Fachbereich Erziehungswissenschaft und Psychologie
der Freien Universität Berlin

Process Modeling in Social Decision Making

Dissertation
zur Erlangung des akademischen Grades
Doktor der Philosophie (Dr. phil.)
Doctor of Philosophy (Ph.D.)

vorgelegt von
Jolene H. Tan
Berlin, 29 Februar 2016

Erstgutachter
Prof. Dr. Gerd Gigerenzer

Zweitgutachterin
Prof. Dr. Katja Liebal

Tag der Disputation
24. Oktober 2016

Zusammenfassung

Das Forschungsinteresse an der Frage wie die Vorteile gesellschaftlicher Kooperation genutzt und gleichzeitig die Risiken einer Ausbeutung durch andere vermieden werden können ist in den letzten Jahrzehnten signifikant gestiegen. Dennoch wissen wir sehr wenig darüber wie gesellschaftliche Entscheidungen getroffen werden; insbesondere ist unklar, auf welchen kognitiven Entscheidungsprozessen gesellschaftliche Interaktionen basieren. Das Ziel meiner Dissertation ist es diese wenig erforschten Entscheidungsprozesse, die gesellschaftlichen Interaktionen zu Grunde liegen, systematisch zu studieren. Ich benutze die Theorien „Adapted Mind“ und „Bounded Rationality“, um zu untersuchen wie Menschen evolutionär wiederkehrende Probleme gesell-schaftlichen Lebens unter den realistischen Bedingungen des Zeitdrucks, der unvollständigen Information und begrenzter menschlicher Informations-verarbeitungskapazitäten lösen. Ich kombiniere diese theoretischen Ansätze mit der Methode der Kognitiven Prozessmodellierung, um genaue Vorhersagen über die zugrunde liegenden Entscheidungenprozesse zu generieren und empirisch zu testen. In der Einführung stelle ich die wichtigsten Kontroversen des Forschungsfeldes vor, die den Hintergrund für die darauf folgenden Kapitel bilden. Im ersten Kapitel schlage ich einen theoretischen Rahmen vor, der genutzt werden kann, um zu bestimmen, was ein Kognitives Prozessmodell ist. Dieser theoretische Rahmen enthält eine Menge von notwendigen Bedingungen, die ein Modell erfüllen muss, um als Prozessmodell in Betracht gezogen zu werden. Das „How To“ Format dieses Kapitels kann als Anleitung zur Konstruktion von Prozessmodellen dienen. Das zweite Kapitel erklärt wie Prozessmodelle genutzt werden können, um gesellschaftliche Entscheidungen wie Forgiveness zu studieren. Ich habe zwei Modelle – ein heuritisches Fast-and-Frugal Trees Modell und ein lineares Modell names Franklin's Rule – entwickelt und getestet. Ich habe gefunden, dass beide Modelle meine empirischen Daten ähnlich gut beschreiben und vorhersagen (Richtigkeit von ca. 80 % in Beschreibung und von ca. 70 % in Vorhersage). Das dritte Kapitel erweitert das vorhergehende, indem es untersucht wie die Information über die Basirate der Freundlichkeit des gesellschaftlichen Umfelds genutzt wird, um zu entscheiden, ob man einer Person verzeiht. Ich zeige erstens, dass die Information über die Basisrate in Forgiveness Entscheidungen berücksichtigt wird; und zweitens, dass sich diese als Social Trust, definiert als eine Erwartung darüber, ob andere grundsätzlich gut oder schlecht sind, ausdrücken lässt. Zusammenfassend, erweitert meine Dissertation unser Verständnis über Kooperation, indem es die kognitiven Prozesse, die gesellschaftlichen Interaktionen zu Grunde liegen, präzise definiert und testet.

Summary

Understanding how the benefits of cooperation can be reaped while the risks of exploitation from other individuals can be managed has received significant research attention in the past few decades. However, despite its prominence, little is known about how we make these social decisions; it is unclear what decision processes underlie our interactions with others. My goal in this dissertation was to investigate the decision processes of social interactions. I adopted the perspective of the "adapted mind" and "bounded rationality" in order to investigate how humans solve the evolutionarily recurrent problems of social living under limitations of time, information, and computational ability. I combined these theoretical foundations with the methodology of cognitive process modeling, which enabled me to test fine-grained predictions about the underlying decision processes. In the introduction chapter, I provided a brief overview of some controversies in the field so as to provide the backdrop for the rest of the chapters. In the first chapter, I proposed a framework that can be used to qualify what is a cognitive process model. The framework contains necessary conditions that a model needs to fulfill in order to be considered a process model. The "how to" format of the chapter can serve as a guide for building process models. The second chapter is an exemplification of how process models can be used to study social decisions such as forgiveness. I developed and tested two models—the heuristic-based fast-and-frugal trees, and the linear model Franklin's rule—and found that both models performed similarly well (accuracy of ~80% in description and ~70% in prediction). The third chapter extended the previous by examining how base rate information about the benevolence of the social environment is used in decisions about whether to forgive. I provided evidence that base rate information is used in forgiveness decisions and it is expressed as a level of social trust, a belief about whether people are generally benevolent or malevolent. Taken together, my dissertation advanced understanding about cooperation by specifying and testing the decision processes that underlie social interaction.

Contents

Process Modeling in Social Decision Making

Chapter 0.

Social decision making: From function to process

1. INTRODUCTION

1.1. What is social decision making?

Should I trust? Should I harm or help? Should I forgive or punish? Should I enter or exit a relationship? Some of life's most agonizing decisions are social—they have consequences that impact or depend on the actions of other individuals; they have inputs that are taken from the behavior or inferred mental states of other individuals.

Understanding the decisions that support social life and enable the scale of cooperation prevalent in human societies has received great research attention in the past decade (Nowak, 2006; Tomasello & Vaish, 2011). These decisions have been studied under the banners of altruism, morality, conflict, and competition. They have been investigated in disciplines beyond psychology, including biology, economics, as well as philosophy. However, despite its prominence, little is known about the process of such decisions. How we make decisions about interacting with others is still not clearly understood.

My goal in this dissertation is to shed some light on the processes of social decisions and to take steps towards integrating several theoretical traditions. My approach draws from the view of the mind as a product of evolution containing cognitive adaptations to deal with evolutionarily recurrent and significant tasks (Cosmides, Barrett, & Tooby, 2010; Laland & Brown, 2002). These cognitive adaptations have allowed our ancestors to successfully survive in harsh uncertain environments, and continue to enable modern humans to solve the problems that arise from social living. This view of the "adapted mind" is complementary with the perspective that humans are bounded but rational, with the ability to find good solutions for difficult problems under limitations of time, information, and computational ability (Gigerenzer, Todd, & The ABC Research Group, 1999; Hertwig, Hoffrage, & The ABC Research Group, 2012). I have combined these two perspectives with the methods of cognitive process modeling, which enable the testing of fine-grained predictions about the processes underlying social decisions.

In this chapter, I will provide a brief overview of some of the theoretical and methodological approaches in the study of social decision-making. I will discuss some controversies in the field concerning social rationality that will provide the backdrop for the next few chapters of the dissertation. Finally, I will discuss how the chapters of the dissertation address these controversies and advances understanding of social decision making.

1.2. Searching for social rationality

Odious behavior ("sin") is at the heart of
our most powerful research in social psychology.

—Aronson, 1999

People use moral heuristics […] that lead to
mistaken and even absurd moral judgments.

—Sunstein, 2005

The dominant view of people's social decision-making competencies is unflattering. People are seen as hapless and "predictably irrational" decision makers who cannot help but commit errors (Ariely, 2009; Kahneman, 2012; Nisbett & Ross, 1980; Thaler & Sunstein, 2008). This view, prevalent in social psychology and behavioral economics, was greatly influenced by the heuristics-and-biases program (Gigerenzer, 1991; Gilovich, Griffin, & Kahneman, 2002; Krueger & Funder, 2004). Research in this tradition has been (disparagingly) referred to as "the 'People are Stupid' school of psychology" (Kihlstrom, 2004, p. 36) for its seemingly inexhaustible compilation of biases that cause social actors to stumble. Some prominent examples include the actor-observer bias, the self-serving bias, the trait ascription bias, the projection bias, the base rate fallacy, and of course, the fundamental attribution error (Krueger & Funder, 2004).

But yet, in real life, people mostly make reasonable choices and lead happy productive lives (Diener & Diener, 1996). They choose appropriate and attractive partners, forgive small harms, maintain long-term rewarding relationships, punish cheaters, and help to maintain order in society. How do we reconcile these competencies and accomplishments with the view of the hapless social actor?

One reason for the bleak view is that the heuristics-and-biases paradigm relies on inappropriate normative benchmarks (Gigerenzer, 1991, 1996; Koehler, 1996). It ignores important distinctions in statistics and misunderstands rationality by assuming that there is only one "correct" answer. For example, the classic demonstration that people commit the "conjunction fallacy" (i.e., the "Linda problem") had participants read a description of Linda being "single, outspoken and bright," as well as "deeply concerned with issues of discrimination and social justice" (Tversky & Kahneman, 1983). They then had to judge whether Linda is more probably, A) a bank teller, or B) a bank teller and a feminist. Because the probability of a conjunction of two events—Linda being both a bank teller and a feminist—cannot be greater than that of one of its parts, the correct answer is said to be A. Almost all participants however, responded B, inciting the conclusion of pervasive fallacious reasoning.

Such a conclusion has been criticized on several fronts. Notably, the question was concerned with the probability of a single-event, which is held by frequentist

statisticians to be outside probability theory's jurisdiction of repeated events (Gigerenzer, 1991). Furthermore, it assumes that participants approached the question as one of deductive logic instead of a conversation governed by maxims of cooperativeness (Dulany & Hilton, 1991). In doing so, it ignored "the human capacity for semantic and pragmatic inference" (p. 276, Hertwig & Gigerenzer, 1999) by concluding that the ability to go beyond the explicit to infer the speaker's intended meaning is a defect instead of a hallmark of human intelligence.

Perhaps the paradigm's biggest failing is its explanatory emptiness. It is mostly satisfied with describing various cognitive failures and explaining its occurrence by way of redescription (see Gigerenzer, 1996). For example, the explanation of why most people choose the incorrect answer to the Linda problem has been that they use a "representativeness heuristic" (p. 299, Tversky & Kahneman, 1983). What this heuristic means exactly is not precisely spelt out and no falsifiable process model of the phenomenon has been proposed (Gigerenzer, 1991). With such an explanation, the field is left with no clearer understanding of why such pervasive "irrationality" exists and how it is produced.

To be fair, this approach has been instrumental in disabusing the notion of humans as perfectly rational "demons" unconstrained by time, knowledge, and computational capacities (Gigerenzer et al., 1999; Gintis, 2000; Henrich et al., 2001). Nevertheless, the study of social decisions needs to go beyond asking the pointless and judgmental question of whether people are rational, to asking why people make the decisions they do. What are the cognitive processes that underlie social decisions? What is the adaptive significance of various social decisions? How do these processes lead to fitness-enhancing decisions? To answer these questions, a new theoretical and methodological approach is needed.

1.3. The adapted mind: Bounded but rational

Let us take a moment to imagine what modern physics would look like if it did not embrace Albert Einstein's theory of relativity. Or, what the field of biology would be if it did not incorporate Charles Darwin's theory of evolution by natural selection. How could the leaps in these fields be made without these grand theories?

Psychology, in contrast, has no such theory that unifies large bodies of its work. The state of theorizing in psychology has been lamented as "a patchwork of small territories" with researchers being reluctant to interact with the theories of related topics (p. 734, Gigerenzer, 2010). More injurious than the lack of an integrated theory is the lack of desire to build one (Mischel, 2008; Watkins, 2010). Indeed, modern psychology has drifted far from its roots—Kurt Lewin, one of the earliest psychologists, had the maxim, "there is nothing as practical as a good theory".

In recent years, a candidate for such a unifying theory has emerged. Perhaps reflecting the contemporary trend of dissolving disciplinary borders, this theory is not

limited to psychology, but cuts across the various fields interested in cognition and behavior such as anthropology, primatology, neuroscience, biology, and economics. That candidate is the *theory of the adapted mind*, an application of evolution by natural selection to the study of cognition and behavior (e.g., Buss, 1995; Cosmides & Tooby, 1992; Daly & Wilson, 1999; Tomasello, 1999).

This theoretical approach, which has come to be known as "evolutionary psychology", starts from the assumption that "the mind consists of a set of adaptations designed to solve the long-standing adaptive problems humans encountered as hunter-gatherers" (p. 163, Cosmides & Tooby, 1992). Human's social cognition is thus the result of our ancestors having had to solve the adaptive problems that arise from social life (Cosmides et al., 2010; Kenrick et al., 2009). Many of such problems persist even in modern life; for example, being occasionally harmed is an inevitable part of interacting with others and whether to continue a relationship in its aftermath (i.e., to forgive and reconcile) is a decision that many social animals including modern humans and their hunter-gatherer ancestors would have to make. I will return to the topic of forgiveness and cooperation in Chapters 2 and 3.

At its core, an evolutionary guided study of cognition uses an engineering approach to make testable predictions about the information-processing structure of evolved psychological mechanisms. However, despite its stated emphasis on process, the research products of evolutionary psychology still resembles much of the work that came before; evolutionary psychological approaches are still rarely concerned with specifying the mechanisms of adapted systems, and some theorists have postulated "as-if" processes that resemble the complex demon computations of rational choice theory (e.g., Andrews, Gangestad, & Matthews, 2002; Delton & Robertson, 2012; Jensen & Petersen, 2011).

To specify psychological mechanisms at a better resolution, the evolutionary approach needs to be connected with a theory of decision-making in the real world. In particular, the theory of bounded rationality is a good candidate because it is highly compatible with the ideas of the adapted mind (Cosmides & Tooby, 1994; Gigerenzer et al., 1999). The two theories share the same presupposition that natural selection has endowed agents with cognitive capacities for navigating life's tribulations. Where they differ is on the level of analysis—in contrast to evolutionary approaches that typically focus on functional explanations, bounded rationality focuses on the mechanism. Bounded rationality suggests that the decision-making mechanism (or process) of humans and other animals are likely to take the form of simple heuristics (for example, fast-and-frugal decision trees, which will be discussed in chapter 2) that allow fitness-enhancing decisions to be made under constraints of time, information, and processing capabilities (Gigerenzer, 2010a; Hutchinson & Gigerenzer, 2005). To this end, the bounded rationality approach has amassed a substantial literature showing that simple heuristics work well in many real-world decision tasks and are as good as complicated

decision strategies that are less cognitive plausible (Gigerenzer, Hertwig, & Pachur, 2011). More discussion on the merits of simple heuristics will follow in chapter 2.

1.4. What are process models?

The methodology of process modeling is closely linked to the cognitive revolution in psychology and the study of bounded rationality (Katsikopoulos & Lan, 2011). The cognitive revolution in the 1950s legitimized the study of the mind and highlighted the importance of invoking mental processes to explain behavior (Hastie & Dawes, 2010; Miller, 2003). Herbert Simon, a key player in the cognitive revolution who also pioneered the study of bounded rationality, provided the first demonstration that the mind can be studied using simple programming languages and emphasized that decision making models should incorporate the underlying processes (Gregg & Simon, 1967). Thus, the mind is analogous to a computer executing algorithms, and the challenge for cognitive science was to build models that contain algorithms that capture the processes of decision-making. Process models were also seen by Simon to be a tool of theory integration, he envisioned that process models could be combined "over wider and wider ranges of behaviors until that happy day arrives when we shall have a theory of the whole cognitive man" (Gregg & Simon, 1967, p. 276). Presently, claims of process modeling is commonplace amongst those who follow Simon's lead on bounded rationality closely (Gigerenzer, 2010b; Katsikopoulos, 2014).

Sixty-years after the cognitive revolution, the term "process model" has become widely used in cognitive science[1]. However, it remains unclear what constitutes a process model and how one should be built. While it is uncontroversial that process models are those that "process information" (e.g., Oppenheimer & Kelso, 2015), that description is too general to be useful. In an attempt to resolve these conceptual issues, some have invoked Marr's (1982) three levels of analysis and claimed that process models can be illustrated by the algorithmic level (e.g., Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010). Others, taking a different route, have contrasted process models with the "as-if" models of rational choice theory (Berg & Gigerenzer, 2010). None of these attempts have succeeded in fostering any agreement. This lack of clarity will be addressed in chapter 1.

2. OVERVIEW OF THE DISSERTATION

The three chapters of the dissertation fulfill my goal to shed light on the processes of social decisions as follows:

- Chapter 1, *How to build a process model*, attempts to clarify what process models are using a framework with a list of conditions that a model needs to fulfill

[1] In the last decade, the term "process model" has appeared in approximately 12,400 documents from cognitive psychology and the citations of database-indexed papers using this term have increased steeply (even when controlling for general positive trend).

in order to be considered a process model. This approach stems from the belief that the issue of what is a process model is independent from a second issue about whether process models are good models (Berg & Gigerenzer, 2010; Birnbaum, 2008). Only when there is agreement about what process models are, does the second issue have a chance of being resolved. Thus, the framework can be applied at the point of model construction, including in the absence of data since most model evaluations focus on the ability of models to fit data (e.g., Lewandowsky & Farrell, 2010; Pitt, Myung, & Zhang, 2002; Shiffrin, Lee, Kim, & Wagenmakers, 2008). Furthermore, the "how to" format of the paper will help encourage greater process modeling among psychologists and others interested in modeling decision processes. A version of this chapter was written with Jana Jarecki (University of Basel) and Mirjam Jenny (Max Planck Institute for Human Development).

- Chapter 2, *Error management in forgiveness: A process modeling approach*, is an exemplification of how process models can be used to study decisions in the evolutionary domain. Many evolutionarily-recurrent decisions have been argued to be error management tasks because they are made under uncertainty and feature asymmetric costs and benefits (Haselton & Nettle, 2005; Johnson, Blumstein, Fowler, & Haselton, 2013). According to error management theory, cognitive biases are adaptive and not defective because they promote effective decisions by reducing the likelihood of the more costly error. Nevertheless, how bias is implemented cognitively has not been specified or tested directly (see McKay & Efferson, 2010). This chapter addresses this gap by specifying bias as the decision criterion in signal detection theory, and then using process modeling to test the factors that influence where the bias is set. We developed and tested two models—the heuristic-based fast-and-frugal trees, and the linear model Franklin's rule—that embody the logic of error management but make different assumptions about cognitive implementation. We found that both models performed similarly well (accuracy of ~80% in description and ~70% in prediction), and that the magnitude of the bias could be predicted by error management and signal detection theory. Though this chapter focuses on decisions about forgiveness and cooperation, the general approach of using process models can be potentially applied to other evolutionary-recurrent error management decisions. A version of this chapter was written with Shenghua Luan and Konstantinos Katsikopoulos (both, Max Planck Institute for Human Development).
- Chapter 3, *Assessing the base rate in forgiveness decisions: The adaptive function of social trust*, extends the previous chapter's error management framework of forgiveness by examining how base rate information is used in decisions about whether to forgive. Whether base rate information is used in decision-making is a controversial topic (e.g., Birnbaum, 1983; Gigerenzer, 1991; Gilovich et al., 2002; Koehler, 1996; Welsh & Navarro, 2012). While some

researchers claim the universality and pervasiveness of base-rate neglect and hold it up as an example of a cognitive fallacy (Kahneman & Tversky, 1996), others claim that this conclusion is not warranted given the evidence available (Juslin, Wennerholm, & Winman, 2001; Koehler, 1996). If individuals do make decisions about forgiveness according to error management theory (as demonstrated in chapter 2), then base rate information should be used in the decision as well. This chapter argues that an individual's assessment of the base rate in forgiveness is expressed as a level of social trust, a belief about whether people are generally benevolent or malevolent. Since different individuals are embedded in different environments with varying base rates, individual differences in forgiveness are likely to reflect that difference. I alone wrote this chapter.

3. CONCLUSION

Humans and other social animals are endowed with capacities that enable them to deal effectively with the problems that arise from social living. These capabilities allow them to reap the benefits of cooperation while managing the risks of exploitation from other individuals. Understanding these capabilities requires going beyond investigations of rationality (or irrationality), to ask why individuals make the decisions they do, and what are the processes that produce these decisions. In order to provide satisfactory explanations of how individuals make social decisions, it is necessary to study them at levels that are complementary and build models that reach across levels of analysis (Simon, 1992; Tinbergen, 1963). In this dissertation, I have used insights about the adaptive functions of psychological concepts such as forgiveness (chapter 2) and trust (chapter 3) and applied cognitive process modeling techniques (chapter 1) to specify how multiple pieces of information are combined to produce a decision. It is my hope that this approach will spur more research in this tradition in order to forge greater understanding about how social decisions are made.

REFERENCES

Andrews, P. W., Gangestad, S. W., & Matthews, D. (2002). Adaptationism–How to carry out an exaptationist program. *Behavioral and Brain Sciences*, *25*, 489–553. Retrieved from http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=172185&fileId=S0140525X02000092

Ariely, D. (2009). *Predictably Irrational*. New York: HarperCollins Publishers. doi:10.2501/S1470785309200992

Berg, N., & Gigerenzer, G. (2010). As-if behavioral economics: Neoclassical economics in disguise? *History of Economic Ideas*, *18*(1), 133–165. doi:10.2139/ssrn.1677168

Birnbaum, M. H. (1983). Base rates in Bayesian inference : Signal detection analysis of the cab problem. *The American Journal of Psychology*, *96*(1), 85–94. doi:10.2307/1422211

Birnbaum, M. H. (2008). Evaluation of the priority heuristic as a descriptive model of risky decision making: comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review*, *115*(1), 253–262. doi:10.1037/0033-295X.115.1.253

Buss, D. M. (1995). Evolutionary psychology: A new paradigm for psychological science. *Psychological Inquiry*, *6*(1), 1–30. doi:10.1207/s15327965pli0601_1

Cosmides, L., Barrett, H. C., & Tooby, J. (2010). Adaptive specializations, social exchange, and the evolution of human intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 9007–9014. doi:10.1073/pnas.0914623107

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The Adapted Mind* (pp. 163–228). New York: Oxford University Press. doi:10.1098/rstb.2006.1991

Cosmides, L., & Tooby, J. (1994). Beyond intuition and instinct blindness: Toward an evolutionarily rigorous cognitive science. *Cognition*, *50*, 41–77. doi:10.1016/0010-0277(94)90020-5

Daly, M., & Wilson, M. I. (1999). Human evolutionary psychology and animal behaviour. *Animal Behaviour*, *57*(3), 509–519. doi:10.1006/anbe.1998.1027

Delton, A. W., & Robertson, T. E. (2012). The Social Cognition of Social Foraging: Partner Selection by Underlying Valuation. *Evolution and Human Behavior*, *33*(6), 715–725. doi:10.1016/j.biotechadv.2011.08.021.Secreted

Diener, E., & Diener, C. (1996). Most people are happy. *Psychological Science*, *7*(3), 181–185. doi:10.1111/j.1467-9639.1991.tb00167.x

Dulany, D. E., & Hilton, D. J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition*, *9*(1), 85–110. doi:10.1521/soco.1991.9.1.85

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases.” *European Review of Social Psychology*, *2*(1), 83–115. doi:10.1080/14792779143000033

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*(3), 592–596. doi:10.1037/0033-295X.103.3.592

Gigerenzer, G. (2010a). Moral satisficing: Rethinking moral behavior as bounded rationality. *Topics in Cognitive Science*, *2*(3), 528–554. doi:10.1111/j.1756-8765.2010.01094.x

Gigerenzer, G. (2010b). Personal reflections on theory and psychology. *Theory & Psychology*, *20*(6), 733–743. doi:10.1177/0959354310378184

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: The foundations of adaptive behavior*. (G. Gigerenzer, R. Hertwig, & T. Pachur, Eds.). New York: Oxford University Press. doi:10.1093/acprof:oso/9780199744282.001.0001

Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press. doi:10.1007/s13398-014-0173-7.2

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive*

judgment. Cambridge, UK: Cambridge University Press. doi:10.5465/AMR.2004.14497675

Gintis, H. (2000). Beyond Homo economicus: Evidence from experimental economics. *Ecological Economics*, *35*(3), 311–322. doi:10.1016/S0921-8009(00)00216-0

Gregg, L. W., & Simon, H. A. (1967). Process models and stochastic theories of simple concept formation. *Journal of Mathematical Psychology*, *4*(2), 246–276. doi:10.1016/0022-2496(67)90052-1

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364. doi:10.1016/j.tics.2010.05.004

Haselton, M. G., & Nettle, D. (2005). The paranoid optimist: An integrative evolutional model of cognitive biases. *Personality and Social Psychology Review*, *10*(1), 47–66. doi:10.1207/s15327957pspr1001_3

Hastie, R., & Dawes, R. (2010). *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making*. (R. Hastie & R. M. Dawes, Eds.). SAGE Publications.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Mcelreath, R., … Mcelreath, R. (2001). In search of homo economicus: Behavioral experiments in 15 small-scale societies. *The American Economic Review*, *91*(2), 73–78.

Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: how intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*(4), 275–305. doi:10.1002/(SICI)1099-0771(199912)12:4<275::AID-BDM323>3.0.CO;2-M

Hertwig, R., Hoffrage, U., & The ABC Research Group. (2012). *Simple Heuristics in a Social World*. New York: Oxford University Press. doi:dx.doi.org/10.1093/acprof:oso/9780195388435.003.0001

Hutchinson, J. M. C., & Gigerenzer, G. (2005). Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behavioural Processes*, *69*(2), 97–124. doi:10.1016/j.beproc.2005.02.019

Jensen, N. H., & Petersen, M. B. (2011). To Defer or To Stand Up? How Offender Formidability Affects Third Party Moral Outrage. *Evolutionary Psychology*, *9*(1), 118–136.

Johnson, D. D. P., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology and Evolution*, *28*(8), 474–481. doi:10.1016/j.tree.2013.05.014

Juslin, P., Wennerholm, P., & Winman, A. (2001). High-level reasoning and base-rate use: Do we need cue-competition to explain the inverse base-rate effect? *Journal of Experimental Psychology-Learning Memory and Cognition*, *27*(3), 849–871. doi:10.1037/0278-7393.27.3.849

Kahneman, D. (2012). *Thinking, Fast and Slow*. London: Penguin Books. doi:10.1007/s13398-014-0173-7.2

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*(3), 582–591; discusion 592–596. doi:10.1037/0033-295X.103.3.582

Katsikopoulos, K. V. (2014). Bounded rationality: The two cultures. *Journal of Economic Methodology*, *9427*(January). doi:10.1080/1350178X.2014.965908

Katsikopoulos, K. V., & Lan, C.-H. (2011). Herbert Simon's spell on judgment and decision making. *Judgment and Decision Making*, *6*(8), 722–732.

Kenrick, D. T., Griskevicius, V., Sundie, J. M., Li, N. P., Li, Y. J., & Neuberg, S. L. (2009). Deep Rationality : the Evolutionary Economics of Decision Making. *Social Cognition*, *27*(5), 764–785.

Kihlstrom, J. F. (2004). Is there a "People are Stupid" school in social psychology? *Behavioral and Brain Sciences*, *27*(3), 348– 348.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, *19*, 1–53. doi:10.1017/S0140525X00041157

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *The Behavioral and*

Brain Sciences, *27*(3), 313–327; discussion 328–376. doi:10.1017/S0140525X04000081

Laland, K. N., & Brown, G. R. (2002). *Sense and nonsense: Evolutionary perspectives on human behaviour*. New York: Oxford University Press.

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice*. Thousand Oaks, CA: SAGE Publications.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: W. H. Freeman and Company.

McKay, R. T., & Efferson, C. (2010). The subtleties of error management. *Evolution and Human Behavior*, *31*(5), 309–319. doi:10.1016/j.evolhumbehav.2010.04.005

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144. doi:10.1016/S1364-6613(03)00029-9

Mischel, W. (2008). The Toothbrush Problem. *Observer*. Retrieved from https://aps.psychologicalscience.org/index.php/publications/observer/2008/december-08/the-toothbrush-problem.html

Nisbett, R. E., & Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. New York: Prentice-Hall.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314*, 1560–1563. doi:10.1126/science.1133755

Oppenheimer, D. M., & Kelso, E. (2015). Information Processing as a Paradigm for Decision Making. *Annual Review of Psychology*, *66*(1), 277–294. doi:10.1146/annurev-psych-010814-015148

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, *109*(3), 472–491. doi:10.1037/0033-295X.109.3.472

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical bayesian methods. *Cognitive Science*, *32*(8), 1248–84. doi:10.1080/03640210802414826

Sunstein, C. R. (2005). Moral heuristics. *Behavioral and Brain Sciences*, *28*(4), 531–573. doi:10.1017/S0140525X05000099

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press. doi:10.1086/592313

Tomasello, M. (1999). The human adaptation for culture. *Annual Review of Anthropology*, *28*(1999), 509–529. doi:10.1017/S0140525X0003123X

Tomasello, M., & Vaish, A. (2011). Origins of Human Cooperation and Morality. *Annual Review of Psychology*, *64*(1), 120717165617008. doi:10.1146/annurev-psych-113011-143812

Tversky, A., & Kahneman, D. (1983). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment Amos. *Psychological Review*, *90*(4), 293–315.

Watkins, M. J. (2010). Models as toothbrushes. *Behavioral and Brain Sciences*, *7*(01), 86. doi:10.1017/S0140525X00026303

Welsh, M. B., & Navarro, D. J. (2012). Seeing is believing: Priors, trust, and base rate neglect. *Organizational Behavior and Human Decision Processes*, *119*(1), 1–14. doi:10.1016/j.obhdp.2012.04.001

Chapter 1.

How to build a process model

Abstract. This chapter introduces a framework for building cognitive process models. The goal is to clarify what process models are and to offer guidance about how to build one. This framework was designed to be applicable to models before and also after they are empirically tested. We propose that a process model needs to fulfill the following:

- In addition to the input (i.e., the information entering the cognitive system) and the output (i.e., the decision or behavior of interest), it needs to include at least one intermediate stage (i.e., a cognitive event that transforms the input into the output).
- A clear conceptual scope is provided for the input, output, as well as the intermediate stage.
- Separate and testable predictions for both the output and intermediate stage can be derived.
- The intermediate stage is compatible with current knowledge of human cognition.

Keywords. process model, cognitive model, computational model, Marr's levels

1. INTRODUCTION

One of the earliest mention of the term "process model" was by Gregg and Simon (1967) who advocated the use of models that make detailed assumptions about processes to test inconsistencies in theories. Since then, the term has become widely used in cognitive science. In the last decade, the term "process model" has appeared in approximately 12,400 documents from cognitive psychology; the citations of database-indexed papers using this term have increased steeply even when controlling for a positive citation trend (see Figure 1). There has also been a corresponding growth in interest in process-tracing measures (Schulte-mecklenbeck, Kühlberger, & Ranyard, 2011). Figure 1 shows that in 2013 the articles mentioning "process model" were cited more than those mentioning "formal model", and "agent-based model".

Despite this trend, there is no common understanding of what is a process model and usage of the term in the literature provides little guidance on how process models should be built (Grüne-Yanoff, 2014). What little advice is out there is often too unspecified to be helpful. For example, Lewandowsky and Farrell (2010) recommend that process models need to describe the process in detail, and parameters need to have a psychological interpretation. However, it is unclear what level of detail is required and what constitutes a psychological interpretation. How do researchers interested in

building process models proceed? What can be done to transform non-process models into process models? What kind of data is needed to test these models?

The aim of this chapter is to clarify what process models are using a framework of necessary conditions. This can also be used as a “how to” guide to building process models. We will first provide an overview of some uses of the term in the literature, and then we will present our framework and how it may be applied.

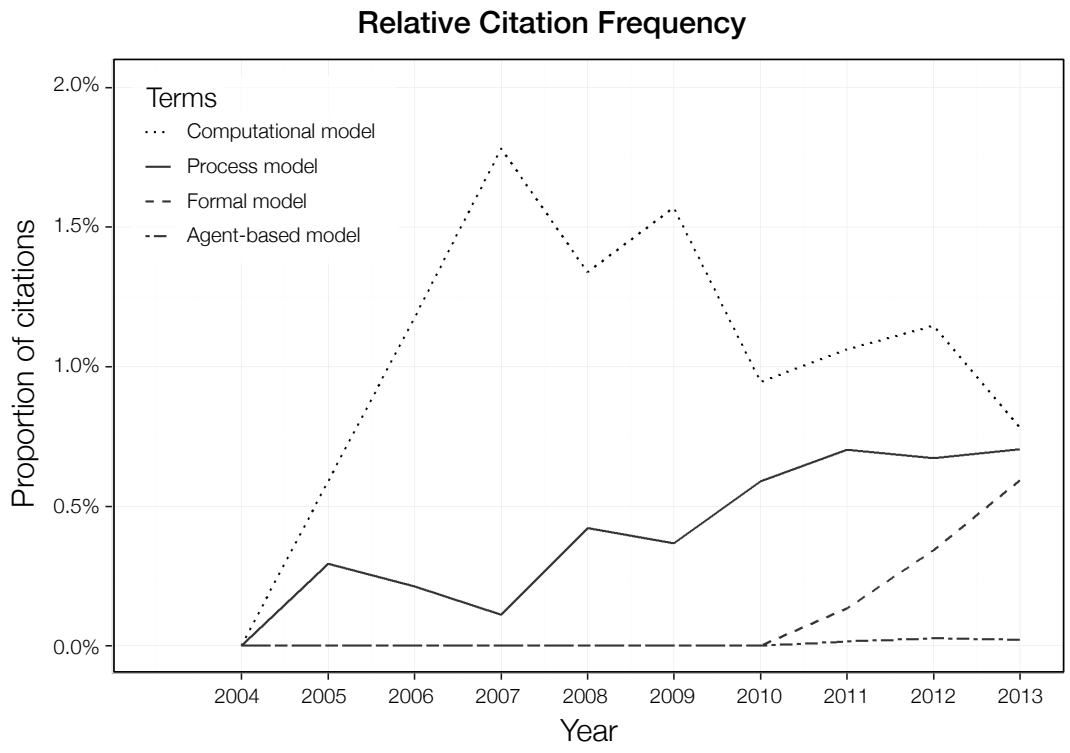

Figure 1. Publications using the term "process model" have been increasing cited over the past decade. The trends shown have been controlled for the general rise in citations. The solid line shows the proportion of citations of articles including the term "process model" AND cognitive science AND judgment and decision making JDM relative to citations including the latter terms but excluding "process model". The dotted lines depict the respective proportions for articles including the term "agent-based model", "formal model", or "computational model" in place of "process model." Cognitive science and JDM were operationalized as "cognitive," "psychology," AND "judgment and decision making" OR "decision making." Source: Web of Knowledge, accessed November 13, 2014. We refrained from comparing with specific mathematical methods, such as "Bayesian model".

1.1. Conceptions of process models

This section discusses what experts think about process models, and also how the term is used in the literature.

1.1.1. Among experts

We surveyed scientists who work with cognitive models and asked them to indicate if they thought some prominent models were process models. One hundred and sixteen models were included in the survey and it was the result of a systematic literature review (see Appendix for details). We had 62 respondents in total, consisting of 35 professors, 16 post-doctoral researchers, and 11 doctoral students. Most had experience in teaching methodology (n = 46), and were familiar with many of the models. Professors, researchers, and students knew and classified on average 50, 49, and 40 models, respectively, indicating that the sample consisted of experts. Although a high proportion agreed that process models are important (n = 51), they did not agree on which models are process models. We analyzed this terms of inter-rater agreement, measured by Fleiss-Cuzick's kappa[2]. We found $\kappa = .27$, indicating low agreement.[3] A split by seniority yielded κ values of .33, .17, and .14 for professors, researchers, and students, respectively. Since the values were all below .60, this indicates that the low agreement was not an artifact of averaging over seniority levels.

This disagreement also suggests that the meta-theories related to process models, such as Marr's (1982) levels of analysis, have not provided a specific enough characterization of the properties of process models. Even though Marr's levels has been widely adopted (Chater, 2009; Jones & Love, 2011; McClelland et al., 2010), it has also been criticized for being difficult to apply (summarized in Griffiths, Lieder, & Goodman, 2015). Process models have been argued to be located at the algorithmic level, which was defined by Marr (1982) as specifying "the algorithm for the transformation" of the input to an output (p. 5), but not much more detail is provided. This is also reflected in our survey responses from the 38 experts familiar with Marr: When asked whether the algorithmic level clarifies what process models are, their opinions were divided between "does not clarify at all" (n = 16) and "clarifies completely" (n = 20) around a "neutral" midpoint (n = 2) on a 7-point Likert-scale.

1.1.2. In the literature

Process models are used with different connotations in the decision-making literature. Do these connotations converge towards a common understanding of the properties of process models? Though none of the authors mentioned below explicitly meant to characterize process models *in general*, their usage of the term forms part of the landscape in which the term is interpreted.

---

[2] Which is a statistic of inter-rater reliability suitable for our data (i.e., dichotomous ratings by more than two judges with an unequal number of judges per item).

[3] $\kappa = 0$ indicates random agreement; $\kappa = 1$ indicates perfect agreement; values above .60 are considered to indicate "good" agreement Fleiss & Cuzick, 1979).

Process models vs. rational models. Rational models are those that provide the utility-maximizing solutions to the statistical problems faced by organisms (Chater, 2009; Griffiths, Vul, & Sanborn, 2012; Lewis, Howes, & Singh, 2014; Sanborn, Griffiths, & Navarro, 2010). They are related to Anderson's (1991) rational analysis and they aim to model the optimal behavior by entering people's goals and capacities into a formal model of the environment.

Rational models are also sometimes contrasted with process models. For example, Lee and Cummins (2004) as well as Bergert and Nosofsky (2007) introduced their papers by comparing rational and process models. The contrast of the two models imply that process models are those that yield solutions that are suboptimal or not guaranteed to optimal, or are those that yield approximately optimal solutions within a fixed margin of error. The latter is also referred to as "rational process models" (see Griffiths et al., 2012; Sanborn et al., 2010). Accordingly, it can be interpreted that a process model is one that predicts suboptimal choices.

Process models vs. "as-if" models. As-if models include input-output transformations that are not claimed to correspond to factual phenomena in the modeled system (Berg & Gigerenzer, 2010; Glöckner & Witteman, 2010; Johnson, Schulte-Mecklenbeck, & Willemsen, 2008). They typically employ mathematical transformations that are chosen for elegance or feasibility and are deliberately free from psychological interpretations; in other words, they describe behavior, but not the processes (Brandstätter, Gigerenzer, & Hertwig, 2006). As-if models rely on Friedman's (1953) essay on positive economics which claims that models should be judged by their ability to make output predictions and not by the realism of their assumptions about the process. Thus, as-if models are often held in opposition to process models (Gigerenzer, Todd, & The ABC Research Group, 1999; Katsikopoulos & Lan, 2011). For instance, Chase, Hertwig, and Gigerenzer (1998) contrasted models that assumed unlimited computational resources with those assuming computational constraints. From this second context, it can be interpreted that process models are characterized by their feasibility: the transformation computations need to be realistic given human mental capacities, or that their parameters need a psychological interpretation (e.g., Berg & Gigerenzer, 2010; Gigerenzer & Goldstein, 1996; Gigerenzer, Hoffrage, & Goldstein, 2008; Gregg & Simon, 1967).

Rational models that yield optimal solutions are sometimes taken to be as-if models with unrealistic computation because optimization procedures are typically onerous and thus require vast mental capabilities. For example, Sanborn et al. (2010) concede that "executing optimal solutions to these problems can be extremely computationally expensive," and thus the challenge for rational modelers is to identify "psychologically plausible mechanisms that would allow the human mind to approximate optimal performance" (p. 1144). Thus, although rational and as-if models can overlap, they need not. In certain situations, especially when data is limited, models that have realistic computation assumptions can outperform as-if models that have

great computational requirements (e.g., Czerlinski, Gigerenzer, & Goldstein, 1999). In this way, as-if models are not necessarily optimal. Furthermore, whether a model can derive the optimal solution hinges on the criteria chosen (Chase et al., 1998; Einhorn & Hogarth, 1981), as well as the task environment (Pleskac & Hertwig, 2014; Todd, Gigerenzer, & The ABC Research Group, 2012). In many real-world environments characterized by uncertainty, the optimal solution is unknowable and thus any claim of optimality is merely wishful thinking (see Brighton & Gigerenzer, 2012).

Process models share formal features. Discussions of process models sometimes invoke formal aspects of modeling. Process models have been related to stochastic computations, like random walk processes (Brown & Heathcote, 2008; Busemeyer & Townsend, 1993; Pike, 1973; Ratcliff, 1978), specifically-developed symbolic languages, like Newell's (1963) Information Processing Language-V. (e.g., Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Gregg & Simon, 1967; Simon & Kotovsky, 1963), or contains continuous time (e.g., Lamberts, Brockdorff, & Heit, 2003). From this, it can be interpreted that process models are characterized by the inclusion of a set of formal properties.

1.1.3. Summary

In sum, process models are used in different areas of the literature and in distinct ways. The first suggests that the decisions predicted by process models have suboptimal performance, the second suggests that computations implemented in the model need to be feasible, while the third suggests that they need to include formal elements such as stochasticity. This brief review corroborates the earlier findings that there is no consensus among experts about the key characteristics that constitute cognitive process.

2. THE FRAMEWORK FOR PROCESS MODELS

Our framework can be applied to cognitive models before they are tested against data as well as after. Crucially, our framework refers to empirical models of human behavior that purport to describe cognitive processing. In the following, the information to be processed is referred to as the input, and the resulting output of interest e.g., behavior or decision is referred to as the output.

In brief, we proposes that in addition to the input and output, process models need to include at least one intermediate stage. The intermediate stages should be compatible with current knowledge about human cognition, and can vary separately from the output. Testable claims can be derived about both the intermediate stages and the output, given the same input. See figure 1.

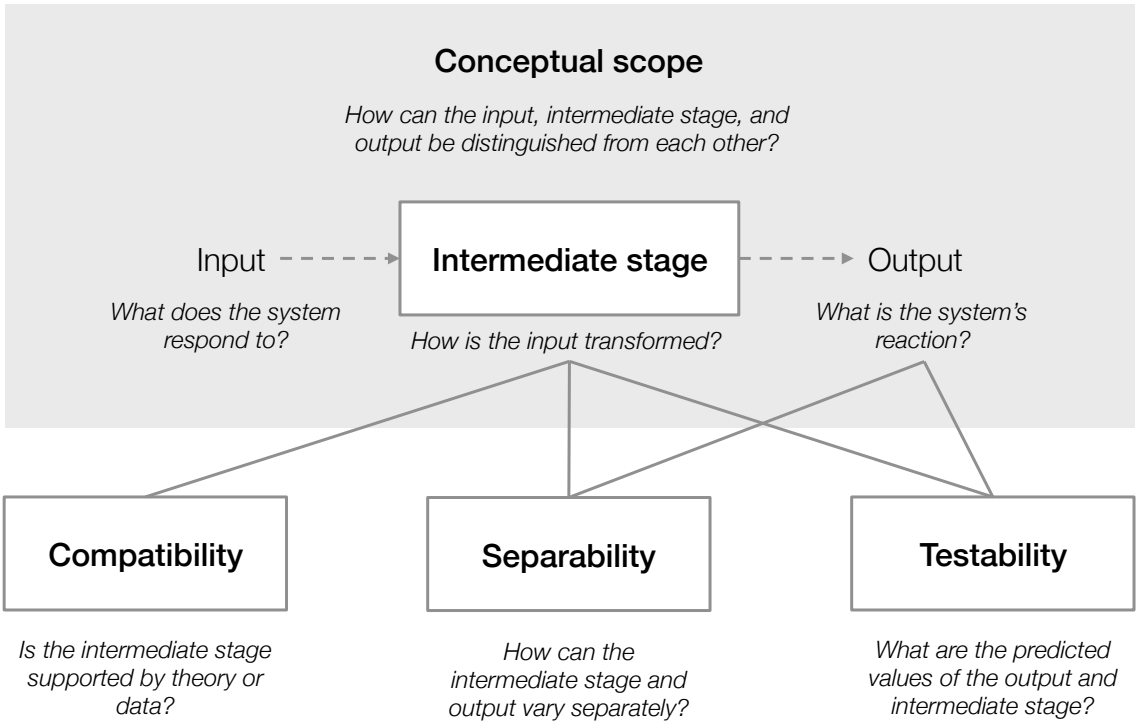

Figure 2. The framework for cognitive process models. The schema shows the requirements for a model of cognition (conceptual scope, input, and output) as well as the additional requirements for a process model (intermediate stage, compatibility, separability, and testability). The solid connecting lines denote the interrelatedness of the requirements.

2.1. Conceptual scope

A model's scope describes the phenomena to which it applies. For non process-models the scope includes only two aspects—the input and the output phenomenon. As an illustration, consider two prominent models of behavior in economic games. The fairness models are concerned with predicting individuals' monetary contributions (i.e., the output) from their social preferences and the value of relative payoffs (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999). The social preference in those models is derived from the output and is the difference between an individual's payoff and the average of everyone else's' payoff. By including only the input (i.e., the payoff structure) and output (i.e., the contribution), the model has a bi-modal scope.

In contrast, the conceptual scope of process models needs to be tri-modal, describing the phenomenon related to the input, the process, as well as the output. Crucially, the scope needs to explicitly differentiate the process from the input and output. For example, Gluth, Rieskamp, and Büchel (2012) offer a tri-modal scope when testing a sequential sampling model of decisions to acquire goods. They differentiate the input (i.e., information about the value of a good) from the process (i.e., the

updating process of evidence accumulation), as well as the output (i.e., the decision to reject or accept).

2.2. Intermediate stages

Our framework requires process models to have at least one intermediate stage specified (in line with Svenson, 1979; Weber & Johnson, 2009). An intermediate stage is the cognitive event or events that, according to the model, occurs after the input but before the output, and is responsible for transforming the input to the output. It can be latent or manifest, continuous or distinct, but must lie within the conceptual scope. Intermediate stages could be, for instance, where an individual looks, which brain region is activated, beliefs about probabilities or causal structures, or how information is compared. This specification is more precise than Marr's (1982) description of the algorithmic level as containing an information "transformation".

For example, cumulative prospect theory (Tversky & Kahneman, 1992) describes risky choice as computation of an utility $u(x, p)$ by multiplying a subjective evaluation of the output and their weight of the probability wp. The formal model does not contain a temporal order; the equation, $u(x, p) = v(x)w(p)$, leaves open whether the mind first weights probabilities or evaluates payoffs. This is an example of a model that contains a transformation but lacks an intermediate stage.

It should be noted that whether a model contains an intermediate stage is independent of whether the proposed transformation is actually carried out in the mind. The question about whether the mind does multiplication in risky choices is an empirical question, not a conceptual one. It is not the inclusion of equations in a model that precludes the existence of an intermediate stage, but the lack of temporal succession. For example, decision field theory (Busemeyer & Townsend, 1993) includes several equations in their model—for instance, a random walk that determines how the propensity to choose one option over the other develops over time—but these equations are proposed to be carried out in a sequence. Thus the model contains clear intermediate stages.

2.3. Compatibility

The information transformation in the intermediate stage of a process model should be compatible with our knowledge about cognitive capacities. Compatibility means relating the hypothesized process to a) at least one supported theory, or b) data about the capabilities of the system. This can be a theoretical argument (e.g., the computations at the intermediate stage are tractable), an empirical argument (e.g., the memory requirements do not exceed known limitations), or a reference to data (e.g., the proposed process is consistent with empirical phenomena).

Thus, compatibility can be seen as an objective specification of the criteria of "plausibility" (e.g., Winkel, Keuken, van Maanen, Wagenmakers, & Forstmann, 2014), since what is deemed plausible tends to be subjective (see Gigerenzer et al., 2008).

This requirement is objective because it can be checked by other researchers. As an illustration, we return again to Busemeyer and Townsend (1993), who linked the computations in decision field theory to findings from approach-avoidance research and choice response-time theories. Similarly, the process hypothesized in the priority heuristic of Brandstätter et al. (2006) assumes that individuals prefer a gamble if its payoff exceeds that of the other gamble by at least 10%, where the threshold of 10% is justified by reference to the culturally embedded number system (i.e., is rooted in an existing theory).

There are two reasons for the compatibility criterion. First, the latent nature of cognitive processing warrants a theoretical or empirical justification. Second, the intermediate stage or stages are under-determined by input—output relations (see Moore, 1997). If multiple possible intermediate stages lead to the same connection of inputs to outputs, one way to distinguish them is to provide a theoretical argument for the compatibility of an intermediate stage.

2.4. Separability

Separability is perhaps the most important aspect of process models. This means that the intermediate stage and the output predict two *separable* dimensions within the conceptual scope. That is, it can happen that the model correctly predicts the output values while incorrectly predicting the intermediate stage values, and vice versa. Without separability, the output and intermediate stage becomes conflated with each other; the criterion prevents equating support for an output prediction with support for hypothesized processes, the logical fallacy of affirming the consequent (Geis & Zwicky, 2011). This is less problematic for output models because the transformation is not part of the model's claim, whereas it is for process models.

Separability can also be seen as a prerequisite for model-based process tracing. A model of the form [input + *attention* → output] uses eye-tracking data as a proxy for attention and therefore it can be constituted as process data. On the other hand, a model with the form [input + *brain activity* → attention] takes eye-tracking data as a measure of the output. From this perspective, the model provides a way to identify what constitutes process data. Furthermore, separability implies that a model in which all parameters are free is an output model. If a linear regression model (e.g., Dawes & Corrigan, 1974) is used to describe choices and the weights are fitted from the choices, there is no separability. Separability implies that the process parameters of the model are not inferred from the outputs.

