire/algorithm measured at Part I were all significantly and strongly correlated with their respective complements in Part II. In addition, all measures of commitment and partner compatibility were strongly positively correlated. Negative affect was also found to be strongly and significantly positively correlated with positive affect. Commitment (Part I) negatively correlated with estimated alternative compatibility (Part I and II), attraction to the alternative and willingness to date the alternative but was not correlated with familiarity with the alternative. Similarly, partner compatibility (Part I) was negatively correlated with estimated alternative and willingness to date the alternative but was not correlated with estimated alternative alternative compatibility (Part I) was negatively correlated with estimated alternative compatibility (Part I) was negatively correlated with estimated alternative compatibility (Part I) was negatively correlated with estimated alternative compatibility (Part I) was negatively correlated with estimated alternative compatibility (Part I) was negatively correlated with estimated alternative compatibility (Part I) was negatively correlated with estimated alternative compatibility (Part I) was negatively correlated with estimated alternative compatibility (Part I) was negatively correlated with estimated alternative compatibility (Part I) was negatively correlated with estimated alternative compatibility (Part I) only) and willingness to date the alternative, but not attraction to

of familiarity with the alternative. And, estimated alternative compatibility (Part I) was positively correlated with attraction to and willingness to date the alternative, but was not significantly correlated with familiarity with the alternative. Perceived accuracy of the partner and alternative compatibility results were significantly correlated with each other, and both estimates were positively correlated with perceptions of the questionnaire/algorithm at Part II.





Note. * p < .05, ** p < .01, *** p < .001.

4.3.2 Efficacy of Random Assignment

An essential methodological element of experimental designs is random assignment. The random assignment of participants to the various levels of the manipulated variable is to "ensure" that participants do not differ significantly on certain relevant characteristics. If a third variable is systematically associated with the manipulated variable, this represents an alternative explanation for differences in the outcome variable between groups (Warner, 2013). To assess the extent to which random assignment was effective in this experiment, chi-square and univariate ANOVAs were conducted on a selection of variables measured during Part I to identify any potential significant differences between the experimental conditions. The variables that were included in this assessment are marked with a cross in Tables 3, 4 and 5.

The numeric variables were assessed with univariate ANOVAs. The assumptions of univariate ANOVA require that the independent variable is categorical, observations are independent, dependent variables are continuous (interval or ratio scale) and approximate a normal distribution, and that conditions have approximately equal variances (homogeneity of variance). The categorical variables were assessed with chi-squares. The assumptions of chi-square require that the observations are independent, group sizes are not extremely unequal, and cells should not have small expected cell frequencies. Specifically, analyses with more than 20% of expected cell counts less than five and expected counts less than one were removed and the analysis was rerun with the remaining categories (Yates et al., 1999). If the implementation of these criteria resulted in only one remaining category, the analysis was not run. While the possibility exists to collapse counts across category options to create larger expected and observed counts, this approach was not employed. The original categories were created because they were considered distinct and collapsing across categories without sufficient justification for which categories to consider similar enough to combine may result in muddled interpretations.

Both univariate ANOVAs and chi-squares require that observations must be independent of one another. While participants were only assigned to a single experimental condition, one potential threat to the assumption of independence is participants whose romantic partners also participate in the study, meaning that both partners from a couple are included in the data set. There were 32 (15.46%) participants who reported their partners also participated in this study (5 in the high

102

threat condition, 7 in the average threat condition, 11 in the low threat condition, and 9 in the control condition. A chi-square test across conditions was not statistically significant ($\chi^2[3] = 1.96$, p = 0.58), meaning the number of participants whose partners participated in the study were distributed evenly across conditions. In other words, random assignment was effective.

Results of the univariate ANOVAs and chi-square analyses revealed no significant differences between experimental conditions (see Tables 3, 4, and 5). Accordingly, random assignment was effective in creating groups of participants that did not differ on the selected demographic variables of interest.

4.3.3 Confirmatory Hypotheses and Moderation Explorations

Each preregistered confirmatory hypothesis was analyzed with a univariate ANOVA. Descriptive statistics are available in Table 5. Bonferroni-Holm adjustments were applied separately to the family of univariate ANOVAs and the family of pairwise comparisons to minimize the probability of Type I error. The Bonferroni-Holm adjustment involved ordering obtained *p*-values from the four confirmatory univariate ANOVAs from smallest to largest. An alpha of 0.05 was divided by the number of analyses and compared to the smallest obtained *p*-value. If the *p*-value was lower than the adjusted alpha, the hypothesis was determined to be statistically significant. The next smallest *p*-value was then compared to 0.05 divided by one less than the total number of analyses. Again, if the obtained *p*-value was less than the adjusted alpha, the hypothesis was said to be statistically significant. The remaining hypotheses were tested in the same way until a hypothesis failed to reach significance. When this occurred, the remaining hypotheses were said to have inadequate strength to be considered statistically significant.

Pairwise comparisons were only conducted when a significant main effect was found. The obtained pairwise comparison *p*-values were not subjected to any adjustment (e.g., Tukey, Scheffe) because the Bonferroni-Holm adjustment was preregistered as the method to limit the possible inflation of Type I error due to multiple testing. Pairwise comparisons that were not preregistered that emerged as significant are reported, however, their respective *p*-values were not adjusted and should be interpreted with caution.

Following each confirmatory analysis, exploratory moderated regression analyses were conducted. For each dependent variable (negative affect, feelings of threat, evaluation of the compatibility questionnaire and algorithm post-manipulation, and perceived accuracy of the alternative compatibility result), the moderated regression analyses were conducted in a hierarchical fashion, with the experimental condition and moderator entered together at the first step, followed by the addition of the interaction term in the second step. This revealed the predictive utility of the interaction terms controlling for the independent influences of the experimental condition and moderator entered together at moderator entered in the first step. A Bonferroni-Holm correction was not utilized for this family of analyses due to its exploratory nature.

For brevity, only significant moderations are described in detail. However, summaries of each step and model for every dependent variable and moderator combination are available for review on the OSF (<u>https://osf.io/34xnt/</u>).

4.3.3.1 Assumption Checks

Assumption checks for the univariate ANOVAs were conducted. The potential violation of the assumption of independence was already identified. Normality was assessed by examining values of skewness and kurtosis while homogeneity of variance was assessed through the Levene's test. Deviations from normality were considered acceptable if the respective Levene's test was non-significant. If the assumption of normality and homogeneity of variance were violated, transformations were employed in attempt to alleviate the violations as detailed in the preregistration. To supplement the manual assumption checks, the 'gylma' package (Global Validation of Linear Models Assumptions; Pena & Slate, 2019) in R was also employed. The package contains a function that assessed skew, kurtosis, heterscedasticity, and the quality of the linear relationships between variables. The results of the Levene's test and the gylma output were considered to determine which type of transformation was most appropriate to use for the analyses. If transformations were not effective in satisfying all assumption checks the transformation with the least violations and/or least severe violation was used in both the confirmatory and exploratory moderation analyses. While efforts were made to correct violated assumptions, remaining violations were not considered a serious cause for concern because univariate ANOVAs are relatively robust as long as the conditions have sample sizes that are not extremely unequal (Field, 2013). A chi-square test assessing whether the count of each condition significantly differed from the count expected if the sample was evenly divided between conditions revealed no such differences, $\chi^2(3) = 1.17$, p = 0.76. Therefore, the condition sample sizes are not considered to be extremely unequal. This analysis was not preregistered but does justify conducting the confirmatory analyses despite one or more assumptions remaining unsatisfied after applying appropriate transformations.

To increase the confidence in the results of a moderated regression analysis, multiple assumptions must be satisfied. Scores on all quantitative variables should approximate a normal distribution, as should the residuals, associations for all pairs of variables should be linear, there should not be any extreme univariate or multivariate outliers, the outcome variable must be quantitative, homoscedasticity should be established, and the independent variables should not be highly correlated with each other (multicollinearity). Normality of the quantitative variables was assessed by examining skewness and kurtosis, creating histograms and QQ plots of the residuals, and conducting Kolmogorov-Smirnov tests of normality. Observations that elicited a Cook's distance value close to one or that emerged as statistically significant at the p < .01 level in the Bonferroni outlier test were flagged as outliers (the former assessed the extent to which each observation affects the regression coefficients, while the latter assessed the extent to which the observation affects the mean). All outcome variables are quantitative. Homoscedasticity and linearity were assessed using the 'gvlma' package, and multicollinearity was assessed by examining the variance inflation factors (VIF), with values exceeding 10 considered problematic (not including the VIF values of interaction terms which naturally have high VIFs; Allison, 2022). Violations of the assumptions are identified in-text where necessary, and the respective steps (if any) taken to alleviate the violations are transparently described. The complete output of the assumption checks for each moderated regression analysis is available on the OSF (https://osf.io/34xnt/).

Preliminary review of skewness and kurtosis levels for the moderator variables revealed issues with normality. Transformations intended to mitigate the violations of normality were enacted prior to checking the assumptions detailed above. Relationship length was moderately positively skewed and commitment and perceived partner compatibility at Part I were moderately negatively skewed. Square root transformations successfully coerced each of these variables to

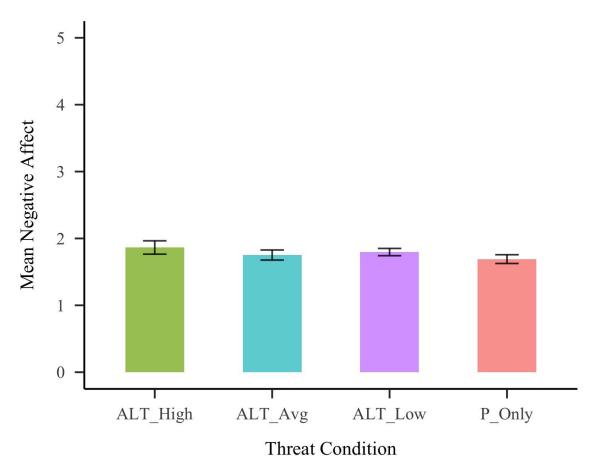
105

have acceptable levels of skewness and kurtosis. The transformed versions of the moderators were used in the following analyses.

4.3.3.2 Negative Affect

Skewness and kurtosis values were within an acceptable range for negative affect, but the Levene's test was violated, F(3, 203) = 4.35, p = 0.005 and gvlma results indicated skewness was abnormal. The scores were subjected to both a square root transformation and a log transformation, but neither resulted in a non-significant Levene's test. However, the results of the Levene's test of the log-transformed scores resulted in the lowest *F*-value, F(3, 203) = 3.20, p = 0.024. Additionally, the gvlma assessment indicated all assumptions were satisfied for log-transformed scores, therefore, these scores were used in the analysis. Negative affect was hypothesized to significantly differ across conditions, specifically with participants in the high-threat condition reporting greater negative affect than participants in the low-threat condition. However, no significant difference emerged, F(3, 203) = 0.90, p = 0.443 (see Figure 9).

Figure 9. Differences in Negative Affect Across Threat Condition



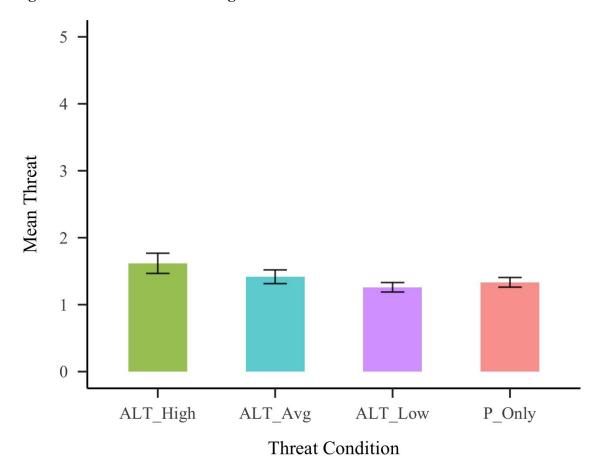
Note. Error bars represent standard error.

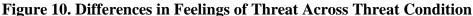
All assumptions were satisfied for the exploratory moderation analyses with negative affect as a dependent variable (see <u>https://osf.io/jhby4/</u> for full output). No statistically significant moderations emerged. Output documents that include detailed results of the moderation analyses are available on the OSF (<u>https://osf.io/dvskj/</u>).

4.3.3.3 Threat

Skewness and kurtosis values were outside of the preregistered acceptable range for the threat item, and the Levene's test was violated, F(3, 202) = 6.23, p < .001. The gvlma assessment indicated skewness, kurtosis and linearity were violated. The scores were subjected to a square root transformation and a log transformation to determine which transformation would be most effective in satisfying the assumptions. Both transformations decreased the extent of skew and kurtosis but the Levene's tests remained violated, with the log-transformed scores eliciting the lowest *F*-value, F(3, 202) = 6.16, p = 0.001. Additionally, the log-transformed scores elicited the

fewest violations of assumptions according to the gvlma assessment, with skewness and linearity remaining violated. Consequently, the log-transformed scores were used for the analysis. Feeling "threatened" was hypothesized to significantly differ across conditions, specifically with participants in the high-threat condition reporting greater feelings of threat than participants in the low-threat condition. However, no significant difference emerged, F(3, 202) = 1.72, p = 0.163 (see Figure 10).





The assumption checks for the exploratory moderations revealed notable violations (see https://osf.io/ae2bc/). For each moderation model, skewness was abnormal and all models except that of relationship length also demonstrated violations of linearity. These violations were observed through the gvlma analyses, Kolmogorov-Smirnov tests and visual inspection of the histograms and QQ plots of the residuals. Floor effects in reports of feeling threatened may be the cause of the violations. Moderations were run as planned, but the results should be

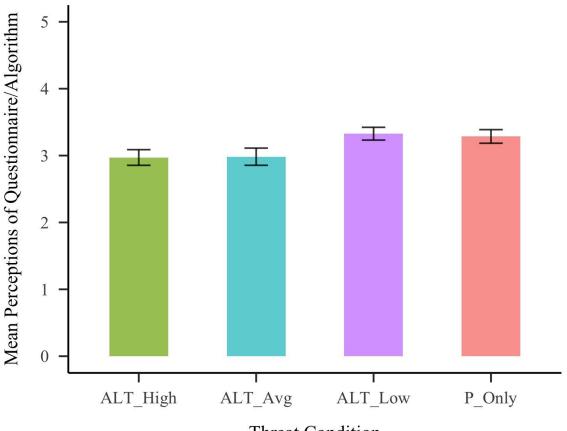
Note. Error bars represent standard error.

interpreted with caution. No significant moderations emerged. Output documents that include detailed results of the moderation analyses with the threat item as a dependent variable are available on the OSF (https://osf.io/u8a5j/).

4.3.3.4 Perceptions of the Compatibility Questionnaire/Algorithm Post-Manipulation

Skewness and kurtosis values were within acceptable range for the perception of the compatibility questionnaire and algorithm. The Levene's test for homogeneity of variance was not violated (F[3, 200] = 0.70, p = 0.551) and all assumptions were satisfied according to the gvlma assessment. Perceptions of the questionnaire and algorithm were hypothesized to significantly differ across conditions, specifically with participants in the high-threat condition reporting greater negative perceptions than participants in the low-threat condition. A significant effect emerged with a medium effect size, F(3, 200) = 3.00, p = 0.032, $\eta^2 = 0.043$ (see Figure 11). Pairwise comparisons revealed the hypothesized difference between the high- and low-threat conditions, t(200) = -2.28, p = 0.024, Cohen's d = 0.453. Two additional significant differences also emerged between the conditions. Participants in the high threat condition reported significantly greater negative perceptions compared to participants in the control condition (t[200] = -2.03, p = 0.044, Cohen's d = 0.401) and participants in the average-threat condition reported significantly greater negative perceptions compared to participants in the low-threat (t[200] = -2.19, p = 0.029, Cohen's d = 0.438). Lastly, a difference between the average-threat and control conditions approached significance (t[200] = -1.94, p = 0.054, Cohen's d = 0.386), with those in the average-threat condition reporting greater negative perceptions.

