



University of Pennsylvania
ScholarlyCommons

Publicly Accessible Penn Dissertations

2016

A Practical Theory of Promising

Molly Beth Sinderbrand

University of Pennsylvania, molsi@sas.upenn.edu

Follow this and additional works at: <https://repository.upenn.edu/edissertations>



Part of the [Ethics and Political Philosophy Commons](#)

Recommended Citation

Sinderbrand, Molly Beth, "A Practical Theory of Promising" (2016). *Publicly Accessible Penn Dissertations*. 2019.

<https://repository.upenn.edu/edissertations/2019>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/edissertations/2019>
For more information, please contact repository@pobox.upenn.edu.

A Practical Theory of Promising

Abstract

In this dissertation, I defend a “social-rights” account of promises. I argue that the essential common feature of promising practices is the transfer of a right to a future action from the promisor to the promisee. The transfer of the right relies on a social practice – specifically, a social norm. Using empirical evidence, I show that the practice of promising fits the definition of a social norm. I then use the theory of social norms to investigate why we are motivated to keep promises. The fact that a single theory can describe both normative foundations and motivation gives credence to the theory: the data on behavioral motivation supports the metaethical theory, and the metaethical theory can help to explain the motivation. My theory also has practical and theoretical applications; I focus on two particular cases. On the theoretical side, I discuss implications for contemporary social contract theories. On the practical side, I evaluate successful and unsuccessful efforts to eliminate female genital cutting (FGC) in various parts of Africa that rely on promises. My theory can explain why these programs are successful in certain circumstances, but not in others. Thus, my theory can explain and inform the creation and implementation of behavioral interventions.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Philosophy

First Advisor

Cristina Bicchieri

Keywords

Ethics, Promising, Social Norms

Subject Categories

Ethics and Political Philosophy | Philosophy

A PRACTICAL THEORY OF PROMISING

Molly Beth Sinderbrand

A DISSERTATION

in

Philosophy

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

Cristina Bicchieri

S. J. Patterson Harvie Professor of Social Thought and Comparative Ethics

Graduate Group Chairperson

Samuel Freeman, Avalon Professor of the Humanities and Graduate Chair in Philosophy

Dissertation Committee

Alexander Guerrero, Assistant Professor of Philosophy, University of Pennsylvania

Adrienne Martin, Akshata Murty '02 and Rishi Sunak Associate Professor of Philosophy, Politics,

and Economics and George R. Roberts Fellow, Claremont McKenna College

A PRACTICAL THEORY OF PROMISING

COPYRIGHT

2016

Molly Beth Sinderbrand

This work is licensed under the
Creative Commons Attribution-
NonCommercial-ShareAlike 3.0
License

To view a copy of this license, visit

<http://creativecommons.org/licenses/by-nc-sa/3.0/>

For Jocelyn

ACKNOWLEDGMENT

There are many individuals who, in various ways, contributed to this dissertation. I would first like to thank Cristina Bicchieri for taking me on as a graduate student, making this project possible, and for all she has done to promote my philosophical and professional development. I am deeply indebted to her for her advice and guidance in research and in life. Adrienne Martin also helped me considerably in forming and reforming my topic, and made sure I started my research and writing with rigor, and in a timely fashion. Alex Guerrero provided excellent comments and critique, and often pushed me to take different philosophical perspectives to make my work stronger.

I owe an incredible amount of gratitude to my friends and colleagues Thomas Noah and Aditi Chaturvedi. Thomas provided consistently useful and insightful comments, criticism, and general discussion. His collaboration and discussion on related projects helped me to hone my ideas for this one as well. Aditi's perspective, ideas, and philosophical support have helped me work through many ideas in this dissertation as well. Both of their friendship has been invaluable throughout this process.

I would also like to thank the Mekor Habracha community and the Graduate Student Network for allowing me opportunities to discuss and present my work to non-philosophers, and more making my time as a graduate student enjoyable. Many members of these communities have helped me refine my thoughts and ideas, supported me through the dissertation process, and provided new and interesting

perspectives. In particular, I am indebted to Mirah Kriger and Sonia Gollance, two of my favorite study partners.

I could not have completed this dissertation without the unending support of Sandra Natson, Sarah Gish-Kraus, and Gerri Winters. Sandy's positive outlook on life, stories, and sense of humor have made graduate school a pleasant experience, and I always appreciate our chats. Sarah's intelligence, organization, and seemingly limitless compassion and desire to help have come to my rescue numerous times. No matter how busy, she always seems to make time, even just for a quick hello. Lastly, Gerri's friendship has been a continual source of support throughout my graduate career. She has made my time at Penn a positive experience, and I am grateful for that.

Finally, I would like to thank my family, without whom I would literally not be here. This includes the entire Sumi clan, headed by my role model and matriarch Jean Sumi, and, of course, my east-coast family, Janet and David Sinderbrand and Thelma Sinderbrand. My parents, Maryann Sumi and Carl Sinderbrand, have supported me through all of my endeavors, academic and otherwise, and were always eager to discuss my work. No one could ask for better parents and role models. My siblings, Ian Calef and Jake and Sarah Sinderbrand, have provided much needed emotional support as well, in many different forms. I am so grateful that I am their sister. My in-laws, Marcy and Moshe Gerstenhaber, as well as Lee and Michael Gerstenhaber, have also helped me through this long process of graduate

school by providing interesting perspectives and discussions, and emotional support. I'm happy – and lucky – to be part of the Gerstenhaber family as well.

Most importantly, I am grateful to Jonathan Gerstenhaber for being my partner in all things, for challenging me and pushing me to grow as a person and as a philosopher, and for his love and support. He made this dissertation possible, not to mention survivable. And, of course, I would like to thank Jocelyn Gerstenhaber for providing me with reasons and motivation to finish this dissertation. It is dedicated to her in honor of all of the promises I will make to her.

ABSTRACT

A PRACTICAL THEORY OF PROMISING

Molly B. Sinderbrand

Cristina Bicchieri

In this dissertation, I defend a “social-rights” account of promises. I argue that the essential common feature of promising practices is the transfer of a right to a future action from the promisor to the promisee. The transfer of the right relies on a social practice – specifically, social norms. Using empirical evidence, I show that the practice of promising is supported by social norms. I then use the theory of social norms to investigate why we are motivated to keep promises. The fact that a single theory can describe both normative foundations and motivation gives credence to the theory: the data on behavioral motivation supports the metaethical theory, and the metaethical theory can help to explain the motivation. My theory also has practical applications. In the final chapter, I evaluate successful and unsuccessful efforts to eliminate female genital cutting (FGC) in various parts of Africa that rely on promises. My theory can help explain why these programs are successful in certain circumstances, but not in others. Thus, my theory can explain and inform the creation and implementation of behavioral interventions.

TABLE OF CONTENTS

| | |
|---|------------|
| ACKNOWLEDGMENT | IV |
| ABSTRACT..... | VII |
| INTRODUCTION | 1 |
| I. MOTIVATIONS..... | 1 |
| <i>I.A. EMPIRICAL MOTIVATION</i> | <i>2</i> |
| <i>I.B. NORMATIVE MOTIVATION</i> | <i>4</i> |
| <i>I.C. PRACTICAL MOTIVATION</i> | <i>5</i> |
| II. ASSUMPTIONS AND METHODOLOGY | 5 |
| III. OUTLINE OF CHAPTERS | 8 |
| CHAPTER 1: WHAT ARE YOU DOING WHEN YOU MAKE A PROMISE AND MEAN IT? | 11 |
| I. INTRODUCTION | 11 |
| II. THE SOCIAL PRACTICE VIEW | 13 |
| <i>II.A. SEARLE AND LEWIS</i> | <i>13</i> |
| <i>II.B. RAWLS.....</i> | <i>18</i> |
| <i>II.C. CREATING REASONS AND WILLING AN OBLIGATION (SHEINMAN AND HART)</i> | <i>21</i> |
| III. THE AUTHORITY-TRANSFER VIEW | 27 |
| <i>III.A. OWENS AND HOOKER</i> | <i>27</i> |
| <i>III.B. AUTHORITY AND SOCIAL PRACTICES.....</i> | <i>30</i> |
| <i>III.C. REDUCING RAWLS AND ADOPTING AUTHORITY-TRANSFER.....</i> | <i>36</i> |
| IV. THE ASSURANCE VIEW | 40 |
| <i>IV.A. NARVESON AND THE UTILITY OF EXPECTATIONS</i> | <i>41</i> |
| <i>IV.B. ROSS, SCANLON, AND UTILITARIAN CIRCULARITY AND AMBIGUITY.....</i> | <i>46</i> |
| <i>IV.C. THOMSON AND RIGHTS</i> | <i>53</i> |
| V. CONCLUSION: INDEPENDENT MORAL PRINCIPLES? | 56 |
| CHAPTER 2: SOCIAL RIGHTS, MORAL RIGHTS, AND MORALIZED SOCIAL RIGHTS | 58 |
| I. INTRODUCTION | 58 |