The reason for the separability criterion is so that models can be refined and updated after they have been tested against data. The predictions of random walk models, for instance, refer to the separate dimensions of choice distributions and reaction time distributions. Early random walk models predicted response times for correct choices (summarized in Ratcliff & Tuerlinckx, 2002), but the data typically showed faster response time for errors than for correct choices (Ratcliff & Smith, 2006;

Ratcliff, Van Zandt, & McKoon, 1999). If the model was not separable, the researchers would not been able to test response time predictions separately and would have just fitted them, and this discrepancy would have gone unnoticed (a similar point was made earlier by Gregg & Simon, 1967). Instead, the process component of earlier models was falsified by data, leading to their refinement. Thus, separability allows for the process in the model to be empirically tested instead of being an assumption.

2.5. Testability

Testability refers to a model having claims about the output and the intermediate stages that can be tested with data. While output testability is nothing new for empirical models, process models need to yield additional testable hypotheses for any intermediate stages encompassed by the scope. The psychological constructs in the intermediate stages should be specified precisely so that it could be operationalized and tested by other researchers.

2.6. Summary and comments

Taken together, a process model needs to have a clear scope and contain at least one psychologically motivated intermediate-stage that occurs after the input but before the output. The process indicated by the intermediate stage needs to be compatible with current scientific knowledge of mental capacities. The model also needs to yield separate predictions for the processes and for the behavior, allowing both to be empirically disentangled.

It should be noted that our framework does not claim that process models are better models than output models. The goal is to clarify, not evaluate; a model that fulfills our criteria may still be a "bad" one (for a discussion of what is a good model, see Myung, Pitt, & Kim, 2003). We think that the discussion about whether a model is good should be kept distinct from whether it is a process model.

The rationale for including these four requirements is that they are independent of the formal notation of models stochastic vs. deterministic, verbal vs. statistical, parallel vs. serial, etc. Secondly, these requirements are data-focused which allow process models to be validated by traditional model testing procedures (e.g., Myung et al., 2003), and allows for a process model to be a "bad" model. Thirdly, the requirements in our framework allows us to ignore matters of optimality discussed earlier, as optimality is a matter that cannot be resolved at the point of model construction.

These requirements have several noteworthy interdependencies that are illustrated in Figure 2 as solid connecting lines. First, the intermediate stage is a prerequisite for compatibility, separability, and testability. We believe that requiring at least one intermediate stage is uncontroversial. This means that models like cumulative prospect theory Kahneman & Tversky, 1979; Tversky & Kahneman, 1992, and weighted additive models do not, at least in their simple form, qualify as process models in our framework. Second, the separability element is connected to the intermediate stage

and the output. This is because the output prediction needs to be separated from the intermediate-stage prediction. Third, testability connects to the intermediate stage and the output, because a process model—as opposed to an output model—requires predictions for both the output and intermediate stage.

In addition to these dependencies, the conceptual scope acts as a constraint—models can only be tested within the specified scope. For example, to predict binary choices, the take-the-best heuristic model (e.g., Gigerenzer & Goldstein, 1996) proposes as the intermediate stage that cues are looked up in order of validity. What is outside of the scope is how the cognitive system computes the validity and thus, according to our framework, the testability of the take-the-best model does not include validity computation. For that, a different model or an extension is needed.

3. A CASE STUDY

We will illustrate the framework by referencing the work of Fischbacher et al. (2013) that proposed a generic lexicographic heuristic model of a responder's decisions in a mini-ultimatum game (henceforth, LEX model). In brief, their model is concerned with the decision to accept or reject an offer in the mini-ultimatum game, an experiment used in behavioral game theory. Participants are paired up with one being assigned the role of a proposer, and the other, the responder. The proposer is given an endowment and has to decide whether to share part of it with the responder. The responder then decides whether to accept or reject the proposer's offer. If the proposer accepts, the allocation is implemented; if rejection is chosen, both participants receive nothing. Even though game theory predicts that responders will accept any offer above zero, it has been repeatedly found that responders frequently reject low offers that are less than half the initial endowment (e.g., Henrich et al., 2001). It is this "anomaly" that is the phenomenon of interest.

The LEX model describes individual decisions in the ultimatum game in terms of three discrete steps (Fischbacher et al., 2013):

> Step 1: If a responder is allocated more than the proposer, the responder will accept immediately.
>
> Step 2: If the responder is allocated less than the proposer, the responder will accept if the offer was "kind". According to Fischbacher et al. (2013, p. 465), "An offer is said to be unkind if it is smaller than the counterfactual allocation."
>
> Step 3: If the offer is considered unkind, the responder will consider if he/she would have made the same offer if the roles were reversed. If yes, the responder will accept, if no, reject.

We will discuss whether this model is a process model using the framework proposed earlier.

Conceptual Scope. A tri-modal conceptual scope is given if the authors define not only the input and the output, but also the intermediate stage or stages. In this case, the LEX model aims to predict the responder's decision whether to accept or reject in the mini-ultimatum game (i.e., the output). The input is given as the proposer's offer. More specifically, the inputs are the three attributes of the offer and their respective binary values: Is the responder's payoff larger than the proposer? Is the proposer's offer kind? Is the proposer's offer what the responder would have made? The intermediate stage is the order of information processing which, in the LEX model, corresponds to a lexicographic consideration of the cues. The tri-modal conceptual scope is thus clear.

Intermediate Stage. A process model requires at least one intermediate stage in terms of an event between input and output. The LEX model specifies three distinct mental events, as described above: thinking about the relative magnitude of the payoff, thinking about kindness, and comparing the offer to one's own potential offer. The LEX model thus includes at least one intermediate stage.

Separability. Separability holds if, given some input values, the values predicted for the output and the intermediate stages can vary independently. In the LEX model, the data that reflects the intermediate stages is reaction time. They can vary independently of the data that reflect the output, which is the decision. Accordingly, the model might predict the decisions correctly, but not the reaction times or vice versa. Measurement separability is therefore fulfilled.

Testability. Model claims are testable if the model allows specific predictions to be made for the output and intermediate stage or stages, such that data can contradict them. The specification of the LEX model allows two predictions to be made. It predicts the responder's decision to reject or accept in the mini-ultimatum game (i.e., output prediction), and that response latency will increase as the number of cues the responder considers increases (i.e., intermediate stage prediction). Both predictions are precise and lie within the scope of the model. Therefore, testability is also fulfilled.

Compatibility. The model is compatible if the intermediate stage or stages can be supported by theory or data related to the conceptual scope. Fischbacher et al. argues that the intermediate stage is compatible because research has found that individuals use similar lexicographic strategies to make decisions (e.g., Gigerenzer & Goldstein, 1996). Therefore, a link to supported theories is fulfilled.

In sum, we conclude that Fischbacher et al.'s (2013) LEX model offers a clear conceptual scope and fulfills the requirements of a process model.

4. DISCUSSION

In the present chapter, we proposed a framework that characterizes process models *in general*. We proposed that process models are descriptive models with several interrelated properties. It needs to include at least one intermediate stage between input and output, as well as have a clear conceptual scope spanning not only the input and output, but also the intermediate stage. It also needs to lead to testable hypotheses within its scope for the output and also the intermediate stage. Furthermore, a process model needs to allow the data reflecting the output to vary independently of the data reflecting the intermediate stage. Finally, the proposed intermediate stage needs to be compatible with current knowledge about cognition within the stated conceptual scope.

While conceptual clarity about the meaning of frequently used terms is desirable in its own right, clarity also facilitates the advancement of the area of interest. Even though many arguments have been made about the advantages of process models (such as Berg & Gigerenzer, 2010; Gregg & Simon, 1967) and there have also been a corresponding interest in process modeling as illustrated by the citation trend mentioned at the start of the chapter, the field faces a shortage of good process models. We think that a key factor is this lack of clarity, and is thus where we hope that the contribution of this chapter lies. We discuss some implications of our framework below.

4.1. What constitutes process data?

It is interesting to note that despite their overlap, discussions about process models are often divorced from that of process-tracing. Even though process data is theoretically required to test a process model (Johnson et al., 2008; Schulte-mecklenbeck et al., 2011), it is unclear what counts as process data. For example, eye-movements may be process data (e.g., Lemonnier, Brémond, & Baccino, 2014; Orquin & Mueller Loose, 2013), but it may also be considered output data (e.g., Reichle, Rayner, & Pollatsek, 2003). The separability and testability criteria of our framework can help to identify process data as those that support the intermediate stage proposed. This makes it easier to bridge the divided fields of process tracing research and cognitive modeling.

4.2. What is plausible?

Even though the assertion that process models need to be “plausible” is meant as a constraint on the kinds of process models that a researcher can propose, it has instead resulted in the researcher being given a great degree of freedom. This is because what is plausible is a subjective criteria; a proposal that the mind forms a prior probability distribution and then updates it into a posterior distribution may be plausible to one researcher but not another (e.g., Jones & Love, 2011). Our framework addresses this issue in a more objective way by reframing plausibility as compatibility.

This also allows for what is considered plausible to be updated as the field's knowledge increases, and for future researchers to look back at past process models and evaluate their compatibility by referencing what was known at that time.

In addition, the compatibility criterion also fosters theory integration by encouraging cognitive models to be connected to theories or evidence about downstream cognitive processes (e.g., attention, working memory) that constitute the cognitive capabilities of the individual and constraint the space of possible transformation processes.

4.3. Advancing debates

Our framework has implications for two ongoing debates about process modeling. The first debate is normative and asks whether process models are more useful than other models. This perspective can be summed as follows: given that the mind is the object of interest, models should incorporate the real mental processes of phenomena in order to provide a genuine explanation (see Berg & Gigerenzer, 2010). Along this line, some scholars argue that process models provide more realistic models of the mind than structural models of judgment outputs (e.g., Svenson, 1979), or economic as-if models (e.g., Berg & Gigerenzer, 2010; Gigerenzer, 2010). Others claim that rational models describe the mind better than process or mechanistic models (e.g., Chater, 2009).

The second debate is about model classification. What counts as a process model of choices? For example, do connectionist network models describe processes (e.g., McClelland et al., 2010), or functions (e.g., Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010)? Is the recognition heuristic a process model (e.g., Goldstein & Gigerenzer, 2002; Pohl, 2011)? Do quantum probability models provide insights into cognitive processes (e.g., Busemeyer, Pothos, Franco, & Trueblood, 2011)? What process data can be predicted by the priority heuristic (e.g., Ayal & Hochman, 2009; Brandstätter et al., 2006; Pachur, Hertwig, Gigerenzer, & Brandstätter, 2013)?

To solve these debates, the first step is to have clarity about what process models are. Only when there is agreement about the characteristics of process models can there be compelling arguments made regarding whether they are desirable for the purposes of the researcher, and whether a model that seeks to be a process model provides the explanation that it advertises.

4.4. Conclusion

The rise in use of modeling techniques is one of the most exciting trends in cognitive science. Modeling allows cognitive processes to be specified and tested at a resolution far greater than ever before. In particular, process modeling can foster greater understanding by testing theories and integrating diverse perspectives in order to build a full picture of human functioning. If the field is to take advantage of the explanatory potential of process models, there needs to be clarity about what process

models *are*. We hope that our framework contributes by providing a common ground for discussions between researchers who share interest in process explanations but have backgrounds in different paradigms, so that better process models in the field will be built.

REFERENCES

Anderson, J. R. (1991). The place of cognitive architectures in a rational analysis. In K. van Len (Ed.), *Architectures for Intelligence*. Hillsdale, New Jersey: Erlbaum.

Ayal, S., & Hochman, G. (2009). Ignorance or integration: The cognitive processes underlying choice behavior. *Journal of Behavioral Decision Making*, *22*(4), 455–474.

Berg, N., & Gigerenzer, G. (2010). As-if behavioral economics: Neoclassical economics in disguise? *History of Economic Ideas*, *18*(1), 133–165. doi:10.2139/ssrn.1677168

Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take-the-best models of decision making. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(1), 107–29. doi:10.1037/0278-7393.33.1.107

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, *90*(1), 166–193. doi:10.1257/aer.90.1.166

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychology Review*, *113*(2), 409–432. doi:10.1037/0033-295X.113.2.409

Brighton, H., & Gigerenzer, G. (2012). Are rational actor models "rational" outside small worlds? *Evolution and Rationality: Decisions, Co-Operation and Strategic Behavior*, 84–109. doi:http://dx.doi.org/10.1017/CBO9780511792601.006

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178. doi:10.1016/j.cogpsych.2007.12.002

Busemeyer, J. R., Pothos, E. M., Franco, R., & Trueblood, J. S. (2011). A quantum theoretical explanation for probability judgment errors. *Psychological Review*, *118*(2), 193–218. doi:10.1037/a0022542

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459. doi:10.1037//0033-295X.100.3.432

Chase, V. M., Hertwig, R., & Gigerenzer, G. (1998). Visions of rationality. *Trends in Cognitive Sciences*, *2*(6), 206–214.

Chater, N. (2009). Rational and mechanistic perspectives on reinforcement learning. *Cognition*, *113*(3), 350–364. doi:10.1016/j.cognition.2008.06.014

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In *Simple Heuristics That Make Us Smart.s smart* (pp. 97–118). New York: Oxford University Press.

Dawes, R. M., & Corrigan, B. (1974). Lineal models in decision making. *Psychological Bulletin*, *81*(2), 95–106. doi:10.1037/h0037613

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, *19*(1), 1–31. doi:10.1146/annurev.ps.32.020181.000413

Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review*, *86*(5), 465–485. doi:10.1037/0033-295X.86.5.465

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 817–868.

Fischbacher, U., Hertwig, R., & Bruhin, A. (2013). How to model heterogeneity in costly punishment: Insights from responders' response times. *Journal of Behavioral Decision Making*, *26*, 462–476. doi:10.1002/bdm.1779 How

Friedman, M. (1953). The methodology of positive economics. In *Essays in Positive Economics*. Chicago: University of Chicago Press. doi:10.1017/CBO9780511581427

Geis, M. L., & Zwicky, A. M. (2011). On invited inferences. *Linguistic Inquiry*, *2*(4), 561–566.

Gigerenzer, G. (2010). Personal reflections on theory and psychology. *Theory & Psychology*, *20*(6), 733–743. doi:10.1177/0959354310378184

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. doi:10.1037/0033-295X.103.4.650

Gigerenzer, G., Hoffrage, U., & Goldstein, D. G. (2008). Fast and frugal heuristics are plausible models of cognition: Reply to Dougherty, Franco-Watkins, and Thomas (2008). *Psychological Review*, *115*(1), 230–239. doi:10.1037/0033-295X.115.1.230

Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press. doi:10.1007/s13398-014-0173-7.2

Glöckner, A., & Witteman, C. (2010). Beyond dual-process models: A categorisation of processes underlying intuitive judgement and decision making. *Thinking & Reasoning*, *16*(1), 1–25. doi:10.1080/13546780903395748

Gluth, S., Rieskamp, J., & Büchel, C. (2012). Deciding when to decide: Time-variant sequential sampling models explain the emergence of value-based decisions in the human brain. *The Journal of Neuroscience*, *32*(31), 10686–10698. doi:10.1523/JNEUROSCI.0727-12.2012

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75–90. doi:10.1037/h0092846

Gregg, L. W., & Simon, H. A. (1967). Process models and stochastic theories of simple concept formation. *Journal of Mathematical Psychology*, *4*(2), 246–276. doi:10.1016/0022-2496(67)90052-1

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364. doi:10.1016/j.tics.2010.05.004

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229. doi:10.1111/tops.12142

Griffiths, T. L., Vul, E., & Sanborn, A. N. (2012). Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science*, *21*(4), 263–268. doi:10.1177/0963721412447619

Grüne-Yanoff, T. (2014). *What are process models and what are they good for?* Unpublished manuscript.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Mcelreath, R., … Mcelreath, R. (2001). In search of homo economicus: Behavioral experiments in 15 small-scale societies. *The American Economic Review*, *91*(2), 73–78.

Johnson, E. J., Schulte-Mecklenbeck, M., & Willemsen, M. C. (2008). Process models deserve process data: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review*, *115*(1), 263–273. doi:10.1037/0033-295X.115.1.263

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *The Behavioral and Brain Sciences*, *34*(4), 169–188. doi:10.1017/S0140525X10003134

Katsikopoulos, K. V., & Lan, C.-H. (2011). Herbert Simon's spell on judgment and decision making. *Judgment and Decision Making*, *6*(8), 722–732.

Lamberts, K., Brockdorff, N., & Heit, E. (2003). Feature-sampling and random-walk models of individual-stimulus recognition. *Journal of Experimental Psychology - General*, *132*(3), 351–378. doi:10.1037/0096-3445.132.3.351

Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: unifying the "take the best" and the "rational" models. *Psychonomic Bulletin & Review*, *11*(2), 343–352. doi:10.3758/BF03196581

Lemonnier, S., Brémond, R., & Baccino, T. (2014). Discriminating cognitive processes with eye movements in a decision-making driving task. *Journal of Eye Movement Research*, *7*(4), 1–14.

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice*. Thousand Oaks, CA: SAGE Publications.

Lewis, R. L., Howes, A., & Singh, S. (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science*, *6*, 279–311. doi:10.1111/tops.12086

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: W. H. Freeman and Company.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*(8), 348–356. doi:10.1016/j.tics.2010.06.002

Moore, A. W. (1997). The undetdetermination/indeterminacy distinction and the analytic/synthetic distinction. *Erkenntnis*, *46*(1), 5–32. doi:10.1023/A:1005382611551

Myung, J. I., Pitt, M. A., & Kim, W. (2003). Model evaluation, testing and selection. In K. Lamberts & R. Goldstone (Eds.), *Handbook of Cognition* (Vol. 1862, pp. 1–45). SAGE Publications Ltd.

Newell, A. (1963). Documentation of IPL-V. *Communications of the ACM*, *6*(3), 86–89. doi:10.1145/366274.366296

Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190–206. doi:10.1016/j.actpsy.2013.06.003

Pachur, T., Hertwig, R., Gigerenzer, G., & Brandstätter, E. (2013). Testing process predictions of models of risky choice: A quantitative model comparison approach. *Frontiers in Psychology*, *4*, 1–22. doi:10.3389/fpsyg.2013.00646

Pike, R. (1973). Response latency models for signal detection. *Psychological Review*, *80*(1), 53–68. doi:10.1037/h0033871

Pleskac, T. J., & Hertwig, R. (2014). Ecologically Rational Choice and the Structure of the Environment. *Journal of Experimental Psychology: General*, *143*(5), 2000–2019. doi:10.1037/h0042769

Pohl, R. (2011). On the use of recognition in inferential decision making: An overview of the debate. *Judgment and Decision Making*, *6*(5), 423–438.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. doi:10.1037/0033-295X.85.2.59

Ratcliff, R., & Smith, P. L. (2006). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–367. doi:10.1037/0033-295X.111.2.333

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*(3), 438–481. doi:10.3758/BF03196302

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, *106*(2), 261–300. doi:10.1037/0033-295X.106.2.261

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*, 445–476. doi:10.1017/S0140525X03000104

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*(4), 1144–1167. doi:10.1037/a0020511

Schulte-mecklenbeck, M., Kühlberger, A., & Ranyard, R. (2011). The role of process data in the development and testing of process models of judgment and decision making, *6*(8), 733–739.

Simon, H. A., & Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review*, *70*(6), 534–546.

Svenson, O. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance*, *23*(1), 86–112. doi:10.1016/0030-5073(79)90048-5

Todd, P. M., Gigerenzer, G., & The ABC Research Group. (2012). *Ecological rationality: Intelligence in the world*. New York: Oxford University Press.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of

uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. doi:10.1007/BF00122574

Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology*, *60*, 53–85. doi:10.1146/annurev.psych.60.110707.163633

Winkel, J., Keuken, M. C., van Maanen, L., Wagenmakers, E.-J., & Forstmann, B. U. (2014). Early evidence affects later decisions: Why evidence accumulation is required to explain response time data. *Psychonomic Bulletin & Review*, *21*(3), 777–784. doi:10.3758/s13423-013-0551-8

Appendix: Model review process

The list of 116 models used in our survey was derived from a systematic literature search. We first identified relevant articles that tested or developed models, and then extracted the models tested in these articles, and selected the most prominent judgment and decision-making models.

Step 1: Identification of Articles. Specifically, we searched for important or new articles proposing a model "important" meant articles that had been cited more than 100 times, "new" meant ones that were published in 2004 or later, respectively in two databases: Google Scholar and ISI Web of Science. We combined the fixed search term "model of" with synonyms for "decision making." The precise search phrase reads "model of * decision" | "model of * decisions" | model of * choice" | "model of * choices" | "model of * preference" | "model of * preferences" | "model of * inference" | "model of * inferences" where | denotes the Boolean OR and * can be any word.

To ensure source relevance, we restricted our search to Judgment and Decision Making Journal, Psychological Review, Journal of Experimental Psychology JEP: General; JEP Learning, Memory, and Cognition, and JEP Human Perception and Performance. This yielded 433 results. From these, we first selected all articles testing cognitive models and excluded, for example, articles proposing measurement scales or mentioning but not testing models. Of these, the first author JBJ selected articles on judgment and decision-making, excluding, for example, articles purely about perception. As this step was somewhat open to interpretation, we randomly selected a subset of 50 articles and checked the reliability of the categorization to fields by cross-coding them coders were JBJ, JHT; this analysis showed a high level of agreement Cohen's kappa = .831. We therefore kept the selection by JBJ.

This procedure identified articles from the judgment and decision-making literature that dealt with models. These articles served as the sources for the models.

Step 2: Model Extraction from the Articles. We extracted the names of all models tested in the articles and looked up the original sources determined by their earliest occurrence in a peer-reviewed journal or book. Importantly, model selection ignored labels such as "process model" or "computational model."

This resulted in a total of 172 individual models. To ensure that the models were still relevant, we included models that had been cited more than 388 times in total or more than an average of 6.6 times per year 388 is the 66th percentile cutoff of all citations, and 6.6 is the 33rd percentile cutoff of average citations per year. We were left with 116 models, listed below.