Figure 11. Differences in Perceptions of the Questionnaire/Algorithm Across Threat Condition



Threat Condition

Note. Error bars represent standard error. Higher scores indicate greater positive perceptions of the questionnaire/algorithm.

All assumptions of moderated regression were satisfied (see <u>https://osf.io/bh95v/</u>). Output documents that include detailed results of the moderation analyses with perception of the algorithm/questionnaire as a dependent variable are available on the OSF(<u>https://osf.io/vnqhx/</u>). Two significant moderations emerged (see Table 6).

 Table 6. Model Comparisons of Moderated Multiple Regression Analyses with Evaluation

 of the Compatibility Questionnaire/Algorithm as the Outcome Variable

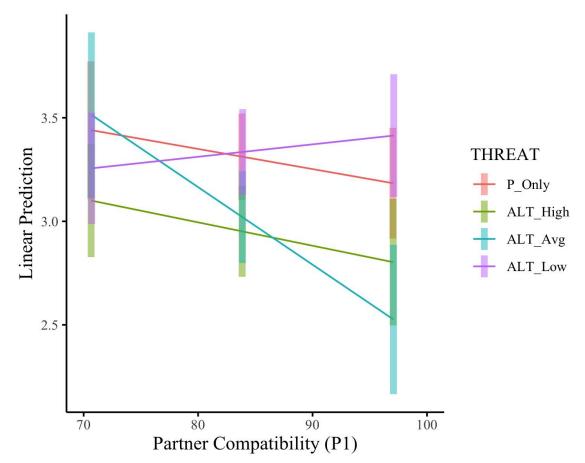
Moderation Models	R^2	Adj. <i>R</i> ²	F	р	ΔR^2	ΔF	р
Partner Compatibility							
Step 1	0.067	0.048	3.55	0.008			
Step 2	0.106	0.074	3.34	0.002	0.039	2.93	0.035
Alternative Compatibility							
Step 1	0.054	0.035	2.81	0.023			
Step 2	0.097	0.064	2.96	0.004	0.043	3.05	0.030
			Mo	derators			
	Pa	rtner Compati	bility	Alt	ernative C	ompatibi	lity

Model Terms	b	Std. Error	t	b	Std. Error	t
Step 1: Main effects						
(Intercept)	2.99	0.17	17.87***	2.98	0.20	14.57***
High Threat	-0.36	0.16	-2.35*	-0.31	0.16	-1.98*
Average Threat	-0.33	0.16	-2.14*	-0.28	0.16	-1.78
Low Threat	-0.002	0.15	-0.02	0.069	0.15	0.46
Moderator	0.084	0.037	2.24*	0.005	0.003	1.65
Step 2: Moderating effect						
(Intercept)	2.96	0.26	11.30***	2.79	0.33	8.57***
High Threat	-0.46	0.41	-1.11	-0.74	0.50	-1.48
Average Threat	-1.05	0.46	-2.29*	0.81	0.51	1.60
Low Threat	0.58	0.39	1.50	0.32	0.49	0.64
Moderator	0.095	0.069	1.38	0.008	0.005	1.54
High Threat x Moderator	0.022	0.10	0.22	0.007	0.008	0.86
Average Threat x Moderator	0.19	0.12	1.60	-0.019	0.009	-2.26*
Low Threat x Moderator	-0.15	0.096	-1.53	-0.004	0.008	-0.52
				~ ~		

Note. Partner compatibility was square root transformed. * p < .05. ** p < .01. *** p < .001. *b* represents unstandardized coefficients.

Perceived partner compatibility at Part I emerged as a significant moderator (see Figure 12). Examination of the simple slopes revealed a significant difference between the average-threat condition and the low-threat conditions (difference = 0.332, t(196) = 2.89, p = 0.022), such that participants who reported moderate levels of compatibility with their partners at Part I perceived the questionnaire and algorithm to be relatively similar across conditions. However, participants who reported higher levels of compatibility with their partners at Part I perceived the questionnaire and algorithm more positively if they were in the low-threat condition compared to those in the average-threat condition.

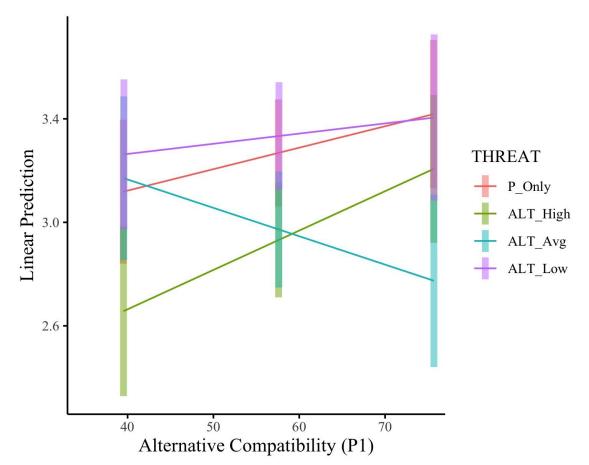
Figure 12. Partner Compatibility (PI) As a Moderator of the Influence of Threat Condition on Evaluations of the Compatibility Questionniare/Algorithm



Note. The raw partner compatibility percentages are displayed rather than the transformed percentages for ease of interpretation. Higher percentages indicate higher perceived compatibility. Percentages are shown at the mean and one standard deviation above and below the mean.

Perceived alternative compatibility at Part I emerged as a significant moderator (see Figure 13). Examination of the simple slopes revealed a significant difference between the high-threat condition and the average-threat conditions (difference = 0.026, t(194) = 2.94, p = 0.019). Participants who reported lower levels of compatibility with their alternatives at Part I perceived the questionnaire and algorithm more positively at Part II if they were in the average-threat condition compared to those in the high-threat condition. In contrast, participants who reported higher levels of compatibility with their alternatives at Part I perceived the questionnaire and algorithm their alternatives at Part I perceived the questionnaire and algorithm their alternatives at Part I perceived the questionnaire and algorithm their alternatives at Part I perceived the questionnaire and algorithm their alternatives at Part I perceived the questionnaire and algorithm their alternatives at Part I perceived the questionnaire and algorithm their alternatives at Part I perceived the questionnaire and algorithm their alternatives at Part I perceived the questionnaire and algorithm more positively if they were in the high-threat condition compared to those in the average-threat condition compared to those in the high-threat condition compared to those in the high-threat condition compared to those in the high-threat condition compared to those in the average-threat condition.

Figure 13. Alternative Compatibility (PI) As a Moderator of the Influence of Threat Condition on Evaluations of the Compatibility Questionniare/Algorithm

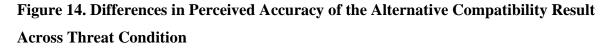


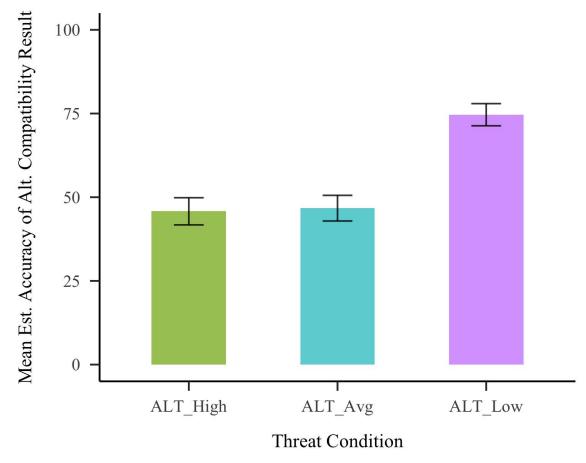
Note. Higher percentages indicate higher perceived compatibility. Percentages are shown at the mean and one standard deviation above and below the mean.

4.3.3.5 Perceived Accuracy of Alternative Compatibility Result

Skewness and kurtosis values were within acceptable range for the perceived accuracy of the "algorithm-generated" compatibility result for the participants' alternatives and the Levene's test for homogeneity of variance was not violated, F(2, 147) = 1.78, p = 0.172. Results of the gvlma analyses confirmed that all assumptions were satisfied. Perceived accuracy of the alternative compatibility result was hypothesized to significantly differ across conditions, specifically with participants in the high-threat condition reporting lower accuracy ratings than participants in the low-threat condition. A significant difference between conditions emerged, F(2, 147) = 20.16, p < .001, with a large effect size $\eta^2 = 0.215$ (see Figure 14). Pairwise comparisons revealed the hypothesized difference between the high- and low-threat conditions, t(147) = -5.35, p < .001, Cohen's d = 1.097. An additional significant difference emerged between the average-threat and low-threat conditions, such that participants in the average-threat condition reported significantly

lower accuracy ratings than participants in the low-threat condition, t(147) = -5.53, p < .001, Cohen's d = 1.062.





Note. Error bars represent standard error.

All assumptions of moderated regression were satisfied (see <u>https://osf.io/sh2tj/</u>). Output documents that include detailed results of the moderation analyses with perceived accuracy of the alternative compatibility result as a dependent variable summarized below are available on the OSF (<u>https://osf.io/mexyb/</u>). Five significant moderations emerged (see Table 7).

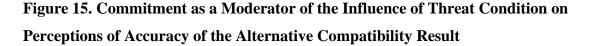
Moderation Mod	lels		R^2		Adj. <i>k</i>	R^2	F		p		ΔR^2		ΔF		р
Commitment															
Step 1			0.216		0.199)	13.41		<.001						
Step 2			0.252		0.226	5	9.68		< .001		0.036		3.43		0.035
Attraction to Al	lternative														
Step 1			0.215		0.199)	13.35		< .001						
Step 2			0.265		0.240)	10.40		< .001		0.050		4.91		0.009
Openness to Da	ting Alter	native													
Step 1	-		0.225		0.209)	14.10		< .001						
Step 2			0.276		0.251		10.98		< .001		0.051		5.11		0.007
Partner Compa	tibility														
Step 1	-		0.221		0.205	5	13.82		<.001						
Step 2			0.256		0.231		9.93		<.001		0.035		3.41		0.036
Alternative Con	npatibility	y													
Step 1	- •		0.238		0.222	2	15.09		< .001						
Step 2			0.347		0.324	ŀ	15.19		< .001		.109		11.93		< .001
-								Moderato	rs						
							On	enness to	Dating						
		Commitr	nent	Attrac	ction to A	lternative	υp	Alternati		Part	ner Com	patibility	Altern	ative Cor	npatibility
Model Terms	b	Std. Error	t	b	Std. Error	t	b	Std. Error	t	b	Std. Error	t	b	Std. Error	t
Step 1:															
Main effects															
(Intercept)	49.78	8.76	5.69***	46.41	10.44	4.45***	37.83	7.66	4.94***	52.79	6.91	7.65***	32.20	7.90	4.08***
High Threat	-0.87	5.39	-0.16	-0.92	5.42	-0.17	-0.71	5.36	-0.13	-0.62	5.37	-0.12	-2.52	5.33	-0.47
Low Threat	27.96	5.23	5.34	27.92	5.24	5.33***	28.10	5.20	5.40***	28.12	5.22	5.39***	27.50	5.17	5.32***
Moderator	-2.39	6.19	-0.39***	0.082	2.51	0.033	2.74	2.05	1.34	-1.56	1.49	-1.05	0.27	0.12	2.19
Step 2:															
Moderating															
effect															
(Intercept)	52.15	15.51	3.36***	13.01	17.67	0.74	16.07	12.41	1.30	43.01	12.52	3.43***	29.47	12.26	2.40*
High Threat	-39.48	22.88	-1.73	20.11	24.23	0.83	11.73	16.94	0.69	-6.32	16.56	-0.38	-36.59	17.05	-2.15*
Low Threat	41.08	19.40	2.12*	95.91	23.28	4.12***	75.76	16.61	4.56***	55.43	15.79	3.51***	70.49	16.83	4.19**
Moderator	-4.26	11.82	-0.36	8.70	4.46	1.95	9.43	3.65	2.59*	0.96	3.09	0.31	0.32	0.21	1.53
High Threat x Moderator	29.70	17.27	1.72	-5.21	6.30	-0.83	-3.75	5.02	-0.75	1.27	3.97	0.32	0.56	0.21	2.02*

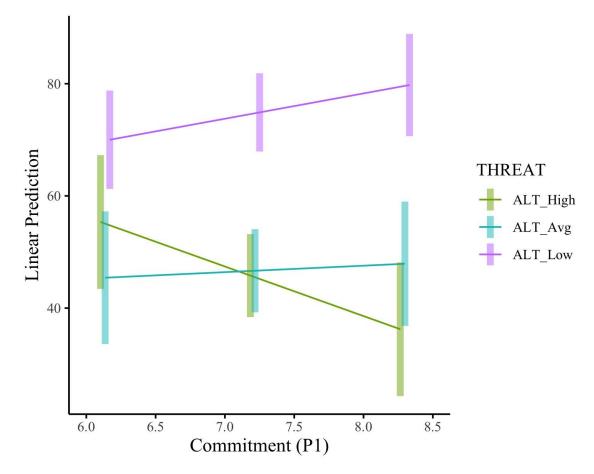
 Table 7. Model Comparisons of Moderated Multiple Regression Analyses with Perceived Accuracy of the Alternative Compatibility Result as the Outcome Variable

Low Threat x	10.12	14.62	0.60	-17.68	5.90	-3.00**	-14.83	/ 01	-3.02**	-6.90	2 0 1	1 0 1	0.77	0.20	-2.69**	
Moderator	-10.13	14.02	-0.69	-17.08	5.90	-5.00	-14.85	4.91	-5.02	-0.90	3.81	-1.81	-0.77	0.29	-2.09	

Note. Commitment and partner compatibility were square root transformed. * p < .05. ** p < .01. *** p < .001. *b* signifies unstandardized coefficients.

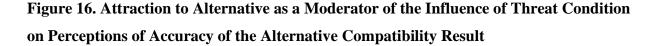
When commitment was explored as a moderator, examination of the simple slopes revealed a significant difference between the high-threat condition and the low-threat conditions (difference = -39.83, t(144) = 2.61, p = 0.027), such that participants who had "lower" levels of commitment rated the accuracy of the alternative compatibility result fairly similarly across conditions. However, of participants who reported high commitment, those who were shown to have low compatibility with their alternatives perceived the alternative compatibility result to be significantly more accurate, compared to participants who were shown to have high compatibility with their alternatives, who perceived the result to be significantly less accurate (see Figure 15).

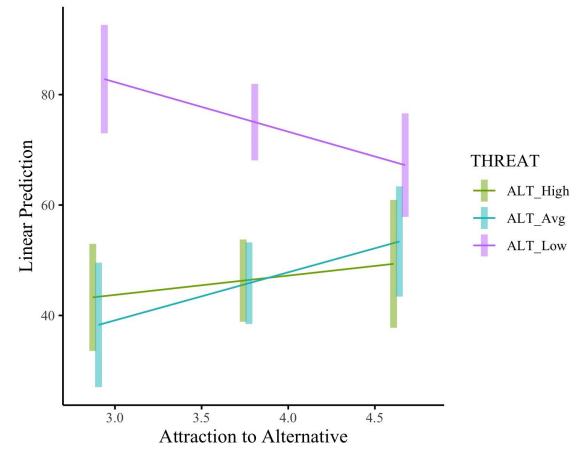




Note. The raw commitment scores are displayed rather than the transformed scores for ease of interpretation. Higher scores indicate higher reported commitment. Scores are shown at the mean and one standard deviation above and below the mean.