| | |
|--|----------------|
| II. WHAT DOES IT MEAN TO HAVE A RIGHT? | 59 |
| III. SOCIAL NORMS AND SOCIAL RIGHTS | 69 |
| IV. PROMISING IS SUPPORTED BY SOCIAL NORMS... MOSTLY | 75 |
| V. IS PROMISING SUPPORTED BY MORAL NORMS? | 83 |
| <i>V.A. IN-PRINCIPLE UNCONDITIONALITY</i> | 84 |
| <i>V.B. UNIVERSALITY: APPLICATION AND UBIQUITY</i> | 91 |
| VI. MORALIZATION | 100 |
| <i>VI.A. TWO ANALOGIES: ALLEGIANCE AND MODESTY</i> | 104 |
| CHAPTER 3: RIGHTS-BASED MOTIVATION FOR PROMISE-KEEPING | 110 |
| I. INTRODUCTION | 110 |
| II. PROMISES ARE KEPT AND BREACHES ARE PUNISHED | 111 |
| <i>II.A. KEEPING PROMISES IN COMMONS DILEMMAS AND TRUST GAMES</i> | 112 |
| <i>II.B. PUNISHMENT OF BREACH</i> | 119 |
| <i>II.C. CONCLUSIONS FROM EXPERIMENTS</i> | 127 |
| III. EXPLANATIONS OF PROMISING | 128 |
| <i>III.A. EXPECTATIONS AND GUILT-AVERSION</i> | 129 |
| <i>III.B. COMMITMENTS</i> | 140 |
| <i>III.C. SOCIAL NORMS AND RIGHTS</i> | 146 |
| CHAPTER 4: APPLYING A THEORY OF PROMISES TO SOCIAL NORM CHANGE | 150 |
| INTRODUCTION | 150 |
| WHEN IS A PROMISE NOT A PROMISE? A SOCIAL-RIGHTS VIEW OF PROMISING | 153 |
| <i>CONTENT: WHAT DO PROMISSORY RIGHTS APPLY TO?</i> | 158 |
| <i>APPLICATION: WHEN DO PROMISSORY RIGHTS APPLY, AND TO WHOM?</i> | 159 |
| <i>MECHANISM AND STRENGTH: HOW DO PROMISSORY RIGHTS APPLY?</i> | 161 |
| <i>EXISTENCE: DOES PROMISING APPLY?</i> | 164 |
| THE NORM CHANGE PROCESS AND THE IMPORTANCE OF PUBLIC COMMITMENTS | 167 |
| <i>THE NORM-CHANGE PROCESS</i> | 168 |
| <i>EXAMPLES OF PUBLIC COMMITMENTS: THE NEED FOR TRUST</i> | 174 |
| PUBLIC COMMITMENTS AS PROMISES: PUTTING THE THEORIES TOGETHER | 178 |
| <i>PUBLIC COMMITMENTS ARE PROMISES, AND THAT IS WHY THEY WORK</i> | 179 |
| <i>CONCLUSIONS: USING PROMISES IN PRACTICE</i> | 188 |
| CONCLUSION | 192 |
| PROMISING FOR SOCIAL CHANGE | 192 |

| | |
|---|------------|
| PROMISING AND POLITICAL PHILOSOPHY | 193 |
| PROMISING AND RELATED PHENOMENA | 195 |
| CLOSING THOUGHTS..... | 196 |
| REFERENCES..... | 197 |

INTRODUCTION

This dissertation seeks to give an account of promising that will inform both normative discourses and behavioral research and interventions concerning promises and contracts. I present a theory of promising that not only describes the normative basis of the obligation to keep a promise, but also describes the behavioral motivation behind promising. It thus exists at the intersection of several fields and sub-fields, including behavioral economics and decision theory, moral psychology, and philosophy, and in particular social and political philosophy, applied ethics, metaethics, and normative ethics.

I call this theory of promising a “social-rights” account because it states that promising is essentially a transfer of social rights: rights supported by social norms. But before I begin presenting the theory itself, I want to give a short introduction to the motivations and assumptions behind the project, as well as an overview of the methodology and argument structure.

I. MOTIVATIONS

The impetus for creating a theory of promising that can explain empirical data and provide a basis for normative theorizing about promising is threefold. First, the empirical data on promissory behavior runs counter to rational choice-based predictions, and, as I will show in chapter three, the current theories that purport to explain the data are inadequate. Second, on the normative side, much of the debate about promising ignores or disregards empirical work on promising.

This leaves room for a theory of promising that can take work in the social sciences into account while still informing normative theory. Lastly, I believe that an appropriately explanatory and predictive theory of promising can be used to change other harmful practices.

I.A. EMPIRICAL MOTIVATION

I began this project because I wanted to explain the apparently (though not actually) counter-preferential behavior that often occurs in strategic game experiments in which players are allowed to promise. Take, for instance, a basic one-shot trust game with promising. In a trust game, one player acts as the investor and one acts as the trustee. The investor may choose to give her allotment of money to the trustee. If she does so, the trustee may “invest” the money to multiply it by some amount. He may then split the (much larger) return, allowing both parties to profit, or keep the entire amount. The players are generally anonymous, or will not play or meet again.

In this one-shot game, assuming that players are trying to maximize their expected monetary utility, we would expect the investor to simply keep the money. Her reasoning would be simple: if she invests, the trustee, trying to maximize his utility, will keep all of the money, and she will end up with nothing. Something is better than nothing, so she should not invest. She will not see him again, so she does not have to worry about a social sanction, like annoyance or awkwardness, for not giving him anything.

We would also expect that a promise from the trustee that he will split the return evenly would have no effect on the investor's behavior. Since she will not see him again, she has no way of enforcing that promise. Hence, she would not believe his promise, and would not invest. But experimental data shows that, when players are allowed to promise, trustees promise, investors believe them, and trustees follow through. This result runs counter to the prediction from rational choice theory. And this data is robust: promises are made and kept in other types of behavioral games,¹ across various communication methods, and even when interacting with automated players.²

We may be tempted to explain this result with an error theory, instead of creating a new theory of promising. When subjects exchange promises, it "places agents in a familiar context... where people who make promises tend to keep them." Subjects will expect others to comply, even without an enforcement mechanism, out of something like habit. They also receive contextual cues – like gestures and tone – that support the judgment that the promises are credible, or that others are trustworthy.³ Subjects are simply mistaken in making, trusting, and fulfilling these promises.

¹ See, for instance, the work of Dawes et al. 1977 on social dilemmas. Communication allows for promising, which in turn increases cooperation.

² Bicchieri et al. 2010.

³ Bicchieri et al. 2010, 128.

But even this error theory requires a background theory of promising that tells us how promises create this habit of trust, and why subjects do not correct their errors in strategic contexts in which they are allowed time to contemplate the best course of action. Hence, we still need a theory of promising that can explain why promises are so powerful, even if their power is just creating errors. And, as I will argue, though there have been several theories that have tried to explain these data,⁴ none of them have been sufficient.

I.B. NORMATIVE MOTIVATION

An important, and sometimes ignored, aspect of creating a theory that can explain a phenomenon is to have a clear idea of what that phenomenon is. Hence, a theory of promising that can explain experimental data must also explain what a promise is and what it does. To answer this question, I looked to the normative ethical and metaethical literature on promising. However, in exploring this literature, I noticed that there was almost no interaction between the metaphysical and metaethical accounts of promising and the experimental data and explanations of promising.⁵

⁴ See, for example, Vanberg 2008, Ederer and Stremitzer 2013, and Charness and Dufwenberg 2006.

⁵ Most of the debate in the normative ethical (and metaethical) literature revolves around a debate between conventionalism and expectationalism. Essentially, the disagreement is about whether conventions are necessary for there to exist a duty to keep one's promise. But it does not address whether conventions are actually used, or whether promises are kept, even if they should be. See, for example, Shiffrin 2008.

Thus, by creating a theory of promising that takes both the normative and the descriptive into account, I am hoping to bridge the gap between them, and allow them to reinforce each other. The fact that a single theory can describe both normative foundations and motivation gives credence to the theory: the data on behavioral motivation supports the metaethical theory, and the metaethical theory can help to explain the motivation. Metaethical theory can help to tell us what a promise is, which helps explain why promises are kept. Experimental data can help to determine under which circumstances a promise can be made, and how to induce individuals to keep their promises and act ethically.

I.C. PRACTICAL MOTIVATION

Through working with UNICEF program developers and NGO workers in a summer UNICEF course on social norms at the University of Pennsylvania, I learned of another use of promises: changing harmful social norms. There are several successful intervention programs that use group promises to change harmful practices, like female genital cutting or child marriage. However, there are other interventions that use promises that are less successful.⁶ By creating a theory of promises that can explain why promises are sometimes successful and sometimes not, we may be able to design better interventions, or at least better predict when a promise would work and when it would not.

II. ASSUMPTIONS AND METHODOLOGY

⁶ See Mackie and LeJeune 2009.

There are three basic methodological assumptions underlying this dissertation. First, it is unapologetically naturalistic; second, it is behavioristic; and third, it is individualistic. I will explain each of these assumptions.

I assume naturalism in both the ethical and the metaphysical sense. As a methodology, naturalism states that we can learn about ethics and metaphysics by studying the natural world, using the methods of natural and social science. In other words, to explain phenomena, ethical and otherwise, we do not need to posit any entities beside those that exist in the natural world, or are reducible to these natural entities.

In ethics, the argument against using naturalism indiscriminately is that it is invalid to derive any normative claim from purely descriptive premises. This is the naturalistic fallacy, often attributed to Hume, but proposed more recently by G.E. Moore. However, we can still use a somewhat naturalistic methodology and use descriptive premises to inform our normative conclusions without committing the naturalistic fallacy. We must simply admit normative premises as well. In this dissertation, I am only committed to naturalism in this weaker sense. Normative conclusions can be informed by natural facts without being reducible to natural facts.

There is a second sense in which I assume a naturalist methodology. Instead of studying promising in a normative sense – what we should do with regard to promises – I am studying it in a descriptive or behaviorist sense. That is, I am

exploring what individuals actually do with regard to promises, not what they should do. Although I do believe that what individuals actually do might inform our theories about what they should do in some ways, I do not believe that these questions are reducible to each other. I am simply studying an ethical phenomenon in the anthropologist's sense of "ethics", not the philosopher's. It may perhaps be called "behavioral ethics", as opposed to normative ethics.

This point about behavioral ethics relates to my other methodological assumptions, behaviorism and individualism. Individualism is the belief that the basic unit of analysis is the individual, not entire social entities like communities or nations. This is not to say that social entities are not important, or that an individualist cannot use anthropological or sociological data that concerns social entities. According to early economist Carl Menger, this "exact orientation" of research "does not deny the unity" of social entities. "It tries, rather, to explain the origins and functions of these unified structures in an exact way, to explain how these 'real entities' come about and how they function."⁷ The idea is only that these social entities explainable or analyzable in terms of individuals and individual behavior. Individualism is what grounds the use of data about individuals and their psychological motivation to study complex social phenomena.

⁷ Menger 1985, 143

Behaviorism is the belief that we can study social and psychological phenomena by looking at behavior.⁸ It is what grounds the use of observations of behavior, either in experiments or in the field, to study psychological motivation. Since both individualism and behaviorism these have long pedigrees in the social sciences,⁹ I will assume both of them for my arguments in this dissertation.