Models included in the survey

1. ACT-IF, Adaptive Control of Thought in Information Foraging (Pirolli & Card, 1999)
2. Additive Difference Model (Morrison 1962; Tversky, 1969)
3. Additive Trade-off Model between Informativeness and Accuracy (Yaniv & Foster, 1995)
4. Additive-utility Model of Delay Discounting (Killeen, 2009)
5. ALM, Associative Learning Model (DeLosh, Busemeyer, McDaniel, 1997; Busemeyer, Byun, DeLosh, McDaniel, 1997)
6. Anchoring and Adjustment Model (Kahneman & Tversky, 1982; Einhorn & Hogarth, 1985)
7. ASCM, Adaptive Strategy Choice Model of algebraic strategies (Siegler & Shipley, 1995)
8. Associative Accumulation Model (Bhatia, 2013)
9. Attractor Model of Visual Discrimination (Wang, 2002)
10. Availability Heuristic (Tversky & Kahneman, 1973)
11. Beta Delta Preference Model of Temporal Discounting (Laibson, 1997)
12. Biased Encoding Model associative storage network model of memory-based judgment (Hastie 1980)
13. Brunswik's Lens Model (Brunswik, 1956)
14. BSR, Bayesian Sequential Risk Taking Model (Wallsten, Pleskac, & Lejuez, 2005; Pleskac, 2008)
15. Causal Bayes Nets (Spirtes, Glymour, & Scheines, 1993; Pearl, 2000)
16. Complement Model of Charitable-giving (Bernheim, 1994)
17. Conditional Probability Model (Oaksford, Chater, & Larkin, 2000)
18. Constructed-Choice Model (Krantz & Kunreuther, 2007)
19. Constructionist Theory of Inference Generation (Graesser, Singer, & Trabasso, 1994)
20. CPT, Cumulative Prospect Theory (Tversky & Kahneman, 1992)
21. Delta-Rule Model (Rescorla & Wagner, 1972)
22. Denrell's Experience Sampling Model (Denrell, 2005)
23. Dimensional Overlap Model (Kornblum, Hasbroucq, & Osman, 1990)
24. Dimensional Weight Model (Birnbaum & Stegner, 1979; Tversky, Sattath, & Slovic, 1988)
25. Discrete-Slot Model of Working Memory (Zhang & Luck, 2008, 2009)
26. DM, Diffusion Model (Ratcliff, 1978)
27. Dual Process Model of Deductive Inference (Verschueren, Schaeken, & d'Ydewalle, 2005)
28. EBA, Elimination by Aspects (Tversky, 1972)
29. EBM, Frequency-sensitive Exemplar Model (Nosofsky, 1988)
30. EBRW, Exemplar-based Random Walk Model (Nosofksy & Palmeri, 1997)
31. EGCM-RT, Extended Generalized Context Model for Reaction Times (Lamberts, 2000)
32. EGCM, Extended Generalized Context Model (Lamberts, 1998)
33. EW, Equal Weighting Model (Dawes, 1979)
34. Exemplar Model (Medin & Schaffer, 1978)
35. Exemplar-Based Network Model (Nosofsky, Kruschke, McKinley, 1992)
36. Exponential Strategy Selection Model (Rieskamp & Otto, 2006)
37. Extension of the Leaky, Competing Accumulator Model (Usher & McClelland, 2004)
38. FA Model, Fractional Adjustment Model (Weber, Shafir, & Blais, 2004)
39. Feedforward inhibition model (Shadlen & Newsome , 2001)
40. Forgetting Strategy Selection Model (Rieskamp & Otto, 2006)
41. FSDT, Fuzzy Signal Detection Theory (Hancock, Masalonis, & Parasuraman, 2000)
42. GCM, Generalized Context Model (Nosofsky, 1986)
43. General Linear Classifier (Medin & Schwanenflugel, 1981; Ashby & Gott, 1988)
44. GQC, General Quadratic Classifier (Ashby & Gott, 1988; Ashby 1992)
45. gRAT, Generalized Version of a Rational Model/WADD (Nosofsky & Bergert, 2007)
46. gTTB, Generalized Version of Take-the-Best (Nosofsky & Bergert, 2007)
47. Herrnstein's Matching Law (Herrnstein, 1961)
48. HyGene (Thomas, Dougherty, Sprenger, & Harbison, 2008)
49. Hyperbolic Discounting Model (Elster, 1979)
50. Imagination Strategy Selection Model (Rieskamp & Otto, 2006)
51. Incongruity-biased Encoding Model (Hastie & Kumar 1979; Hastie, 1980, 1984; Scrull, 1981)
52. Increasing Probability Model (Wallsten, Pleskac, Lejuez, 2005)
53. Independence Model of Memory and Judgment (Anderson & Hubert, 1963; Anderson, 1981)
54. Independent Race Model (Logan & Cowan, 1984)
55. Integrated System Model of Attention and Decision Making (Smith & Ratcliff, 2009)
56. LBA, Linear Ballistic Accumulator Model (Brown & Heathcote, 2008)
57. LCA, Leaky, Competing Accumulator Model (Usher & McClelland, 2001)
58. Leaky Accumulator Model with Relative Criteria (Ratcliff & Smith, 2004)
59. Leaky Accumulator Model (Ratcliff & Smith, 2004)
60. Lexicographic Semiorder (Luce, 1956; Tversky, 1969)
61. Linear Decision-boundary Model (Ashby & Townsend, 1986)
62. Linear Regression
63. LISA, Learning and Inference with Schemas and Analogies (Hummel & Holyoak, 1996; 1997)
64. Least Mean Square Network Model/Configural Cue Adaptive Network Model (Gluck & Bower, 1988)
65. Matching Heuristic (Dhami & Ayton, 2001)
66. MDFT, Multialternative Decision Field Theory (Roe, Busemeyer, & Townsend, 2001)
67. MIN, Minimalist (Gigerenzer & Goldstein, 1966; Gigerenzer et. al, 1999)

68. Minimum-distance Classifier (Ashby & Townsend, 1986)
69. Mixture Model of Transitive Preferences (Regenwetter, Dana, & Davis-Stober, 2011)
70. MMN, Max-minus-next Diffusion Model (Ratcliff & McKoon, 1997; McMillen & Holmes, 2006)
71. Mutual Inhibition Model (Usher & McClelland, 2001)
72. Naive Bayes Classifier (Czerlinski, Gigerenzer, & Goldstein, 1999)
73. Nonstationary Process Increasing Probability Model (Wallsten, Pleskac, Lejuez, 2005)
74. OU, Ornstein-Uhlenbeck Diffusion Model (Busemeyer & Townsend, 1993)
75. PCS, Parallel Constraint Satisfaction Model for Probabilistic Decision Tasks (Glöckner & Betsch, 2008)
76. PH, Priority Heuristic (Brandstätter, Gigerenzer, & Hertwig, 2006)
77. Preference for Sequences Model (Loewenstein & Prelec, 1993)
78. Present-value Comparison Model (Ainsly, 1992)
79. Pretree, Preference Tree (Tversky & Sattath, 1979)
80. Priority Model (Rieskamp, 2008)
81. Probabilistic Contrast Model (Cheng & Novick, 1990)
82. Prototype Model (Reed; 1972)
83. PT, Prospect Theory (Kahneman & Tversky, 1979)
84. Quantum Judgment Model (Busemeyer, Pothos, Franco, & Trueblood, 2011)
85. r-Model (Hilbig, Erdfelder, & Pohl, 2010)
86. RAM, Rank Affected Multiplicative Model (Birnbaum & Stegner, 1979; Birnbaum & McIntosh, 1996)
87. Random Walk Model (Stone, 1960; Laming, 1968; Link & Heath, 1975)
88. Recruitment Model (LaBerge, 1992)
89. RELAC, Reinforcment Learning of Cognitive Strategies (Erev & Barron, 2005)
90. RH, Recognition Heuristic (Goldstein & Gigerenzer, 1999, 2002)
91. Rule Competition Model (Busemeyer, & Myung, 1992)
92. Rule-based Categorization Models (Nosofsky, Clark, & Shin, 1989)
93. RULEX, Rule-plus-exception Model (Nosofsky, Palmeri, & McKinley, 1994)
94. SAMBA, Selective Attention, Mapping, and Ballistic Accumulation Model (Brown, Marley, Donkin, & Heathcote, 2008)
95. SDT, Signal Detection Theory (Tanner & Swets, 1954; Swets, Tanner, & Birdsall, 1961)
96. SEMAUT, Subjective Expected Multi-Attribute Utility Model (Savage, 1954)
97. SS Power Model (Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008)
98. SSL, Strategy Selection Learning Theory (Rieskamp & Otto, 2006)
99. Stationary Process Model (Wallsten, Pleskac, Lejuez, 2005)
100. Story Model of juror decision making (Pennington & Hastie, 1986, 1988)
101. Structural Equation Model
102. Subjective Expected Utility Model (von Neumann & Morgenstern, 1947)
103. SUSTAIN, Supervised and Unsupervised Stratified Adaptive Incremental Network (Love & Medin, 1998; Love, Medin & Gureckis, 2004)
104. SVM, Sequential Value Matching (Johnson & Busemeyer, 2005)
105. Target Model (Wallsten, Pleskac, & Lejuez, 2005)
106. TAX, Transfer of Attention Exchange Model (Birnbaum & Stegner, 1979)
107. Three-stage model (Hasbroucq & Guiard, 1991)
108. Tradeoff Model of Intertemporal Choice (Scholten & Read, 2010)
109. TTB, Take-the-best (Gigerenzer & Goldstein, 1996)
110. UCIP, Unlimited Capacity Independent Parallel Processing Model (Townsend & Wenger, 2004)
111. Utility Functions
112. WADD, Weighted Additive Model (Payne, Bettmann, & Johnson, 1993; Keeney & Raiffa, 1976)
113. Warm Glow Model of Charitable-giving (Andreoni, 1990)
114. Weighting Model
115. Wiener Diffusion Model (Stone, 1960; Laming, 1968; Link & Heath, 1975)
116. Wyer and Srull's Storage Bin Model (Wyer & Srull, 1986)

Chapter 2.

Error management in forgiveness: A process modeling approach[4]

Abstract. Whether to forgive is a key decision supporting cooperation. Like many other evolutionarily recurrent decisions, it is made under uncertainty and requires the trade-off of costs and benefits. This decision can be understood as an error management task: Forgiving is adaptive if a relationship with the "harmdoer" will be fitness enhancing and not adaptive if a relationship with the harmdoer will be fitness reducing. The decision is biased toward lowering the likelihood of the more costly error; depending on the context, either erroneously not forgiving or forgiving may be more costly. Building on this, two cognitive models of the forgiveness decision were developed. We examined how well these models described participants' forgiveness decisions in hypothetical scenarios and predicted their decisions in recalled real-life incidents and found that the models performed similarly and generally well—around 80% in describing and 70% in prediction. Moreover, this modeling approach allowed us to estimate the decision bias of each participant, and we found that in general it was biased as prescribed by error management theory. In addition to testing mechanistic models of the decision, which has been largely absent in research on error management theory, our study also contributes to forgiveness research by applying a novel experimental method that investigates both hypothetical and real-life decisions. These models and experimental methods could be used to study other evolutionarily recurrent problems, advancing understanding of how they are solved.

Keywords. cooperation, forgiveness, fast-and-frugal trees, franklin's rule, error management theory, signal detection theory, process models

1. INTRODUCTION

"Forgiveness is the bridesmaid; cooperation is the bride."

Michael McCullough, *Beyond Revenge: The Evolution of the Forgiveness Instinct*

Recurrent cooperative relationships are widespread in humans and other social animals (Dugatkin, 2002). Because such relationships are often threatened by harm arising from conflicts of interests, communication errors, or mere random

[4] This chapter is a slightly different version of my journal article titled, "A signal-detection approach to modeling forgiveness decisions", published in *Evolution and Human Behavior*. The original article is accessible at http://dx.doi.org/10.1016/j.evolhumbehav.2016.06.004

noise, choosing an appropriate action in the aftermath of harm is a crucial evolutionarily recurrent problem that social animals would have evolved mechanisms to solve (Aureli, Cords, & van Schaik, 2002). Nevertheless, harm can, but need not, result in revenge and the termination of cooperation; agents may instead choose to forgive the "harmdoer" and to continue the relationship (McCullough, Kurzban, & Tabak, 2013). Because forgiveness can be fitness enhancing by maintaining cooperation over time (Godfray, 1992), deciding whether to forgive is a key decision of cooperation.

Understanding cooperation among nonkin has received significant research attention in the last few decades. The first part of this endeavor has been to explain why an agent performs costly actions to benefit another. Inspired by work on reciprocal altruism (Trivers, 1971) and game theoretical insights (Axelrod, Hamilton, Series, & Mar, 2008; Boyd & Richerson, 1992), researchers have made major advances in understanding how cheaters are curbed so that cooperation can be beneficial (Kurzban, Burton-Chellew, & West, 2015). The second part has been to clarify how agents cooperate, which has been referred to as creating "high-resolution maps" of the intricate proximate phenotypic processes (Cosmides & Tooby, 1992). Unlike understanding why, process understanding is still nascent. It is an open question what computational rules are used to process information in decisions about forgiveness and cooperation.

Like most evolutionarily recurrent tasks, decisions about forgiveness are made under uncertainty and feature asymmetric costs and benefits (McCullough et al., 2013). One way of conceptualizing how this asymmetry may shape the decision process[5] is through the lenses of signal detection (Green & Swets, 1966) and error management theory (EMT; Haselton & Buss 2000). Because errors are inevitable in uncertain environments and have different costs, EMT posits that adapted systems of cognition are biased to guide behavior toward incurring the less costly error (Haselton & Nettle, 2005). Thus, biases are design features rather than defects and are calibrated by the relative effects of errors on fitness. In this light, decisions about forgiveness can be investigated as error management tasks and the decision process can be expected to resemble that of similar tasks. This perspective allows us to hypothesize about the characteristics of the decision process and the contexts in which biases toward or against forgiveness would occur. For example, given the same harm situation, we would expect agents to be biased toward forgiving those with whom they have fitness interdependencies and biased against

[5] There has been debate about whether bias occurs on the perceptual (cognitive) or decision (behavioral) level (e.g., Marshall, Trimmer, Houston, & McNamara, 2013; McKay & Efferson, 2010). Nevertheless, recent experimental work on the prototypical example of error management, that is, gender differences in the perception of sexual interest, has identified bias as being on the level of decision (Perilloux & Kurzban, 2015). While the debate is far from resolved, our study focuses on testing how bias is expressed at the decision level.

forgiving others with whom they have unrewarding relations (McCullough et al., 2013).

The focus of our study is on forgiveness decisions. We specify the structure of the task and its possible cognitive solutions. We investigate how forgiveness decisions can be described and predicted by two models: a compensatory weighting-and-adding linear model and a noncompensatory fast-and-frugal heuristic. Both models incorporate the essentials of EMT but make different assumptions about cognitive implementation. The parameters estimated in these models allow us to test predictions regarding the impact of error costs on the direction and magnitude of bias. Beyond forgiveness decisions, this modeling approach can potentially be used to understand how agents solve other evolutionarily recurrent problems under uncertainty (e.g., Johnson, Blumstein, Fowler, & Haselton 2013).

1.1. Forgiveness as error management

1.1.1. The framework

Forgiveness functions to maintain relationships after conflict and enable continued cooperation between the victim and the harmdoer (Burnette, McCullough, Van Tongeren, & Davis, 2012; McCullough, Fincham, & Tsang, 2003). From this perspective, forgiving is adaptive if the harmdoer is an ally but not a foe. We use the term ally to refer to an agent with whom a relationship will result in more fitness benefits than costs, and foe as one with whom a relationship will result in more costs than benefits (McCullough et al., 2013). Table 1 displays the four possible outcomes of forgiveness decisions: Correct decisions are when an ally is forgiven (*true positive*) and a foe is not (*true negative*) and incorrect decisions are when a foe is forgiven (*false positive*) and an ally is not (*false negative*). With a true positive, the victim gains the net benefits from the relationship with the ally, and with a false negative, the victim misses out on those benefits. On the flip side, with a false positive, the victim faces the net exploitation costs of the relationship with the foe, and with a true negative, the victim is spared those costs.

Informed by EMT and signal detection theory, we assume that there are two subprocesses involved in the decision of whether to forgive: judging the strength of evidence that the harmdoer is an ally and setting an appropriate bias, or decision criterion. Forgiveness is chosen when the evidence strength exceeds the decision criterion (Figure 1). Setting a liberal criterion means forgiving even when the evidence is weak, indicating a bias toward forgiving, whereas a conservative criterion means forgiving only when the evidence is strong, indicating a bias against forgiving.

Table 1.
Possible Outcomes of Forgiveness Decisions

Decision	Nature of the harmdoer	
	Ally	Foe
Forgive	True positive	False positive
Do not forgive	False negative	True negative

Note. An ally denotes an agent with whom a relationship will bring more fitness benefits than costs, whereas a foe is the reverse.

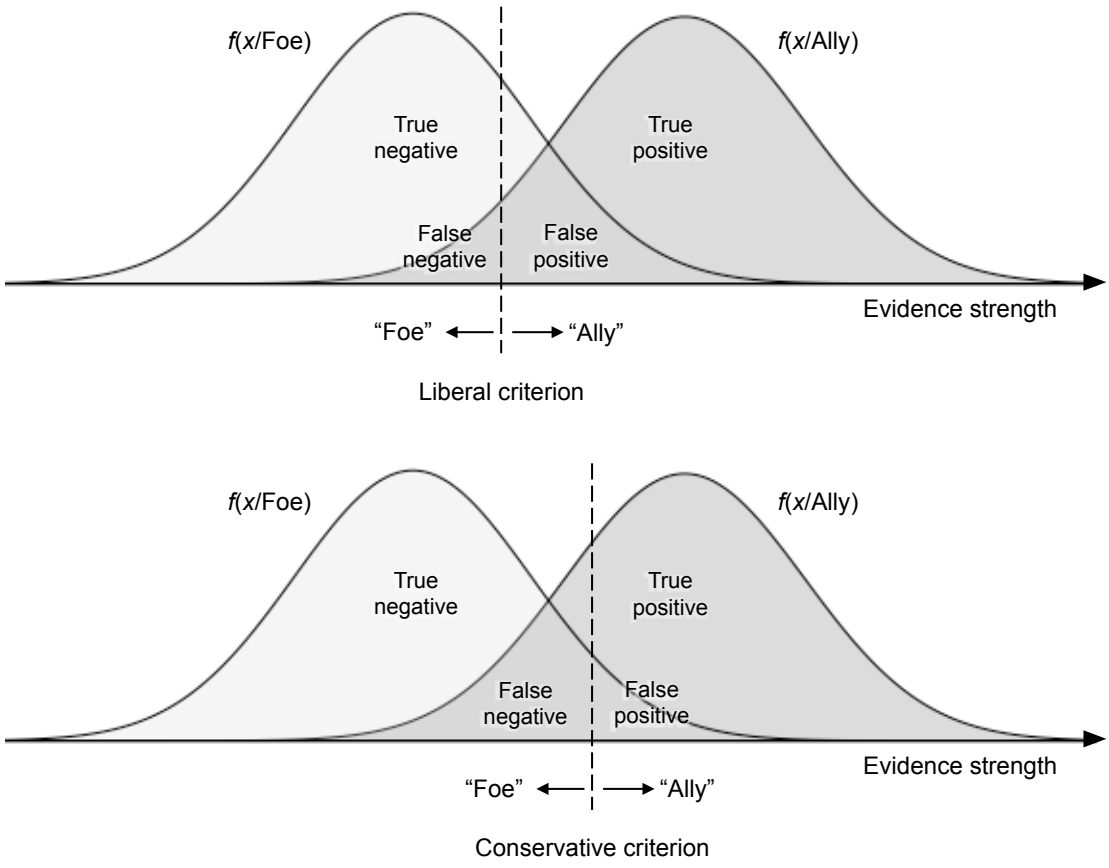

Figure 1. The assumed frequency distributions of allies, *f*(x/Ally), and foes, *f*(x/Foe), on the continuum of evidence strength. Given that *f*(x/Ally) and *f*(x/Foe) are fixed, a liberal criterion (top) reduces the probability of false negatives at the expense of a higher probability of false positives, and a conservative criterion (below) has the opposite effect.

When an agent has a high prosocial concern for the other's welfare relative to its own, the agent is more likely to make sacrifices and provide fitness benefits to the other (e.g., Struthers, Eaton, Santelli, Uchiyama, & Shirvani, 2008; Tooby & Cosmides, 2008). Thus, the greater the harmdoer's inferred prosocial concern for the victim, the stronger the evidence that the harmdoer is an ally. Nevertheless, strong evidence is no guarantee that the harmdoer is an ally. This is because the evidence is inferred from current observations and is imperfectly linked to the future. Furthermore, the evidence is likely to be perceived with some noise. Thus, there is inherent uncertainty in the decision, which is illustrated in Figure 1 by the overlapping frequency distributions of allies and foes.

Given this uncertainty, where should the decision criterion be set? In other words, how strong must the evidence strength be for the harmdoer to be forgiven? The selection of the criterion reflects a trade-off: Assuming that the two distributions are fixed, a liberal criterion reduces the likelihood of a false negative (i.e., not forgiving an ally) at the expense of increasing that of a false positive (i.e., forgiving a foe), and a conservative criterion has the opposite effect. To lower the total cost of errors, a liberal criterion should be adopted when false negatives are costlier, and a conservative one when false positives are costlier.

1.1.2. Predictors of forgiveness

Since forgiveness decisions are made in a wide variety of contexts that vary in cost–benefit asymmetry, decision makers need to judge the evidence strength and cost of errors from information in the environment. In this section we review some predictors examined in the present study (summarized in Table 2).

Table 2. Predictors of Forgiveness Examined in the Present Study

Subprocess	Predictor	Description
Judging evidence strength	Intent to harm	Harmdoer had the goal of reducing the offended's fitness.
	Blame for harm	Harmdoer caused or could have prevented the offended's fitness loss.
	Sincere apology	Harmdoer's reparative gesture communicated remorse and repentance.
Selecting decision criterion	Relationship value	Potential fitness gains from resuming interaction with harmdoer.
	Exploitation risk	Potential fitness costs of resuming interaction with harmdoer.

Evidence strength. Judging the harmdoer's prosocial concern or the strength of the evidence that the harmdoer is an ally requires insight into the mental state of the harmdoer. To this end, the victim may consider the harmdoer's intent to harm, whether the harmdoer was to blame for the harm, and whether a sincere apology was offered. These three cues or predictors were taken from a meta-analysis of forgiveness involving 175 studies (Fehr, Gelfand, & Nag, 2010). They were the variables with the strongest main effect on forgiveness within the category of variables related to making sense of the harm and the harmdoer. Of the three, intent had the strongest effect on forgiveness, followed by apology and blame. In addition to being well studied, the level of abstraction of these cues makes them relevant across a wide range of forgiveness contexts.

With an intent to harm, the harmdoer is inferred to have the goal of reducing the victim's fitness, or at the very least, to be indifferent to the impact the action would have on the victim's welfare (Malle & Knobe, 1997; Struthers et al., 2008; Weiner, 1995). Intention to harm is thus a strong cue that the harmdoer is likely to repeat the harm and that the strength of the evidence that the harmdoer is an ally is low (Petersen, Sell, Tooby, & Cosmides, 2012).

The concept of blame is closely related to attributions of responsibility and accountability (Weiner, 1995). Blame is assigned when a harmdoer's actions directly led to the harm done or when the harmdoer could have prevented the harm (Alicke, 2000). Blame is generally less indicative than intent, because a harmdoer who caused the harm could have done so accidentally rather than out of malice. Nevertheless, blame indicates a propensity to harm and thus is a cue that weakens the evidence strength.

An apology is a reparative gesture offered by the harmdoer, and when sincere, it communicates remorse and repentance (Dhami, 2012; Schlenker & Darby, 1981). It is an attempt by the harmdoer to be seen as benevolent, worthy of forgiveness, and it is in general an effective technique for promoting relationship repair (Fehr & Gelfand, 2010). A sincere apology can also be seen as a promise by the harmdoer to increase future prosocial concern (Sell, 2011), as well as an indication that the harmdoer highly values the relationship with the victim (Ohtsubo & Yagi, 2015). In general, a sincere apology is a cue that strengthens the evidence that the harmdoer is an ally.

Decision criterion. When assessing the cost of a false negative, the victim may consider the perceived relationship value (RV) of the harmdoer—that is, how beneficial a future relationship with the harmdoer will be (Burnette et al., 2012)—and the greater the RV, the greater the cost of not forgiving if the harmdoer is an ally. Additionally, the cost of a false positive can be informed by the perceived exploitation risk (ER) of the harmdoer—that is, how much harm the harmdoer may

cause in future interactions—and the greater a harmdoer's ER, the greater the cost of incorrectly forgiving.

In sum, when a harmdoer is perceived to have high RV and low ER ($H_{RV}\_L_{ER}$), the decision criterion should be liberal to reduce the likelihood of a false negative; in contrast, when the RV is low and the ER is high ($L_{RV}\_H_{ER}$), the decision criterion should be conservative to lower the likelihood of a false positive. Indeed, it has been found that individuals who perceived their harmdoers as "$H_{RV}\_L_{ER}$" were more likely to forgive than those who perceived their harmdoers as "$L_{RV}\_H_{ER}$," even after controlling for evidence strength variables, such as the intention to harm (McCullough, Luna, Berry, Tabak, & Bono, 2010).

1.2. Modeling forgiveness decisions

How are the many predictors integrated into a decision about whether to forgive? We propose two models that make different assumptions about how the mind estimates the evidence strength and implements the criterion: Franklin's rule (FR), a compensatory weighting-and-adding linear model, and fast-and-frugal trees (FFTs), a noncompensatory simple heuristic.

1.2.1. A linear model: Franklin's rule

Linear models have been the archetypical models of human judgment and decision making (Gigerenzer & Murray, 1987; Hammond, 1996; Harries, Evans, Dennis, & Dean, 1996). They assume that cues are weighted by their importance and summed up to form a continuous judgment of the evidence. This judgment is then compared to a criterion to arrive at a decision. Due to this procedure, linear models are compensatory: An undesirable value in one cue (e.g., an intent to harm) can be compensated for by a desirable value in another (e.g., a sincere apology), so that a positive decision (e.g., "forgive") may still be made.

FR is a linear model that deals with discrete-valued cues (Dhami, 2003; Goldstein & Gigerenzer, 2002) and assumes that cues are integrated with the following formula:

$$E = \sum_{m=1}^{M} \frac{w_m x_m}{\sum w_m} \quad (1)$$

where E is the overall evidence strength, M is the total number of cues, w_m is the positive-valued weight of each cue, between 0 and 1, and x_m is a binary value of a cue: It is "1" if the cue is positively related to evidence strength and is present (e.g., there is a sincere apology) or if the cue is negatively related to evidence strength and is absent (e.g., there is no intent to harm); and it is "0" if a positive cue is absent (e.g., there is no sincere apology) or if a negative cue is present (e.g., there is an intent to harm). After forming the evidence strength E, which is a number between 0

and 1 (inclusive), a decision criterion x_{DC} is applied to make the decision; that is, forgive if $E > x_{DC}$, and not forgive if otherwise.

Linear models, such as FR, have been supported as valid descriptive models for various tasks of judgment and decision making (e.g., Anderson, 1971; Brehmer, 1994; Hammond, 1996). However, such models make cognitive demands decision makers that can be taxing (Payne, Bettman, & Johnson, 1993). Alternatively, decision makers may use heuristics that search cues sequentially and lead to decisions without considering all information.

1.2.2. A heuristic: Fast and frugal trees

Heuristics are simple decision strategies that can be implemented with little computation and information (Gigerenzer, Hertwig, & Pachur, 2011; Gigerenzer, Todd, & The ABC Research Group, 1999). They are also referred to as rules of thumb and have been proposed to underlie much of human and animal decision making (Hutchinson & Gigerenzer, 2005). FFTs are a particularly suitable heuristic for forgiveness decisions. This heuristic is lexicographic and assumes that relevant cues are looked up in order of importance and a decision is made once a cue value is in favor of one option (Martignon, Katsikopoulos, & Woike, 2008).