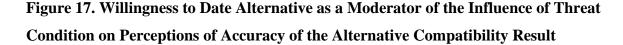
When participant attraction to their alternative was explored as a moderator, examination of the simple slopes revealed a significant difference between the average-threat and low-threat groups (difference = 17.68, t(144) = 3.00, p = 0.009). Participants who reported being more attracted to their alternative rated the accuracy of the alternative compatibility result similarly across condition. However, of participants who reported being less attracted to their alternative, those who were shown to have low compatibility with their alternatives perceived the accuracy of the result to be significantly higher, compared to participants who were shown to have average compatibility with their alternatives, who perceived the accuracy of the result to be significantly lower (see Figure 16).

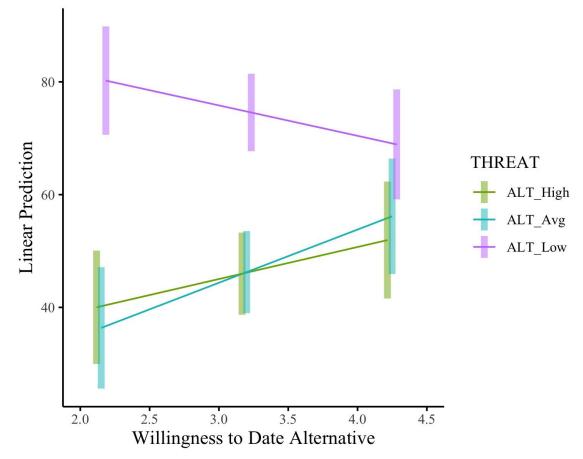




Note. Higher scores indicate higher reported attraction to the alternative. Scores are shown at the mean and one standard deviation above and below the mean.

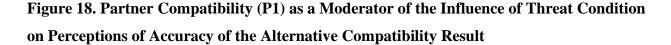
When participants' openness to dating their alternative was explored as a moderator, examination of the simple slopes revealed a significant difference between the average-threat and low-threat groups (difference = 14.83, t(144) = 3.02, p = 0.008), and a difference between the high-threat and low-threat groups that approached significance (difference = 11.08, t(144) =2.32, p = 0.056). Participants who reported being more willing to date their alternative rated the accuracy of the alternative compatibility result similarly across condition. However, of participants who reported being less willing to date their alternative, those who were shown to have low compatibility with their alternatives perceived the alternative compatibility result to be significantly more accurate, compared to participants who were shown to have high or average compatibility with their alternatives, who perceived the result to be significantly less accurate.

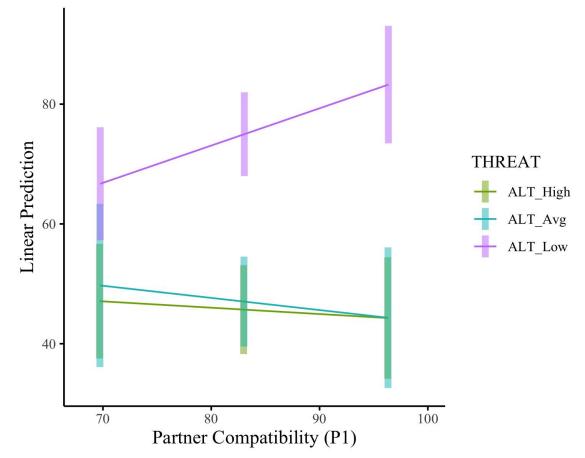




Note. Higher scores indicate higher reported willingness to entering a romantic relationship with the alternative. Scores are shown at the mean and one standard deviation above and below the mean.

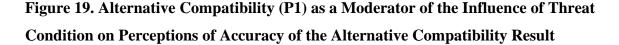
When perceived partner compatibility at Part I was explored as a moderator, examination of the simple slopes revealed a significant difference between the high-threat and low-threat groups (difference = 8.17, t(144) = 2.44, p = 0.042). Participants who reported more moderate levels of compatibility with their partners rated the accuracy of the alternative compatibility result similarly across conditions. However, of participants who reported high levels of compatibility with their partners, those who were shown to have low compatibility with their alternatives perceived the accuracy of the result to be significantly higher, compared to participants who were shown to have high compatibility with their alternatives, who perceived the accuracy of the result to be significantly higher, who perceived the accuracy of the result to be significantly higher, who perceived the accuracy of the result to be significantly higher.

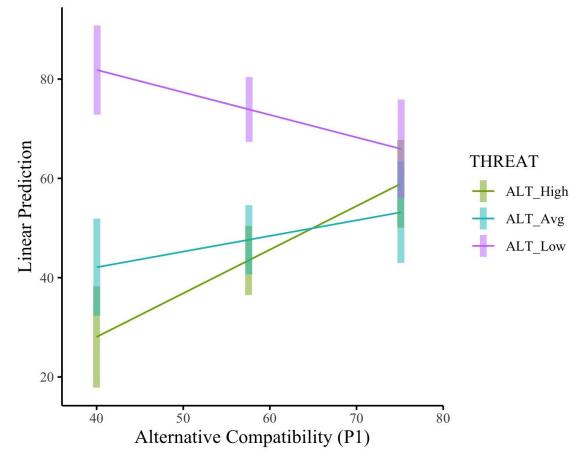




Note. The raw partner compatibility percentages are displayed rather than the transformed percentages for ease of interpretation. Higher percentages indicate higher perceived compatibility. Percentages are shown at the mean and one standard deviation above and below the mean.

When perceived alternative compatibility at Part I was explored as a moderator, examination of the simple slopes revealed a significant difference between the high-threat and low-threat groups (difference = 1.33, t(143) = 4.88, p < .001) and the average-threat and low-threat groups (difference = 0.769, t(143) = 2.69, p = 0.022). Participants who reported higher levels of compatibility with their alternatives at Part I rated the accuracy of the alternative compatibility result similarly across conditions. However, of participants who reported low levels of compatibility with their alternatives, those who were shown to have low compatibility with their alternatives, those who were shown to have low participants who were shown to have high or average compatibility with their alternatives, who perceived the result to be significantly with their alternatives, who perceived the result to be significantly less accurate (see Figure 19).





Note. Higher percentages indicate higher perceived compatibility. Percentages are shown at the mean and one standard deviation above and below the mean.

4.3.3.6 Conclusions of Confirmatory Analyses

The *p*-values of the four confirmatory univariate ANOVAs were ordered from smallest to largest and compared to the adjusted alpha to infer statistical significance (see Table 8). Hypotheses 1 and 2 were not supported. Though Hypothesis 3 emerged as significant below 0.05, once the Holm-Bonferroni adjustment was applied it did not meet the adjusted cut-off for statistical significance and was thus not considered to be supported. Hypothesis 4 was supported, with the predicted difference between the high and low threat groups emerging in the expected direction. An additional group difference emerged as well, with those in the average threat group reporting the alternative compatibility results to be significantly less accurate than those in the low threat group.

Hypothesis	U	nivariate ANO	VAs	Pa	Pairwise Comparisons			
	Adjusted alpha	Obtained <i>p</i> -value	Supported or failed	Adjusted alpha	Obtained <i>p</i> -value	Supported or failed		
H4: Est. accuracy of alternative compatibility result	0.0125	< .001	Supported	0.05	< .001	Supported		
H3: Perceptions of questionnaire & algorithm	0.0167	0.032	Failed					
H2: "Threat" item	0.025	0.163	Failed					
H1: Negative Affect	0.05	0.443	Failed					

Table 8.	Conclus	ions of	Confirmato	ory A	Analyses
----------	---------	---------	------------	-------	----------

4.3.4 Qualitative Analyses

Immediately following exposure to the compatibility results and before any quantitative measures were administered, participants were provided space to share their initial reactions in a free response format. Participants were shown the following prompt: "Take a moment to reflect on how you feel about your results. What are your initial thoughts, feelings, and/or physiological reactions? Please be as descriptive and honest as possible in your response." Only 12 (5.80%) participants did not respond to the prompt.

The responses were coded using an inductive qualitative content analysis approach (Elo & Kyngäs, 2008; Cole, 1988). Content analysis was used to condense and summarize participants' reactions to their compatibility results into categories that could be used to make inferences

about how individuals responded to the threat manipulation. The inductive approach was selected over the deductive approach due to the lack of sufficient preexisting knowledge about how participants respond to a situation in which an attractive alternative may be considered a threat. However, because I was aware of research (such as that described in the introduction to this chapter) that previously identified possible responses that can be expected following a threat, this knowledge likely influenced how I identified, coded, and categorized the responses.

Content analysis has three major phases: preparation, organization, and reporting (Elo & Kyngäs, 2008). During the preparation phase, I decided to rely heavily on the actual, surface level content of the responses as opposed to making assumptions about what the participants were thinking or feeling. However, the participants' experimental condition was available during coding to provide clues to contextualize their responses. During the organization phase, I read through the responses and made observations of patterns or recurring topics which I then formed into codes, sub-themes, and larger themes. I then reread the responses and categorized portions of responses or full responses into each code. Lastly, frequency counts and exemplars of each code from each threat condition (if applicable) were extracted. This information is presented in Table 9 and Table 10 respectively and is further discussed below.

			Threat C	ondition		
Theme	Code	ALT_High $(N = 48, 1 \text{ missing})$	ALT_Avg (N = 48, 3 missing)	ALT_Low $(N = 54, 3 \text{ missing})$	$\begin{array}{c} P_Only\\ (N=57, \ 5 \ missing) \end{array}$	Total $(N = 207, 12 \text{ missing})$
	Positive Happy, content	2 (0.97%)	6 (2.90%)	8 (3.86%)	6 (2.90%)	22 (10.63%)
	Negative Disappointed, sad, or shocked	8 (3.86%)	10 (4.83%)	12 (5.80%)	8 (3.86%)	38 (18.36%)
	Anxious, insecure	0 (0.00%)	1 (0.48%)	2 (0.97%)	1 (0.48%)	4 (1.93%)
Affective	Uncomfortable	4 (1.93%)	2 (0.97%)	0 (0.00%)	1 (0.48%)	7 (3.38%)
Affe	Anger, disgust	2 (0.97%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (0.97%)
	Neutral Surprised, interested (valence unknown)	15 (7.25%)	14 (6.76%)	14 (6.76%)	13 (6.28%)	56 (27.05%)
	Indifferent, unaffected	6 (2.90%)	8 (3.86%)	7 (3.38%)	8 (3.86%)	29 (14.01%)
	Physiological Response	3 (1.45%)	1 (0.48%)	1 (0.48%)	0 (0.00%)	5 (2.42%)
	Accuracy Accurate	5 (2.42%)	12 (5.80%)	28 (13.53%)	17 (8.21%)	62 (29.95%)
	Inaccurate	24 (11.59%)	22 (10.63%)	25 (12.08%)	22 (10.63%)	93 (44.93%)
Cognitive	Defensive Rationalization	10 (4.83%)	14 (6.76%)	14 (6.76%)	15 (7.25%)	53 (25.60%)
Cogr	Criticism Questionnaire	7 (3.38%)	5 (2.42%)	6 (2.90%)	4 (1.93%)	22 (10.63%)
	Algorithm	6 (2.90%)	8 (3.86%)	3 (1.45%)	4 (1.93%)	21 (10.14%)
	Self	3 (1.45%)	5 (2.42%)	1 (0.48%)	1 (0.48%)	10 (4.83%)

Table 9. Frequency Counts of Codes Reported Within Each Group and Within the Full Sample

			Threat Cond	lition	
Theme	Codes	ALT_High $(N = 48, 1 \text{ missing})$	ALT_Avg (<i>N</i> = 48, 3 missing)	ALT_Low $(N = 54, 3 \text{ missing})$	P_Only $(N = 57, 5 \text{ missing})$
	Positive Happy, content	Having a compatibility score over 75% for two completely different beings seems like a good thing to me	Happy, relieved that my current partner is better than the other person	I felt secure with the results of the study	I am content with the results.
ee ve	Negative Disappointed, sad, or shocked	I am a little shocked. I understood my attraction for the "potential" person, but I	A bit disappointed my partner and I aren't more compatible	36% for my potential partner was quite shocking	but a little sad it's only "moderately compatible"
		didn't believe and almost don't believe it! I have more compatibility with them than my current partner.	I'm shocked at how similar the results are between the two (partner & potential partner)	I am surprised that I am only moderately compatible with my husband!	I'm kind of shocked to see the results
	Anxious, insecure		At first, I felt insecure because I thought "only 70%?" as if it were an indication that our relationship is not going to last because of this	seeing that we were only moderately compatible triggered a bit of anxiety that maybe our relationship is lacking something	I was really worried for a moment thinking that I needed to be more compatible or something
Affective	Uncomfortable	I felt kind of weird about the potential partner stuff It's mildly distressing at first	I also feel a bit icky about the naming of those – partner and potential partner- made me feel like I wascomparing the two		When I saw the results, I was a bit nonplussed
	Anger, disgust	I felt a bit disgusted actually	I feel uncomfortable with the results and unsatisfied		
	Neutral Surprised, interested (valence unknown)	I feel upset and defensive I am surprised at the difference between my compatibility with my actual partner versus with my potential partner	I am surprised that the results were similar for current and prospective partner	I am somewhat surprised that the compatibility with [alternative] is this low as several individuals in my	I initially was surprised that we weren't more compatible I was surprised that our
		I was surprised at both results	It surprises me that my partner is low	friend circle remark that we would be a better match	compatibility score was as high as it was

Table 10. Exemplars Extracted from the Content Analysis Divided by Condition

I'm surprised by how accurate it was

Indifferent, unaffected	About what I expected Doesn't change how I view my partner	I'm not bothered by the results I am not too surprised! So I do not really mind the results too much	I felt nothing emotionally or physically upon seeing these results I know that it doesn't change anything in our relationship	The results don't mean a lot to me I am not surprised I don't know how to feel
Physiological Response	My heart dropped haha! Mouth hanging open, furrowed brow, narrowed eyes My initial feelings were being taken aback and taking a deep breath	My heart sank a little	I laughed out loud when I saw [the results]	
Accuracy Accurate	I would say they are relatively true in terms of how accurate they are to our relationship dynamic There are parts of my results that did not surprise me, particularly in "Communication and Trust" and "Interests and Personality." I felt like those were fairly accurate	I was surprised that some of my scores were higher for the non- partner person, but overall, they scored less compatibility which adds up, in my opinion My partner and I being 76% seems accurate	I am already aware of the compatibility of my relationship and felt this reflected the dynamics well For the other person, I agree we would be less romantically compatible	About moderately compatible sounds rights The results are overall realistic