III. OUTLINE OF CHAPTERS

The dissertation is comprised of four chapters: the first two contain arguments for the two elements of the theory, and the last two provide independent support for accepting the theory. The first chapter shows that promises are rights-based. The second shows that these rights are specifically social rights, or that promises are supported by social norms. The third supports that theory by showing that it can explain experimental data better than any other current theory. The fourth supports the theory by showing that it can be useful for designing social-change interventions.

In the first chapter, I argue for a rights-based theory of promising. I outline three different accounts of promising – that is, what you are doing when you make a promise. These accounts are gleaned from the normative-ethical literature on promising. Although these theories mainly attempt to explain the normative basis of promising, they all presuppose a view of what promising is. I argue that the best

⁸ Sellers 1963, 22.

⁹ See, for instance, Weber 1968, Arrow 1994, Menger 1985, and Hayek 1948.

interpretations of the two of the accounts – the social-practice view and the assurance view – both reduce to a version of the third, the authority-transfer view. The authority-transfer account essentially views promises as transfers of rights to particular actions. I conclude that, when you make a promise, you are transferring a right to your action. In the next chapter, I discuss what type of right you are transferring.

In the second chapter, I set up a framework for answering the question of what type of right you transfer when you make a promise. I distinguish between essentially moral rights – rights based on fundamental moral principles – and essentially social rights, which are based on social rules, but may be moralized. I show how promising behaviors fulfill the definition of a social norm, but has some moral elements, rejecting the idea that it is supported by simply moral norms. Lastly, I connect the idea of a moralized social norm to the notion of a right, and conclude that promising transfers a moralized social right. I propose a few ways of differentiating an essentially moral right from an essentially social right, although it is a contentious distinction. From this discussion, I conclude that when you promise, you are essentially transferring a moralized social right over your action to another individual.

From this account in chapters 1 and 2, we can draw two related conclusions: promises transfer rights, and the normative basis of promise-keeping lies is the fact that the promisee has acquired a right. In the third chapter, I argue that, since

promises are supported by social norms, we can also conclude that promise-keepers are motivated by a conditional preference to follow the social rules of promising. A social norms approach, by itself, explains motivation to adhere to a social norm, and this motivation is testable. I test my account of promising by showing that it can account for the experimental data on promise-keeping. Further, I argue that the experimental data support my view of promising: not only can my view account for all of the data, but it is the best explanation of it.

In final chapter, I explore the application of this account in social and political change. Though I focus on the use of public commitments in the abandonment of FGC, this theory can be applied to norm change more generally to secure legal and human rights. I use the idea of public commitments to show how my account of promising can be used in behavioral interventions, specifically those concerned with FGC. Programs aimed at abandoning FGC often use a public commitment ceremony in which all community members can witness the commitment of others to end the practice. I argue that the success of such programs will depend on whether or not the public commitment was a promise. My theory of promising shows that norms supporting promising may be culturally specific, and may be more or less relevant in different cultures. Analyzing the strength of the norms supporting promising, their cultural significance, and their particular form will allow programs to create better interventions, and will allow policy- and grant-makers to predict which interventions are likely to work.

CHAPTER 1: WHAT ARE YOU DOING WHEN YOU MAKE A PROMISE AND MEAN IT?

I. INTRODUCTION

Before exploring why individuals keep promises -- even when it appears irrational to do so -- it is important to analyze the behaviors surrounding promising itself. That is, it is important to be clear about what you are doing when you make a promise and mean it. Within the normative-ethical literature on promises, there are three prominent views about what you are doing when you make a promise. First, you may be triggering a set of rules that governs the social practice of promising, thereby obligating yourself. On this view, promises are constituted by a set of social or linguistic rules, and when you make a promise, you are indicating that you are taking part in the practice and following its rules. The rules of the practice create obligations (or claims), and give you reason to do the promised action. When you make a promise, you are using a social practice to create an obligation or to give yourself a reason to perform an action. Proponents of this view include John Searle, David Lewis, and John Rawls. I will call this the 'social practice view'.

Second, you might be transferring a moral or social right to the performance of an action. That is, you may be giving the promisee authority over a specific action

of yours, perhaps in exchange for authority over an action of hers. By ‘social right’¹⁰, I mean a claim, power, or other rights-bundle that you have because of certain social practices or rules, just as a legal right derives from legal rules, a moral right derives from moral rules, and a claim within a game derives from the rules of the game. For instance, if there is a social practice of tipping servers at restaurants, servers have a social right to tips. This rights-based view of promising has been espoused by David Owens, and to a lesser extent Brad Hooker, and I will call it the ‘authority-transfer view’.

Lastly, you might be offering the promisee assurance, inviting her to rely on your performance of an action, or deliberately creating an expectation in her of the action’s performance. In this group of views, a promise is essentially an invitation to form legitimate expectations which the promisee can use to plan her own actions. This group of theories includes those of T.M. Scanlon, J.J. Thomson, W.D. Ross, and Jan Narveson. I will call this view the ‘assurance view’.

In this chapter, I will explore each of these positions in turn. I will also argue that, when you promise and mean it, you are transferring authority over your action. The general strategy will be as follows. Primarily, I will attempt to show that the best way to understand each of the other views reduces them to this authority-transfer view of promising. But even if we do not accept these versions of the social-practice and assurance views, any version of these views is at least compatible with

¹⁰ Social rights in this sense may also be called ‘informal rights,’ to distinguish them from formal legal rights.

the authority-transfer view. I then argue that, given independent reasons we have for accepting the authority-transfer view, all things being equal, we should favor the authority-transfer interpretation of promises.

As a caveat, I also want to be clear about a question that that I am not addressing in this chapter: I am not asking what you are saying when you make a promise. That is, I will not discuss the meaning of the term ‘promise,’ or its use, except in relation to those philosophers who reduce my question to this one. I am avoiding this question for several reasons. I am looking at a behavioral phenomenon, so it should be clear that we cannot *only* analyze that way we use the term “promise” -- I can make a promise without saying “I promise”. This is especially true for non-English speakers who engage in promising. My reason for not addressing language at all is simply this: there are better ways of getting at moral intuitions than through linguistic intuitions. Linguistic intuitions do not necessarily reveal truth. In the words of H.L.A. Hart, “Plainly we cannot grapple adequately with this issue if we see it as one concerning the properties of linguistic usage.”¹¹ And so, for this chapter, the issue of promising language will be, for the most part, ignored.

II. THE SOCIAL PRACTICE VIEW

II.A. SEARLE AND LEWIS

¹¹ Hart 1961, 209

According to John Searle, when you make a promise and mean it, you are uttering a sentence predicating a future act with the intention a) of doing that act, b) of putting yourself under the obligation to do that act, and c) of the promisee recognizing that obligation in virtue of her knowledge of the meaning of a sentence.¹² In essence, you are putting yourself under an obligation in virtue of the meaning of your utterance, and the promisee's understanding of the meaning of the utterance. I will explain this further.

The meaning of an utterance -- in this case, a promise -- is determined by the rules of the language. You are only taking on an obligation because the rules of the language specify that the term 'I promise' means that you are taking on an obligation. Here, Searle differentiates between two types of rules, regulative and constitutive. Regulative rules exist independently of the activity; they govern it. Constitutive rules constitute an activity; the activity is logically dependent on the rules.¹³ According to Searle, language rules are constitutive. Languages are constituted by a set of (implicit) semantic and syntactic rules. They are logically dependent on these rules.

Rules of promising are, in Searle's account, simply a subset of linguistic rules. A promise is a performance of a particular type of speech act, and performances of speech acts are logically dependent on linguistic rules.¹⁴ And linguistic rules are

¹² Searle 1969, 60

¹³ Searle 1969, 34

¹⁴ This is Searle's main claim. He states, "The semantic structure of a language may be regarded as a conventional realization of a series of sets of underlying constitutive rules, and that

social or conventional rules. So the obligation to keep a promise is *social* since it is based on the meaning of the utterance, which is conventional. You have invoked constitutive rules of promising.

Thus, we can reword the original answer to the question to take into account these social and linguistic constitutive rules of promising. Originally, we said that Searle's account of making a promise and meaning it states that a promise is a predication of a future act with the intent to obligate yourself, and to get the promisee to recognize the obligation. We can specify this to say that you intend to obligate yourself, and intend that obligation to be recognized, *in virtue of the meaning of the speech act*. The obligation that you intend exists in virtue of the meaning of the utterance. The meaning of the utterance is determined by the constitutive rules of the language.

Hence, when you make a promise and mean it, you are simply taking part in, or correctly following the rules of, a social practice. It is similar to a game of baseball. When you play baseball (and mean it!) you are simply following the rules of a particular social practice. Like baseball, promising is constituted by certain rules, in this case linguistic rules. When you make a promise, you are invoking or triggering those rules; you are participating in the practice. You are following rules.

speech acts are acts characteristically performed by uttering expressions in accordance with these sets of constitutive rules" (37).

David Lewis has a similar view of what you are doing when you make a promise and mean it, but he adds the concept of truthfulness. Essentially, he is adding to Searle's view both a definition of a social practice or convention and a specification of the convention that applies to promising behavior. He explains what it means to follow the constitutive rules of promising and of language in general. For Lewis, when you make a promise and mean it, not only are you following the rules of your language, but you are also participating in a convention of truthfulness upon which those rules depend, and you are participating because it is in your self-interest.

To begin, Lewis outlines what it means to have a convention within a population. For Lewis, a convention is a solution to an imperfect coordination problem. It always us to coordinate our behavior to reach an equilibrium. A Behavioral regularity R in population P for situation S is a convention iff, in almost every instance of S in P:

1. Almost everyone conforms to R;
2. almost everyone expects almost everyone else to conform to R;
3. almost everyone has similar preferences with regard to possible action combinations;
4. almost everyone prefers that "any one more" conform to R, so long as almost everyone conforms to R; and

5. almost everyone prefers that "any one more" conform to R', so long as almost everyone else does, where R' is a behavioral regularity incompatible with R.¹⁵

In addition, these conditions are all common knowledge within S. The key features of a convention, then, are conformity within a society, expectations of conformity, the preference to conform as long as (almost) everyone else does, and arbitrariness -- there could be a different behavior that solves the same problem.