FFTs are a special class of decision trees that have m + 1 decision exits, with one exit for each of the first m – 1 cues and two exits for the last cue (Luan, Schooler, & Gigerenzer, 2011). The exit in an FFT indicates the decision option (e.g., “forgive” or “do not forgive”), and an exit occurs when the condition set on a cue is met (e.g., if there is no intent to harm, then forgive). In contrast to linear models, FFTs are noncompensatory models of decision making: Desirable values on cues lower in the search hierarchy cannot overturn the decision following an undesirable value on the cue that is searched earlier. We use the FFT shown in the top panel of Figure 2 as an illustration. In this FFT, the cues are examined in the order of intent, blame, and apology, and a decision is made as follows:

Step 1: Did the harmdoer intend to harm? If no, forgive; if yes, next step;

Step 2: Was the harmdoer to be blamed? If yes, do not forgive; if no, next step;

Step 3: Did the harmdoer apologize sincerely? If yes, forgive; if no, do not forgive.

With this FFT, all harmdoers who had no intent to harm will be forgiven, regardless of the values of the other two cues.

The exit structures of FFTs—that is, the decision exits associated with the top m – 1 cues—correspond to different decision criteria (Luan et al., 2011). For example, with three cues, there are $2^3 - 1 = 4$ possible exit structures, shown in the lower panel in Figure 2. The trees are listed from the most liberal on the left to the most conservative on the right. Within the context of forgiveness, we refer to

them as the loving tree, less-loving tree, less-spiteful tree, and spiteful tree, respectively. Unlike the decision criterion in FR, which can theoretically take any value within the range, the number of exit structures limits the number of possible criteria in FFTs.

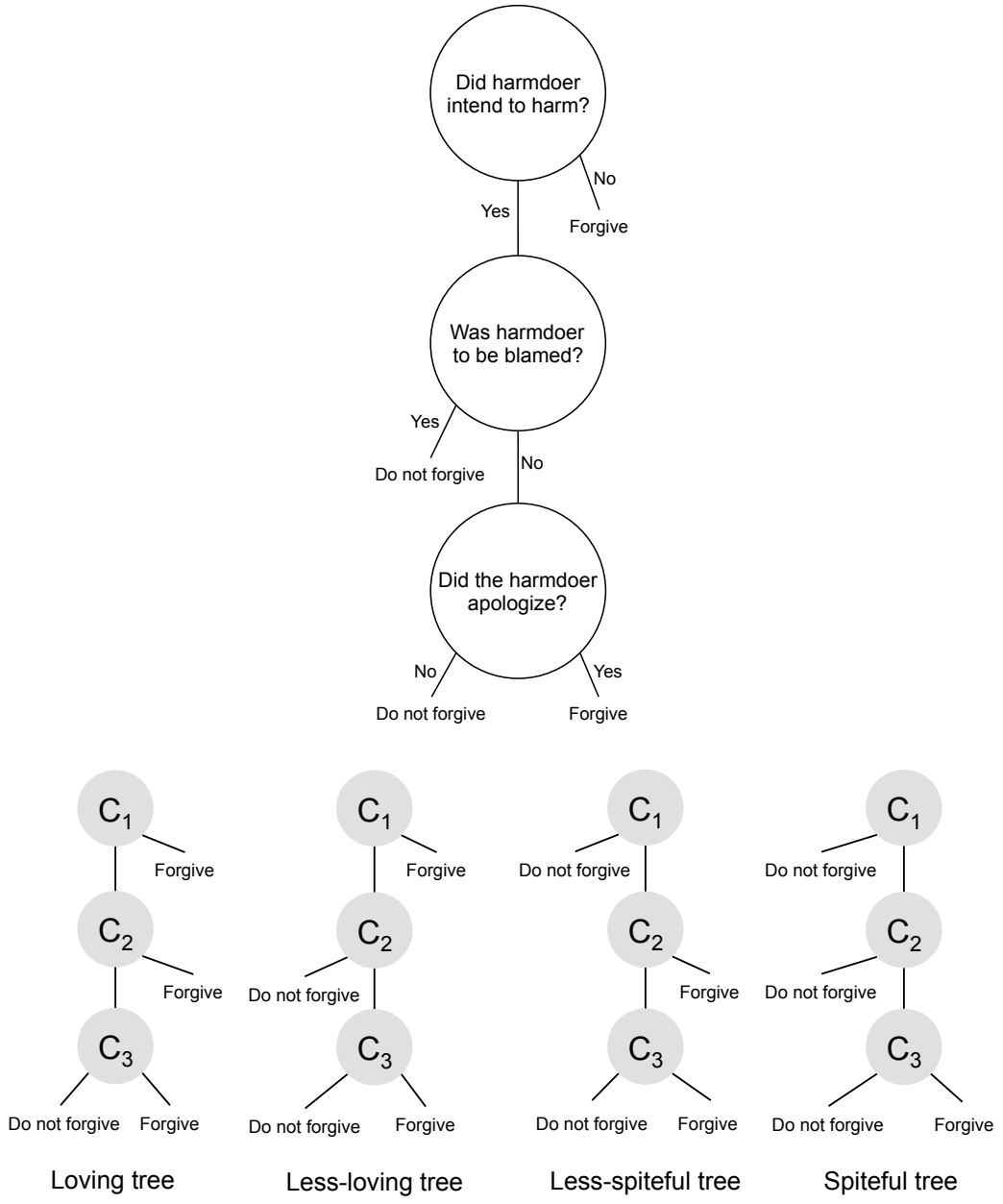

Figure 2. (Top) Illustration of how an offended individual might make a forgiveness decision using a fast-and-frugal tree (FFT). (Below) The four exit structures of an FFT, with C_1, C_2, and C_3 representing three evidence-strength cues in a fixed order. From left to right, the FFTs become less and less liberal and more and more conservative, and we name each FFT according to its expected decision tendency or bias.

As with the decision criterion, the evidence strength is also represented in a discrete form in an FFT. Specifically, a harmdoer's evidence strength can be expressed as a cue profile such as [1, 0, 0], where each number indicates whether the value of a cue strengthens ("1") or weakens ("0") the evidence and the numbers are presented in the order of how cues are searched. For example, if the cue order is intent, blame, apology, a cue profile of [1, 0, 0] indicates that there was no intent, the harmdoer was to be blamed, and there was no sincere apology; there are in total 23 = 8 possible cue profiles given three cues (see Table 3 for cue profiles used in the present study). Because FFTs are noncompensatory, a cue profile with 1s appearing more to the left (e.g., [1, 0, 0]) represents stronger evidence in comparison to another profile with 1s appearing later (e.g., [0, 1, 1]). This means that using an FFT with a particular cue order, it would be easier to "forgive" and harder to "not forgive" a harmdoer with the former profile than with the latter.[6]

Like other noncompensatory heuristics (Brandstätter, Gigerenzer, & Hertwig, 2006; Bröder, 2002), FFTs have been supported as valid descriptive models in a variety of domains (Luan et al., 2011). However, studies that compared FR and FFTs have mixed findings. For example, bail judges' decisions were better described by FFTs (Dhami & Ayton, 2001; Dhami, 2003) and traffic judges' decisions were better described by FR (Leiser & Schatzberg, 2008), whereas physicians' decisions about drug prescriptions were equally well described by both (Dhami & Harries, 2001).

1.2.3. Summary

In sum, FR is a compensatory linear model whereas FFT is a noncompensatory heuristic. Despite their differences in the assumptions made about the cognitive implementation of the subprocesses, both are well suited as decision models that embody EMT principles. Furthermore, both have been supported as good descriptive models for decisions in other domains, suggesting that they are plausible models for forgiveness decisions, as well.

2. THE PRESENT STUDY

Viewing the decision of whether to forgive as an error management task, we sought to answer two main questions: (1) How well can forgiveness decisions be described and predicted by the two models, FFTs and FR, and (2), is the selection

[6] [6] With a profile [1, 0, 0], the four FFTs from left to right in the lower panel of Figure 2 will make the decisions "forgive," "forgive," "not forgive," and "not forgive" at the 1st, 1st, 3rd, and 2nd cue, respectively. For the profile [0, 1, 1], the four FFTs will make the decisions "not forgive," "forgive," "not forgive," and "not forgive" at the 1st, 2nd, 1st, and 1st cue, respectively. Each of the four FFTs searched the same number or fewer cues to make a "forgive" decision and more cues to make a "not forgive" decision for profile [1, 0, 0] than [0, 1, 1], showing that the evidence for "forgive" is stronger in the former.

of the decision criterion influenced by variables that provide information on the cost of errors?

In our study, we first asked participants to recall a hurtful incident, provide details about the incident and the harmdoer, and report whether they had chosen to forgive. We then measured the subjective importance of each of the three evidence-strength cues. After that, participants made hypothetical decisions about the same harmdoer but with varying values for the evidence-strength cues. The hypothetical decisions enabled us to estimate the decision criterion adopted by each participant. Finally, with all parameters either measured or estimated, we applied the two models to predict the decision made by each participant in the recalled incident.

Previous research investigated either recalled (e.g., Finkel, Rusbult, Kumashiro, & Hannon, 2002) or hypothetical forgiveness decisions (e.g., Shackelford, Buss, & Bennett, 2002; Struthers et al., 2008), but never both together. There also has been no study that compared how forgiveness decisions can be described by different models, let alone with the hypothetical–recalled method used in our study. We refer to our method as ecological cross-validation, because it is a variation of the statistical method of cross-validation (e.g., Zucchini, 2000), and the parameters are fitted in hypothetical trials and used to predict real-life decisions.

EMT's key insight is that individuals set the decision criterion to reduce the likelihood of the more costly decision error. For forgiveness decisions, we postulate that the perceived RV and ER of the harmdoer should inform the cost of false negatives and false positives, respectively. Following this line of reasoning, we hypothesized that the decision criterion would be more liberal when RV exceeds ER and more conservative when it is the reverse. Testing this hypothesis was made possible by our modeling approach that allowed for the decision criteria to be estimated and compared.

We tested this hypothesis in two ways. First, we took the difference between RV and ER as reported by each participant (RV – ER, i.e., the rating of RV minus that of ER) and used it as an index of the relative cost of errors. Next, we derived the accuracy of a model with a specific decision criterion (e.g., the loving tree) in fitting the hypothetical decisions made by the participant and calculated the correlation between RV – ER and the accuracies across all participants. The hypothesis would be supported if RV – ER positively correlated with the accuracy of the models with the liberal criteria and negatively correlated with those with conservative criteria. This is based on the assumption that accuracy is an indication of the likelihood that the decision criterion specified in the model was adopted.

Second, we focused on two groups of participants whose relative cost of errors should differ markedly: those who rated the RV with the harmdoer as high

and the ER as low ($H_{RV}\_L_{ER}$) and those who rated the RV as low and the ER as high ($L_{RV}\_H_{ER}$). The hypothesis would be supported if more $H_{RV}\_L_{ER}$ participants were found to adopt a liberal criterion than $L_{RV}\_H_{ER}$ participants and more $L_{RV}\_H_{ER}$ participants were found to adopt a conservative criterion than $H_{RV}\_L_{ER}$ participants.

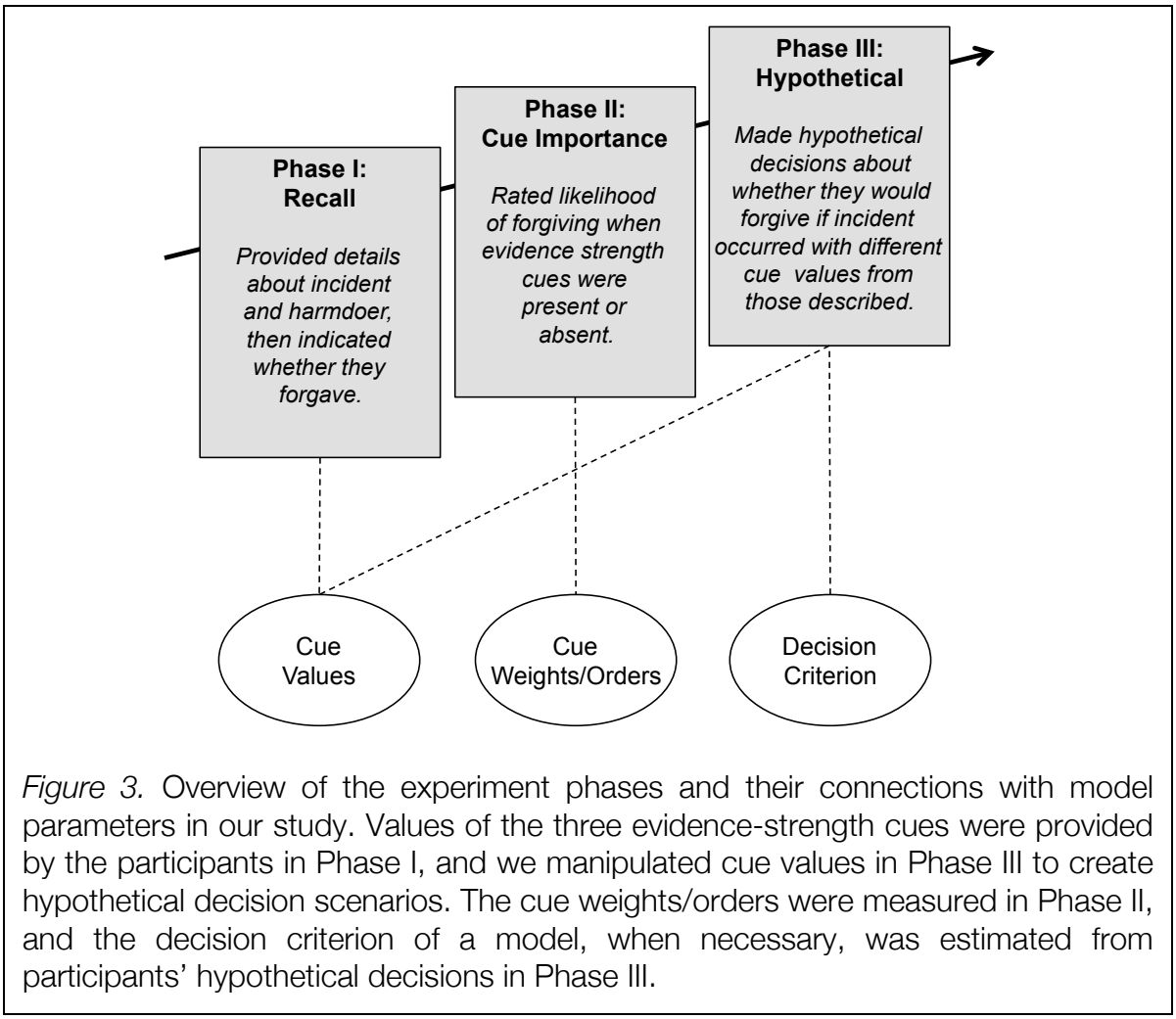

Figure 3. Overview of the experiment phases and their connections with model parameters in our study. Values of the three evidence-strength cues were provided by the participants in Phase I, and we manipulated cue values in Phase III to create hypothetical decision scenarios. The cue weights/orders were measured in Phase II, and the decision criterion of a model, when necessary, was estimated from participants' hypothetical decisions in Phase III.

2.1. Method

2.1.1. Participants

Two hundred forty-nine participants (51.8% female, M_{age} = 33.7 years, age range = 18–70 years) residing in the United States took part in this study. All were recruited via Amazon's Mechanical Turk and were remunerated U.S. $1.00 for their participation. Eleven participants (4 female) were dropped as they admitted that they had not paid full attention or were confused at some point. We originally planned on recruiting 250 participants.

2.1.2. Procedure and materials

Participants went through three phases in the study (see a summary in Figure 3) and provided demographic information before the study was concluded.

Phase I: Recall. Participants were asked to recall an incident in the last 6 months in which they had "felt wronged, let down, betrayed, or hurt" by a friend, romantic partner, or colleague, and spent 1–2 min writing about it. The harmdoer's initials were recorded and used in subsequent sections to refer to the harmdoer. Participants then responded to questions about the incident, provided details about their relationship with the harmdoer, and stated whether they forgave the harmdoer.

Evidence-strength cues, including the perception of the harmdoer's intention to hurt, how much the harmdoer was to blame, and the extent to which a sincere apology was offered, were measured using materials adapted from previous studies (Aquino, Tripp, & Bies, 2001; McCullough et al., 2003; McCullough, Worthington, & Rachal, 1997). All three cues were measured on a scale from 1 (*not at all*) to 7 (*completely*). Because the models we tested typically use binary cues as input, we dichotomized the cues by taking values above the mid-point (i.e., > 4) as present and the rest as absent. Variables that indicate the costs of errors, that is, RV and ER, were measured on a 7-point scale developed by Burnette and colleagues (2012). Items include "Our relationship is very rewarding to me" (RV) and "I feel like he/she might do something bad to me again" (ER).

The recalled forgiveness decision was measured both as a dichotomous yes–no item (i.e., "Have you forgiven [initials of harmdoer]?") and as a continuous value using the subscales of avoidance and revenge in the Transgression-Related Interpersonal Motivations Scale (McCullough et al., 1998). Items include "I'm going to avoid him/her" (avoidance) and "I want to see him/her hurt and miserable" (revenge).

Phase II: Measuring cue importance. For each of the three evidence-strength cues, participants rated the likelihood of forgiving the harmdoer when the cue was present and when it was absent, independent of other cues. The likelihoods were reported on a sliding scale from 0 (*definitely not forgive*) to 100 (*definitely forgive*), and the absolute difference between the two was taken as the subjective importance of a cue.

To increase reliability, we created two statements for the presence/absence of each cue. For example, the presence of blame was framed in one as "You blame the person and you feel that he/she has wronged you" and in another as "You feel that the person has victimized you and you blame him/her." Each participant rated 12 statements in total, 3 Cues × 2 Values (present and absent) × 2 Versions, in a random order. For each cue, its measured importance did not change between the two versions (see details in Results).

After data collection was concluded, we dropped 3 participants whose likelihoods of forgiving did not differ when a cue was present or absent, and another 14 whose reported likelihoods were in the reverse directions (e.g., more likely to forgive when the blame cue was present than when it was absent). Both were indications that the participants were not attentive to the task or did not fully understand the instructions.

Phase III: Hypothetical decisions. Participants indicated whether they would forgive the harmdoer (yes–no) if the recalled incident had unfolded differently from how it was described in the hypothetical scenarios. In these scenarios, we systematically manipulated the values of the three evidence-strength cues by combining the statements in Phase II. There were six cue–value combinations or profiles[7] (see Table 3) with two versions of each; thus, participants responded to a total of 12 scenarios in randomized order.

We dropped participants who forgave in all 12 trials (n = 25) or none (n = 18), because this showed that they either were insensitive to the cues or did not fully understand the task instructions. After all data-checking procedures, we were left with 181 participants who were included in the main analyses described below.

2.2. Results

2.2.1. Harmdoers and offenses

Harmdoers were friends (44.4%), romantic partners (34.8%), colleagues (18.2%), family members (7.7%), or "others" (2.2%). They were female in 52.5% of all the incidents recalled and were the same gender as the participant in 50.8% of the incidents. Excluding family members, the length of the relationship with the harmdoer ranged from 6 days to 40 years, with a median of 5.1 years. Offenses included infidelity (e.g., "My long-time girlfriend cheated on me"), physical assault (e.g., "She was drunk and acted very aggressively toward me"), cancelled appointments (e.g., "He was supposed to give me a ride to an important professional event, but cancelled at the last minute"), and lying (e.g., "My colleague lied about me to the management in order to save herself"), among others. Excluding one participant who indicated that the incident was still ongoing, the reported length of time since the offense had occurred ranged from 1 day to 9.9 years, with a median of 3.2 months.

[7] A full-factorial design with three cues would have yielded eight cue profiles. However, two cue profiles yield scenarios where no offense was committed: The first is when the harmdoer apologized but neither had the intent to hurt nor was blamed for the harm, and the second is when the harmdoer did not apologize, but again, neither had the intent to hurt nor was blamed for the harm. To avoid confusion, these two cue profiles were not included.

2.2.2. Importance of evidence-strength cues

The measured importance of each cue did not differ significantly between the two versions of statements (see above), all ps > .05; thus, we took the average of the two versions as each cue's subjective importance. Overall, intent had the highest subjective importance (*M* = 64.4, *SD* = 23.3), followed by blame (*M* = 49.1, *SD* = 25.5) and apology (*M* = 48.3, *SD* = 25.4). Similarly, intent had on average the highest rank (1.5), with apology (2.2) and blame (2.3) ranked lower.

Table 3. Forgiveness Rates Across Cue Profiles in Recalled and Hypothetical Decisions

Cue profile[a]	Intent (–)[b]	Blame (–)[b]	Apology (+)[b]	Frequency in recalled incidents (*n* = 181)	Forgiveness rate	
					Recalled (%)	Hypothetical (%)
1	N	Y	Y	28	82.1	89.0
2	Y	Y	Y	31	48.4	61.6
3	N	Y	N	54	44.4	56.4
4	Y	Y	N	49	24.5	10.8
5	Y	N	Y	0	–	68.5
6	Y	N	N	1	100	30.4
7	N	N	Y	4	100	–
8	N	N	N	14	78.6	–

[a] The upper half of the table contains the cue profiles that were each reported by more than 10% of the participants in the recalled incidents. They are listed in descending order of forgiveness rates. In the lower half of the table, there are two cue profiles (7 and 8) that were not included in the hypothetical decisions.

[b] For intent, a value of "Y" indicates that the harmdoer had the intent to harm. Because this cue is negatively associated with evidence strength, the cue value was coded as 0 in the Franklin's rule model; the same applies for the other negative cue blame. However, for apology, a value of "Y" indicates that the harmdoer offered a sincere apology. This cue is positively associated with evidence strength and thus its value was coded as 1 in the FR model.

2.2.3. Cue profiles in recalled incidents

We classified each recalled incident by its cue profile, which is the combination of binary evidence-strength cue values, and report the frequency of each cue profile in Table 3. To the best of our knowledge, ours is the first study to examine how the three evidence-strength cues are distributed and related to forgiveness decisions in real life.

In general, the cue profiles were unevenly distributed, with four reported by less than 10% of the participants. An examination of the presence of each cue

revealed that blame was present in an overwhelming majority of the incidents (162 of 181) and was much more common than intent (81) or apology (63). Moreover, there was a curious relationship between blame and intent: When intent was present, blame was also present in all but one case; however, when intent was absent (100 cases), blame was still present in the majority of them (82). Thus, it appears that intent is a sufficient but not necessary condition for blame. When intent is absent, other factors may drive evaluations of the blameworthiness, such as culpable control and foresight (e.g., Alicke, Buckingham, Zell, & Davis, 2008).

2.2.4. Forgiveness rates

Among 181 participants, 49.7% reported that they had forgiven the harmdoer for the recalled incident. Forgiveness was most common for friends (59.7%) and romantic partners (58.7%) and least common for colleagues (27.3%) and family members (21.4%). When participants forgave, they rated their motivation for avoidance and revenge lower than when they did not forgive, Welch's $t(173.26) = 2.16$, $p = .03$ and $t(138.89) = 2.20$, $p = .03$, respectively. This supports the validity of measuring forgiveness as a dichotomous construct.

Table 3 shows that forgiveness rates varied for harmdoers with different cue profiles, and among the four more frequent profiles, their rankings in forgiveness rate were similar in the recalled and the hypothetical decisions. Furthermore, the rankings also appear sensible. For example, we expected harmdoers with the intent–blame–apology profile [N, Y, N] to be forgiven more frequently than those with the profile [Y, Y, N], because intent is absent in the former but present in the latter, and this is indeed what was observed.

2.2.5. Decision criterion variables

The average ratings of the perceived RV and ER of the harmdoer were close to the mid-point of the 7-point scale, $M = 4.1$, $SD = 1.6$ for RV and $M = 4.2$, $SD = 1.5$ for ER. RV and ER are negatively correlated $r = -.36$, $p < .001$, suggesting that harmdoers perceived as being high in RV were also likely to be perceived as being low in ER, and vice versa (see Figure 4). For both RV and ER, we treated ratings above 4.5 as "high" and below 3.5 as "low," because of the ambiguity around the mid-point 4. Due to the negative correlation between RV and ER, more participants were classified as $L_{RV}\_H_{ER}$ ($n = 33$) and $H_{RV}\_L_{ER}$ ($n = 36$) than $H_{RV}\_H_{ER}$ ($n = 18$) and $L_{RV}\_L_{ER}$ ($n = 13$). We also took the difference between the two, RV – ER, as an index of the relative cost of the two errors. As discussed, we assumed that the greater the value of RV – ER, the more likely that the cost of a false negative would outweigh that of a false positive. As shown in Figure 4, the distribution of RV – ER is roughly normal and has a mean close to zero ($M = 0.1$, $SD = 2.6$).

2.2.6. Model performance

Following the ecological cross-validation approach, we first tested how well FFTs and FR fit participants' decisions in the hypothetical scenarios and then how well the models predicted participants' decisions in the recalled incidents. Figure 3 summarizes how we measured or estimated the parameters required in each model, and more details of our modeling method are provided in the Appendix.