Inaccurate	Based on these results I do not believe that they are accurate I don't really believe the results are accurate in my opinion due to my wanting to be with my current partner way more than I would every want to be with the one I am not dating I was surprised to see greater compatibility with my potential partner because I don't think we have that much in common or get along as well compared to my current partner	Feel it is inaccurate. We are a really great couple. We have challenges but really seem to work well at solving together. I also see myself as more compatible with my partner, specifically in terms of values, than the results seem to suggest I would say attraction for both of them are much higher than the results say	All of the scores were lower than expected. Particularly, with my romantic (not hypothetical) partner. I feel that compatibility with my current partner should be highest in trust and honesty I thought I had higher compatibility with the person who isn't my partner, but the numbers showed the opposite	I also didn't think attraction would be that highnot that I don't find my partner attractivebut that it would be higher than the communication/trust result I have been with my partner for almost 3 years and can easily envision a future with them, and therefore expected a very high to almost perfect compatibility
Defensive Rationalization	My potential partner I know actively wants to date me so that may have increased their results since I think they have high interest in me. My current partner and I have been together almost X years and we are long distance so I can see how I got these results I can understand how I have chemistry with the potential partner, but I also know it would not work out for practical reasons because at the end of the day people only need to be compatible enough, there's no use dwelling on every minute different path our life could have taken that may be better	We do have our differences that may be considered "incompatible" but, so far, I think they don't negatively affect our relationship and might even strengthen it with our own interests/hobbies/values But we are also in a long-distance relationship so that has to be an influential factor and we have only been together a short time Sometimes our differences make the relationship more interesting!	After that I thought an average of 70 something is a great score, too. Because for a relationship to work, people shouldn't be 100% compatible. We additionally are continuing to figure out each other's communication styles and are always developing new interests I don't think compatibility as a numeric is the only thing that matters and is not the sole influence of whether or not we will grow as a couple, and be happy together in the future	But after thinking about the types of questions that were asked, I realized we likekly valued things in a pretty difference order. And it's something to accept in a partner. I also think that if you completely agree on all beliefs in life there is no challenge that helps you grow together. I understand even in relationships that work well it is unlikely to ever be fully compatible and am glad it was at least moderate and not low

Criticism Questionnaire	I don't think it's possible to assess compatibility of a relationship based on the questions asked in this survey I don't know how the questions asked could have computed to compatibility. I think more things could have been taken into account in terms of what the relationship with the current partner is like. I feel as if the questions that were asked did not do my partner and I justice	I don't think the survey really captured the intricacies and dynamic elements of relationship that's been ongoing for 3+ years. Some of the questions in the survey were far too broad to answer accurately. Some questions felt irrelevant to compatibility (astrological signs) or just too general (many people probably look for similar traits in relationship such as trust)	I don't know how accurate they are simply because I was not asked enough questions about our day-to- day life, how we met I think good marriages changes over time and rely on more factors than those accounted for I question the ability of this study to fully capture compatibility if both partners are not present.	I don't think the survey gave enough or the proper questions to accurately judge this It's hard to determine attraction/compatibility with 1 partner completing a multiple- choice survey
Algorithm	Doesn't really change anything cause this algorithm doesn't know anything about me or my relationship or my partner I question if an algorithm can	I feel like an algorithm cannot really determine to totality of my compatibility with another It's hard to understand compatibility in terms of	I'm curious how it was calculated, the results don't feel like something that should be able to be measured in a percentage	I feel like the test's definition of compatibility is the same person whereas mine is whether the two people make sense in a relationship
	determine something like that Lots of compatibility measures simply cannot be measured!	percentages. What does it mean to be 76% compatible? It's just an algorithm	I feel like the algorithm significantly underestimated my compatibility with my partner	It's wrong it doesn't work. I would like to learn more about the program's methods
Self	When completing answers for the other person (not my partner) I might have been overzealous about my answers because I don't know the other person as well as I know my partner	My answers regarding my "hypothetical partner" were mostly guesswork	it may be due to some hesitancy when answering questions indicating someone else's feelings. I wouldn't want to be too assertive with how I thought the other individuals felt	After doing some honest reflection, I think the lower compatibility has to do with me and my personality. There are things I would like to work on with respect of myself that I think would impact my compatibility in any relationship

Note. Minor spelling mistakes and grammar were corrected. Initials of the partner or alternative provided in the responses were changed to [partner/alternative] to protect anonymity.

Overall, 424 responses or response fragments were coded. Two major themes emerged: affective responses (38.45%) and cognitive evaluations (61.56%).

Affective responses were divided into positive (13.50% of the affective responses), negative (31.29%), neutral (52.15%) and physiological responses (3.07%). Positive affect was less common and was encapsulated by fewer positively valenced emotions than negative affect. The negative affect sub-theme included a higher number of distinct emotions, with disappointment, shock, and sadness being most commonly reported in contrast to anxiety, discomfort and anger. Participants in the high threat condition reported the least frequent amount of positive affect, with higher and similar frequencies reported for the remaining groups. Notably, when the negative affect codes were collapsed together, participants in the high, average, and low threat conditions reported similar frequencies of negative affect, with participants in the partner-only condition reporting the least frequent reports of negative emotions. Anger was only reported in the high threat condition, while anxiety and insecurity were present in every condition except the high threat condition. Participants in the partner-only condition reported the least amount of interest or surprise in reaction to their results, while the frequency of participants who reported feeling indifferent or unaffected was relatively equal across conditions.

Cognitive evaluations were further divided into subthemes based on content related to accuracy (59.39% of the cognitive evaluation responses), criticism (20.31%), and defensive rationalizations of the results (20.31%). Participants who commented on the level of accuracy of their results often considered one or both results to only be partially accurate. Defensive rationalization was identified as a sub-theme and code that demonstrated engagement in "mental gymnastics" to justify or explain away the results. Some participants mentioned that being 100% compatible is not realistic or positive, suggesting an attempt to soothe any experienced discomfort after being told they were only moderately compatible with their partners. For example, a participant in the average threat condition said, "we do have our differences that may be considered 'incompatible' but, so far, I think they don't negatively affect our relationship and might even strengthen it with our own interests/hobbies/values." Lastly, criticism that could be considered both destructive and constructive was observed to be directed toward the questionnaire, algorithm, and the participant themselves. For example, a participant responded

129

that "some questions felt irrelevant to compatibility," which is a constructive critique, while another responded that "it's just an algorithm," indicating skepticism or hostility towards the quality and validity of the compatibility questionnaire and "algorithm." Relatedly, some participants questioned their original responses to the questionnaire in Part I (e.g., "I wouldn't want to be too assertive with how I thought the other individuals felt"), while others seemed to reflect on their own characteristics (e.g., There are things I would like to work on with respect of myself that I think would impact my compatibility in any relationship").

It appears that reports of accuracy varied according to the degree of threat, such that participants in the high threat condition were least likely to comment that their results were accurate, followed by comparable reports in the average threat and partner-only conditions, and lastly the low threat condition. Similarly, participants in the high threat condition reported that the results were inaccurate more frequently than did participants in the average and low threat conditions, with participants in the partner-only condition commenting least frequently on the inaccuracy of the results. Interestingly, the percentage of participants who engaged in defensive rationalization was comparable across conditions. Lastly, criticism directed toward the questionnaire, algorithm, and the participant themselves were relatively equally frequent across conditions, with criticism towards the self being less common overall.

Finally, I explicitly looked for evidence in the responses of primary and secondary cognitive appraisals (Lazarus & Folkman, 1984). Approximately 10% of the participants (equally spaced throughout each condition) responded in a manner that reported an initial emotional experience followed by a period of reflection that indicated deeper processing and consideration of the results. For example, a participant in the low threat condition responded, "I immediately feel a bit sad and disappointed...Taking a pause and second thought on these results, I wonder if I jumped too quickly to see moderate [compatibility with partner] as a lack of perfection," while a participant in the high threat condition reported, "It's mildly distressing at first...[but] at the end of the day people only need to be compatible enough." While only identified in a minority of responses, these results offer preliminary evidence that when faced with a threat, reactions occur in waves, with different waves potentially characterized by different cognitive and affective responses.

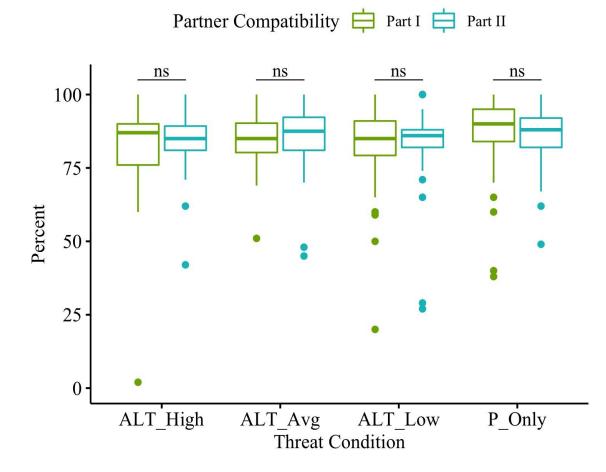
4.3.5 Additional Exploratory Analyses

Additional analyses were conducted that were not preregistered. Assumptions were not assessed and statistical significance was assumed with *p*-values equal to and below 0.05. Due to the number of tests conducted the results should be interpreted with caution because of the risk of inflated Type I error. Readers interested in exploring the data further can access the data files and codebooks on the OSF (https://osf.io/5g78j/).

4.3.5.1 Differences in Measures Between Part I and Part II

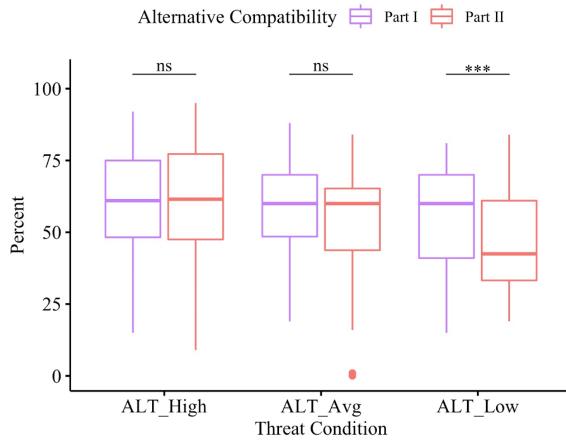
Some measures were administered at both Part I and Part II: estimated partner compatibility, alternative compatibility, commitment, and perceptions of the questionnaire/algorithm. A series of paired *t*-tests were conducted to determine whether any differences were observed within each threat condition between Part I and Part II, with *p*-values adjusted with a Bonferroni correction to account for multiple tests. When differences between estimates of partner compatibility were compared, no significant differences emerged. In other words, exposure to the compatibility results did not significantly affect participants' perceived compatibility with their partners (see Figure 20).

Figure 20. Differences in Estimates of Partner Compatibility Between Part I and Part II Within Condition



However, when differences between estimates of alternative compatibility were compared within threat condition, a significant difference emerged for participants in the low threat condition such that participants perceived their compatibility with their alternatives to be significantly lower than their original estimates after exposure to the compatibility results, t(53) = 3.84, p < .001 (see Figure 21).

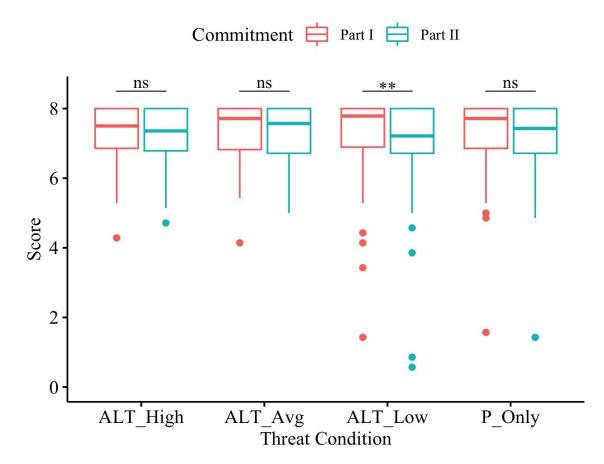
Figure 21. Differences in Estimates of Alternative Compatibility Between Part I and Part II Within Condition



Note. The partner-only control condition is not shown because participants in this condition were only asked to report their perceived alternative compatibility at Part I.

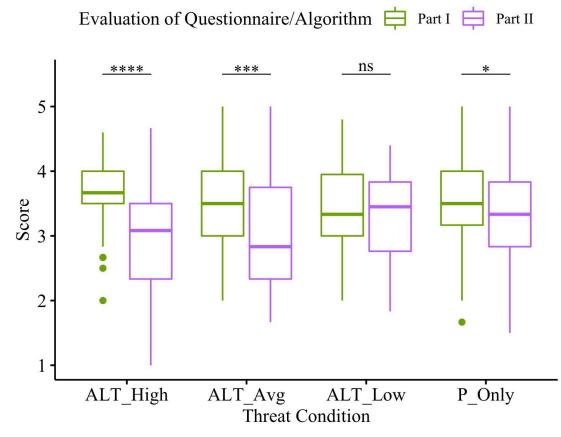
Another significant difference emerged for participants in the low threat condition regarding their reported levels of commitment. Participants reported lower levels of commitment after being exposed to their compatibility results in Part II than they reported during Part I, t(53) = 3.05, p = .004. There were no significant differences in commitment in any condition other than the low threat condition (see Figure 22).

Figure 22. Differences in Commitment Between Part I and Part II Within Condition



Lastly, when evaluations of the compatibility questionnaire/algorithm were explored between Part I and Part II, multiple significant changes were observed. Specifically, participants in the high threat (t[48] = 5.85, p < .001), average threat (t[48] = 3.62, p < .001) and partner-only control condition (t[53] = 2.01, p = .050) reported significantly lower perceptions of the quality of the compatibility questionnaire algorithm after viewing their compatibility results (see Figure 23).

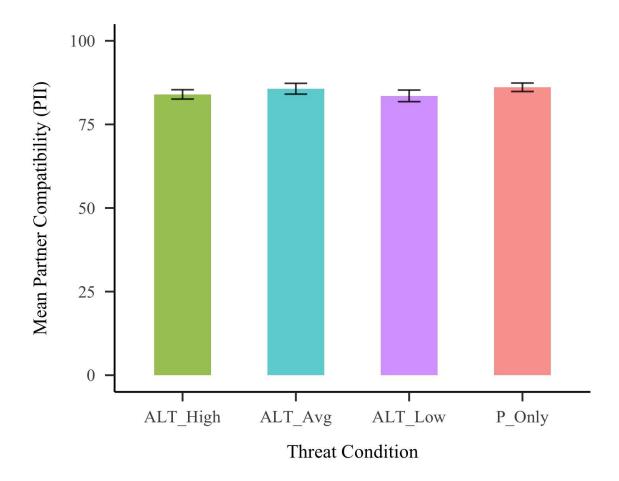
Figure 23. Differences in Evaluations of the Questionnaire/Algorithm Between Part I and Part II Within Condition



4.3.5.2 Estimated Partner and Alternative Compatibility in Part II

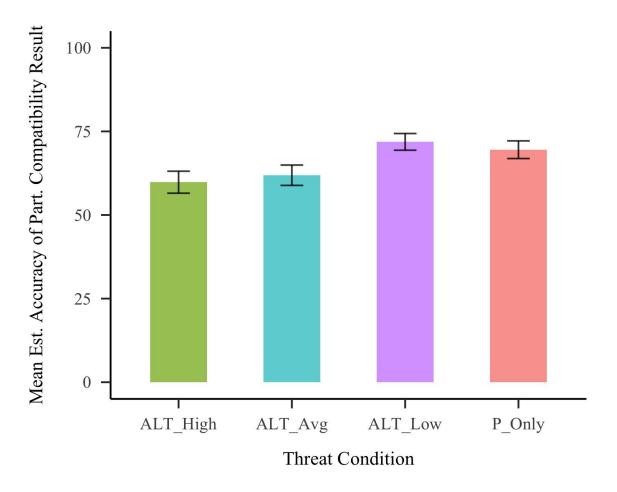
Analyses were conducted to explore whether estimates of partner and alternative compatibility estimates differed across threat condition after participants viewed their compatibility results. In addition, perceived accuracy of the partner compatibility result was explored. A one-way ANOVA revealed no significant differences in estimated partner compatibility across threat condition, F(3, 202) = 0.71, p = 0.548 (see Figure 24).