Lewis goes on to argue that language falls under this definition of conventions. In particular, language conventions rely on a convention of truthfulness -- languages only work if we expect others to be truthful, and if we prefer to be truthful as long as they are.¹⁶ Otherwise, we could not communicate. If others are not truthful, we have no chance of communication, so it is better for me to stop being truthful. Since we all generally have a preference to communicate, we will all generally have a conditional preference to be truthful.

The convention of truthfulness also supports the convention of promising. When you make a promise, in Lewis's view, you are making a statement that can only be made true by the performing of an action.¹⁷ You are already under the (conventional) obligation to be truthful, an obligation which exists because it is in your self-interest as a participant in the convention of truthfulness. So, in this sense, when you make a promise and mean it, you are opting to participate in a convention

¹⁵ Lewis 1969, 78

¹⁶ Lewis 1969, 178

¹⁷ Lewis 1969, 187

of truthfulness whose rules tell you to perform an action, and make it in your self-interest to perform that action. When you make a promise and mean it, you are giving yourself a self-interested reason to perform the action.

Lewis's explanation of what you are doing when you make a promise and mean it picks up where Searle's leaves off. Searle explains promising in terms of following the constitutive rules of the language, and Lewis expands on this idea by specifying what type of rules they are, what they rely on, and what sort of reasons they give us to follow them. Making a promise and meaning it is opting to participate in rule-constituted behavior, and to follow those rules, based on the meaning of your utterance. But Lewis adds that you are following those rules because, in making that utterance, you are giving yourself a reason to follow them. You have a reason to follow the rules because you want to coordinate, or reach the optimal payoff in the coordination game, and following the rules allows you to coordinate. The reason comes from your own desire to coordinate. Taking Lewis and Searle together, when you make a promise and mean it, you are giving yourself a reason to perform the action based on the meaning of your utterance.

II.B. RAWLS

Rawls's account is similar to those of Searle and Lewis, but it does not rely on language conventions or the meaning of promises. Instead, Rawls simply appeals to the convention of promising itself. Essentially, when you make a promise and mean it, you are putting yourself under a social and moral obligation to perform a

particular act by invoking a set of socially determined rules that constitute the act of promising. As with Lewis's and Searle's accounts, when you promise, you are engaging in or invoking a social practice that is defined by following certain rules.

In "Two Concepts of Rules," Rawls uses promising as an example or test case for distinguishing between 'summary rules' and 'practice rules'. Summary rules approximately map onto Searle's regulative rules. Summary rules are created by summarizing past decisions, and are created in order to facilitate future decision-making.¹⁸ Like Searle's regulative rules, summary rules come about after the fact -- they regulate behavior that already exists without them. An example of summary rules is rules of etiquette. Etiquette regulates behavior that already exists without it, like eating and drinking.

Practice rules, on the other hand, are rules that are logically prior to the behavior they apply to. They are akin – if not identical – to Searle's constitutive rules; they define particular institutions or practices. According to Rawls, promises are constituted by practice rules. This claim is supported by the idea that any promising behavior is logically dependent on pre-existing rules that require that promises be kept: "unless there were already the understanding that one keeps one's promises as part of the practice itself there couldn't have been any cases of promising."¹⁹ There is no reason to make a promise – or to keep it – unless there is

¹⁸ Rawls 1955, 22

¹⁹ Rawls 1955, 30

also an expectation that it be kept. This expectation comes from the knowledge that keeping promises is obligatory, and this obligation in turn comes from the social practice.

Rawls's account of promises as defined by practice rules also overlaps with Lewis's account of promises as conventions, though it predates the theories of both Lewis and Searle. We have practice rules in part because, in Rawls's words, the "attempt to coordinate behavior by trying to foresee how others will act is bound to fail."²⁰ We need practice rules to predict how others will act so that we can coordinate our own behavior. Hence, social practices like promises are created at least in part to solve coordination games, just like Lewisian conventions. We have practice rules because it is in our interest to coordinate, and practice rules allow us to predict the behavior of others.

In Rawls's view, then, promising is a social practice or convention in that it a) is constituted by practice rules, and b) helps us to coordinate our behavior by allowing us to create expectations. When you make a promise and mean it, you are creating expectations by creating an obligation for yourself. You are creating an obligation by invoking the social rules of promising, which stipulate that uttering particular sentences under particular conditions puts you under an obligation to perform particular acts. When you make a promise, you are giving yourself an obligation by invoking the rules of promising.

²⁰ Rawls 1955, 24

II.C. CREATING REASONS AND WILLING AN OBLIGATION (SHEINMAN AND HART)

In this discussion of the social practice view, I have described the act of making a promise in two ways: creating reasons to perform the promised action and creating an obligation to do the promised action. In this section, I would like to describe both of these ideas in further depth, in the end combining them to create a final social-practice view of promising: when you make a promise and mean it, you are giving yourself practice-based reasons to do what you have promised, at least one of which stems from the practice-based obligation to keep the promise.

To begin, Hanoch Sheinman has argued for a view of promising as creating practice-based reasons to perform a particular act. For Sheinman, when you make a promise and mean it, you are giving yourself a reason to do what is promised. This reason is based on the social practice of promising -- for instance, you might lose face if you don't do what is promised, or you might harm your relationship. I will explore his argument in more depth, and show how it fits with the social-practice views of Rawls, Lewis, and Searle.

For Sheinman, a (perfect) promise has three conditions. First, the promisee must give herself a self-interested motivating practice reason to perform an action by communicating to the promisee an intention to give herself this reason. The promisee thus believes that the promise is reason-giving for the promisor. Second, the action must be in the promisee's interest. Third, both the promisor and the promisee believe these conditions hold; that is, they believe that it was in fact a

promise.²¹ Basically, according to Sheinman, a promise gives the promisor reason to do what is promised by invoking the rules of the practice.²²

The basic argument is this: in order for a promisee to believe that the promisor will keep her promise, she has to believe that the promisor has overriding reasons to do what was promised. And this reason cannot be based on the promisee's belief -- that would be a circle. And so the promisee must believe that the promisor has independent reasons to keep the promise.

But where do these independent reasons come from? It cannot be that the promisor has overriding reasons independent of the promise. Otherwise, there would be no reason to make a promise. Promises work by giving us reasons to keep the promise to counterbalance the self-interested reasons not to do what has been promised.²³ Even when we have other reasons to do what's promised, we are still aware of, and motivated by, our opposing self-interest.²⁴ For instance, if I say I will meet you for lunch, I might have many reasons to do so – the pleasure of your company, good food, and so on. If you ask me to promise to meet you for lunch, it will be because you fear that those reasons are not overriding. You believe I need another reason, a reason that comes from the promise itself. The promise itself must be reason-giving.

²¹ Sheinman 2008, 298

²² Sheinman 2008, 299

²³ Sheinman 2008, 306

²⁴ Sheinman 2008, 307

There are several ways in which a promise might be reason-giving. I may fear the damage done to our relationship when I break a promise, or the loss of trust if the community were to think of me as a promise-breaker. These reasons are all based on promising as a social practice – if it were not a social practice, there could be no such sanctions. Hence, Sheinman not only argues for why promises must be reason-giving social practices, but also shows us how they might give us reasons.

Searle's, Lewis's, and Rawls's accounts fit nicely into Sheinman's picture. In Searle's and Lewis's views, when you make a promise and mean it, you are giving yourself reason to perform the action based on the meaning of your utterance. But it is not always clear what that reason would be, or how it would be motivated, apart from allowing you to coordinate. And it is not clear that coordination is always the goal when we make specific promises. Sheinman specifies these theories by showing how the practice of promising might be reason-giving: you may fear the sanctions on promise-breaking.

On Rawls's account, when you make a promise and mean it, you are giving yourself an obligation by invoking the rules of promising. This obligation is a social obligation – it is an obligation based on the rules of promising. Sheinman's idea of social sanctions fleshes out what that obligation consists in. When you invoke the rules of promising, you are creating an obligation in that you risk sanctions if you do not follow the rules. The obligation is reason-giving, on this reading, because of the risks involved in breaking it. Hence, we can see Rawls's, Searle's, and Lewis's views

as versions of the view that, when you make a promise and mean it, you are giving yourself a (usually) overriding practiced-based reason to do what you have promised.

The second way to describe these views of promising uses the concept of willing an obligation. This idea comes from Hume. In trying to determine what a promise is, or upon what faculty it depends, Hume rules out all of the other options. It is not a resolution because a resolution alone cannot create an obligation. It is not a desire because often we do not desire what we promise. It is not the willing of the action because the will can only influence present actions.²⁵ It must thus be the willing of an obligation.²⁶

The problem, however, is that it seems impossible to will an obligation. For Hume, morality depends on the sentiments. A change in obligation implies a change in sentiment; willing a new obligation thus implies willing a new sentiment. But willing a new sentiment is absurd, and so a promise cannot be the willing of an obligation.²⁷ Even if we do not accept Hume's sentimentalist theory, the idea of willing an obligation seems to be rather odd – obligations do not seem to be the kind of thing that we can just decide to have. How, then, are obligations willed or formed?

²⁵ Hume 2000, 331

²⁶ Hume 2000, 332

²⁷ Hume 2000, 332

H.L.A. Hart discusses voluntary obligation, especially in relation to promising, in his paper "Legal and Moral Obligation". Hart makes an analogy to legal obligation: in order to create (or will) a legal obligation, there must be rules that tell you how to create that obligation.²⁸ Once you have those rules, willing an obligation is just a matter of following them. But then where do these previous rules of the system come from? In Hart's view, they exist insofar as they "refer to the *actual practice* of the particular social group whose legal system is under consideration."²⁹ The actual practice that upholds these rules includes both the recognition by the group that deviation warrants criticism, as well as the recognition that the words of the legal authority are reason-giving.³⁰ The prior acceptance of the rules for creating legal obligations is what allows us to create these obligations in any given circumstance.

The same can be said about other sorts of obligations. If we want to be able to will obligations into existence, we need prior rules that tell us how to do so. These prior rules allow us to circumvent Hume's problem with willing an obligation -- we are not creating obligations out of thin air, but rather using accepted social rules to create them. Promising is an example of willing an obligation through prior social rules. The rules of promising, if they are recognized by the social group, allow us to create social (and perhaps moral) obligations to perform a particular act.