Estimated criterion. For both FFTs and FR, the decision criterion that most accurately fitted each participant's hypothetical decisions was the one that they were estimated to have adopted. We found that majority of the participants were estimated to have adopted a liberal criterion: For FFTs, 128 of 181 participants were estimated to have adopted either the loving (73 participants) or the less-loving (55) trees; and for FR, 124 were estimated to have applied the criteria of x_{DC} = .10 (74) or x_{DC} = .40 (50). This suggests that participants were generally more lenient in deciding whether to forgive[8] and may reflect the common view of forgiveness being a moral good (e.g., McCullough, 2008).

Hypothetical decisions. Using the estimated decision criterion for each participant, the average fitting accuracy of the hypothetical decisions of each model was fairly high: 81.3% (*SD* = 11.6%) for FFTs and 80.4% (*SD* = 11.6%) for FR. The accuracies of both models were substantially higher than the base rate of decisions to forgive (i.e., 52.8%). Comparing the models' accuracy at the participant level, we found that the two models were equally accurate for 142 of the 181 participants; FFTs were better than FR for 24 participants and FR were better than FFTs for the remaining 15 participants.

The main reason why FFTs and FR performed so similarly is that the two models produced the same decision outputs for 87.5% of all the hypothetical decisions. Moreover, the congruence of the two models with a similar decision criterion was even higher (see Table 4). For example, when the decision criterion was the most liberal (i.e., the loving tree and FR with x_{DC} = .10), the congruence rate was 97.1%; and the average congruence rate across the four criteria was 94.9%. These results suggest high levels of model mimicry between FFTs and FR and indicate that both models could capture the participants' decision processes in the hypothetical decisions.

Recalled decisions. For each participant, we fixed the decision criterion for each model with what the participant was estimated to have used. We then used the models with the fixed criterion to predict the recalled decision. Compared to its fitting accuracy in the hypothetical decisions, each model predicted less accurately:

[8] At first glance, the high prevalence of lenient decision criteria may seem contradictory to the low prevalence of forgiveness decisions (about 50%) in our study. However, decisions are the joint output of the values of the evidence-strength cues and the decision criterion. In particular, the recalled incidents and the hypothetical trials both featured cue profiles of relatively low evidence strength.

69.0% (*SD* = 42.8%) for FFTs and 69.1% (*SD* = 43.7%) for FR. This drop in accuracy was expected in model testing using cross-validation procedures (Brandstätter et al., 2006; Gigerenzer & Brighton, 2009), and with an accuracy slightly lower than 70%, each model still predicted recalled decisions substantially better than a naïve model that predicted not forgiving in all recalled incidents because that decision was more prevalent (50.3%).

At the participant level, both models predicted the recalled decisions of 118 participants correctly and those of 49 participants incorrectly. For the remaining 14 participants, FFTs and FR each correctly predicted the decisions of 7 participants. As with the hypothetical decisions, the high congruence between the two models' decision outputs was the reason why they predicted at almost the same level of accuracy (see Table 4).

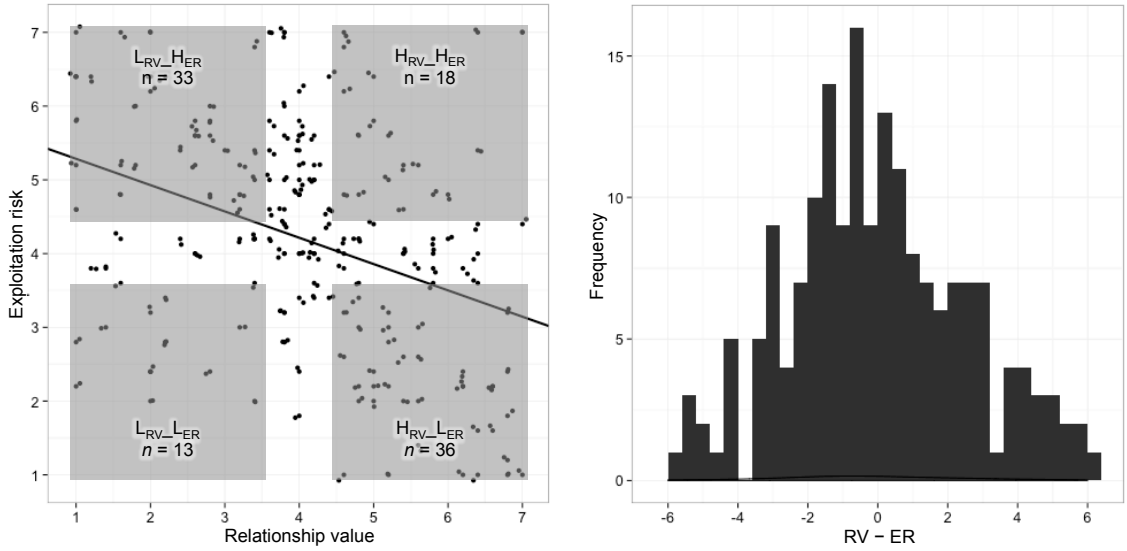

Figure 4. (Left) A scatterplot of the reported relationship value (RV) and exploitation risk (ER) of a harmdoer by all participants. The grey squares highlight four groups of participants who rated harmdoers' RV and ER as "high" (i.e., ratings above 4.5) and/or "low" (i.e., below 3.5). For example, "$L_{RV}\_H_{ER}$" is the group of participants who reported simultaneously a low rating of RV and a high rating of ER of a harmdoer. The right panel displays the frequency distribution of RV – ER (i.e., the rating of RV minus that of ER) across all participants.

Table 4. Congruence of Fitted and Predicted Decisions Between Fast-and-Frugal Trees and Franklin's Rule

Decisions	Estimated decision criterion[a]	Fixed decision criterion				
		Most liberal	Less liberal	Less conservative	Most conservative	M
Hypothetical	87.5%	97.1%	91.3%	92.6%	98.5%	94.9%
Recalled	92.1%	98.3%	91.7%	95.0%	97.2%	95.6%

[a] The model with the decision criterion that led to the highest fitting accuracy in the hypothetical decisions of each participant.

2.2.7. Decision criterion selection

We tested our hypothesis that the decision criterion used would depend on the perceived RV and ER of the harmdoer in two ways.

RV – ER and fitting accuracy. We focused on the hypothetical decisions and examined the correlation between RV – ER and the fitting accuracy of the models fixed with each of the four decision criteria. As shown in Figure 5, RV – ER correlated positively with the accuracy of the models with the most liberal criteria (i.e., the loving tree and FR with $x_{DC} = .10$) and negatively with the accuracy of the models with the most conservative criteria (i.e., the spiteful tree and FR with $x_{DC} = .90$). The negative correlations also held when the decision criteria of the models were .60), but the correlations were no longer positive when the decision criteria of the models were less liberal (i.e., the less loving tree and FR with $x_{DC} = .40$). Because fitting accuracy indicates the likelihood that a certain criterion was adopted, these results show that in general, the greater the cost of false negatives relative to that of false positives (i.e., RV > ER), the more likely that the decision criteria adopted were liberal and less likely that they were conservative.

$H_{RV}\_L_{ER}$ and $L_{RV}\_H_{ER}$ participants. We also examined the decision criterion estimated to have been adopted by two groups of participants: $H_{RV}\_L_{ER}$, those who rated the RV with the harmdoer as high (i.e., > 4.5) and ER as low (i.e., < 3.5), and $L_{RV}\_H_{ER}$, whose ratings on the two variables were the reverse. Recall that we hypothesized that the former group should be more likely to adopt the most liberal criterion and less likely to adopt the most conservative criterion than the latter.

We calculated the proportion of participants who were estimated to adopt either the most liberal or the most conservative criterion for each model. The results are presented in Figure 6. For FFTs, $H_{RV}\_L_{ER}$ participants were more likely to adopt the most liberal criterion than $L_{RV}\_H_{ER}$ participants (69.4% vs. 36.4%), $\chi 2(1, 68) = 6.30$, $p = .01$, but $L_{RV}\_H_{ER}$ participants were more likely to adopt the most conservative criterion than $H_{RV}\_L_{ER}$ participants (24.2% vs. 8.3%), although the latter difference was not statistically significant, $\chi 2(1, 68) = 2.17$, $p = .14$; very similar

results were also found for FR. The general pattern of these results is consistent with our hypothesis.

Summary. Our hypotheses were based on EMT's argument that adaptive cognitive systems are biased in the direction of reducing the likelihood of the more costly error. In the context of forgiveness, we assumed that the costs of false negatives and false positives could be represented, at least in part, by the perceived RV and ER of the harmdoer. We tested our hypothesis in two ways and found support for it in both. We conclude that error management concerns were involved in the decision about whether to forgive as investigated in our study.

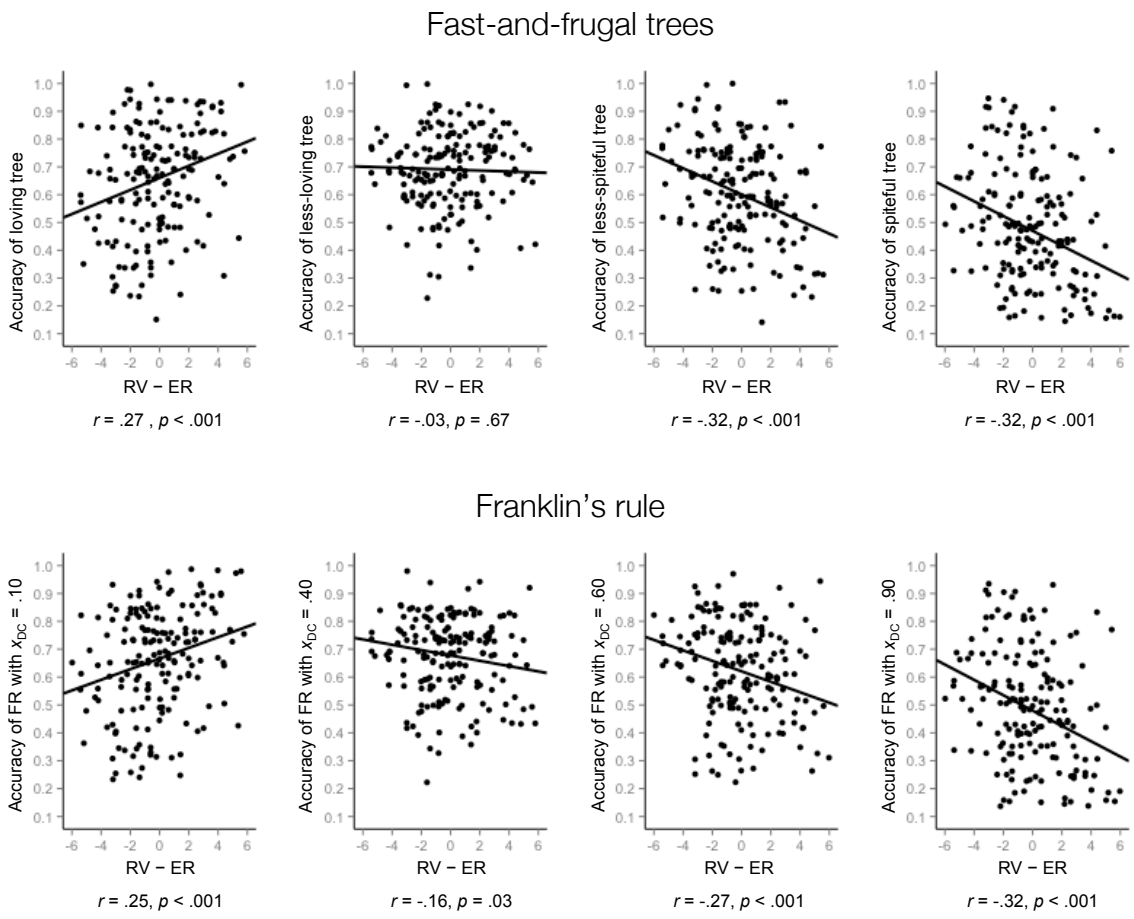

Figure 5. Scatterplots showing the association between RV – ER (i.e., relationship value minus exploitation risk) and the fitting accuracy of each of the fast-and-frugal trees and the Franklin's rule (FR) models that assumed a fixed decision criterion for all participants. Each plot contains the best fitting regression line with the bivariate Pearson's r reported below.

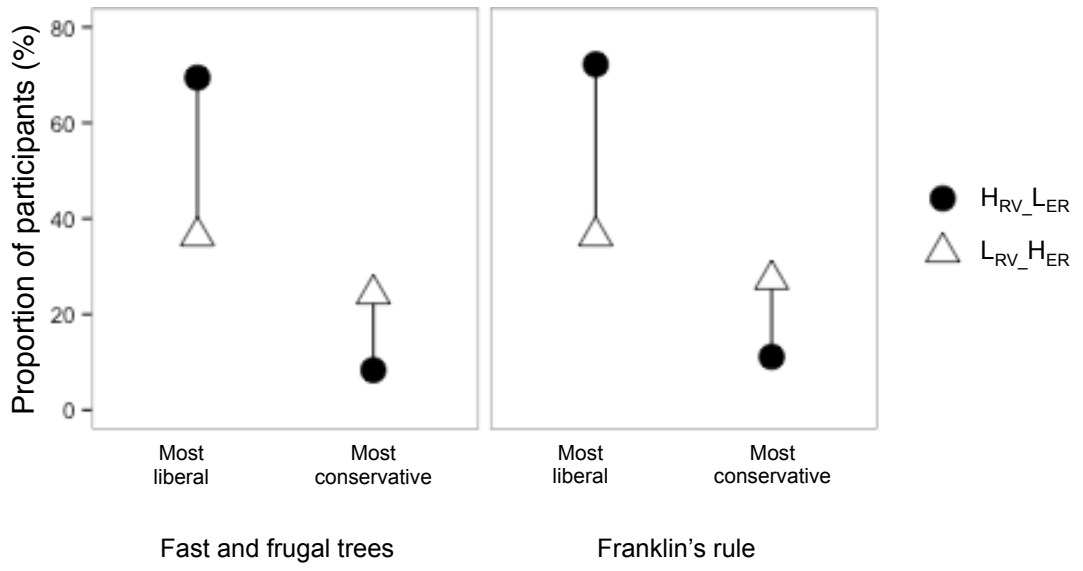

Figure 6. The proportions of participants whose decision criteria were estimated to be either the most liberal or the most conservative in each type of model. The participants are from two contrasting groups, $H_{RV}\_L_{ER}$ (n = 36), who gave a high rating of relationship value (RV) and a low rating of exploitation risk (ER) for the harmdoer, and $L_{RV}\_H_{ER}$ (n = 33), whose ratings were in the reverse directions. The lines signify the differences in proportion.

3. GENERAL DISCUSSION

EMT has been influential in the study of adaptive behavior and has led to insights about biases across many domains (Johnson et al., 2013). However, empirical EMT studies have rarely investigated how biases are implemented cognitively (e.g., McKay & Efferson, 2010) or how biases are related to the integration of evidence-strength cues in decision making. We addressed this gap in the literature by developing and testing mechanistic models that embody the logic of EMT and make explicit assumptions about the cognitive implementations of the two subprocesses. The two models we tested, FFTs and FR, were both descriptive of forgiveness decisions even though they made different assumptions about cognitive implementation. We showed that the decision biases adopted in forgiveness decisions generally followed the qualitative patterns predicted by EMT. In the following, we discuss some issues related to the findings of our study and argue for the applicability of our modeling and experimental approaches in other evolutionarily recurrent task domains.

3.1. The ecological cross-validation approach

One novel methodological contribution of our study is the ecological cross-validation approach, in which we estimated a key parameter of a model (i.e., the decision criterion) in hypothetical decisions and applied it to predict real recalled

decisions. This approach enabled us to assess the model's ability to predict the object of interest, that is, people's decisions made in everyday interactions, which is often absent in other modeling studies. While this approach was able to produce ecologically valid and practically relevant outcomes, a disadvantage is that it likely ignored other relevant factors that impact real-life events. In the context of our study, for example, a harmdoer may have previously demonstrated a very high (or a very low) prosocial concern for the victim in other incidents through other cues than those we had measured, reducing a model's ability to predict the victim's decision correctly.

As such, this is one reason why there was about a 10% drop in model accuracy between fitting hypothetical decisions (≈ 80%) and predicting recalled decisions (≈ 70%) for both FFTs and FR. Another reason is that we estimated the decision criterion based on a very limited sample of hypothetical decisions (i.e., 12 for each participant). Previous studies have shown that smaller samples typically result in a poorer precision of parameter estimations, making out-of-sample predictions less accurate (e.g., Gigerenzer & Brighton, 2009). In light of these modeling difficulties (i.e., predicting noisy real-life decisions on the basis of limited data), we consider a near 70% prediction accuracy an achievement and take it as evidence in favor of the overall soundness of our models. In any case, it sets a first benchmark against which other models can be compared.

3.2. Model mimicry

In both fitting and prediction, FR and FFTs achieved remarkably similar levels of accuracy, which was driven mainly by the high congruence rates between the two models' decision outputs (see Table 4). There may be several reasons for this result: First, both models implement the two subprocesses of EMT and have parameters corresponding to them. Because they were derived from the same principles, the resulting similar performance is not that surprising. Second, our study was not designed to test which model would be a better descriptive one; that would require a different methodology (such as Garcia-Retamero & Dhami, 2009). Third, due to limited data and the possibility that participants switch strategies across time, model mimicry happens frequently in model testing and it has largely been accepted in the literature (e.g., Regenwetter, Dana, Davis-Stober, & Guo, 2011). Finally, it has been shown that linear models and lexicographic heuristics lead to very similar choices when there are a few binary cues (Katsikopoulos, 2013).

3.3. Relationship between evidence strength and decision criterion

In signal detection theory (Green & Swets, 1966), from which many concepts of EMT originate, factors affecting evidence strength are typically assumed to be orthogonal to those affecting the decision criterion. This assumption, however, may not hold in real life. Nevertheless, whatever the ecological reasons for the correlation, they should not affect the predictions made by our models and the

conclusions of our study, because orthogonality was not a requirement in our models.

Though the variables of evidence strength and decision criterion may not be completely orthogonal, it is still helpful to treat them as separate because of their distinct effects. For instance, a subordinate may have no other choice but to forgive a superior at work regardless of the strength of the evidence that the superior is an ally, because maintaining that relationship is so vital that the criterion for forgiving is very low (e.g., Aquino et al., 2001). Similarly, the harmdoer may be a close associate of a highly valued third party (e.g., a sibling's spouse) and the cost of a false negative may also include the loss of benefits with that third party (e.g., Descioli & Kurzban, 2011; Pietraszewski & German, 2013). At the other extreme, there are cases, such as premeditated murders, where the evidence strength is so weak that forgiving becomes impossible no matter where the decision criterion is set (e.g., Daly & Wilson, 1988).

3.4. Connections with other systems

The forgiveness system does not exist in isolation but connects with other systems for inputs and outputs. Depending on the decision, the output may be passed on to the reconciliation (Worthington, 2006) or the revenge (Petersen et al., 2012) system, and whether reconciliation or revenge is achieved is then managed by a separate system of self-control that regulates behavior to achieve the desired relational outcome (Balliet, Li, & Joireman, 2011). Connecting forgiveness with these output systems allows for decision and behavior to be disentangled: An agent may decide to forgive but fail to reconcile with the harmdoer because of poor self-control, among other reasons.

A necessary input to the forgiveness system is one that builds a representation of the harmdoer's prosocial concern for the victim, which in our study was the strength of the evidence that the harmdoer is an ally. This conceptualization is similar to the monitored welfare trade-off ratio (WTR) of a relationship partner that an agent may use in social decision making (Tooby, Cosmides, Sell, Lieberman, & Sznycer, 2008). The higher a partner's WTR, the more likely she or he is to make sacrifices to benefit the agent in the future, and thus the more likely that a continued relationship with the partner will bring more fitness benefits than costs. Whether a measure of WTR can replace measures of evidence strength in our models is a topic we plan to explore in future studies.

3.5. Many variables of forgiveness

We investigated five variables related to forgiveness in this study (Table 2), although more than 25 have been examined in a meta-analysis (Fehr et al., 2010). Besides practical concerns, we wanted to limit the number of variables investigated because decision makers are known to be frugal with information use and do not

search for and consider all possible variables when deciding (Dhami & Ayton, 2001; Gigerenzer et al., 1999).

Furthermore, the perspective of the current EMT framework suggests that while decision makers need information that informs both the evidence strength and cost of errors, only a few within each category may be sufficient. In this way, the framework can also organize the many variables of forgiveness according to the subprocess of the decision they impact. One implication of this organization is that information search will be directed to the category where information is lacking. For example, a decision maker who has a good estimate of the harmdoer's prosocial concern (i.e., evidence strength) will be less likely to seek out additional information about their intent and will instead be motivated to seek out information about the cost of errors such as the ER to set the criterion.

3.6. Conclusion

Our use of cognitive models to investigate forgiveness decisions builds on forgiveness decisions as an error management task and provides insights on the process of how such decisions are made. Though the principal topic investigated in this paper is forgiveness, the general theme is how decision makers handle uncertainty and solve problems across contexts that vary in the fitness consequences of the two errors. Forgiveness is one of many natural problems faced by agents that can be understood from the perspective of EMT (e.g., Johnson et al., 2013; Oaten, Stevenson, & Case, 2009). Notably, this also includes other social decision problems closely related to forgiveness and cooperation; error management concerns are also likely to be implicated in decisions about revenge, social exchange, and the maintenance of coalitions. Our attempt is thus a demonstration of the potential of EMT to spark investigations of the structure of adaptive cognitive mechanisms and open the proverbial "black box" that has remained closed despite decades of research.

REFERENCES

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*(4), 556–574. doi:10.1037/0033-2909.126.4.556

Alicke, M. D., Buckingham, J., Zell, E., & Davis, T. (2008). Culpable control and counterfactual reasoning in the psychology of blame. *Personality and Social Psychology Bulletin*, *34*(10), 1371–1381. doi:10.1177/0146167208321594

Anderson, N. H. (1971). Integration theory and attitude change. *Psychological Review*, *79*(3), 171–206. doi:10.1037/h0021465

Aquino, K., Tripp, T. M., & Bies, R. J. (2001). How employees respond to personal offense: the effects of blame attribution, victim status, and offender status on revenge and reconciliation in the workplace. *The Journal of Applied Psychology*, *86*(1), 52–59. doi:10.1037/0021-9010.86.1.52

Aureli, F., Cords, M., & van Schaik, C. P. (2002). Conflict resolution following aggression in gregarious animals: A predictive framework. *Animal Behaviour*, *64*, 325–343. doi:10.1006/anbe.2002.3071

Axelrod, R., Hamilton, W. D., Series, N., & Mar, N. (2008). The evolution of cooperation. *Science*, *211*(27), 1390–1396. doi:10.1086/383541

Balliet, D., Li, N. P., & Joireman, J. (2011). Relating trait self-control and forgiveness within prosocials and proselfs: Compensatory versus synergistic models. *Journal of Personality and Social Psychology*, *101*(5), 1090–1105. doi:10.1037/a0024967

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, *13*, 171–195. doi:10.1016/0162-3095(92)90032-Y

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychology Review*, *113*(2), 409–432. doi:10.1037/0033-295X.113.2.409

Brehmer, B. (1994). The psychology of linear judgement models. *Acta Psychologica*, *87*(2-3), 137–154. doi:10.1016/0001-6918(94)90048-5

Bröder, A. (2002). Take the best, Dawes' rule, and compensatory decision strategies: A regression-based classification method. *Quality & Quantity*, *36*, 219–238. doi:10.1023/A:1016080517126

Burnette, J. L., McCullough, M. E., Van Tongeren, D. R., & Davis, D. E. (2012). Forgiveness results from integrating information about relationship value and exploitation risk. *Personality and Social Psychology Bulletin*, *38*(3), 345–356. doi:10.1177/0146167211424582

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The Adapted Mind* (pp. 163–228). New York: Oxford University Press. doi:10.1098/rstb.2006.1991

Daly, M., & Wilson, M. I. (1988). Homocide and human nature. In *Homocide* (5th ed.). New Brunswick, New Jersey: Transaction Publishers.

Descioli, P., & Kurzban, R. (2011). The company you keep: Friendship decisions from a functional perspective. In J. I. Krueger (Ed.), *Social Judgment and Decision Making* (1st ed., pp. 209–225). New York: Psychology Press.