Figure 24. Differences in Partner Compatibility at Part II Across Threat Condition



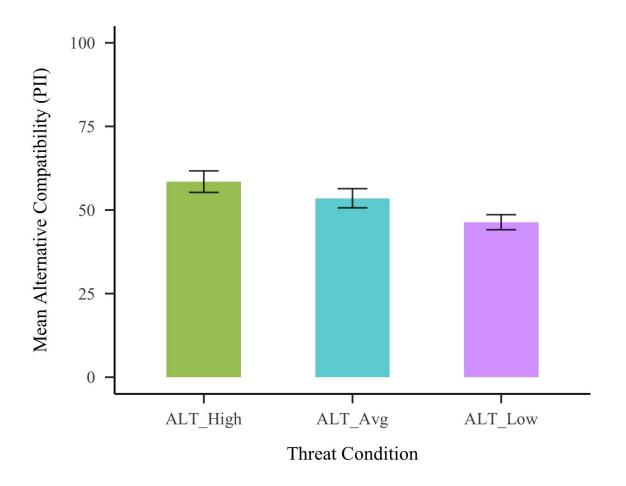
A one-way ANOVA revealed a significant difference in the perceived accuracy of the partner compatibility result (F[3, 203] = 4.13, p = 0.007, $\eta^2 = 0.058$), such that participants in the high threat condition perceived the accuracy of the result to be significantly lower than those in the low threat condition (t[203] = -2.96, p = .018, Cohen's d = -.588; see Figure 25). Two comparisons approached significance. Those in the high threat condition perceived the partner compatibility result to be less accurate than the partner-only control condition (t[203] = -2.42, p = .077, Cohen's d = 0.474) and participants in the average threat condition perceived their results to be less accurate than those in the low threat condition (t[203] = -2.45, p = .077, Cohen's d = -0.476).

Figure 25. Differences in Estimates of Partner Compatibility Across Threat Condition



A one-way ANOVA was conducted to determine whether estimates of alternative compatibility in Part II differed between threat conditions after participants were exposed to their compatibility results. Results revealed a significant difference in estimated alternative compatibility (*F*[2, 147] = 4.96, p = 0.008, $\eta^2 = 0.063$), such that participants in the high threat condition estimated their compatibility with their alternatives significantly higher than those in the low threat condition, t(147) = 3.12, p = .006, Cohen's d = .619 (see Figure 26).

Figure 26. Differences in Alternative Compatibility at Part II Across Threat Condition



4.3.6 Results Summary

Participants in the high threat condition were predicted to report greater feelings of negative affect (H1), threat (H2), and more negative perceptions of the questionnaire/algorithm (H3) than participants in the low threat condition. In addition, participants in the high threat condition were expected to perceive the accuracy of the alternative compatibility result to be lower than participants in the low threat condition (H4). After adjusting the threshold that determines statistical significance for multiple testing, only H4 was supported such that participants in the low threat condition perceived the accuracy of the alternative compatibility result to be significantly more accurate than those in the high threat (planned comparison) and average threat (unplanned comparison) conditions.

Exploratory moderation analyses were conducted to determine whether personal beliefs or relationship characteristics may influence the relationship between threat condition and the

selected outcome variables. Specifically, relationship length, commitment (Part I), perceived partner and alternative compatibility (Part I), and initial perceptions of the questionnaire/algorithm (Part I) were explored as potential moderators. No significant moderations were found for the measures of negative affect or threat. Perception of the questionnaire/algorithm was significantly moderated by partner compatibility and alternative compatibility. Specifically, participants in the low threat condition had significantly more positive perceptions of the questionnaire/algorithm than participants in the average threat condition only when initial partner compatibility was perceived to be high. Initial alternative compatibility moderated perceptions of the questionnaire/algorithm such that when the alternative compatibility result more closely matched the participants' initial estimations, perceptions of the questionnaire/algorithm were more positive. This was observed for only the high and average threat conditions. Perceived accuracy of the alternative compatibility result was significantly moderated by commitment such that only when initial commitment was high, participants in the low threat condition perceived the result to be significantly more accurate than those in the high threat condition. Attraction and willingness to date the alternative both emerged as significant moderators as well, with participants in the low threat condition perceiving the result to be significantly more accurate than those in the average threat condition only when initial attraction and willingness to date the alternative were low (again indicating better alignment between the results and participants' initial beliefs). Lastly, both initial partner and alternative compatibility significantly moderated perceived accuracy of the alternative compatibility result. Only when initial partner compatibility was high did participants in the low threat condition report greater perceived accuracy than those in the high threat condition. In contrast, only when initial alternative compatibility was low did participants in the low threat condition perceive greater accuracy than those in the high and average threat conditions.

Qualitative content analysis of participants' initial reactions after seeing the compatibility results revealed that participants reported affective and cognitive threat responses. Participants rarely reported physiological reactions and positive affective responses were less common than negative affective responses. Neutral or indifferent responses were reported by over half the sample. While more distinct negative emotions were reported (compared to freely elicited positive emotions), the majority of the negative emotions involved feeling disappointed, shocked or sad. Anger, disgust, insecurity and anxiety were rarely reported. Participants in the high threat

condition reported the least amount of positive emotion but reported a comparable amount of negative emotion as those in the other conditions. Notably, participants in the partner-only condition reported the least amount of negative affect. Regarding cognitive evaluations, approximately 60% of the sample commented on the accuracy of their results, while 40% of the sample criticized the questionnaire, algorithm or themselves, or demonstrated effort to make sense of the results through logical thinking and critical reflection. Participants in the high threat condition were least likely to comment that the results were accurate and most likely to comment the results were inaccurate. Participants across conditions engaged in comparable levels of defensive rationalization and criticism. Finally, an additional search for evidence of a dual component cognitive appraisal process revealed 10% of the sample demonstrated an initial, often more emotional reaction followed by a greater sense of acceptance and relief following further consideration.

Exploratory analyses revealed some differences between measures administered at Part I and Part II. Participants in the low threat condition reported lower commitment and estimated alternative compatibility at Part II than Part I. Evaluations of the compatibility questionnaire/algorithm were significantly lower at Part II than Part I for participants in the high, average, and partner-only conditions. Estimates of partner compatibility did not significantly change in any condition between Part I and Part II. Relatedly, at Part II estimates of partner compatibility did not differ across conditions, but participants in the high threat condition believed the accuracy of their revealed partner compatibility result to be significantly lower than that of participants in the low threat condition. Finally, participants in the high threat condition reported estimating their compatibility with their alternative significantly higher than that of participants in the low threat condition.

4.4 Discussion

The primary core assumption of the sub-theory of devaluation of alternatives is that attractive alternative partners are threatening. This assumption is essential to the theory because without the perception of a threat, relationship maintenance mechanisms such as devaluation should not be observed. Despite its importance, existing research does not corroborate this assumption. The

current study aimed to test this assumption using a novel operationalization of threat and multiple measures intended to capture threat perception and responses.

Past research has shown that threats may elicit negative affect (Tomaka et al., 1993; Millburn; 1977; Hodgins et al., 2010) and stimulate physiological processes believed to be associated with threat responses (Hodgins et al., 2010; Bitsios et al., 1996; Henderson et al., 2017). In this study, negative affect was measured in three ways: the well validated negative affect index of the PANAS (Watson et al., 1988), a single item that asked how threatened participants felt, and via a free-response question that asked participants to reflect on their thoughts, feelings and physiological responses. I hypothesized that differences across threat conditions would be found such that participants in the high threat condition would experience greater feelings of threat and negative affect than participants in the low threat group, but no such differences emerged in the quantitative measures. This pattern of findings was further supported by the content analysis of the qualitative question. While the participants in the high threat condition reported feeling positive emotions less frequently, negative emotions were experienced relatively equally across conditions, with disappointment, shock and sadness emerging most often. Only four participants in the entire sample reported feeling anxious or insecure, feelings that may be most comparable to feeling threatened, and none of these participants were in the high threat condition. Of the five participants who reported physiological responses, three were in the high threat condition. These results suggest that while emotional responses were evoked to a certain degree, intense experiences of negative emotions and their complementary physiological reactions were quite rare. Though participants were explicitly asked to report any physiological reactions they experienced, this may be difficult to do and future research may benefit from using heart rate monitors, eye trackers, or electrodermal devices to measure actual physiological responses immediately following exposure to a threat. In general, the manipulation of being provided with information that you are more compatible with a person who is not your partner (high threat condition) and/or being told that you have moderate compatibility with your partner (all conditions) seemed to evoke more prominent feelings of disappointment rather than threat. These results fail to corroborate the claim that attractive alternatives are threatening or perceived as threatening.

Aside from affective responses, freely elicited cognitive evaluations were identified and categorized. Notably, approximately 20% of the sample provided criticism of the questionnaire or algorithm. These results may be interpreted in different ways. First, the negative valence of these cognitive evaluations may suggest that participants did feel threatened to a certain extent, with their evaluations representing criticism or a negative behavioral response after viewing their compatibility results. Past research has suggested that hostile reactions or enhanced negative reactions often follow threats that are perceived as uncredible or ambiguous (Millburn, 1977). But, these evaluations may also serve a different purpose. Instead of representing the perception of a threat, they may represent a threat response that serves an adaptive purpose. More negative evaluations of the questionnaire/algorithm and lower ratings of accuracy may be considered appropriate operationalizations of devaluation. By enacting these responses, participants may be attempting to alleviate the disappointment they felt upon seeing their results. By devaluing the quality and accuracy of the results, they may become less credible in the eyes of the participants, and therefore seen as less threatening. Interestingly, the presence of such criticism was relatively equal across experimental conditions, meaning it was not affected by the level of experienced threat. However, quantitative results showed that perceptions of the questionnaire/algorithm were significantly lower after viewing the compatibility results for the high threat, average threat, and partner-only conditions. Only participants who were shown to have low compatibility with their alternatives (i.e., lowest threat) did not significantly adjust their perceptions of the questionnaire/algorithm from Part I to Part II. These results suggest mixed evidence for the observation of devaluation following exposure to threatening alternatives. From a methodological standpoint, these results suggest that utilizing a mixed methods approach may be beneficial when measuring devaluation in the future.

Additional potential evidence of devaluation was found as well. As the degree of threat increased perceptions of the questionnaire/algorithm became less positive and perceived accuracy decreased. However, these associations were both moderated by initial estimates of partner and alternative compatibility. When estimated partner compatibility was high, participants in the low threat condition had more positive perceptions and accuracy ratings. Alternative compatibility seemed to moderate the associations such that when original estimates of alternative compatibility were closer to the result displayed in the threat condition (e.g., initial alternative compatibility was estimated to be 40% and the low threat condition alternative compatibility

result was 36%), perceptions of the questionnaire and accuracy were higher. Similar moderations of accuracy of the alternative compatibility result were found with commitment and attraction to and willingness to date the alternative, such that only when initial reports of commitment were high and the alternative measures were low did participants in the low threat condition perceive the results to be more accurate. Generally, when the displayed results matched the participants' original perceptions of their partners and alternatives, they perceived the questionnaire/algorithm more positively and considered the results to be more accurate.

Estimates of partner compatibility remained unchanged after exposure to the compatibility results, but participants in the low threat condition estimated their compatibility with their alternatives to be significantly lower than they did at Part I. The average of estimated alternative compatibility at Part I was just over 50%, while the alternative compatibility result in the low threat condition was 36%, meaning these participants' estimates may have been influenced by this anchor. Lastly, a comparison of commitment between Part I and Part II revealed a peculiar result. Commitment significantly decreased only for participants in the low threat condition. Perhaps participants in the low threat condition did not reach the threshold of activation for the relationship maintenance system, and thus did not engage in behaviors that could protect or enhance feelings of commitment, whereas participants in the high and average threat conditions did cross the threshold. Even so, it is curious that such a result emerged in the low threat condition and not in the partner-only condition, which suggests that the partner compatibility result could not explain the difference in commitment. And, with the participants' alternative revealed to be low in compatibility, the participants should not experience temptation to leave their current relationships. However, despite the significant difference, commitment at Part II on average in the low threat condition was still well above the midpoint of the scale, indicating that while commitment might have decreased, this decrease may not be of practical importance.

One methodological design element that may have influenced responses was the order in which they were presented. The free-response question was prompted first, followed by the quantitative measures. While this may have allowed participants to be unrestricted and unbiased in their open-ended responses, it may also have allowed them time to process and reflect on their feelings, consequently altering them from their original state prior to completing the quantitative items. Support for this observation comes from the detection of dual cognitive appraisals

(Lazarus & Folkman, 1984) in the qualitative responses. Approximately 10% of the sample reported an initial heightened emotional reaction followed by deeper processing that seemed to result in a state of acceptance and less intense emotions. This process may have lessened the intensity of emotions that participants reported experiencing when they were shown the quantitative measures. The detection of dual stage cognitive appraisal has notable implications for how past research should be interpreted that has used brief visual exposures to attractive alternatives as the operationalization of threat. While immediate evaluations may be measured best with quick and simple ratings of physical attractiveness, perhaps participants should be given time to recognize and process the emotions evoked from exposure to an attractive alternative, as this later state may be more realistic rather than the seemingly fleeting nature of the initial response.

Past research that has tested the assumption that attractive alternative partners are threatening have only employed implicit measures of threat perception (Lydon et al., 2008; Plant et al., 2010). This study used explicit measures based on participants self-reports in both quantitative and qualitative formats. While implicit measures are beneficial in that unconscious responses may be identified, it is important to utilize a variety of measures to assess the robustness of the effect as well as determine the extent to which individuals are aware of their own feelings and behaviors. In comparing the results, past research with implicit measures found inconsistent evidence for threat perception, with Lydon et al.'s (2008) research finding women implicitly perceived threat to a greater extent whereas Plant et al. (2010) found greater perception in men. While gender differences were not explored in the current study, results showed that explicit measures of affective responses to the threat (i.e., negative affect and feelings of threat) did not differ between threat conditions. Though feelings of threat may be perceived implicitly, participants may be unable to verbalize it, or may feel less threatened after having time to process the threat and their ability to cope with it.

In addition, two of the three studies that have tested threat perception have used static images as operationalizations of a threatening alternative, which severely lacks external validity. While Lydon et al. (2008) instructed participants to think of a real person, the extent to which the chosen person was considered an alternative was unmeasured. The current study addressed these limitations by asking participants to identify a person they considered an alternative to their

current dating partner and assessed the extent to which they were attracted to and willing to date the alternative. Therefore, participants were faced with a real alternative, and one they considered to be attractive and a likely partner, representing a real threat. Further, this study employed a novel operationalization of threat – simulating compatibility results with the partner and the alternative. Compared to brief exposures of attractive photographs of strangers, this manipulation contained characteristics that created the opportunity to stimulate greater feelings of threat. First, compatibility quizzes on internet sites such as Buzzfeed are highly popular, so a compatibility questionnaire/algorithm "created" by university psychologists was likely regarded as credible and valid. This was demonstrated by average perceptions of the questionnaire/algorithm at Part I emerging around the midpoint of the scale (i.e., neutral/positive). Participants also provided explicit relationship-relevant information about themselves, their partners, and an actual potential alternative partner. Unlike past research that only measured commitment between the participant and partner, or measured the quality of alternatives in general, this study collected more comprehensive information and asked participants to identify a single potential alternative who was ideally attractive and single, which 70% of the participants successfully did. Notably, some participants in an open-ended response question in Part I mentioned that just the act of filling out the questionnaire about a potential alternative was uncomfortable. When participants received their "results", they were provided with easily interpretable compatibility information broken down by relationship-relevant categories and with an overall score. Displaying results side-by-side facilitated direct comparisons between the participants' partner and potential alternative in the experimental conditions.