²⁸ Hart 1958, 87

²⁹ Hart 1958, 88

³⁰ Hart 1958, 90

The idea of willing an obligation concords with Searle's, Lewis's, and Rawls's views of promising. It is straightforwardly aligned with Rawls's view: when you promise, you invoke practice rules to create an obligation to perform an act. For Searle and Lewis, it is not as clear whether you are giving yourself reasons by creating an obligation or through some other means. But creating a social obligation through the rules of the practice of promising would certainly be reason-giving. So Searle and Lewis can certainly accommodate this view, even if it is not the whole story.

That leads to a further point. The view that promising is willing an obligation seems to be a specification of the view that promising is giving yourself practice-based reasons to perform a particular act. For certainly obligations are reason-giving, and, if we adopt Hart's view, the obligations must be created by a social practice. In addition, the idea that when you promise, you will an obligation seems to fit well with our moral psychology. When we promise, we certainly think that we have an obligation to do what is promised.

However, this obligation may not be the only reason-giving feature of promising. For instance, it is unclear whether sanctioning and maintaining relationships are included in the obligation to keep promises, or whether they are side-effects that provide their own reasons. Hence, we can summarize the social practice view as follows: when you make a promise and mean it, you are giving

yourself practice-based reasons to do what you have promised, at least one of which stems from the practice-based obligation to keep the promise.

III. THE AUTHORITY-TRANSFER VIEW

III.A. OWENS AND HOOKER

According to David Owens, when you make a promise and mean it, you are transferring authority to the promisee. More specifically, you are transferring control over some action of yours to the promisee – she now has the right to demand the performance of the action, or a claim against you. In this section, I will detail what is meant by authority transfer, why it might be a plausible view of promises, and what it entails, or rather does not entail.

In general, we consider ourselves to have authority over our own actions. We, and usually only we, have the right to decide what we will do. This authority implies a negative right: all other things equal, we have a claim of non-interference against others. But, like many other rights, this claim-right can be waived, transferred, or overridden. Included in the authority over our own actions is the authority to waive or transfer our claim-rights. Hence, the authority I have over my own actions includes a claim to non-interference, all other things equal, the power to waive my claim to my action, and the power to transfer the claim to my action. Owens assumes this is relatively undisputed – it should be clear to most that we have rights to our actions.

A promise, then, is the transferring of the claim over one's action. According to Owens, "promising is a mechanism by which one person actively assumes authority over another."³¹ Unlike other claim-generating actions – those that generate the need for compensation or reciprocity, for instance – the authority transfer in a promise is active. The promisor actively gives up her claim to the action, and the promisee actively accepts the authority.³²

Owens gives several reasons to support the idea that when you promise and mean it, you are transferring authority over your actions. First, the authority-transfer view allows us to distinguish promises from other forms of commitment. Owens distinguishes between two types of interests a promisee might have that would be satisfied by promising. The promisee might have an information interest: having information about another's future actions will help her to form accurate expectations and plan her own actions. But she may also have a separate interest in having authority over another's actions. Owens claims that humans have such an authority interest -- they want authority over the actions of others for its own sake, not just to make plans or coordinate action.³³

If we view promises as means for actively transferring authority over an action, then they would be able to fulfill this authority interest. The fulfillment of the authority interest, in turn, allows us to distinguish between promises and other

³¹ Owens 2006, 72

³² Owens 2006, 73

³³ Owens 2006, 52

commitments or expectation-generating acts. On the surface, promising looks a lot like prediction, testimony, and the expression of intention. All of these generate expectations about one's future actions. But the authority-transfer view allows us to distinguish between these types of commitments: promising serves the authority interest, which prediction, testimony, and intent all serve the information interest. That is why promises are more binding: they generate claims on behalf of the promisee. Hence, this view allows us to explain what is particular about promising.

Second, the authority-transfer view can explain why one would make a promise, not just why one would accept it. Usually, when you transfer authority through a promise, you are exchanging authority over a less important action for authority over a more important one.³⁴ Many of the promises we make are reciprocal – 'I'll buy this round if you buy the next,' 'I won't cheat if you don't' – so we rely on the performance of actions by other as much as they rely on us. The authority view shows how this exchange is possible. It is more valuable for me to have authority over your action – perhaps buying the next round – than it is for me to have authority over my action – in this case, buying the first round. And so, just like exchanging money and property, we exchange authority over actions.

There may be other reasons why we would want to transfer authority. At times we may want to exchange authority for goods or services, as we do with promissory notes, credit cards, and other loans. At other times, It may be that we do

³⁴ Owens 2006, 70

not trust our future selves to uphold our present ideals or intentions. I may promise to go to your party not because I expect an authority transfer in return, but because I fear that, when the time comes, I may be tempted to stay home. As in Sheinman's view, in this case, the promise gives me an extra reason to perform the action. In the authority-transfer view, this reason comes from the claim I have given you.

Lastly, and most importantly, according to Brad Hooker, the authority-transfer view is consistent with many pre-theoretical accounts of promising. Hooker lists several properties of promising, all of which are satisfied by the authority-transfer view. First, he states that breaking a promise wrongs the promisee, and it does so even if there is no loss in welfare. In the authority-transfer view, the wrong of breaking a promise is the failure to fulfill a claim, regardless of welfare.³⁵ Second, the duties generated by promises are self-imposed and backward-looking. The claim that promises are the active transfer of authority over a future action can accommodate this idea as well.³⁶ Third, the promisor's duty has a correlated claim on behalf of the promisee. This is clearly satisfied by Owens's view – the promise is characterized first and foremost by the promisee's claim.³⁷ The authority-transfer view, then, satisfies Hooker's pre-theoretical characteristics of promising.

III.B. AUTHORITY AND SOCIAL PRACTICES

³⁵ Hooker 2011, 8

³⁶ Hooker 2011, 9

³⁷ Hooker 2011, 11

A further benefit of the authority-transfer view is that it is consistent with several other views and, in some cases, can explain these views. In particular, this view is consistent with the idea that promising is a social practice, and it can encompass the social-practice view of promising discussed earlier. In this section, I will discuss the relation of authority-transfer to social practices using the analogy of property transfer. In the next section, I will show how we can explain Rawls's social practice view in particular using the authority-transfer view.

In Owens's formulation of the social practice view, it is unclear what type of claim you are transferring when you make a promise. There are several types of claims in several types of rule-based systems, some of which may overlap. For instance, one may have legal claims within a legal system, one may have moral claims within a moral system, and perhaps one may have social claims within a social system. One claim may fulfill multiple categories, as well. A divorced parent may have a legal claim against her child's other parent for child support, and a moral claim against him for help raising the child. She may even have a moral claim to child support because it is morally right to follow a justified law. In addition, she may have a social claim against him if she lives in a community that recognizes that both parents should support and child, and that is willing to sanction those who do not. The claim is (perhaps) the same, but exists and is justified simultaneously by several different systems.

In Owens's authority-transfer view, it is not clear whether the promisee's newly-transferred claim is a moral claim or a social claim, or perhaps both.³⁸ I will argue that it is plausible that the claim is a moral claim that is based on a social claim, and thus the authority-transfer account can be compatible with the idea that promising is a social practice. To show that it is plausible that the claim transferred in a promise is social, I will use an argument by analogy with another type of plausibly social claim – that is, property claims.

In the case of property, the most powerful argument that property claims are primarily social, or that any moral practice of property presupposes a social practice, comes from David Hume. For Hume, property is an institution that developed in order to facilitate coordination and cooperation. Property rights were created to coordinate how individuals divide and consume resources in a world where resources are scarce. But there are many systems of property that can secure the end of coordinating scarce resources, and these systems may have different virtues – some may be more egalitarian, some more efficient, and some simpler. Taken together, these three features of property – its status as a human institution, its function of coordination, and its arbitrariness – satisfy the criteria of a Lewisian convention. So, for Hume, property is conventional; it is a social practice.

Hence, the claims generated by the social practice of property are primarily social, though they may be legal and moral as well. Recall that, according to Hart,

³⁸ It is clearly not a legal claim. The legal analogue to the promise is the contract.

any legal practice or obligation presupposes a social practice. The same is true of property. For legal property to exist, there must be either a pre-existing social practice of property that legitimizes the legal rules, or there must be a social practice of obeying the legal authorities who determine property rights. Hume seems to think that, generally, property rights develop as a social practice before they are codified into law.

The moral aspect of property, according to Hume, presupposes the social as well. Hume's moral system is based on sentiments. If property is a social creation, and thus not natural, then there can be no natural sentiment by which we approve of right actions regarding property. The approval of property must be developed as we witness the success of the system of property in coordinating our actions. Hence, the existence of moral property rights also presupposes the existence of social (or perhaps legal) property rights.

The same point can be made without using Hume's sentimentalist moral theory, instead substituting the idea of arbitrariness. Property rights are arbitrary in that there may be many systems of property, all of which coordinate the use of scarce resources. The idea is that property rights are bundles – they may include combinations of claims to exclusive use of certain resources, powers to transfer resources, liberties to use resource without excluding others, and so on. The bundles may differ in different systems, but also may differ across various types of resources within any particular system. For instance, in American law, I have the

right to use my land and exclude others from using it, but in Swedish law, I only have the right to use. On the other hand, in Swedish law, I could exclude others from using the food and shelter on my land.

The question, then, is where the moral theory enters. Since property rights are arbitrary in the sense that there are many possible systems that succeed in coordinating use of scarce resources, we cannot say that property rights are moral full-stop. We would be applying the moral theory at too high a level. It would be similar to claiming that it is moral to follow the law – surely it depends upon which law we are following. We can only say that *particular sets* of property rights are moral, although it may be that many or all of them are. Particular sets of property rights are determined by social practice. Hence, we can only apply the moral theory once we have a social practice of property to which to apply it.

We have seen that there is at least one form of right – that is, property rights – in which moral claim-rights plausibly presuppose social claim-rights. We can now analogize from control over physical resources to control over actions. It is at least plausible that rights over actions, and especially powers to transfer control over actions, have the same sort of social basis as property.