Dhami, M. K. (2003). Psychological models of professional decision making. *Psychological Science*, *14*(2), 175–180. doi:10.1111/1467-9280.01438

Dhami, M. K. (2012). Offer and acceptance of apology in victim-offender mediation. *Critical Criminology*, *20*(1), 45–60. doi:10.1007/s10612-011-9149-5

Dhami, M. K., & Ayton, P. (2001). Bailing and jailing the fast and frugal way. *Journal of Behavioral Decision Making*, *14*(2), 141–168. doi:10.1002/bdm.371

Dhami, M. K., & Harries, C. (2001). Fast and frugal versus regression models of human judgement. *Thinking & Reasoning*, *7*(1), 5–27. doi:10.1080/13546780042000019

Dugatkin, L. A. (2002). Cooperation in animals: An evolutionary overview. *Biology and Philosophy*, *17*, 459–476. doi:10.1023/A:1020573415343

Fehr, R., & Gelfand, M. J. (2010). When apologies work: How matching apology components to victims' self-construals facilitates forgiveness. *Organizational Behavior and Human Decision Processes*, *113*(1), 37–50. doi:10.1016/j.obhdp.2010.04.002

Fehr, R., Gelfand, M. J., & Nag, M. (2010). The road to forgiveness: A meta-analytic synthesis of its situational and dispositional correlates. *Psychological Bulletin*, *136*(5), 894–914. doi:10.1037/a0019993

Finkel, E. J., Rusbult, C. E., Kumashiro, M., & Hannon, P. A. (2002). Dealing with betrayal in close relationships: Does commitment promote forgiveness? *Journal of Personality and Social Psychology*, *82*(6), 956–974. doi:10.1037/0022-3514.82.6.956

Garcia-Retamero, R., & Dhami, M. K. (2009). Take-the-best in expert-novice decision strategies for residential burglary. *Psychonomic Bulletin & Review*, (16), 163–169.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*(1), 107–143. doi:10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: The foundations of adaptive behavior*. (G. Gigerenzer, R. Hertwig, & T. Pachur, Eds.). New York: Oxford University Press. doi:10.1093/acprof:oso/9780199744282.001.0001

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press. doi:10.1007/s13398-014-0173-7.2

Godfray, H. C. J. (1992). The evolution of forgiveness. *Nature*, *355*(6357), 206–207. doi:10.1038/355206a0

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75–90. doi:10.1037/h0092846

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Hammond, K. R. (1996). *Human judgement and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice.* New York: Oxford University Press.

Harries, C., Evans, J., Dennis, I., & Dean, J. (1996). A clinical judgement analysis of prescribing decisions in general practice. *Le Travail Humain*, *59*(1), 87–109.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective in cross-sex mind reading. *Journal of Personality and Social Psychology*, *78*(1), 81–91. doi:dx.doi.org/10.1037/0022-3514.78.1.81

Haselton, M. G., & Nettle, D. (2005). The paranoid optimist: An integrative evolutional model of cognitive biases. *Personality and Social Psychology Review*, *10*(1), 47–66. doi:10.1207/s15327957pspr1001_3

Hutchinson, J. M. C., & Gigerenzer, G. (2005). Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behavioural Processes*, *69*(2), 97–124. doi:10.1016/j.beproc.2005.02.019

Johnson, D. D. P., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology and Evolution*, *28*(8), 474–481. doi:10.1016/j.tree.2013.05.014

Katsikopoulos, K. V. (2013). Why Do Simple Heuristics Perform Well in Choices with Binary Attributes? *Decision Analysis*, *10*(4), 327–340. doi:10.1287/deca.2013.0281

Kurzban, R., Burton-Chellew, M. N., & West, S. A. (2015). The evolution of altruism in humans. *Annual Review of Psychology*, *66*(1), 575–599. doi:10.1146/annurev-psych-010814-015355

Leiser, D., & Schatzberg, D.-R. (2008). On the complexity of traffic judges' decisions. *Judgment and Decision Making*, *3*(8), 667–678.

Luan, S., Schooler, L. J., & Gigerenzer, G. (2011). A signal-detection analysis of fast-and-frugal trees. *Psychological Review*, *118*(2), 316–338. doi:10.1037/a0022684

Malle, B. F., & Knobe, J. (1997). The Folk Concept of Intentionality. *Journal of Experimental Social Psychology*, *121*(33), 101–121. doi:10.1006/jesp.1996.1314

Marshall, J. A. R., Trimmer, P. C., Houston, A. I., & McNamara, J. M. (2013). On evolutionary explanations of cognitive biases. *Trends in Ecology & Evolution*, *28*(8), 469–473. doi:10.1016/j.tree.2013.05.013

Martignon, L., Katsikopoulos, K. V., & Woike, J. K. (2008). Categorization with limited resources: A family of simple heuristics. *Journal of Mathematical Psychology*, *52*, 352–361. doi:10.1016/j.jmp.2008.04.003

McCullough, M. E., Fincham, F. D., & Tsang, J.-A. (2003). Forgiveness, forbearance, and time: The temporal unfolding of transgression-related interpersonal motivations. *Journal of Personality and Social Psychology*, *84*(3), 540–557. doi:10.1037/0022-3514.84.3.540

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2013). Cognitive systems for revenge and forgiveness. *The Behavioral and Brain Sciences*, *36*(1), 1–15. doi:10.1017/S0140525X11002160

McCullough, M. E., Luna, L. R., Berry, J. W., Tabak, B. A., & Bono, G. (2010). On the form and function of forgiving: Modeling the time-forgiveness relationship and testing the valuable relationships hypothesis. *Emotion*, *10*(3), 358–376. doi:10.1037/a0019349

McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington, E. L., Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *Journal of Personality and Social Psychology*, *75*(6), 11–12.

McCullough, M. E., Worthington, E. L., & Rachal, K. C. (1997). Interpersonal forgiving in close relationships. *Journal of Personality and Social Psychology*, *73*(2), 321–336. doi:10.1037/0022-3514.73.2.321

McKay, R. T., & Efferson, C. (2010). The subtleties of error management. *Evolution and Human Behavior*, *31*(5), 309–319. doi:10.1016/j.evolhumbehav.2010.04.005

Oaten, M., Stevenson, R. J., & Case, T. I. (2009). Disgust as a Disease-Avoidance Mechanism. *Psychological Bulletin*, *135*(2), 303–321. doi:10.1037/a0014823

Ohtsubo, Y., & Yagi, A. (2015). Relationship value promotes costly apology-making: testing the valuable relationships hypothesis from the perpetrator's perspective. *Evolution and Human Behavior*, *36*(3), 232–239. doi:10.1016/j.evolhumbehav.2014.11.008

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York, UK: Cambridge University Press.

Perilloux, C., & Kurzban, R. (2015). Do men overperceive women's sexual interest? *Psychological Science*, *26*(1), 70–77. doi:10.1177/0956797614555727

Petersen, M. B., Sell, A., Tooby, J., & Cosmides, L. (2012). To punish or repair? Evolutionary psychology and lay intuitions about modern criminal justice. *Evolution and Human Behavior*, *33*(6), 682–695.

Pietraszewski, D., & German, T. C. (2013). Coalitional psychology on the playground: Reasoning about indirect social consequences in preschoolers and adults. *Cognition*, *126*(3), 352–363. doi:10.1016/j.cognition.2012.10.009

Regenwetter, M., Dana, J., Davis-Stober, C. P., & Guo, Y. (2011). Parsimonious testing of transitive or intransitive preferences: Reply to Birnbaum (2011). *Psychological Review*, *118*(4), 684–688. doi:10.1037/a0025291

Schlenker, B. R., & Darby, B. W. (1981). The use of apologies in social predicaments. *Social Psychology Quarterly*, *44*(3), 271–278. doi:10.2307/3033840

Sell, A. N. (2011). The recalibrational theory and violent anger. *Aggression and Violent Behavior*, *16*(5), 381–389. doi:10.1016/j.avb.2011.04.013

Shackelford, T. K., Buss, D. M., & Bennett, K. (2002). Forgiveness or breakup: Sex differences in responses to a partner's infidelity. *Cognition & Emotion*, *16*(2), 299–307. doi:10.1080/02699930143000202

Struthers, C. W., Eaton, J., Santelli, A. G., Uchiyama, M., & Shirvani, N. (2008). The effects of attributions of intent and apology on forgiveness: When saying sorry may not help the story. *Journal of Experimental Social Psychology*, *44*, 983–992. doi:10.1016/j.jesp.2008.02.006

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

Tooby, J., Cosmides, L., Sell, A., Lieberman, D., & Sznycer, D. (2008). Internal regularory variables and the design of human motivation: A computational and evolutionary approach. In A. J. Elliot (Ed.), *Handbook of Approach and Avoidance Motivation* (pp. 251–270). Taylor & Francis Group.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*, 35–57. doi:10.1086/406755

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct.* New York: Guilford Press.

Worthington, E. L. (2006). *Forgiveness and reconciliation: Theory and application*. New York: Routledge.

Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology*, *44*(1), 41–61. doi:10.1006/jmps.1999.1276

Appendix: Modeling Procedure

We modeled the forgiveness decisions using Franklin's rule (FR) and fast-and-frugal trees (FFTs), with four decision criterion values within each model. In the FR model, we examined xDC values of .10, .40, .60, and .90, which correspond respectively to the decision criteria of the loving tree, the less-loving tree, the less-spiteful tree, and the spiteful tree (listed from liberal to conservative; see Figure 2).

In addition to the decision criterion, the other parameters needed to implement a model on each participant's decisions are the cue values and the cue weights/orders (see Figure 3). The cue values in the recalled decision were taken from participants' reports of the incident (e.g., whether they perceived that the harmdoer was to blame), whereas the values in the hypothetical scenarios were given to the participants with statements. For example, the following was used in the hypothetical decision phase to indicate that the harmdoer intended to harm and was blamed for the harm, but did not offer a sincere apology: "You blame <harmdoer's initials> and you feel that he/she has wronged you. You feel that <harmdoer's initials> has victimized you and you blame him/her. The next time you meet, <harmdoer's initials> chats but does not mention the incident."

The cue weights/orders were estimated from a participant's responses in the absent/present procedure (see Phase II in Figure 3). For FR, the importance of each cue was divided by the sum of the importance of all three cues, and this was used as the weight for each cue. For FFTs, the importance of each cue informed its cue order. When cues were tied in importance, one was randomly selected to be ordered first.

We first examined the hypothetical decisions and compared them with the predictions of an FR or an FFT model with a particular decision criterion (i.e., with one particular x_{DC} or exit structure). We then chose the x_{DC} or exit structure, among the four examined in each model category, which maximized a model's accuracy for a participant in the hypothetical decisions. If there was more than one maximizing value, we chose one randomly. This procedure allowed us to estimate the decision criterion for each model that each participant was likely adopting. With all parameters fixed at their measured or estimated values, we applied each model to predict the recalled forgiveness decision of each participant.

Whereas the random selection among the best fitting criteria had no impact on a model's accuracy in the hypothetical decisions, it could impact the model's prediction accuracy in the recalled decision. This is because a model's prediction accuracy may be affected by the criterion is chosen (e.g., a loving tree leads to a correct prediction for a participant but a less loving tree does not, although both fit the participant's hypothetical decisions equally well). To mitigate this effect, we repeated the random selection 100 times and took the average as the model's prediction accuracy.

Chapter 3.

Assessing the base rate in forgiveness decisions: The function of social trust

How do individual differences in the *tendency to forgive* impact the decision about whether to forgive? This chapter builds on the error management framework of forgiveness and proposes that individual differences in the tendency to forgive reflect differences in the social environment. Since environments vary in the base rate of allies and foes, decisions about whether to forgive should incorporate that information. Base rate assessments will likely be expressed as *social trust*—a belief about whether people are generally benevolent or malevolent. An individual's level of social trust is in turn prompted by life history events, and constitutes a part of their life history strategy. It was hypothesized and found that a higher level of social trust and a greater tendency to forgive is associated with a slow life history strategy. The implications of these associations as well as the relation between life history theory and error management theory will be explored in this chapter.

Keywords. forgiveness, individual differences, social trust, life-history strategies, base-rate, error management

1. INTRODUCTION

At first glance, the error management framework of forgiveness decisions introduced in the previous chapter does not leave room for individual differences. Faced with the exact same incident and harmdoer (i.e., with the same evidence strength and the same cost of errors), different individuals should be expected to make the same decision. Yet, individuals differ in their forgiveness rates and they also exhibit stable individual differences in the tendency to forgive across situations and time (Balliet, Li, & Joireman, 2011; Fehr, Gelfand, & Nag, 2010). While the previous chapter was focused on the cost of errors, according to error management theory (Haselton & Nettle, 2005) and signal detection theory (Green & Swets, 1966), the selection of the decision criterion is also influenced by the *base rate*. In the context of forgiveness, the base rate refers to the relative proportion of allies to foes in the decision maker's social environment. The term *allies* is used to refer to agents with whom a relationship will bring more fitness benefits than costs, while the term *foes* refer to the reverse (as in the previous chapter). Since different individuals are embedded in different environments with varying base rates, individual differences in forgiveness tendencies may reflect that difference.

In this light, I will argue in this chapter that one of the factors influencing an individual's tendency to forgive (i.e., *trait forgiveness*) is the base rate of allies in their social environment. Nevertheless, individuals are unlikely to have a numerical base rate in mind; instead, base rate assessments will likely take the form of a belief about whether people are generally benevolent or malevolent, and be expressed as a level of *social trust* (e.g., Nannestad, 2008; Yamagishi, 2011). This chapter will explore the role of social trust in forgiveness from the perspective of life history theory. Initial evidence in support of the association between trait forgiveness, social trust, and life history variables will also be presented.

1.1. The role of social trust

Conceptualizations of trust fall under two broad kinds—*particularized* and *generalized*—where the former is targeted at a specified individual (e.g., "John"), and the latter is concerned with an unspecified individual (e.g., "people in general") that one does not have information about (Bjørnskov, 2007; Nannestad, 2008). While both kinds of trust have their place in the error management framework of forgiveness[9], the focus of this chapter is on the generalized kind (i.e., social trust) as it is a stable individual difference like trait forgiveness (e.g., Carl & Billari, 2014), and is aligned with the current conception of base rate. Unless otherwise stated, subsequent mentions of trust in this chapter refer to the generalized kind.

Trust, according to Delhey and Newton (2005), is "the belief that others will not deliberately or knowingly do us harm, if they can avoid it, and will look after our interests, if this is possible" (p. 311). Similarly, Yamagishi (2011) defines trust as "the default expectation of another person's trustworthiness in the absence of information about that person" (p. 114), and argues that trust acts as a "bias" that can be beneficial in certain environments. These definitions conceptualize trust as an assumption about whether people in general are more likely to provide benefits or to inflict costs. Thus, an individual's degree of social trust can be seen as a belief about the base rate of allies in their social environment[10]. Furthermore, the theorized impact of social trust as a bias is analogous to that of the decision criterion in the error management framework. As discussed in the previous chapter, a liberal decision criterion indicates a bias towards forgiving while a conservative criterion indicates a bias against forgiving.

If trust does indicate the base rate, then it should also have some validity in predicting whether people in general are more likely to be malevolent or benevolent. Indeed, there is some evidence that supports the assertion. For example, aggregate levels of trust across countries are correlated with indicators of the countries'

[9] Particularized trust can potentially be seen as a component of evidence strength. For example, when a decision maker has high trust in a specific individual (e.g., John), this suggests that the evidence strength that John is an ally is likely to be high (i.e., a relationship with John will likely result in more fitness benefits than costs).

[10] A similar argument was made regarding stereotypes and base-rate beliefs (see Gigerenzer, 1991).

trustworthiness, such as levels of corruption and homicide (see Bjørnskov, 2007; Elgar & Aitken, 2011; Nannestad, 2008; Uslaner, 2004). Levels of trust is also predictive of the likelihood of wallets returned in an experiment (Dufwenberg & Gneezy, 2000).

Having high levels of trust is associated with greater prosocial and cooperative behavior, as well as a variety of positive outcomes such as better health and greater happiness (Bekkers, 2012; Evans & Krueger, 2014; Helliwell, Huang, & Wang, 2014). It is also sometimes referred to as "social capital" because nations with greater trust tend to also have more efficient public institutions, higher economic growth rates, and have better management of crises (Bjørnskov, 2007; Helliwell et al., 2014; Putnam, 2001). In sum, trust is essential to maintaining the wide spread cooperation amongst non-kin found across societies and thus is a crucial part of understanding how individuals make social decisions and solve the recurrent problems of social living.

1.2. Trust and the forgiveness framework

The association between trust and forgiveness has rarely been explored even though the two concepts overlap considerably (for exceptions, see Molden & Finkel, 2010; Wieselquist, 2009). For instance, incidents that engender forgiveness decisions are often described colloquially as "breaches of trust" (e.g., Boon & Sulsky, 1997), and a widely used measurement instrument of forgiveness includes distrust as an indication of lack of forgiveness (e.g., McCullough et al., 1998). Reconciliation, the relational and behavioral outcome of forgiveness, is also sometimes described as a "restoration of violated trust" (e.g., Fincham, 2000).

Perhaps it is because the terms are often used interchangeably that their association is seen as obvious and unworthy of investigation. Nevertheless, the topics of forgiveness and trust have generated two separate and substantial bodies of research that have largely ignored each other. It has even been argued that research in forgiveness has limited implications for that of trust (e.g., Haselhuhn, Kennedy, Kray, Van Zant, & Schweitzer, 2015). Despite the evident conceptual closeness, it is not known where the similarities and differences between trust and forgiveness lie, and to what extent the results from one can be applied to the other. Conceiving trust within the error management framework of forgiveness is a first step in integrating these related but separate streams of research.

From the perspective of the error management framework of forgiveness, trust is analyzed based on its function or current utility (e.g., Bateson & Laland, 2013; Tinbergen, 1963). If trust is an indication of the base rate of allies, then it should be the product of an intelligent assessment of the social environment and be discriminately held. This function of trust is likely to be the same across many other decisions of cooperation that also have fitness outcomes that are impacted by the base rate of allies and foes (e.g., Balliet & Lange, 2013).

In this light, trust affects the forgiveness decision as follows: When it is high, it is an indication that the environment has more allies than foes, and therefore the harmdoer is more likely to be an ally. In this environment, the probability of committing a false negative error (i.e., not forgiving an ally) is greater than that of committing a false positive (i.e., forgiving a foe). Thus, forgiving is more likely to promote fitness than not forgiving, and there should be a greater tendency towards forgiveness. The opposite conclusion will hold if trust is low. Whereas error management discussions typically focus on the *cost* of errors, the present discussion about the base rate shifts the focus to its *probability*. Both the cost and probability of errors should impact where the decision criterion is set.

1.3. The rationality of stable traits

The question, "what function does generalized trust serve?" can be formulated more broadly as, "What good are stable traits?" I will argue that stable traits, such as social trust or the tendency to forgive, are adaptive strategies that help an individual navigate uncertain environments.

One reason for seeing traits as adaptations[11] is that they are also found in many other animals. A growing body of evidence shows that primates (Freeman & Gosling, 2010; Herrmann, Hare, Cissewski, & Tomasello, 2011), sheep (Sibbald, Erhard, McLeod, & Hooper, 2009), geese (Kurvers et al., 2009), fishes, bees and many other species (Bergmüller & Taborsky, 2010) exhibit "consistent individual differences in the average behavior across time and contexts" (Dingemanse, Kazem, Réale, & Wright, 2010, p. 81). Another reason is that stable traits have consequences for many important life outcomes such as quality of relationships, level of happiness, quality of parental care etc. (Ozer & Benet-Martínez, 2005; Prinzie, Stams, Deković, Reijntjes, & Belsky, 2009), indicating that they represent different approaches to solving fitness-relevant problems.

Thus, a personality trait can be viewed as an adaptive strategy for dealing with uncertainty about *future* environments (Herrmann et al., 2011; Tooby & Cosmides, 1990; Wolf, van Doorn, Leimar, & Weissing, 2007). Different environments trigger different adaptive strategies[12], likely by interacting with the individual's existing phenotype such as gender or body size, which motivate behaviors that will be adaptive in the forecasted future environment (Buss, 2009). In this way, traits that promote an individual's fitness are those that have a good fit between the environment that it is calibrated for (i.e., the forecasted future environment) and the environment in which the trait is expressed (see also Todd, Gigerenzer, & The ABC Research Group, 2012).

[11] Not all individual differences are adaptations; some are by-products or noise resulting from the evolutionary process.

[12] This perspective, however, does not suggest determinism as individuals possess some flexibility to adjust their life-history strategies to match the current environment (Brumbach et al., 2009).

One way of characterizing how adaptive strategies are shaped is through life history theory, a framework that explains the existence of individual variation within species[13] in terms of the tradeoffs that need to be made between the allocation and capture of energy throughout their lifespan[14] (Buss, 2009; Funder, 2001; Nettle, 2006; Tooby & Cosmides, 1990). Given that energy budgets are limited, individuals need to make allocations between different activities that contribute towards their fitness. In humans, for instance, allocating energy to pursuing new romantic relationships leaves less energy to be spent on parenting efforts or gaining resources.

The tradeoffs that an individual makes across different domains cluster together and constitute a life-history strategy (LHS) that can fall on a continuum from fast to slow. Early-life environments that are harsh and unpredictable, such as those characterized by poverty and high mortality rates, tend to trigger the adoption of faster LHS (Brumbach, Figueredo, & Ellis, 2009). Individuals who adopt a fast LHS tend to allocate more resources toward social exploitation, as well as achieving sexual variety and earlier reproduction (Griskevicius, Delton, Robertson, & Tybur, 2011; Olderbak & Figueredo, 2010). In contrast, individuals who adopt a slow LHS tend to allocate energy towards maintaining relationships amongst kin, kith, and long-term partners. This approach to studying traits in humans has generated a large body of research that has linked LHS to Big 5 personality traits (Nettle, 2006), dark triad traits like Machiavellianism (Jonason, Koenig, & Tost, 2010), as well as domain-specific risk taking propensities (Jarecki & Wilke, 2015; Wang, Kruger, & Wilke, 2009).

In general, individuals that adopt a slow LHS tend to have fewer children, and invest more in each child than those with a fast LHS (Brumbach et al., 2009). Slow LHS individuals also tend to live longer and commit more to long-term relationships, as well as engage in more long-term planning (Figueredo et al., 2006). They also tend to engage in more prosocial behavior and experience greater social support (Figueredo et al., 2006). With particular relevance to the current topic, individuals pursuing a slow LHS tend to have greater social trust than those pursuing a fast LHS (Petersen & Aarøe, 2015), reiterating the role that trust plays in supporting cooperative relationships.

[13] Life history theory was originally used to explain variation between species but has recently been extended to explain variation within species (Figueredo, Vásquez, Brumbach, & Schneider, 2007).

[14] The theory also specifies how changes in life stage trigger strategy changes, such as how tradeoffs change when an individual becomes a parent (e.g., Wang et al., 2009).

2. THE PRESENT RESEARCH

The thesis of this chapter is that trait forgiveness and social trust are adaptive responses to the base rate of allies in the environment. In other words, that both of these individual differences are strategies that help manage uncertainty in the social environment. Because the fitness outcomes of the forgiveness decision are uncertain—it is unclear if forgiving will result in more fitness benefits or costs (see previous chapter for more details)—assumptions about the base-rate manages this uncertainty and promotes adaptive decisions.

I expect that individuals who are higher in trait forgiveness and social trust to be those that pursue slower LHS. This leads to the following hypotheses:

H_1: Trait forgiveness and social trust are positively related.

H_2: Greater trait forgiveness is associated with a slower LHS.

H_3: The relationship between LHS and trait forgiveness is mediated by social trust.

2.1. Method

2.1.1. Participants

One hundred twenty-one participants (51.2% female, M_{age} = 36.34 years, age range = 20–67 years) residing in the United States took part in this study. All were recruited via Amazon's Mechanical Turk and were remunerated U.S. $0.30 for their participation. We originally planned on recruiting 120 participants. All participants passed the attention checks and were included in the analyses.

2.1.2. Materials and procedure

Trait forgiveness was measured using Brown's (2003) tendency to forgive scale. The scale included 4 items measured on a scale from 1 (*completely disagree*) to 7; α (*completely agree*), with Cronbach's α = .85. Items include, "I have a tendency to harbor grudges," as well as "I tend to get over it quickly when someone hurts my feelings".

Social trust was measured using an adaptation of the items used in the General Social Survey. Six items were used and measured on a scale from 1 (*completely disagree*) to 7 (*completely agree*); α = .89. Items include, "Generally speaking, most people can be trusted," as well as "People are just looking out for themselves".

Two measurements of LHS were used, each reflecting a dominant approach in the literature. First, Figueredo et al.'s (2005) Mini-K, a psychometric scale to measure the cognitive and behavioral correlates of LHS. This scale included 20 items measured on a scale from -3 (*disagree strongly*) to 3 (*agree strongly*); α = .84. Items include, "I try to understand how I got into a situation to figure out how to

handle it," and "I am often in social contact with my friends." A higher score is indicative of a slower LHS.

Following the approach of Wang et al. (2009) as well as Hill, Ross, and Low (1997), four other life history indicators were also measured. They are gender, romantic-relationship status, parental status, as well as subjective life expectancy. In general, faster LHS tend to be adopted by individuals who are male, are in a committed romantic relationship, are not parents, or who have less expectation of living till old age. The effect of each of these indicators will be examined separately as it is unclear what their cumulative effects are.

Participants responded to the trait forgiveness, social trust, and Mini-K measures, in randomized order. They then provided information about the other life history indicators as part of the demographics, before the study was concluded.