The manipulation used in the current study was multifaceted and created multiple opportunities to create feelings of threat within the participants. One unexpected finding was the prevalence of feelings of disappointment with the partner compatibility result which was constant across conditions and set at 76%. The partner compatibility result was not intended to be the cause of any notable discomfort or disappointment, but it may have influenced participants' responses to the outcome measures to an equal or greater extent than the alternative compatibility results. Interpreting the partner compatibility as low or inadequate may be threatening in itself. While not an external threat like attractive alternatives, low compatibility with one's partner may be more appropriately considered an internal threat, a threat that exists within one's relationship.

Therefore, this would still be considered a source of threat that may stimulate relationship maintenance behaviors.

Participants in this study were mainly heterosexual women. While not uncommon for samples collected from university populations, more diverse samples may provide a richer and more generalizable understanding of how alternative partners are perceived. In addition, the compatibility questionnaire was designed for this study and did not undergo pilot testing. Some participant responses about the quality of the questionnaire were constructive and valid and should be changed if this paradigm is to be used by other researchers. For example, many participants indicated feeling skeptical that the questionnaire was only one-sided (i.e., that the participant was asked about their partners' and alternatives' preferences and opinions) and that there were questions about astrological signs. If this methodology is replicated or adapted in the future, researchers should aim to recruit individuals from more diverse backgrounds, adapt the questionnaire to have less "guesswork" in terms of partner and alternative preferences, and opt to set the partner compatibility result to be higher so as to avoid any distracting negative emotions related to the partner when the true focus is the alternative. Improvements to the manipulation may increase the perceived credibility and validity of the questionnaire and results, which could both increase the level of experienced threat and decrease the number of critiques participants could make, giving them fewer justifiable elements of the study to devalue.

Despite the addition of the current study, research about how attractive alternatives are perceived is still lacking in both quantity and quality. Future research efforts should continue to diversify both how threat is manipulated and how threat perception is measured. Inconsistent results in threat perception have been found with implicit measures in the past, and the current study found evidence of affective responses to threat only in qualitative response formats. Evidence found in the current study of processing of threats in two phases indicate that researchers may want to employ measures of threat perception at multiple time points as well, as initial reactions may differ notably from reactions after participants are allowed to consider the severity of the threat and their coping abilities.

The present research was specifically designed to test the primary assumption that attractive alternative partners are perceived as threats. Results showed that participants in different threat

conditions did not differ in explicit, quantitative measures of negative affect or threat. Inclusion of qualitative results revealed more nuanced responses that did show that participants were emotionally affected. The largest differences between threat conditions involved participants' cognitive evaluations of the questionnaire/algorithm and the accuracy of the results, with those in the low threat condition reporting greater positive perceptions and higher perceptions of accuracy. However, whether cognitive evaluations such as these are considered evidence of threat perception or devaluation (a strategy that can result in maintenance/protection or enhancement) remains up for debate. Much is left to be discovered through efforts to test the primary assumption of the sub-theory of devaluation of alternatives.

Chapter 5

5 General Discussion

Devaluation of alternatives is often identified as a key relationship maintenance strategy that is specifically employed to combat the threat of attractive alternatives. However, since the theory was proposed by Johnson and Rusbult (1989), few efforts have been dedicated to evaluating the progression of the theory and assessing the quantity and quality of the existing evidence that claims to corroborate it. The present research fulfilled this need by formalizing the theory of devaluation of alternatives using a metatheoretical approach that involved categorization, deconstruction, reconstruction via theory mapping, and evaluation using Lakatos's and Salmon's principles of "money in the bank" and "the damn strange coincidence" (Chapter 2). *P*-curve meta-analyses were used to determine whether *p*-hacking or selective reporting could be detected in a selection of the literature (Chapter 3). These processes revealed many of the theoretical assumptions were understudied and lacked sufficient corroboration, and evidence of selective reporting was not detected. Guided by these findings, the assumption that attractive alternatives are threatening was selected and tested using a novel experimental manipulation (Chapter 4).

Theories exist for the purpose of providing a broad structure of knowledge that integrates empirical observations and facilitates predictions (Gawronski & Bodenhausen, 2015; Gray, 2017). However, theories in social psychology are often reiterated and relied upon in simplified forms that omit underlying key assumptions and complexities. Consequently, even results from well-designed research may only provide weak corroborative evidence if they are informed by verbal theoretical models rather than formal theoretical models. Chapter 2 alleviated this issue by formalizing the theory of devaluation of alternatives. The processes of categorization and deconstruction involved returning to the origins of the theory in Johnson and Rusbult (1989). The initial proposal of the theory and the conceptual definitions of key terms were extracted from the original article and underlying assumptions were identified and categorized a part of the hard core or peripheral belt. Notably, these processes revealed that devaluation of alternatives is a specific situation that is nested within a larger theoretical framework of external threat-induced relationship maintenance. In their original article Johnson and Rusbult explicitly recognized that other maintenance strategies existed that could be effective in combating the threat of attractive

alternatives; devaluation is the strategy that is most known and tested in this context. During reconstruction, theory maps of devaluation of alternatives and the broader model of relationship maintenance in response to external threats were created. Mapping these models together revealed that commitment was theorized to play a significant role in devaluation, but the role of commitment was less explicit in the broader model. This resulted in the addition of a peripheral assumption to the broader theory of relationship maintenance. Finally, devaluation was evaluated by providing a brief summary of the evolution of research on the theory and by taking an indepth look at the quantity, quality and robustness of existing evidence that tested each of the assumptions of devaluation of alternatives. This process revealed that all of the assumptions lack adequate corroboration. Research that assessed the assumption that devaluation may occur when the relationship maintenance system was activated was found to have the most "money in the bank." However, this assumption was not considered to contain verisimilitude because most of the corroborating research used only a small range of operationalizations of attractive alternatives and devaluation, those being photographs of attractive individuals and ratings of physical attractiveness, respectively. Devaluation of alternatives has thus far only been narrowly tested and there is more theoretical space in which to expand research on this phenomenon.

Chapter 3 provided a more systematic and quantitative approach to evaluating the literature. *P*-curve analyses did not detect evidence of *p*-hacking or selective reporting in a sample of the devaluation literature. These analyses, however, were not as robust as one may prefer, and became inconclusive with the exclusion of a few studies with the lowest obtained *p*-values. While these analyses should alleviate some concern about the presence of questionable research practices in this area of research, only cautious conclusions can be made about the strength and quality of the existing evidence. It seems likely that an effect exists, but more research is needed to make more confident conclusions about the evidentiary value of this literature.

Overall, formalizing and evaluating the theory of devaluation of alternatives was both beneficial and necessary. This work explicitly dictated the components of the theory and clearly identified the gaps in our current understanding of the phenomenon that future researchers may use to inform their own empirical pursuits in this area. Assertive claims are often made about the threatening nature of attractive alternatives and the subsequent elicitation of devaluation to protect one's relationship. However, a close inspection of the existing evidence suggests that

such claims are unwarranted and misleading. This conclusion may be unsettling and may reflect our typical regard of statistically significant p-values as signals of practically significant or true effects, without critical consideration of the methods and measures used to detect them or without evaluating the existing body of evidence as a whole. Further, this may also reflect the consequences of conflating statistically significant hypothesis tests with support for a theory, an embodiment of the logical fallacy of affirming the consequent (Meehl, 1967). When this error is made by multiple independent scholars and goes uncorrected, scientific progress and advancement is stymied. To correct course, researchers should engage in theory formalization and evaluation endeavors to assess the current state of their respective theories of interest. In addition, researchers should engage with the philosophy of science literature, or collaborate with scholars in that field, to "return to our roots" and be reminded of what the true purpose of science is – to increase our knowledge of the world through building better theories and subjecting them to strict tests.

Knowledge from the formalization and evaluation processes informed the development of a study specifically designed to test a core assumption of the theory that was found to have inadequate empirical support. Despite attractive alternatives being repeatedly touted as threats, very few studies have directly tested this claim. In the present study, coupled participants completed compatibility questionnaires with their current partner and an actual person in their lives that they were attracted to and could see themselves dating if they were not with their partner. Participants were shown manipulated compatibility results that showed moderate compatibility with their partners (76%) and varied the level of compatibility with their potential alternatives as either high (88%), average (73%) or low (36%), with a control condition only being shown their compatibility with their partner. Past research has shown that negatively valenced emotions are elicited following exposure to threat (Hodgins et al., 2010), but the present results did not detect different levels of negative affect or feelings of threat following exposure to the compatibility results. Open-ended responses following threat exposure revealed that around half of the participants felt indifferent and unaffected by their results, while disappointment was the most frequently reported negative emotion. Participants were also given the opportunity to provide evaluations of the compatibility questionnaire/algorithm and the perceived accuracy of the results. Generally, participants had more positive perceptions and considered the results to be more accurate when they aligned with their initial beliefs about their

compatibility with their partners and alternatives. And, as predicted, participants in the low threat condition perceived the accuracy of their alternative compatibility result to be more accurate than those in the high threat condition. Further analysis of the open-ended responses revealed a variety of cognitive evaluations regarding accuracy, defensive rationalizations of the results, and criticism directed towards the questionnaire, algorithm and to a lesser extent, the self. And, evidence of a two-part cognitive appraisal process (Lazarus & Folkman, 1984) emerged, such that initial reactions appeared to be more emotionally intense but seemed to decrease after deeper consideration of the credibility of the threat and the quality of one's relationship.

Overall, these findings suggest that threat perception in response to attractive alternatives is a complex process. While past research has shown that threat perception has been detected with implicit measures (Lydon et al., 2008; Plant et al., 2010), the same was not found in this study when explicit, quantitative self-report measures were employed. This study was the first to my knowledge to provide participants with the opportunity to report their thoughts and feelings in an open-ended format, allowing for rich descriptions of their emotional states and thought processes to be analyzed. Use of these measures greatly expanded our knowledge of how coupled individuals perceive attractive alternatives and how they cope with potentially disappointing or threatening information.

Perhaps a more provocative interpretation may be that attractive alternatives are not considered by coupled individuals to be threatening at all. The primary core assumption of relationship maintenance is that external stressors, such as attractive alternatives are threatening. As it stands, this assumption must be satisfied for relationship maintenance processes such as devaluation to occur. However, the present results failed to find strong evidence of responses that indicated the perception of a threat. The results did indicate, though, that participants engaged in devaluation of the source of the threat (e.g., rating the questionnaire/algorithm to be of lower quality and the alternative compatibility result to be less accurate). Qualitative responses also offered similar evidence of criticism directed towards the source of the threat. This suggests that perhaps perceiving attractive alternatives as threats is not a necessary precursor to relationship maintenance behaviors. While more research is needed, if more evidence is found that falsifies this assumption the sub-theory of devaluation of alternatives will need to be revised. In addition, research will need to be conducted to determine whether other currently accepted external

threats, such as attractive alternatives for one's partner, are perceived as threatening. Perhaps attractive alternatives for oneself are the exception, a supposed threat that turned out not to be a threat. But perhaps other supposed threats are also not actually perceived as threatening. If this is the case, the broader theory of relationship maintenance in response to threats may need to be revised as well. This research has potentially large implications for how we understand and model relationship maintenance processes.

Taken together, the research presented herein demonstrates the effectiveness and utility of formalizing theory and designing studies specifically with the intention of directly testing and refining specific assumptions of the theory. Social psychology is a theory-rich discipline; however our empirical pursuits would be much improved by dedicating effort towards formalizing and evaluating the current state and verisimilitude of our existing theories. As this research has demonstrated, theories that are thought to be well-founded may not be. While such revelations may be discouraging, the process of formalization helpfully illuminates the areas in which our research efforts should be directed so that we may reembark on our continuous journey towards the truth.

References

- Agnew, C. R., & VanderDrift, L. E. (2015). Relationship maintenance and dissolution. In M. Mikulincer, P. R. Shaver, J. A. Simpson, & J. F. Dovidio (Eds.), *APA handbook of personality and social psychology, Volume 3: Interpersonal relations* (pp. 581–604). American Psychological Association. https://doi.org/10.1037/14344-021
- Allison, P. (2022, May 12). When can you safely ignore multicollinearity? Statistical Horizons. https://statisticalhorizons.com/multicollinearity/

Allport, G. W. (1954). The nature of prejudice. Addison-Wesley.

- Bazzini, D. G., & Shaffer, D. R. (1999). Resisting temptation revisited: Devaluation versus enhancement of an attractive suitor by exclusive and nonexclusive daters. *Personality and Social Psychology Bulletin*, 25(2), 162–176. <u>https://doi.org/10.1177/0146167299025002003</u>
- Berscheid, E., & Walster, E. (1974). Physical attractiveness. In *Advances in experimental social psychology*, *Volume 7* (pp. 157–215). Elsevier. https://doi.org/10.1016/S0065-2601(08)60037-4
- Betzig, L. (1989). Causes of conjugal dissolution: A cross-cultural study. *Current Anthropology*, *30*(5), 654–676. <u>https://doi.org/10.1086/203798</u>
- Birnbaum, G. E., Mizrahi, M., Kovler, L., Shutzman, B., Aloni-Soroker, A., & Reis, H. T. (2019).
 Our fragile relationships: Relationship threat and its effect on the allure of alternative mates. *Archives of Sexual Behavior*, 48(3), 703–713. https://doi.org/10.1007/s10508-018-1321-5
- Birnbaum, G. E., Svitelman, N., Bar-Shalom, A., & Porat, O. (2008). The thin line between reality and imagination: Attachment orientations and the effects of relationship threats on sexual

fantasies. *Personality and Social Psychology Bulletin*, *34*(9), 1185–1199. https://doi.org/10.1177/0146167208319692

- Birnbaum, G. E., Weisberg, Y. J., & Simpson, J. A. (2011). Desire under attack: Attachment orientations and the effects of relationship threat on sexual motivations. *Journal of Social and Personal Relationships*, 28(4), 448–468. https://doi.org/10.1177/0265407510381932
- Bitsios, P., Szabadi, E., & Bradshaw, C. M. (1996). The inhibition of the pupillary light reflex by the threat of an electric shock: A potential laboratory model of human anxiety. *Journal of Psychopharmacology*, 10(4), 279–287. <u>https://doi.org/10.1177/026988119601000404</u>
- Borsboom, D. (2013, November 20). Theoretical Amnesia. Open Science Collaboration Blog. http://osc.centerforopenscience.org/2013/11/20/theoretical-amnesia/

Bowlby, J. (1969/1982). Attachment and loss, Volume 1, Attachment. New York: Basic Books.

- Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *The Journal of Abnormal and Social Psychology*, *52*(3), 384–389. <u>https://doi.org/10.1037/h0041006</u>
- Buss, D. M. (1998). Sexual strategies theory: Historical origins and current status. *Journal of Sex Research*, *35*(1), 19-31. <u>https://doi.org/10.1080/00224499809551914</u>
- Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review*, 100(2), 204–232. <u>https://doi.org/10.1037/0033-</u> 295X.100.2.204
- Cole, F. L. (1988) Content analysis: Process and application. *Clinical Nurse Specialist*, 2, 53-57. https://doi.org/10.1097/00002800-198800210-00025

- Cole, S., Trope, Y., & Balcetis, E. (2016). In the eye of the betrothed: Perceptual downgrading of attractive alternative romantic partners. *Personality and Social Psychology Bulletin*, 42(7), 879–892. <u>https://doi.org/10.1177/0146167216646546</u>
- Cooper, R., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, *17*(4), 297–338. <u>https://doi.org/10.1080/026432900380427</u>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <u>https://doi.org/10.1037/h0040957</u>
- de Lenne, O., Wittevronghel, L., Vandenbosch, L., & Eggermont, S. (2019). Romantic relationship commitment and the threat of alternatives on social media. *Personal Relationships*, 26(4), 680–693. <u>https://doi.org/10.1111/pere.12299</u>

Deutsch, M., & Krauss, R.M. (1965). Theories in social psychology. Basic Books.