We will start by supposing that Owens is right in claiming that promising is the transfer of authority over one's action. We can extrapolate from that claim that there must be a system by which we transfer authority over actions, again following Hart's reasoning. It is certainly plausible that the rules for transferring authority

over action can vary across social systems, or, in other words, that the rights-bundles regarding actions vary. It is in fact possible that promising itself varies across social systems, not just in the way we promise, but in the kinds of actions that can be promised, sanctions surrounding violation, and the conditions attached to the promise. I may not, for instance, be able to promise you my labor (in the form of selling myself into slavery), but may be able to promise you my labor in exchange for goods.

Given that it is plausible that systems of rights-bundles regarding actions are arbitrary, it is also plausible that they are conventional, in the Humean and Lewisian sense. The arbitrariness of the systems, or the possible variation across societies, indicates that these systems are social institutions, much like property. In addition, the conventional allocation of rights over actions can have a coordinating function. Hume's example of the rowers illustrates this well. In order to properly row a boat, the two rowers must coordinate their oar-strokes. In Hume's example, the rowers naturally create a convention for coordinating their strokes; however, this coordination can also be achieved by particular allocations of rights over actions. For instance, if there were an authority with the right to dictate the actions of both rowers, she could easily coordinate their strokes. Like the allocation of resources, the allocation of control over action can have a coordinating effect.

Since it is plausible that systems of rights over actions are arbitrary, institutional, and have coordinating function, it is plausible that these systems are

Lewisian conventions, and thus social practices. Since, in Owens's view, promises are the transferring of authority over actions, and since the transferring of authority over actions makes up a subset of the rights over actions, it is plausible that promises are social practices. Thus, Owens's authority-transfer view is compatible with the idea that promises are social practices.

III.C. REDUCING RAWLS AND ADOPTING AUTHORITY-TRANSFER

Interpreted the right way, the authority-transfer view of promising can encompass the Rawlsian social-practice conception of promising. I focus on the Rawlsian view because it seems to be the best-developed – and most discussed – of the three social practice views I have described. We saw earlier that in Rawls's view, when you make a promise and mean it, you are giving yourself an obligation by invoking the practice-rules of promising. There is a side question about whether the obligation you give yourself is moral, and if it is, how the social obligation becomes moral, or morally justified. I will try to show that if the obligation is moral, it is based on the claims – or rights – of the promisee. Rawls's view can then fit into the authority-transfer view.

According to Rawls, if the social practice is morally justified, the obligation becomes a moral obligation. That is, if it is morally required that we have this social practice (or something like it), then the rules become morally rules, and the obligation becomes a moral obligation. But it is unclear how the moral justification of the social practice transforms the social rules into moral rules.

There are several answers to this question. Chris Melenovsky suggests using the Principle of Legitimate Expectations. When someone is warranted by the rules of a justified practice to have legitimate expectations of another, she has a valid claim against that person.³⁹ Hence, you are creating a moral claim on behalf of the promisee by invoking the practice rules. If the practice of promising is justified, then simply invoking the rules causes the promisee to form legitimate expectations, which creates a claim. The rules are not themselves moral, but rather the expectations (and thus claims) that they create.

Tamar Schapiro has a different explanation, focusing on shared agency instead of legitimate expectations. Though she primarily focuses on the social practice of negotiation, her view can also be applied to promising. On her account, a practice gives us moral claims on one another because it is in part constituted by shared agency. Participants see themselves and others as agents who can follow the rules from their own consciences, and the practice promotes some common good. There is then a moral principle that states that if a social practice is constituted by shared agency, it is wrong to use the practice to serve one's own unilateral will – it makes the practice into a sham.⁴⁰ Breaking a promise subverts the practice of promising in such a way. It takes a socially beneficial practice, an expression of a shared will, and transforms it into something that serves the individual's will. Since

³⁹ Melenovsky (forthcoming).

⁴⁰ Schapiro 2003, 344

it is wrong to subvert the practice in such a way, we all have claims against each other to have promises kept.

In both of these views, however, the social obligation to keep promises becomes a moral obligation because the practice gives us moral claims on one another. We can amend the Rawlsian picture of what you are saying when you make a promise to include these moral claims. Taking Rawls, Melenovsky, and Schapiro together, when you make a promise and mean it, you are giving someone a moral claim against you by giving yourself a social obligation, which comes from invoking the rules of promising. Giving someone a moral claim against you for the performance of an action – no matter how you do it – seems to be the same thing as transferring authority over that action.

Promissory obligations, by any account, should have correlated rights on behalf of the promisee. Since the social practice views can account for promissory obligations, they should be able to accommodate at least a minimal account of rights,⁴¹ and so promises will include some sort of transfer of rights. The question is whether we consider the duty to be prior to the right, or vice versa. It may also be that neither is prior -- that both exist, and mutually entail each other. Does the promisee's right give the promisor an obligation, does the promisor's obligation give the promisee a right, or both?

⁴¹ This is not to say it is a *satisfactory* account of rights. For more discussion on this, see Gilbert 2004.

For Rawls, it seems that the right is required for understanding the duty, and thus the right is prior. For Searle and Lewis, on the other hand, it is not clear if the right is indeed prior to the obligation – neither seems to be prior. In Rawls’s view, the claim is prior to the obligation because the practice is also a moral one; were it only a social practice, the promissory obligation might be prior to the right. Hence, if Searle and Lewis want to claim that promising is a moral practice as well as a social one, they would need similar arguments to those that Melenovsky and Schapiro apply to Rawls. They simply do not discuss the moral practice.⁴²

The discussion of promising as a moral practice, and not just a social one, lends credence to the Rawlsian picture of the social-practice view. Essentially, if the practice is moral, Searle’s and Lewis’s views would need to prioritize rights. And we should independently reject the view that the practice is not moral because it contradicts the common view of what it means to promise.

But suppose, despite this advantage, we reject the Rawlsian picture of promising. We should still accept the authority-transfer interpretation of the social-practice view. Even if we reject Rawls’s view, the social-practice view and the authority-transfer view are compatible. We simply need a way to decide between them. As Hart states when discussing competing concepts of law, “If we are to make a reasoned choice between these concepts, it must be because one is superior to the other in the way in which it will assist our theoretical inquiries, or advance and

⁴² However, Searle does claim at one point that promising is not necessarily moral, but rather oftentimes merely conventional (188).

clarify our moral deliberations, or both.”⁴³ The authority-transfer view, as we have seen, assists our theoretical inquiries because it has more explanatory power -- it helps us explain why we make promises. It clarifies our moral deliberations in that it can help us distinguish between promises and other commitments in a meaningful way. This gives us reason to favor the authority-transfer interpretation of the social-practice view, even if we do not adopt Rawls’s view.

IV. THE ASSURANCE VIEW

So far, we have discussed the social-practice and authority-transfer views. Of these two accounts, I have argued that we should adopt the authority-transfer view. The final account I want to discuss is what I call the ‘assurance view’, as put forward by Jan Narveson, J.J. Thomson, and T.M. Scanlon. The essence of this view is that when you make a promise and mean it, you are providing assurance. What providing assurance means differs according to each theory, but in general has to do with deliberately creating expectations. I will describe each of Narveson, Thomson, and Scanlon’s theories, and then combine them into a single account of what you are doing when you make a promise and mean it according to the assurance view. I will then make a similar argument for an authority-transfer interpretation of the assurance view as I did for the social-practice view. I will argue that the assurance view is best explained in terms of rights, but even if we do not adopt this specific

⁴³ Hart 1961, 209

version of the assurance view, we should still adopt an authority-transfer interpretation.

IV.A. NARVESON AND THE UTILITY OF EXPECTATIONS

Jan Narveson begins his account of promising with describing what is wrong about promise-breaking, and then uses this account of wrongness to reach the essential features of promising. As a utilitarian, he must show how breaking a promise diminishes overall utility. There are several ways in which breaking a promise can lower the utility of the promisee. First, the promisee has likely structured her expectations, and thus plans, around the promised action actually occurring. In this case, breaking a promise diminishes utility by keeping the promisee from being able to best plan how to maximize her own utility.

Second, beyond the ability to plan and predict, people like to have correct expectations. We simply like to have our expectations met, whether or not they allow us to plan our own actions and satisfy our own, independent preferences. In Narveson's words, "In promising, it is generally agreed, I give you to expect that I will do the promised thing. And in defaulting, I disappoint this expectation, a disappointment which is additional to the loss of the promised benefit, whatever it may have been."⁴⁴ Disappointing expectations affects utility in this additional sense

⁴⁴ Narveson 1971, 214

because making a promise creates an opportunity for an expectation to be met, or for an increase in utility.⁴⁵

Third, not only do we like to have our expectations met, and not only do promises create opportunities for utility to be raised or lowered, but those expectations and opportunities are also created deliberately. Thwarting an expectation that one deliberately created in another causes an even greater decrease in the other's utility. We like to have our expectations met, but it is even more important to us when those expectations are created deliberately.

From this, we can tease out several features of what you are doing when you make a promise and mean it. You are creating a situation in which utility could be raised or lowered, and you are doing so not just by creating expectations, but by deliberately creating expectations. But then the question becomes how you manage to create the expectations. This is where Narveson departs from Rawls – for Rawls, expectations are created by engaging in the social practice of promising. We use the pre-existing social practice to create obligations or reasons for action. Narveson wants to reject this use of the social practice because he wants to claim that the obligation to keep a promise is prior to the social practice of promising – expectations are prior to the social practice.⁴⁶

⁴⁵ Narveson 1971, 226

⁴⁶ Narveson 1971, 227

In order to explain how these expectations are generated, Narveson substitutes social practices with assurance. In his view, cannot be that, in promising, you simply believe I am obligated. That would be circular: you would not believe I am obligated unless you had formed an expectation, and you cannot form an expectation until you believe I am obligated. This is where he brings in the concept of assurance. I assure you that I will perform the promised action, and you then form an expectation because expecting is the linguistically appropriate response to assurance.⁴⁷ Once you form the expectation, I am obligated to fulfill it.

Hence, when you make a promise and mean it, you are offering the assurance necessary for the promisee to gain a specific type of expectation. The expectation is particular in that expected utility for its fulfillment is higher because it was deliberately created. This expectation, in turn, gives you a particularly strong obligation because you now have the ability to raise (or lower) utility so drastically.