2.2. Results

As hypothesized, trait forgiveness (M = 4.04, SD = 1.38), social trust (M = 4.38, SD = 1.09), and LHS as measured by the Mini-K (M = .93, SD = .80) were positively correlated with each other (see table 1). However, none of the life history indicators were related to trait forgiveness or social trust.

With the exception of parental status, the life history indicators were all related to the Mini-K. Individuals who had a greater subjective life expectancy were more likely to score higher on the Mini-K suggesting that they were more likely to adopt a slow LHS, r = .24, p = .008. Those in a committed romantic relationship (M = 1.17, SD = .74) scored higher on the Mini-K than those who were not (M = 1.15, SD = .75), $t(202)$ = -6.81, $p < .001$. Contrary to the typical finding in the literature, females (M = 1.20, SD = .75) had lower scores on the Mini-K than males (M = .64, SD = .74), $t(202)$ = -6.81, $p < .001$. Bonferroni corrections for multiple comparisons were done for the p-values of the t-tests.

The relationship between Mini-K and trait forgiveness was mediated by social trust (see Figure 1). The standardized indirect effect was (.36)(.40) = .14. We tested the significance of this indirect effect using bootstrapping procedures. The bootstrapped unstandardized indirect effect was .14, and the 95% confidence interval ranged from .03, .29, suggesting that the effect was non-zero.

Taken together, the results support all three hypotheses. It was hypothesized and found that trait forgiveness and social trust are positively related, suggesting that individual differences in the tendency to forgive may reflect an assumption of the base rate of allies (reflected by social trust) in the individual's environment. Furthermore, LHS as measured by the Mini-K and trait forgiveness were positively related, and this relationship is mediated by social trust, supporting the argument that social trust is associated with the adoption of a slow LHS. However, no such relationship with the other life history indicators such as life expectancy was found.

Since LHS as measured by the Mini-K is related to both social trust and trait forgiveness, this suggests that being trusting and forgiving are two traits that are clustered with those others of a slow LHS. Even though three of four of the life history indicators were related to Mini-K, one of them (gender) was in the opposite direction to that of the general trend in the literature (e.g., Figueredo et al., 2006; Hill et al., 1997; Wang et al., 2009).

Table 1.
Pearson correlation matrix

	Social trust	Mini-K
Trait forgiveness	.43 (< .001)	.28 (.002)
Social trust	-	.26 (.004)

Note. *p*-values are indicated in parentheses.

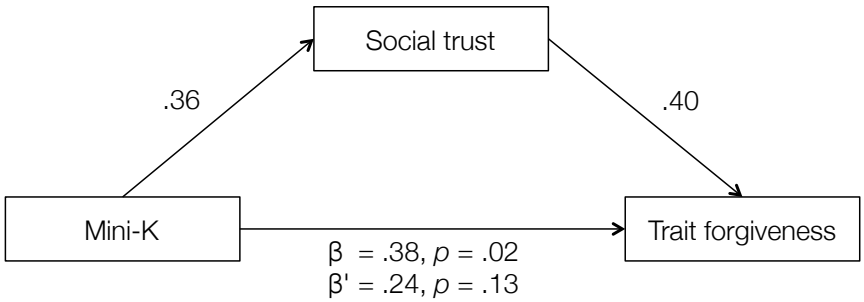

Figure 1. Mediation model of life history strategy (as measured by the Mini-K), social trust, and trait forgiveness.

3. DISCUSSION

How do individual differences in forgiveness promote fitness? This question was explored in this chapter from the perspective of the different tradeoffs that different individuals have to make. Since making tradeoffs feature in many evolutionarily recurrent tasks, it is no surprise that two prominent theories in evolutionary psychology and biology—life history theory and error management theory—specify how individuals manage tradeoffs. Whereas life history theory is concerned with how tradeoffs in energy allocation guide broad behavioral strategies (Figueredo et al., 2006; Hill et al., 1997), error management theory is concerned

with how tradeoffs in the cost of errors shape specific decisions (Haselton & Nettle, 2005; Johnson, Blumstein, Fowler, & Haselton, 2013).

The present chapter drew from both these theories to specify how individual differences in the tendency to forgive and the degree of social trust feature in the decision about whether to forgive. To the best of knowledge, this is the first to outline the connections between these two theories. From the perspective of life history theory, both of these traits support a behavioral strategy that manages uncertainty in the social environment. Uncertainty in the error management framework of forgiveness is about the fitness outcome of the decision (i.e., the net benefits or costs that result from continuing a relationship with the harmdoer), and both of these traits manage this uncertainty by assuming a certain base rate of allies in the environment. This assumption then influences whether the decision criterion should be liberal or conservative.

The results from the study support the hypotheses that trait forgiveness correlates with social trust, and that social trust mediates the relationship between life-history strategy and trait forgiveness. More broadly, this association highlights how environmental information (such as the base rate) is incorporated judiciously in social decision-making. Implications and some future directions are discussed below.

3.1. Error management framework of forgiveness and cooperation

The goal of this chapter was to incorporate individual differences in the tendency to forgive into the error management framework of forgiveness proposed in the previous chapter. Establishing the association between trait forgiveness and social trust is the first step to understanding how base rate information is implicated in the forgiveness decision. The next step is to employ the cognitive modeling techniques used in the previous chapter to examine how base rate information affects the decision criterion. If the argument is correct, then individuals who have higher social trust should be more likely to adopt a liberal criterion.

Since social trust is implicated in other decisions and behaviors related to cooperation, this suggests that information about the base rate may be used in decisions beyond forgiveness. Because many evolutionary recurrent social decisions tasks can be understood as error management tasks (e.g., Haselton & Galperin, 2013; Johnson et al., 2013), social trust may be a variable that connects many of them.

The present investigation on base rates also extends error management theory by highlighting how the decision criterion is impacted not only by error costs, but also by error *likelihood*. Many other error management tasks are likely to be similarly affected by base rate information which, depending on the task, may be indicated by psychological variables other than trust.

3.2. The controversy of base-rate neglect

Do individuals consider the base rate when making decisions? This question has been the source of debate amongst researchers interested in the rationality of cognition (Birnbaum, 1983; Gigerenzer, 1991; Gilovich, Griffin, & Kahneman, 2002; Koehler, 1996; Welsh & Navarro, 2012).

On one hand, it has been claimed that "the genuineness, the robustness, and the generality of the base-rate fallacy are matters of established fact" (Bar-Hillel, 1980, p. 215). On the other hand, there is also a large body empirical evidence that individuals, humans as well as pigeons, incorporate base rate information in their decision process (Ajzen, 1977; Fantino, Kanevsky, & Charlton, 2005; Gigerenzer, 1991; Koehler, 1996). These typically fall under two categories: those that present base-rate information in more comprehensible formats (e.g., using natural frequencies; Gigerenzer, 1991; Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000), and those where base-rates are learnt intuitively (e.g., Juslin, Wennerholm, & Winman, 2001; Manis, Dovalina, Avis, & Cardoze, 1980).

The current chapter's analysis of social trust as reflecting base rate information speaks to this controversy by proposing a third category. It suggests that base rate information is not only used in decision-making, but can be expressed as an individual difference that is adapted to a particular environment. Crucially, the focus on social trust suggests that base rate information in natural and evolutionarily recurrent tasks may be manifested as emotions or attitudes, rather than as an abstract knowledge. Furthermore, that there may be some degree of innate preparedness to learn base rates that are evolutionarily relevant (e.g., Marks & Nesse, 1994; Oaten, Stevenson, & Case, 2009).

3.3. Addressing conceptual issues in trust

The proposal to view trust as an indication of base rate may address some of the conceptual issues that beset the topic of trust (see Nannestad, 2008; Yamagishi, 2011). First, there are plural conceptions of trust; it has been conceived as a decision, an emotion, a belief, and also a behavior (e.g., Bohnet & Zeckhauser, 2004; Eckel & Wilson, 2004; Evans & Krueger, 2014; Jones, 1996; Sztompka, 1999). Second, there have also been debates about whether trusting or being trustworthy is the basis of cooperation (e.g., Hardin, 2002; Putnam, 2001), as well as confusion about the particularized and generalized kinds of trust mentioned earlier. There is also disagreement about the normativity of trust such as whether it is rational or whether high trusters are gullible (e.g., Schlenker, Helm, & Tedeschi, 1973; Yamagishi & Kikuchi, 1999). To date, there is no overarching theory that can provide a principled approach to resolving these controversies (Nannestad, 2008).

From the perspective of the error management framework, social trust is more likely to be an emotion or a belief that influences decision and/or behavior. Being trusting makes it more likely that an individual will engage in cooperative actions

with another agent, but whether this generates net fitness benefits depends on whether the other agent is trustworthy. The social trust as base-rate proposition suggests that being trusting is the result of a rational assessment that other agents in the environment are trustworthy, suggesting that both are likely to correlate on the aggregate level. This can thus potentially explain why it is difficult to disentangle the two empirically (e.g., Hardin, 2002). Lastly, since particularized trust concerns a specific agent (e.g., "John"), it can be taken as a component of the strength of evidence that the agent is an ally. When social (generalized) trust is high, the decision criterion is likely to be liberal and John is likely to be forgiven even if evidence strength is not high (i.e., the individual does not trust John very much). This example demonstrates the distinct role that each kind of trust plays in the forgiveness decision.

3.5. Conclusion

What good are stable individual differences in forgiveness and trust? This question has been asked in this chapter and one answer has been offered: These traits reflect assumptions about the social environment and they help an individual manage the uncertainty of the outcomes of social decisions. Having a greater trait forgiveness and a higher level of social trust motivates the individual to repair relationships following conflict with the expectation that stable long-term relationships will be fitness-enhancing.

References

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology*, *35*(5), 303–314. doi:10.1037/0022-3514.35.5.303

Balliet, D., & Lange, P. a. M. Van. (2013). Trust, punishment, and cooperation across 18 societies: A meta-analysis. *Perspectives on Psychological Science*, *8*(4), 363–379. doi:10.1177/1745691613488533

Balliet, D., Li, N. P., & Joireman, J. (2011). Relating trait self-control and forgiveness within prosocials and proselfs: Compensatory versus synergistic models. *Journal of Personality and Social Psychology*, *101*(5), 1090–1105. doi:10.1037/a0024967

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*(3), 211–233. doi:10.1016/0001-6918(80)90046-3

Bateson, P., & Laland, K. N. (2013). Tinbergen's four questions: An appreciation and an update. *Trends in Ecology and Evolution*, *28*(12), 712–718. doi:10.1016/j.tree.2013.09.013

Bekkers, R. (2012). Trust and volunteering: Selection or causation? Evidence From a 4 year panel study. *Political Behavior*, *34*, 225–247. doi:10.1007/s11109-011-9165-x

Bergmüller, R., & Taborsky, M. (2010). Animal personality due to social niche specialisation. *Trends in Ecology and Evolution*, *25*(9), 504–511. doi:10.1016/j.tree.2010.06.012

Birnbaum, M. H. (1983). Base rates in Bayesian inference : Signal detection analysis of the cab problem. *The American Journal of Psychology*, *96*(1), 85–94. doi:10.2307/1422211

Bjørnskov, C. (2007). Determinants of generalized trust: A cross-country comparison. *Public Choice*, *130*(1-2), 1–21. doi:10.1007/s11127-006-9069-1

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior and Organization*, *55*, 467–484. doi:10.1016/j.jebo.2003.11.004

Boon, S. D., & Sulsky, L. M. (1997). Attributions of blame and forgiveness in romantic relationships: A policy-capturing study. *Journal of Social Behaviour and Personality*, *12*(1), 19–44.

Brown, R. P. (2003). Measuring individual differences in the tendency to forgive: Construct validity and links with depression. *Personality and Social Psychology Bulletin*, *29*(6), 759–771. doi:10.1177/0146167203252882

Brumbach, B. H., Figueredo, A. J., & Ellis, B. J. (2009). Effects of harsh and unpredictable environments in adolescence on development of life history strategies. *Human Nature*, *20*, 25–51. doi:10.1007/s12110-009-9059-3

Buss, D. M. (2009). How can evolutionary psychology successfully explain personality and individual differences? *Perspectives on Psychological Science*, *4*(4), 359–366. doi:10.1111/j.1745-6924.2009.01138.x

Carl, N., & Billari, F. C. (2014). Generalized trust and intelligence in the United States. *PLoS ONE*, *9*(3), 1–10. doi:10.1371/journal.pone.0091786

Delhey, J., & Newton, K. (2005). Predicting cross-national levels of social trust: Global pattern or Nordic exceptionalism? *European Sociological Review*, *21*(4), 311–327. doi:10.1093/esr/jci022

Dingemanse, N. J., Kazem, A. J. N., Réale, D., & Wright, J. (2010). Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology and Evolution*, *25*(2), 81–89. doi:10.1016/j.tree.2009.07.013

Dufwenberg, M., & Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior*, *30*, 163–182. doi:DOI: 10.1006/game.1999.0715

Eckel, C. C., & Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior and Organization*, *55*, 447–465. doi:10.1016/j.jebo.2003.11.003

Elgar, F. J., & Aitken, N. (2011). Income inequality, trust and homicide in 33 countries. *European*

Journal of Public Health, *21*(2), 241–246. doi:10.1093/eurpub/ckq068

Evans, A. M., & Krueger, J. I. (2014). Outcomes and expectations in dilemmas of trust. *Judgment and Decision Making*, *9*(2), 90–103.

Fantino, E., Kanevsky, I. G., & Charlton, S. R. (2005). Teaching pigeons to commit base-rate neglect. *Psychological Science*, *16*(10), 820–825.

Fehr, R., Gelfand, M. J., & Nag, M. (2010). The road to forgiveness: A meta-analytic synthesis of its situational and dispositional correlates. *Psychological Bulletin*, *136*(5), 894–914. doi:10.1037/a0019993

Figueredo, A. J., Vásquez, G., Brumbach, B. H., & Schneider, S. M. R. (2007). The k-factor, covitality, and personality: A psychometric test of life history theory. *Human Nature*, *18*(1), 47–73. doi:10.1007/BF02820846

Figueredo, A. J., Vásquez, G., Brumbach, B. H., Schneider, S. M. R., Sefcek, J. A., Tal, I. R., … Jacobs, W. J. (2006). Consilience and life history theory: From genes to brain to reproductive strategy. *Developmental Review*, *26*, 243–275. doi:10.1016/j.dr.2006.02.002

Figueredo, A. J., Vásquez, G., Brumbach, B. H., Sefcek, J. A., Kirsner, B. R., & Jacobs, W. J. (2005). The K-factor: Individual differences in life history strategy. *Personality and Individual Differences*, *39*(8), 1349–1360. doi:10.1016/j.paid.2005.06.009

Fincham, F. D. (2000). The kiss of the porcupines: From attributing responsibility to forgiving. *Personal Relationships*, *7*(7), 1–23. doi:10.1111/j.1475-6811.2000.tb00001.x

Freeman, H. D., & Gosling, S. D. (2010). Personality in nonhuman primates: A review and evaluation of past research. *American Journal of Primatology*, *72*(8), 653–671. doi:10.1002/ajp.20833

Funder, D. C. (2001). Personality. *Annual Review of Psychology*, *52*, 197–221. doi:10.1146/annurev.psych.52.1.197

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases.” *European Review of Social Psychology*, *2*(1), 83–115. doi:10.1080/14792779143000033

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, UK: Cambridge University Press. doi:10.5465/AMR.2004.14497675

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Griskevicius, V., Delton, A. W., Robertson, T. E., & Tybur, J. M. (2011). Environment Contingency in Life History Strategies: The Influence of Mortality and Socioeconomic Status on Reproductive Timing. *Journal of Personality and Social Psychology*, *100*(6), 1015–26. doi:10.1037/a0022403

Hardin, R. (2002). *Trust and trustworthiness*. New York: Russell Sage Foundation.

Haselhuhn, M. P., Kennedy, J. A., Kray, L. J., Van Zant, A. B., & Schweitzer, M. E. (2015). Gender differences in trust dynamics: Women trust more than men following a trust violation. *Journal of Experimental Social Psychology*, *56*, 104–109. doi:10.1016/j.jesp.2014.09.007

Haselton, M. G., & Galperin, A. (2013). Error management in relationships. In J. Simpson & L. Campbell (Eds.), *The Oxford Handbook of Close Relationships* (pp. 234–254). Oxford University Press, USA. doi:10.1093/oxfordhb/9780195398694.001.0001

Haselton, M. G., & Nettle, D. (2005). The paranoid optimist: An integrative evolutional model of cognitive biases. *Personality and Social Psychology Review*, *10*(1), 47–66. doi:10.1207/s15327957pspr1001_3

Helliwell, J. F., Huang, H., & Wang, S. (2014). Social capital and well-being in times of crisis. *Journal of Happiness Studies*, *15*, 145–162. doi:10.1007/s10902-013-9441-z

Herrmann, E., Hare, B., Cissewski, J., & Tomasello, M. (2011). A comparison of temperament in nonhuman apes and human infants. *Developmental Science*, *14*(6), 1393–1405. doi:10.1111/j.1467-7687.2011.01082.x

Hill, E. M., Ross, L. T., & Low, B. S. (1997). The role of future unpredictability in human risk-taking. *Human Nature*, *8*(4), 287–325. doi:10.1007/BF02913037

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, *290*, 2261–2262.

Jarecki, J., & Wilke, A. (2015). *Tracing the processes for evolved risk responses*. Berlin.

Johnson, D. D. P., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology and Evolution*, *28*(8), 474–481. doi:10.1016/j.tree.2013.05.014

Jonason, P. K., Koenig, B. L., & Tost, J. (2010). Living a fast life: The dark triad and life history theory. *Human Nature*, *21*(4), 428–442. doi:10.1007/s12110-010-9102-4

Jones, K. (1996). Trust as an affective attitude. *Ethics*, *107*(1), 4–25. doi:10.1086/233694

Juslin, P., Wennerholm, P., & Winman, A. (2001). High-level reasoning and base-rate use: Do we need cue-competition to explain the inverse base-rate effect? *Journal of Experimental Psychology-Learning Memory and Cognition*, *27*(3), 849–871. doi:10.1037/0278-7393.27.3.849

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, *19*, 1–53. doi:10.1017/S0140525X00041157

Kurvers, R. H. J. M., Eijkelenkamp, B., van Oers, K., van Lith, B., van Wieren, S. E., Ydenberg, R. C., & Prins, H. H. T. (2009). Personality differences explain leadership in barnacle geese. *Animal Behaviour*, *78*(2), 447–453. doi:10.1016/j.anbehav.2009.06.002

Manis, M., Dovalina, I., Avis, N. E., & Cardoze, S. (1980). Base rates can affect individual predictions. *Journal of Personality and Social Psychology*, *38*(2), 231–248. doi:10.1037//0022-3514.38.2.231

Marks, I. fM., & Nesse, R. M. (1994). Fear and fitness: An evolutionary analysis of anxiety disorders. *Ethology and Sociobiology*, *15*(5-6), 247–261. doi:10.1016/0162-3095(94)90002-7

McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington, E. L., Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *Journal of Personality and Social Psychology*, *75*(6), 11–12.

Molden, D. C., & Finkel, E. J. (2010). Motivations for promotion and prevention and the role of trust and commitment in interpersonal forgiveness. *Journal of Experimental Social Psychology*, *46*(2), 255–268. doi:10.1016/j.jesp.2009.10.014

Nannestad, P. (2008). What have we learned about generalized trust, if anything? *Annual Review of Political Science*, *11*(1), 413–436. doi:10.1146/annurev.polisci.11.060606.135412

Nettle, D. (2006). The evolution of personality variation in humans and other animals. *American Psychologist*, *61*(6), 622–631. doi:10.1037/0003-066X.61.6.622

Oaten, M., Stevenson, R. J., & Case, T. I. (2009). Disgust as a disease-avoidance mechanism. *Psychological Bulletin*, *135*(2), 303–321. doi:10.1037/a0014823

Olderbak, S. G., & Figueredo, A. J. (2010). Life history strategy as a longitudinal predictor of relationship satisfaction and dissolution. *Personality and Individual Differences*, *49*(3), 234–239. doi:10.1016/j.paid.2010.03.041

Ozer, D. J., & Benet-Martínez, V. (2005). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, *57*, 401–421. doi:10.1146/annurev.psych.57.102904.190127

Petersen, M. B., & Aarøe, L. (2015). Birth weight and adult social trust: Evidence for early calibration of social cognition. *Psychological Science*, *26*(11), 1681–1692.

Prinzie, P., Stams, G. J. J. M., Deković, M., Reijntjes, A. H. A., & Belsky, J. (2009). The relations between parents' Big Five personality factors and parenting: a meta-analytic review. *Journal of Personality and Social Psychology*, *97*(2), 351–62. doi:10.1037/a0015823

Putnam, R. D. (2001). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.

Schlenker, B. R., Helm, B., & Tedeschi, J. T. (1973). The effects of personality and situational variables on behavioral trust. *Journal of Personality and Social Psychology*, *25*(3), 419–427.

doi:10.1037/h0034088

Sibbald, A. M., Erhard, H. W., McLeod, J. E., & Hooper, R. J. (2009). Individual personality and the spatial distribution of groups of grazing animals: An example with sheep. *Behavioural Processes*, *82*(3), 319–326. doi:10.1016/j.beproc.2009.07.011

Sztompka, P. (1999). *Trust: A sociological theory*. Cambridge University Press.

Tinbergen, N. (1963). On aims and methods of Ethology. *Zietschrift Für Tierpsychologie*, *20*, 410–433. doi:10.1163/157075605774840941

Todd, P. M., Gigerenzer, G., & The ABC Research Group. (2012). *Ecological rationality: Intelligence in the world*. New York: Oxford University Press.

Tooby, J., & Cosmides, L. (1990). On the universality of human nature and the uniqueness of the individual: the role of genetics and adaptation. *Journal of Personality*, *58*(1), 17–67. doi:10.1111/j.1467-6494.1990.tb00907.x

Uslaner, E. M. (2004). Trust and corruption. In J. G. Lambsdorf, M. Taube, & M. Schramm (Eds.), *Corruption and the New Institutional Economics*. London: Routledge.

Wang, X. T. (Xiao-T., Kruger, D. J., & Wilke, A. (2009). Life history variables and risk-taking propensity. *Evolution and Human Behavior*, *30*(2), 77–84. doi:10.1016/j.evolhumbehav.2008.09.006

Welsh, M. B., & Navarro, D. J. (2012). Seeing is believing: Priors, trust, and base rate neglect. *Organizational Behavior and Human Decision Processes*, *119*(1), 1–14. doi:10.1016/j.obhdp.2012.04.001

Wieselquist, J. (2009). Interpersonal forgiveness, trust, and the investment model of commitment. *Journal of Social and Personal Relationships*, *26*(4), 531–548. doi:10.1177/0265407509347931

Wolf, M., van Doorn, G. S., Leimar, O., & Weissing, F. J. (2007). Life-history trade-offs favour the evolution of animal personalities. *Nature*, *447*(7144), 581–4. doi:10.1038/nature05835

Yamagishi, T. (2011). Trust as social intelligence. In *Trust: The Evolutionary Game of Mind and Society*. doi:10.1007/978-4-431-53936-0

Yamagishi, T., & Kikuchi, M. (1999). Trust, gullibility, and social intelligence. *Asian Journal of Social Psychology*, *2*(1), 145–161. doi:10.1111/1467-839X.00030

Acknowledgements

This dissertation is the result of research I carried out as a Predoctoral Fellow at the Center for Adaptive Behavior and Cognition (ABC), at the Max Planck Institute for Human Development.

ABC is truly research paradise. I have been lucky to benefit from the support, encouragement, and advice from the members and visitors of the group. I have gained from the time to think and the freedom to question, and I will continue to reap the rewards for the years to come. Witnessing the intellectual rigor upheld at the group has set the standard for how I think science should be conducted and the kind of scientist that I aspire to be.

I would especially like to thank my amazing advisors, Shenghua and Konstantinos. Shenghua, for his unwavering support and friendship; discussions with him have sharpened my ordinary ideas and made them better. Konstantinos, for his big-picture perspective and enthusiasm about my work; having him as a sounding board has been invaluable.

I would also like to thank Gerd and Lael, who moved administrative mountains so that I could be at ABC. I hope that I lived up to the faith that was placed in me. More importantly, I would like to thank them for inspiring me and for being research role models.

ABC would not be the same without the wonderful friends that I have made here—Stojan, Wasilios, Jana, Mirjam, Perke, Elisabeth, Astrid, Özgür, Amit, Henry, Niklas, Hanna, Martin, Monika, Michelle, Roman, Daniel, Christian… and so many more. Our wonderful meandering conversations have been precious gifts.

Finally, I *need* to thank my parents, Lynn and Bernard. For giving me life and the courage to pursue my passions. For their unconditional love and support throughout everything.

Jolene
February 2016

Erklärung zur Dissertation

Hiermit versichere ich, dass ich die vorgelegte Arbeit selbständig verfasst habe. Andere als die angegebenen Hilfsmittel habe ich nicht verwendet. Die Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Jolene Tan

_____

Berlin, February 2016