- DeWall, C. N., Maner, J. K., Deckman, T., & Rouby, D. A. (2011). Forbidden fruit: Inattention to attractive alternatives provokes implicit relationship reactance. *Journal of Personality and Social Psychology*, 100(4), 621–629. <u>https://doi.org/10.1037/a0021749</u>
- Dienes, Z. (2008). Understanding psychology as a science: An introduction to scientific and statistical inference. Palgrave Macmillan.
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

- Etcheverry, P. E., Le, B., Wu, T.-F., & Wei, M. (2013). Attachment and the investment model:
 Predictors of relationship commitment, maintenance, and persistence. *Personal Relationships*, 20(3), 546–567. <u>https://doi.org/10.1111/j.1475-6811.2012.01423.x</u>
- Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS ONE*, 5(4), e10068. <u>https://doi.org/10.1371/journal.pone.0010068</u>
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891–904. <u>https://doi.org/10.1007/s11192-011-0494-7</u>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <u>https://doi.org/10.3758/BF03193146</u>
- Fiedler, K. (2004). Tools, toys, truisms, and theories: Some thoughts on the creative cycle of theory formation. *Personality and Social Psychology Review*, 8(2), 123–131. https://doi.org/10.1207/s15327957pspr0802_5

Field, A. (2013). Discovering statistics using IBM SPSS statistics. Sage.

- Fincham, F. D., & May, R. W. (2017). Infidelity in romantic relationships. *Current Opinion in Psychology*, 13, 70–74. <u>https://doi.org/10.1016/j.copsyc.2016.03.008</u>
- Forscher, B. K. (1963). Chaos in the brickyard. *Science*, *142*(3590), 339–339. https://doi.org/10.1126/science.142.3590.339.a
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502–1505. <u>https://doi.org/10.1126/science.1255484</u>

- Gangestad, S. W., & Thornhill, R. (1997). Human sexual selection and developmental stability. *Evolutionary Social Psychology*, *10*, 169-196.
- Gawronski, B., & Bodenhausen, G. V. (2015). *Theory and explanation in social psychology*. Guilford Publications.
- Gonzaga, G. C., Haselton, M. G., Smurda, J., Davies, M.S., & Poore, J. C. (2008). Love, desire, and the suppression of thoughts of romantic alternatives. *Evolution and Human Behavior*, 29(2), 119–126. <u>https://doi.org/10.1016/j.evolhumbehav.2007.11.003</u>
- Gray, K. (2017). How to map theory: Reliable methods are fruitless without rigorous theory. *Perspectives on Psychological Science*, *12*(5), 731–741. https://doi.org/10.1177/1745691617691949
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82(1), 1–20. <u>https://doi.org/10.1037/h0076157</u>
- Hadden, B. W. (2016). When temptations aren't tempting: The autonomy and derogation of alternatives model (0000-0001-8386-3116) [Doctoral dissertation, University of Houston].
 <u>Published ETD Collection</u>
- Hart, J., Shaver, P. R., & Goldenberg, J. L. (2005). Attachment, self-esteem, worldviews, and terror management: Evidence for a tripartite security system. *Journal of Personality and Social Psychology*, 88(6), 999–1013. <u>https://doi.org/10.1037/0022-3514.88.6.999</u>
- Hazan, C., & Shaver, P. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology*, 52(3), 511–524. <u>https://doi.org/10.1037/0022-3514.52.3.511</u>

- Henderson, R. R., Bradley, M. M., & Lang, P. J. (2018). Emotional imagery and pupil diameter. *Psychophysiology*, 55(6), e13050. https://doi.org/10.1111/psyp.13050
- Hodgins, H. S., Weibust, K. S., Weinstein, N., Shiffman, S., Miller, A., Coombs, G., & Adair, K. C. (2010). The cost of self-protection: Threat response and performance as a function of autonomous and controlled motivations. *Personality and Social Psychology Bulletin*, *36*(8), 1101–1114. <u>https://doi.org/10.1177/0146167210375618</u>
- Hui, C. M., & Molden, D. C. (2014). Maintaining commitment in the presence of alternative opportunities: The role of motivations for prevention or promotion. *Social Cognition*, *32*(6), 571–584. <u>https://doi.org/10.1521/soco.2014.32.6.571</u>
- Johnson, D. J., & Rusbult, C. E. (1989). Resisting temptation: Devaluation of alternative partners as a means of maintaining commitment in close relationships. *Journal of Personality and Social Psychology*, 57(6), 967–980. <u>https://doi.org/10.1037/0022-3514.57.6.967</u>
- Jordan, C. H., Spencer, S. J., Zanna, M. P., Hoshino-Browne, E., & Correll, J. (2003). Secure and defensive high self-esteem. *Journal of Personality and Social Psychology*, 85(5), 969–978. <u>https://doi.org/10.1037/0022-3514.85.5.969</u>

Kahneman, D. (2012). Thinking, fast and slow. Penguin Books.

Karremans, J. C., Dotsch, R., & Corneille, O. (2011). Romantic relationship status biases memory of faces of attractive opposite-sex others: Evidence from a reverse-correlation paradigm. *Cognition*, *121*(3), 422–426. <u>https://doi.org/10.1016/j.cognition.2011.07.008</u> Karremans, J. C., & Verwijmeren, T. (2008). Mimicking attractive opposite-sex others: The role of romantic relationship status. *Personality and Social Psychology Bulletin*, 34(7), 939–950. <u>https://doi.org/10.1177/0146167208316693</u>

Kelley, H. H. (1959). The social psychology of groups. Transaction Publishers.

- Kelley, H. H. (1983). Love and commitment. In H. H. Kelley, Berscheid, E., Christensen, A., Harvey, J. H., Huston, T. L., Levinger, G., ... & Peterson, D. R. (Eds.), *Close relationships*, (pp. 265-314). W.H. Freeman.
- Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal relations: A theory of interdependence*. New York: Wiley.
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196-217. <u>https://doi.org/10.1207/s15327957pspr0203_4</u>
- Koranyi, N., & Rothermund, K. (2012). When the grass on the other side of the fence doesn't matter: Reciprocal romantic interest neutralizes attentional bias towards attractive alternatives. *Journal* of Experimental Social Psychology, 48(1), 186–191. <u>https://doi.org/10.1016/j.jesp.2011.06.012</u>
- Laeng, B., & Falkenberg, L. (2007). Women's pupillary responses to sexually significant others during the hormonal cycle. *Hormones and Behavior*, 52(4), 520-530.

https://doi.org/10.1016/j.yhbeh.2007.07.013

Lakatos, I. (1970). Falsification and the methodology of scientific programs. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge*, (pp. 91-196). Cambridge University Press. <u>https://doi.org/10.1017/cbo9781139171434.009</u>

- Lakatos, I. (1978). Introduction: Science and pseudoscience. In J. Worrall & G. Currie (Eds.), *The methodology of scientific research programmes: Philosophical papers* (pp. 1-7). Cambridge:
 Cambridge University Press. <u>https://doi.org/10.1017/CBO9780511621123.002</u>
- Lakens, D. (2015, March 20). How a p-value between 0.04-0.05 equals a p-value between 0.16-017. *The 20% Statistician*. <u>http://daniellakens.blogspot.com/2015/03/how-p-value-between-004-</u> 005-equals-p.html

Lazarus, R. S., & Folkman, S. (1984). Stress, appraisal, and coping. Springer Publishing Company.

- Lee, B. H., & O'Sullivan, L. F. (2018). Ain't misbehavin? Monogamy maintenance strategies in heterosexual romantic relationships. *Personal Relationships*, 25(2), 205–232. https://doi.org/10.1111/pere.12235
- Lee, B. H., & O'Sullivan, L. F. (2019). Walk the line: How successful are efforts to maintain monogamy in intimate relationships? *Archives of Sexual Behavior*, 48(6), 1735–1748. https://doi.org/10.1007/s10508-018-1376-3
- Leik, R. K., & Leik, S. A. (2021). Transition to interpersonal commitment. In R. L. Hamblin & J. H. Kunkel, *Behavioral theory in sociology* (1st ed., pp. 299–322). Routledge. https://doi.org/10.4324/9781351319201-11
- Linardatos, L., & Lydon, J. E. (2011). Relationship-specific identification and spontaneous relationship maintenance processes. *Journal of Personality and Social Psychology*, *101*(4), 737– 753. <u>https://doi.org/10.1037/a0023647</u>

- Lydon, J. E., Fitzsimons, G. M., & Naidoo, L. (2003). Devaluation versus enhancement of attractive alternatives: A critical test using the calibration paradigm. *Personality and Social Psychology Bulletin*, 29(3), 349–359. <u>https://doi.org/10.1177/0146167202250202</u>
- Lydon, J. E., Meana, M., Sepinwall, D., Richards, N., & Mayman, S. (1999). The commitment calibration hypothesis: When do people devalue attractive alternatives? *Personality and Social Psychology Bulletin*, 25(2), 152–161. <u>https://doi.org/10.1177/0146167299025002002</u>
- Lydon, J. E., Menzies-Toman, D., Burton, K., & Bell, C. (2008). If-then contingencies and the differential effects of the availability of an attractive alternative on relationship maintenance for men and women. *Journal of Personality and Social Psychology*, 95(1), 50–65. https://doi.org/10.1037/0022-3514.95.1.50
- Lydon, J., & Karremans, J. C. (2015). Relationship regulation in the face of eye candy: A motivated cognition framework for understanding responses to attractive alternatives. *Current Opinion in Psychology*, 1, 76–80. <u>https://doi.org/10.1016/j.copsyc.2015.01.011</u>
- Lyubomirsky, S., & Ross, L. (1999). Changes in attractiveness of elected, rejected, and precluded alternatives: A comparison of happy and unhappy individuals. *Journal of Personality and Social Psychology*, *76*(6), 988–1007. <u>https://doi.org/10.1037/0022-3514.76.6.988</u>
- Ma, Y., Xue, W., & Tu, S. (2019). Automatic inattention to attractive alternative partners helps male heterosexual chinese college students maintain romantic relationships. *Frontiers in Psychology*, 10, 1687. <u>https://doi.org/10.3389/fpsyg.2019.01687</u>
- Ma, Y., Xue, W., Zhao, G., Tu, S., & Zheng, Y. (2019). Romantic love and attentional biases toward attractive alternatives and rivals: Long-term relationship maintenance among female chinese

college students. *Evolutionary Psychology*, *17*(4), 147470491989760. https://doi.org/10.1177/1474704919897601

- Ma-Kellams, C., Wang, M. C., & Cardiel, H. (2017). Attractiveness and relationship longevity:
 Beauty is not what it is cracked up to be: Attractiveness and relationship outcomes. *Personal Relationships*, 24(1), 146–161. <u>https://doi.org/10.1111/pere.12173</u>
- Maner, J. K., Rouby, D. A., & Gonzaga, G. C. (2008). Automatic inattention to attractive alternatives: The evolved psychology of relationship maintenance. *Evolution and Human Behavior*, 29(5), 343–349. <u>https://doi.org/10.1016/j.evolhumbehav.2008.04.003</u>
- Maner, J. K., Gailliot, M. T., & Miller, S. L. (2009). The implicit cognition of relationship maintenance: Inattention to attractive alternatives. *Journal of Experimental Social Psychology*, 45(1), 174–179. <u>https://doi.org/10.1016/j.jesp.2008.08.002</u>
- McGovern, M. [Michael McGovern]. (2018). *Paul Meehl Philosophical Psychology Lecture 5/12* [Video]. YouTube. <u>https://www.youtube.com/watch?v=t87bSqjhrng&t=1s</u>
- McNulty, J. K., & Dugas, A. (2019). A dyadic perspective on gratitude sheds light on both its benefits and its costs: Evidence that low gratitude acts as a "weak link". *Journal of Family Psychology*, 33(7), 876–881. <u>https://doi.org/10.1037/fam0000533</u>
- McNulty, J. K., Meltzer, A. L., Makhanova, A., & Maner, J. K. (2018). Attentional and evaluative biases help people maintain relationships by avoiding infidelity. *Journal of Personality and Social Psychology*, *115*(1), 76–95. <u>https://doi.org/10.1037/pspi0000127</u>

- Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*(2), 103–115. <u>https://doi.org/10.1086/288135</u>
- Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1(2), 108–141. https://doi.org/10.1207/s15327965pli0102_1
- Meehl, P. E. (2004). Cliometric metatheory III: Peircean consensus, verisimilitude and asymptotic method. *The British Journal for the Philosophy of Science*, 55(4), 615–643. https://doi.org/10.1093/bjps/55.4.615
- Meyer, M. L., Berkman, E. T., Karremans, J. C., & Lieberman, M. D. (2011). Incidental regulation of attraction: The neural basis of the derogation of attractive alternatives in romantic relationships.
 Cognition & Emotion, 25(3), 490–505. https://doi.org/10.1080/02699931.2010.527494
- Mikulincer, M., Gillath, O., & Shaver, P. R. (2002). Activation of the attachment system in adulthood: Threat-related primes increase the accessibility of mental representations of attachment figures. *Journal of Personality and Social Psychology*, 83(4), 881–895.
 https://doi.org/10.1037/0022-3514.83.4.881
- Milburn, T. W. (1961). The concept of deterrence: Some logical and psychological considerations. *Journal of Social Issues*, 17(3), 3–11. <u>https://doi.org/10.1111/j.1540-</u> 4560.1961.tb01678.x
- Milburn, T. W. (1969). Intellectual history of a research program. In D. G. Pruitt & R. C. Snyder (Eds.), *Theory and research on the causes of war*. Englewood Cliffs, N.J.: Prentice-Hall.

- Milburn, T. W. (1977). The nature of threat. *Journal of Social Issues*, *33*(1), 126–139. https://doi.org/10.1111/j.1540-4560.1977.tb01872.x
- Miller, R. S. (1997). Inattentive and contented: Relationship commitment and attention to alternatives. *Journal of Personality and Social Psychology*, 73(4), 758–766. https://doi.org/10.1037/0022-3514.73.4.758
- Miller, R. S. (2003). On being admired but overlooked: Reflections on "attention to alternatives" in close relationships. *Psychological Inquiry*, 14(3–4), 284–288. https://doi.org/10.1080/1047840X.2003.9682894
- Miller, S. L., & Maner, J. K. (2010). Evolution and relationship maintenance: Fertility cues lead committed men to devalue relationship alternatives. *Journal of Experimental Social Psychology*, 46(6), 1081–1084. <u>https://doi.org/10.1016/j.jesp.2010.07.004</u>
- Miller, S. L., Prokosch, M. L., & Maner, J. K. (2012). Relationship maintenance and biases on the line bisection task: Attractive alternatives, asymmetrical cortical activity, and approach–avoidance motivation. *Journal of Experimental Social Psychology*, 48(2), 566–569. https://doi.org/10.1016/j.jesp.2011.10.012
- Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*(3), 221–229. <u>https://doi.org/10.1038/s41562-018-0522-1</u>
- Norman, D. A., & Shallice, T. (1986). Attention to action. In Davidson, R.J., Schwartz, G.E., & Shapiro, D. (Eds.) *Consciousness and self-regulation (Volume 4*, pp. 1-18). Springer, Boston, MA. <u>https://doi.org/10.1007/978-1-4757-0629-1_1</u>

- O'Boyle, E. H., Banks, G. C., & Gonzalez-Mulé, E. (2017). The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management*, *43*(2), 376–399. https://doi.org/10.1177/0149206314527133
- Ogolsky, B. G., Monk, J. K., Rice, T. M., Theisen, J. C., & Maniotes, C. R. (2017). Relationship maintenance: A review of research on romantic relationships. *Journal of Family Theory & Review*, 9(3), 275–306. <u>https://doi.org/10.1111/jftr.12205</u>
- Overall, N. C., & Sibley, C. G. (2008). Attachment and attraction toward romantic partners versus relevant alternatives within daily interactions. *Personality and Individual Differences*, 44(5), 1126–1137. <u>https://doi.org/10.1016/j.paid.2007.11.006</u>
- Peña, E. A., & Slate, E. H. (2006). Global validation of linear model assumptions. *Journal of the American Statistical Association*, 101(473), 341–354. https://doi.org/10.1198/01621450500000637
- Petit, W. E., & Ford, T. E. (2015). Effect of relationship status on perceptions of physical attractiveness for alternative partners: Effect of relationship status. *Personal Relationships*, 22(2), 348–355. <u>https://doi.org/10.1111/pere.12084</u>
- Plant, E. A., Kunstman, J. W., & Maner, J. K. (2010). You do not only hurt the one you love: Selfprotective responses to attractive relationship alternatives. *Journal of Experimental Social Psychology*, 46(2), 474–477. <u>https://doi.org/10.1016/j.jesp.2009.12.012</u>
- Plant, E. A., & Peruche, B. M. (2005). The consequences of race for police officers' responses to criminal suspects. *Psychological Science*, *16*(3), 180–183. <u>https://doi.org/10.1111/j.0956-7976.2005.00800.x</u>

Poincaré, H. (1905). Science and hypothesis. The Walter Scott Publishing Co., LTD.