There are two basic issues with this formulation of the assurance view that merit further discussion. First, it is unclear how this view differs from a social practice view. Narveson is trying to avoid relying on social practices – the obligation is prior to the practice in his view – but it is still not clear how expectations are created without using the social practice of promising. What, then, separates his view from a social practice view? Second, and relatedly, it is not clear what exactly

⁴⁷ Narveson 1971, 215

assurance is. I will address each of these, and then move on to several other formulations of the assurance view.

I mentioned earlier that Narveson distances himself from Rawls by substituting the concept of assurance for social practices. But he also claims that assurance creates expectations because expecting is the linguistically appropriate response to assurance.⁴⁸ And language, as we have seen in Searle and Lewis, is a convention. Thus, Narveson must be relying on conventions, and his view seems indistinguishable from a social practice view.

The difference between Narveson's assurance view and the social practice views lies in how they rely on social practices. Certainly promising depends, in some sense, on a social practice. How else would we communicate the intention or desire to promise in the first place, and how else would one accept a promise? But this does not mean that social practices are at the foundation of promising. We may offer assurance through conventional means, like language, but the obligation to keep the promise, once assurance is given, is not based on the convention. It is based on the utility derived from the expectations created.

To clarify this last point, Narveson distinguishes the rules of a game from what he calls the ethics of a game. The rules tell you how to play the game, but they do not tell you not to cheat, or to start playing the game in the first place – these are

⁴⁸ Narveson 1971, 215

external.⁴⁹ They are the ethics of the game. In terms of promises, the social practice of promising may tell you how to engage in promising, but further ethical rules are needed to tell you not to cheat and break your promise.

According to Narveson, the ethics of games are non-linguistic and non-conventional, for example, “one ought to do what one has given others to understand that one believes that one is obligated to do,” and “one ought to stick to one's original undertakings.”⁵⁰ These previous rules govern the conventional rules of the practice of promising. They are what makes breaking a promise wrong. And so they are at the foundation of what you are doing when you make a promise and mean it. For the social practice view, these principles are not comprehensible, or not comprehensibly normative, without a social practice. In this view, the obligation is created first by the social practice, whereas in the assurance view, the social practice itself is governed by these principles.

We have established that the assurance view is distinct from the social practice view, but we have yet to discuss exactly what it means to give assurance. The previous discussion about the ethics of the game of promising may be able to shed some light on this. Narveson states that one of these principles is that “one ought to do what one has given others to understand that one believes that one is obligated to do.” It seems, then, that giving assurance has to do with deliberately

⁴⁹ Narveson 1971, 229

⁵⁰ Narveson 1971, 230

creating a belief in another about one's own belief state – namely, that one believes one has an obligation.

This account of assurance skirts the circularity objection seen earlier: if assurance were just believing an obligation to exist, it would rely on a previous expectation, and hence could not create the same expectation. But since assurance relies on second-order beliefs, it does not require previous expectations. Hence, when you make a promise to me, you are leading me to believe that you think you have an obligation. From that belief, I can form an expectation about your action, the thwarting of which significantly decreases my utility. When you make a promise and mean it, then, you are giving assurance, in the sense that you are deliberately causing the forming of a particular second-order belief – the belief in your own belief about your obligation.

IV.B. ROSS, SCANLON, AND UTILITARIAN CIRCULARITY AND AMBIGUITY

While Narveson does address the issue of circularity in expectations, substituting second-order beliefs for first-order ones, his account still has a circularity problem related to its utilitarian basis. He claims that breaking a promise is wrong insofar as it decreases overall utility. It decreases overall utility because deliberately creating expectations causes an increase in potential utility change – my utility has the potential to increase or decrease further when my expectations are created by someone else's promise. The difference in potential utility change is

what separates promising from other expectation-generating action. It is what makes breaking a promise more wrong than simply disappointing someone.

The problem is that the account does not specify why promising increases the potential change in utility. And it seems that reason why I am more likely to be hurt when you break your promise, as opposed to disappointing other expectations, is because I believe that you are obligated to keep your promise. The pre-existing obligation is what makes my expectation strong, distinguishes it from other expectations, and causes the greater potential utility. But, for Narveson, the potential change in utility is what causes the obligation to exist. As Michael Robins states in his objection to Narveson, “the utility of promise keeping presupposes the expectation that promises will be kept,” and the expectation comes from an obligation.⁵¹ This account is thus circular: the existence of the utilitarian obligation depends on there already being an obligation.

Narveson can respond to this objection using the second-order beliefs upon which assurance is based. Since assurance requires only the promisee’s belief in the promisor’s belief in an obligation, we can account for the difference in potential utility using the belief in the promisor’s belief. My utility decreases more when I think you have broken what *you* take to be an obligation to me, regardless of whether I think it is an obligation.

⁵¹ Robins 1976, 324

There are three possible responses to this argument. First, even second-order beliefs are likely to arise from a concurrent first-order belief in an obligation, though perhaps not always. Just as my first-order belief needs some sort of grounding in an actual obligation, so too might my second-order belief. Why should I believe that you believe you have an obligation? In many cases, it will be because I believe you have an obligation. Of course, there may be other reasons – I may know that you are a member of a certain religion, or that you have previously acted as though you have a particular belief, or you may simply say “I now believe I am under an obligation”. But often my belief in your belief will come from that fact that the thing in which you believe actually exists. I believe that you believe that the Supreme Court exercises judicial power, that the soup course comes before the dessert course, and so on, because I believe these things are true. My belief in your belief comes from the fact that we have similar experiences of these things, experiences that signal that these things actually exist.

Second, it is questionable whether these second-order beliefs do cause the drastic change in potential utility that Narveson wants. For example, suppose your mother insists on buying you a Christmas gift every year. She thinks she has a duty to buy you a gift, and you know that she thinks this, but you, as an adult, do not believe she has this duty. Hence, you have come to expect a Christmas gift every year from her. Your father also insists on buying you a Christmas gift, but he *promises* you a specific present. On the other hand, your in-laws also buy you a gift

every year, but you do not believe that they believe they have a duty. They are simply kind people. You thus also expect a gift from them.

Unfortunately, come Christmas time, you do not receive any gifts from any parents. All of them acted deliberately, and had no good excuses for thwarting your expectations. Certainly you have a loss in utility, but it seems that the utility loss will be different for all three cases. Your utility will decrease from the thwarted expectations caused by your in-laws, but it will decrease more from the thwarted expectations caused by your mother. After all, she believed she had a duty toward you. But your utility would likely decrease the most because of your father's broken promise, *because he promised*. It is not straightforwardly true that your second-order belief, as in the mother case, can account for the same utility decrease as the promise illustrated by the father case.

This second response leads to a larger final issue. Even if assurance in this second-order sense can lead to such a drastic change in utility, it may not do so reliably. Recall that we differentiate promises from other expectation-creating acts by their effect on utility. Breaking a promise is worse than thwarting other expectations because it reliably lowers utility more than any other expectation. But it does not seem, taking into account the previous example, that expectations created by second-order beliefs in obligations will always, or even predictably, create greater potential utility than any other sort of expectation. It seems to depend on the circumstance. This is a problem of ambiguity: it is ambiguous where

to draw the line between promises and other expectation-generating acts in terms of wrongness.

The easiest way to solve this problem of utilitarian ambiguity is to abandon utilitarianism and substitute other principles.⁵² W.D. Ross and T.M. Scanlon do just that. In Ross's view, keeping a promise is right "qua being an ensuring of one of the particular states of affairs of which it is an ensuring."⁵³ When you make a promise and mean it, then, you are essentially ensuring that whatever is promised actually comes about. That ensuring, in turn, triggers a duty of fidelity: once you ensure, you are bound by duty to do whatever you have ensured. This duty is a self-evident prima-facie duty. It is intuitive, and always applies unless overridden by some other duty.⁵⁴

This account avoids Narveson's utilitarian circle by allowing the ensuring (and expectations) to trigger a pre-existing principle instead of tying the obligation into the act of assuring. Whereas in Narveson's theory the application of the utilitarian principle is dependent on the existence of an obligation, in Ross's theory the principle of fidelity applies independently anytime one ensures. The principle of fidelity does not require a mysterious difference in potential utility when one promises instead of predicting. Promising triggers the principle of fidelity; predicting does not.

⁵² or perhaps to add mid-level principles with a utilitarian basis

⁵³ Ross 1930, 46

⁵⁴ Ross 1930, 19

At first glance, there are several major problems with Ross's account of promises. First, the concept of ensuring is unclear. I might ensure – in common parlance – without promising. For example, I can ensure that you buy me coffee every Tuesday without promising at all; in this case, it would be a threat. Hence, without a better analysis of this concept, we cannot use it to ground promising. Though this concept is vague, we could likely flesh it out in a way similar to assurance in Narveson's account. Ensuring would be a deliberate causing of expectations the basis of which is a second-order belief. Second, it is unclear how the principle of fidelity is triggered in the first place, especially without an antecedent social practice – is it triggered when we say certain words, or create certain expectations or beliefs? As in Narveson's account, Ross can use social practices, like the social practice of language, to explain how promises are made without making social practices the foundation of his theory.⁵⁵

Lastly, and most importantly, Ross is not clear about what the principle of fidelity states – it is based on intuitions, but it may be that intuitions differ among people. In addition, and for this reason, we may want to reject his reliance on intuitions to support this principle, in which case we would need some other support for a principle of fidelity that can be triggered by a promise. Fortunately, T.M. Scanlon adds both specificity and non-intuitionist support to Ross's account.

⁵⁵ For more on hybrid theories that use social practices without making them foundational, see Kolodny and Wallace 2003.

On Scanlon's account, as on Ross's and Narveson's, when you make a promise and mean it, you are essentially offering assurance. But Scanlon is clearer about what general moral principles underlie the duty to keep promises, and the conditions under which they are triggered. He lays out four moral principles that can apply to keeping promises. First, breaking a promise falls under a general category of wrongness that Scanlon calls "unjustified manipulation", or Principle M.⁵⁶ The principle forbidding unjustified manipulation does not apply only to promises, but to any deliberate manipulation. Scanlon takes this principle to be easily derivable from his contractualist framework -- when deciding on principles, we would agree to the prohibition of unjustified manipulation. But it also seems to be a generally acceptable principle in most moral frameworks, and in common-sense morality. We usually think that unjustified manipulation is immoral.