- Pronk, T. M., Karremans, J. C., & Wigboldus, D. H. J. (2011). How can you resist? Executive control helps romantically involved individuals to stay faithful. *Journal of Personality and Social Psychology*, 100(5), 827–837. https://doi.org/10.1037/a0021993
- Quistberg, K., & Sakaluk, J. K. (2021). Good in theory: A metatheoretical review of social psychological theories of human sexuality [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/pnqfv
- Randall, A. K., & Messerschmitt-Coen, S. (2020). Dyadic coping as relationship maintenance. In Ogolsky, B. G., & Monk, J. K. (Eds.). *Relationship maintenance: Theory, process, and context* (pp. 178-193). Cambridge University Press.
- Ritter, S. M., Karremans, J. C., & van Schie, H. T. (2010). The role of self-regulation in derogating attractive alternatives. *Journal of Experimental Social Psychology*, 46(4), 631–637. <u>https://doi.org/10.1016/j.jesp.2010.02.010</u>
- Rodrigues, David & Lopes, Diniz. (2013). The impact of general and moral commitment in derogating attractive alternatives. In A. Garcia & J. Fitzpatrick (Eds.), *Relações românticas, conjugais e parassociais* (pp. 39-52). UFES.
- Rodrigues, D., Lopes, D., & Kumashiro, M. (2017). The "I" in us, or the eye on us? Regulatory focus, commitment and derogation of an attractive alternative person. *PLOS ONE*, *12*(3), e0174350. <u>https://doi.org/10.1371/journal.pone.0174350</u>

- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <u>https://doi.org/10.1037/0033-2909.86.3.638</u>
- Rusbult, C. E. (1980). Commitment and satisfaction in romantic associations: A test of the investment model. *Journal of Experimental Social Psychology*, *16*(2), 172–186. https://doi.org/10.1016/0022-1031(80)90007-4
- Rusbult, C. E. (1983). A longitudinal test of the investment model: The development (and deterioration) of satisfaction and commitment in heterosexual involvements. *Journal of Personality and Social Psychology*, 45(1), 101–117. <u>https://doi.org/10.1037/0022-3514.45.1.101</u>
- Rusbult, C. E., & Buunk, B. P. (1993). Commitment processes in close relationships: An interdependence analysis. *Journal of Social and Personal Relationships*, 10(2), 175–204. <u>https://doi.org/10.1177/026540759301000202</u>
- Rusbult, C. E., Drigotas, S. M., & Verette, J. (1994). The investment model: An interdependence analysis of commitment processes and relationship maintenance phenomena. In D. J. Canary & L. Stafford (Eds.), *Communication and relational maintenance* (pp. 115–139). Academic Press.
- Rusbult, C. E., Martz, J. M., & Agnew, C. R. (1998). The investment model scale: Measuring commitment level, satisfaction level, quality of alternatives, and investment size. *Personal Relationships*, 5(4), 357–391. <u>https://doi.org/10.1111/j.1475-6811.1998.tb00177.x</u>
- Rusbult, C. E., Olsen, N., & Davis, J. L. (2001). Commitment and relationship maintenance mechanisms. In Harvey, J. H. & Wenzel, A. (Eds.), *Close romantic relationships: Maintenance and enhancement* (pp. 87 – 114). Lawrence Erlbaum Associates.

- Simpson, J. A., Gangestad, S. W., & Lerma, M. (1990). Perception of physical attractiveness: Mechanisms involved in the maintenance of romantic relationships. *Journal of Personality and Social Psychology*, 59(6), 1192–1201. <u>https://doi.org/10.1037/0022-3514.59.6.1192</u>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. https://doi.org/10.1177/0956797611417632
- Simmons, J. P., & Simonsohn, U. (2017). Power posing: *P* curving the evidence. *Psychological Science*, 28(5), 687–693. <u>https://doi.org/10.1177/0956797616658563</u>
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2), 534–547. https://doi.org/10.1037/a0033242
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better p-curves: Making p-curve analysis more robust to errors, fraud, and ambitious p-hacking, a reply to Ulrich and Miller. *Journal of Experimental Psychology: General*, 144(6), 1146-1152.

http://dx.doi.org/10.1037%2Fxge0000104

- Smaldino, P. E. (2017). Models are stupid, and we need more of them. In R. R. Vallacher, S. J. Read, & A. Nowak (Eds.), *Computational Social Psychology* (1st ed., pp. 311–331). Routledge. <u>https://doi.org/10.4324/9781315173726-14</u>
- Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association*, *54*, 30–34. <u>https://doi.org/10.2307/2282137</u>

Tedeschi, J. T. (1970). Threats and promises. In P. Swingle (Ed.), *The structure of conflict*. New York: Academic Press.

Thibaut, J. W., & Kelley, H. H. (1959). The social psychology of groups. John Wiley.

- Tomaka, J., Blascovich, J., Kelsey, R. M., & Leitten, C. L. (1993). Subjective, physiological, and behavioral effects of threat and challenge appraisal. *Journal of Personality and Social Psychology*, 65(2), 248-260. <u>https://doi.org/10.1037/0022-3514.65.2.248</u>
- Trope, Y., & Liberman, N. (2011). Construal level theory. In P. A. M. V. Lange, A. W. Kruglanski, &E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 118–134). London, UK:Sage.
- VanderDrift, L. E., Lewandowski, G. W., & Agnew, C. R. (2011). Reduced self-expansion in current romance and interest in relationship alternatives. *Journal of Social and Personal Relationships*, 28(3), 356–373. <u>https://doi.org/10.1177/0265407510382321</u>
- Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, *13*(4), 411-417. https://doi.org/10.1177/1745691617751884
- Visserman, M. L., & Karremans, J. C. (2014). Romantic relationship status biases the processing of an attractive alternative's behavior. *Personal Relationships*, 21(2), 324– 334. <u>https://doi.org/10.1111/pere.12035</u>

Wallis, S. (2010). Toward a science of metatheory. Integral Review: A Transdiciplinary and Transcultural Journal for New Thought, Research, and Praxis, 6(3). <u>https://ssrn.com/abstract=2322985</u>

- Wallis, S. E. (2014). A systems approach to understanding theory: Finding the core, identifying opportunities for improvement, and integrating fragmented fields. *Systems Research and Behavioral Science*, 31(1), 23–31. <u>https://doi.org/10.1002/sres.2159</u>
- Wang, H., Hahn, A. C., DeBruine, L. M., & Jones, B. C. (2016). Do partnered women discriminate men's faces less along the attractiveness dimension? *Personality and Individual Differences*, 98, 153–156. <u>https://doi.org/10.1016/j.paid.2016.04.024</u>
- Warner, R. M. (2013). Applied statistics: From bivariate through multivariate techniques (2nd ed.)Sage Publications, Inc.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070. <u>https://doi.org/10.1037/0022-3514.54.6.1063</u>

Yates, D., Moore, D., & McCabe, G. (1999). The practice of statistics. New York: Freeman.

Zeigler-Hill, V., Seidman, G., Trombly, D. R. C., Ishaq, J., Shango, R., & Roberts, A. (2020).
 Narcissism and romantic commitment: The role of attitudes toward alternative romantic partners.
 Self and Identity, *19*(4), 473–500. <u>https://doi.org/10.1080/15298868.2019.1635521</u>

Zhang, Q., Maner, J. K., Xu, Y., & Zheng, Y. (2017). Relational motives reduce attentional adhesion to attractive alternatives in heterosexual university students in China. *Archives of Sexual Behavior*, 46(2), 503–511. <u>https://doi.org/10.1007/s10508-016-0759-6</u>

Appendices

Appendix A. Research Ethics Approval



Date: 17 March 2022 To: Prof. Lorne Campbell Project ID: 120098 Study Title: Derogation of Alternatives: An Investigation of the Threat of Attractive Alternatives Short Title: DoA: Threat of Attractive Alternatives Application Type: NMREB Initial Application Review Type: Full Board Full Board Reporting Date: 04/Mar/2022 12:30 Date Approval Issued: 17/Mar/2022 14:40 REB Approval Expiry Date: 17/Mar/2023

Dear Prof. Lorne Campbell

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the above mentioned study, as of the date noted above. NMREB approval for this study remains valid until the expiry date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

This research study is to be conducted by the investigator noted above. All other required institutional approvals and mandated training must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
DoA_Manipulations_Side-by-side	Other Data Collection Instruments	28/Jan/2022	1
DoA_Email_Recruitment_Script_01.28.22	Recruitment Materials	28/Jan/2022	1
DoA_LOI_03.14.22	Implied Consent/Assent	14/Mar/2022	1
DoA_Debrief_All_03.14.22	Debriefing Letter	14/Mar/2022	2
DoA_PartI_Survey_03.14.22	Online Survey	14/Mar/2022	2
DoA_PartII_Survey_03.14.22	Online Survey	14/Mar/2022	2

Documents Acknowledged:

Document Name	Document Type	Document Date	Document Version
DoA_Debrief_Mental_Health_Resources	Other Materials	24/Jan/2022	1
DoA_PartII_Screen_01.28.22	Screening Form/Questionnaire	28/Jan/2022	1
DoA_PartI_Screen_03.14.22	Screening Form/Questionnaire	14/Mar/2022	2

No deviations from, or changes to the protocol should be initiated without prior written approval from the NMREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Ms. Zoë Levi, Research Ethics Officer on behalf of Dr. Randal Graham, NMREB Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

EDUCATION

Expected Summer 2022	Ph.D. , Social Psychology Western University, London, ON, Canada
2018	M.Sc. , Social Psychology Western University, London, ON, Canada
2014	B.Sc. , Psychology & Anthropology University of Oregon, Eugene, OR, USA

PUBLICATIONS

REFEREED JOURNAL ARTICLES

- Koessler, R. B., Kohut, T., & Campbell, L. (2019). When your boo becomes a ghost: The association between breakup strategy and breakup role in experiences of relationship dissolution. *Collabra: Psychology*, *5*(1), 29. <u>http://doi.org/10.1525/collabra.230</u>
- Koessler, R. B., Samani, N. M. V., Charlot, N., & Campbell, L. (2022). The relationship between infidelity risk and oral sex: An independent replication of Pham and Shackelford (2013) and Pham et al. (2013). *Personality and Individual Differences*, 189, 111496. <u>https://doi.org/10.1016/j.paid.2021.111496</u>

PREPRINTS

Koessler, R. B., Kohut, T., & Campbell, L. (2019, November 1). Integration and expansion of qualitative analyses of relationship dissolution through ghosting. [Preprint]. https://doi.org/10.31234/osf.io/3kvdx

CONFERENCES

DATA BLITZES

Charlot, N. H., Koessler, R. B., Samani, N., & Campbell, L. (2020, July). The non-replicable relationship between infidelity risk and oral sex: A replication of Pham and Shackelford (2013) and Pham, Shackelford, and Sela (2013). Accepted as a Data Blitz for the 2020 International Association of Relationship Research Conference. London, England.

POSTER PRESENTATIONS

- Heerey, E. A., Babcock, S. E., Clerke, A. S., & Koessler, R. B. (2020, May). From passive to interactive! Flipping a large introductory classroom. Poster accepted for the 2020 Association for Psychological Science, Chicago, IL, USA. **cancelled due to COVID-19**
- Koessler, R. B., Kohut, T., & Campbell, L. (2019, February). Exploring ghosting: Relationship dissolution in the 21st century. Poster presented at the 2019 Society for Personality and Social Psychology Conference, Portland, OR, USA. <u>https://osf.io/ufw8g/</u>
- Koessler, R. B., Kohut, T., & Campbell, L. (2018, July). When your boo becomes a ghost: Influences of breakup strategy and breakup role on experiences of relationship dissolution. Poster presented at the 2018 International Association of Relationship Research Conference. Fort Collins, CO, USA. <u>https://osf.io/epy23/</u>

AD HOC REVIEWER

Cyberpsychology, Journal of Social and Personal Relationships

INVITED TALKS

- Koessler, R. B. (2020, March). Increasing Research Transparency Using the Open Science Framework. [Lightning talk]. R Ladies of London. London, ON. <u>https://osf.io/z87p9/</u>
- Koessler, R. B. (2020, March). *When Your Boo Becomes a Ghost*. [Guest speaker]. Trinity University, San Antonio, TX. <u>https://osf.io/h5jck/</u>
- Koessler, R. B. (2019, March). *Transparent Research Practices: Preregistration and Open Notebook*. Open Collaborations: Early Career Researcher Edition. [Symposium speaker]. University of Rochester, Rochester, NY. <u>https://osf.io/3jsyf/</u>

RESEARCH EXPERIENCE

 2016 – present Research Collaborator, Love Lab, Western University Advisor: Lorne Campbell, Ph.D.
 2015 Research Assistant, Memory, Attention, and Individual Differences (MAID) Lab, University of Oregon Advisor: Nash Unsworth, Ph.D. & Matthew Robinson, Ph.D.

2014 - 2016	Research Assistant , Personality and Social Dynamics Lab, University of Oregon Advisor: Sanjay Srivastava, Ph.D. & Nicole Lawless-DesJardins, Ph.D.
2014	Research Assistant , Personal Relationships, Interpersonal Stress, and Mindfulness (PRISM) Lab, University of Oregon Advisor: Heidemarie Laurent, Ph.D. & Robin Hertz, Clinical Doctoral Student
2013 - 2014	Research Assistant , Developing Mind Lab, University of Oregon Advisor: Louis Moses, Ph.D. & Caitlin Mahy, Ph.D.

AWARDS AND HONORS

AWARDS

2020	RepliCATS Travel Grant for SIPS 2020, University of Melbourne
2018	Graduate Student Travel Award, International Association of Relationship
	Research Conference
2013-2014	Laurel Scholars Award, University of Oregon
2013-2014	Edith Chambers Scholarship, University of Oregon
2010-2014	Resident Dean's Scholarship, University of Oregon
2010-2011	General Development Scholarship, University of Oregon
HONORS	
2018-2019	Nominee, Graduate Student Teaching Assistant Award
2016-2017	Nominee, Graduate Student Teaching Assistant Award

- 2014
- Nominee, Graduate Student Teaching Assistant Award Phi Beta Kappa, University of Oregon Mortar Board Scholar, University of Oregon 2013-2014
- Dean's List, University of Oregon 2010-2013