Principle M, however, cannot tell us what is distinctly wrong about breaking a promise. For this task, Scanlon brings in several other principles, including Principle D, the principle of Due Care, that states that one should not lead others to form "reasonable but false expectations" about one's actions if it would cause them significant harm, and Principle L (Loss Prevention), which expands Due Care to include negligence.⁵⁷ Again, these principles are justifiable within Scanlon's contractualist framework. And, again, by themselves, they are not distinctive of promises. That is, Principles D and L apply to many situations, including, but not

⁵⁶ Scanlon 1998, 298

⁵⁷ Scanlon 1998, 300

limited to, promising. They cannot isolate the distinctive wrong of breaking a promise, or the distinctive duty to keep it.

To address the specific wrong of breaking a promise, Scanlon adds Principle F, the principle of Fidelity. In his view, at the basis of promising is assurance: the promisee wants to be assured, and wants the assurance to be genuine, perhaps, but not always, so she can plan and make decisions.⁵⁸ This is why we ask for, value, and accept promises, as promisees. Principle F uses the concept of assurance to show what is specifically wrong about breaking a promise. It states: If A “voluntarily and intentionally” creates an expectation in B of an action by A, A knows B wants assurance, A acts to provide this assurance, and A and B have common knowledge of this, then A (morally) must do that action unless B relieves A of it.⁵⁹ Again, this principle can be agreed upon under a contractualist framework.

IV.C. THOMSON AND RIGHTS

It is still unclear, however, exactly what it means to give assurance, and thus how the conditions in Principle F are satisfied. For instance, it is possible that Principle F is triggered only when I say “I promise,” or that it is triggered whenever A knows that B wants an event to occur, and predicts that it will. Through the discussion of what it means to give assurance, I intend to show that the assurance view is compatible with – and best understood in terms of – the authority-transfer

⁵⁸ Scanlon 1998, 303

⁵⁹ Scanlon 1998, 304

view of promises. I will do this by showing that assurance is best understood in terms of claims or rights.

Up until now, I have been assuming Narveson's view of assurance: it is the deliberate creation of a second-order belief in another about one's own obligation. Essentially, the promisor gets the promisee to believe that she believes she has a promissory obligation. This account of assurance is not circular, as we have seen, because it does not rely on a previous obligation, but only on the belief in a belief in an obligation.

Even if we adopt Narveson's account of assurance, we can still discuss how the promisor creates the second-order belief in the promisee. Why should the promisee believe that the promisor believes she has an obligation? Fortunately, J.J. Thomson has an answer to this question. According to Thomson, when you make a promise (or, rather, give your word) and mean it, you are "positively inviting the word-receiver to rely on the word given."⁶⁰ The promisor creates expectations or beliefs in the promisee by inviting reliance. That is, I believe that you believe you have an obligation because you have offered reliance on your action, and once I accept it, I am entitled to that reliance.

Thus, the obligation that I believe you believe you have is parasitic on my believing that you have given me an entitlement to reliance on your action. In other words, I believe that you are offering to transfer a claim and my accepting the offer

⁶⁰ Thomson 1990, 295

completes the transaction. If assurance is cashed out in terms of second-order beliefs, and the second-order beliefs depend on the transfer of rights, then assurance depends on the transfer of rights. If the promisor invites the promisee to rely a proposition, and the promisee accepts it, then the promisee acquires a claim against the promisor to that proposition. In Thomson's powerful language, promising is an offer to "bind oneself," and so when the offer is accepted, the promisor is bound.⁶¹

Once we understand offering assurance as the offer to transfer a right, we can see how the assurance view is compatible with, if not reducible to, the authority-transfer view. The authority-transfer view states that when you make a promise and mean it, you are giving the promisee a claim to the performance of your action. This is exactly what Thomson's invitation to rely does – it gives the promisee a claim over your action once she accepts your offer. If promising is offering assurance, then it is best understood as offering the promisee a claim. And this is the authority-transfer view.

It is important to note, however, that though Thomson's version of the assurance view can fit into the authority-transfer view, and so can the social-practice view, it is not necessarily compatible with the social-practice view. Thomson acknowledges that reliance on your performance of an act can only create an obligation or claim if there is a background of trust. But this does not amount to

⁶¹ Thomson 1990, 302

a social convention governing promising. Though promising uses social practices like language and general trust, that does not imply that promising itself is governed by social practices unless it were reducible to these practices.⁶²

For Thomson, unlike Rawls, making a promise and meaning it does not trigger a social practice or social rules, and does not necessarily generate a social obligation, though it may be necessary to engage in practices like language in order to make a promise. Instead, as in Scanlon, promising seems to trigger a moral principle that you should do what others are relying on you to do, at least when you yourself have created that reliance. Or, more basically, you should fulfill the moral claims that others have on you. And the creation, transfer, and fulfillment of moral claims do not rely on social practices, though these practices may be useful. What we have, then, are two different versions of the authority-transfer view, one that takes social rights as basic, and another that takes moral rights as basic. I will elaborate on this distinction and its implications in the next section.

V. CONCLUSION: INDEPENDENT MORAL PRINCIPLES?

Both the social-practice view and the assurance view can be reduced to different versions of the authority-transfer view. Hence, we can say that when you make a promise and mean it, you are essentially transferring a right over the performance of your action. Where the two versions differ is in what sort of right is transferred. In the social-practice-type authority-transfer view, it is foundationally

⁶² Thomson 1990, 303

a social right; in the assurance-type authority-transfer view, it is foundationally a moral right. In choosing between these, we must essentially decide whether there are necessarily (socially) independent moral principles to ground the obligation to keep promises.

Fortunately, though answer this question is quite important for establishing the basis of the ethical obligation to keep promises, it is not incredibly important for discussions of the motivation to keep promises, or for studying promising behavior. Regardless of whether the right is social or moral, we can use the idea that promises transfer rights to explain promising behavior. In turn, the fact that we can use the authority-transfer account to explain promise-keeping and third-party punishment of promise-breaking and contract-breach supports the authority-transfer account of promising. In the next chapters, I will explore the behavioral data on promise-keeping and contract-breach, especially with regards to punishment. I will argue that the authority-transfer view of promises best explains these data.

CHAPTER 2: SOCIAL RIGHTS, MORAL RIGHTS, AND MORALIZED SOCIAL RIGHTS

I. INTRODUCTION

In the previous chapter, I argued that we should understand promises as a transfer of rights. When you promise, you are essentially giving the promisee a right to your future action, and in turn taking on a duty to perform. However, I did not take a position on whether the right transferred is a moral right or an informal social right. In this chapter, I set up a framework for answering this question. I will argue that promising is a social practice, supported by a set of social norms, and thus the rights transferred are social rights, but that this practice has become moralized, so that the social rights take on a moral character as well.

Before I argue that promising is the transfer of moralized social rights, I will clarify two key points about social and moral rights. First, I will discuss what it means to have a right – what rights are, and what is entailed in holding them – and why it is important to think of promises as rights-transfers. I will then discuss the norms upon which rights are based, namely, social norms and moral norms. I will show how the norms supporting promising fulfill the definition of a social norms to a limited degree. I will explain the divergence using the idea that promising is supported by a set of moralized social norms, rejecting the idea that it is simply a moral phenomenon. Lastly, I will connect the idea of a moralized social norm to the

notion of a right, and conclude that promising is the transfer of a moralized social right. I will propose a few ways of differentiating an essentially moral right from an essentially social right.

II. WHAT DOES IT MEAN TO HAVE A RIGHT?

In his work *Fundamental Legal Conceptions as Applied in Judicial Reasoning*, W.N. Hohfeld developed a schema for characterizing the concepts used legal discourse about rights. I will first describe this schema, then argue that it is useful not just for characterizing legal rights, but also for characterizing any system of rights or rules, including moral and social norms. This framework will aid the analysis of promises because it will show us how rights are connected to – and dependent upon – norms (or rule-based obligations), and how vastly these rights can vary in their form, depending on the content of the norm.

In Hohfeld's schema, rights are bundles of 'incidents', or sub-rights that can be combined to make up more complex rights. Each of these incidents has a corresponding duty, or lack thereof. Incidents are first divided into two levels, first-order and higher-order. First-order incidents include claim-rights and privileges; higher-order incidents include powers and immunities. Since higher-order incidents are essentially rights that govern first-order incidents, and work much the same way, I will focus on the first-order incidents. A claim-right is a claim to a specific good or action from a specific agent or set of agents. A privilege is the opposite of a claim-right: it is the absence of someone else's claim-right to a good or action. According

to Hohfeld, “A [claim-]right is one's affirmative claim against another, and a privilege is one's freedom from the right or claim of another.”⁶³

Furthermore, each of these incidents also has a corresponding duty owed to the right-holder. In Hohfeld's terminology, a claim-right is correlated with a duty, while a privilege is correlated with a 'no-right'.⁶⁴ If a claim right takes the form 'X has a claim on Y to Z', then the corresponding duty has the form 'Y has a duty to X to Z'. If a privilege takes the form 'X has a privilege with regard to Z', then a no-right has the form 'Y has a no-right with regard to Z'. Essentially, the existence of a claim-right entails the existence of a duty; the existence of claim-right-holder entails the existence of at least one duty-bearer. The existence of a privilege entails the non-existence of a particular claim-right, and thus the non-existence of a particular duty. If I have a claim on you, you have a duty to me; if I have a privilege to an action or good, you cannot have a claim to it, and I cannot have a duty.

Hohfeld illustrates his concepts of claim-rights, privileges, and their corresponding duties using the example of rights over land. If I have a right to exclude you from my land –a claim on you to the action of not entering my land – you have a duty not to enter. But I might also have a positive right, or privilege to use the land myself. In this case, my privilege means that I do not have the same duty not to enter as you have.⁶⁵ It is important to note, as well, that these two

⁶³ Hohfeld 1964, p. 60

⁶⁴ Hohfeld 1964, p. 36

⁶⁵ Hohfeld 1964, p. 39

incidents, conceptually, may come apart. I may have a claim on you to stay off my land without having the privilege to enter the land myself. Whatever right I have to this land may include either or both of these incidents. Higher-order incidents may also be bundled with these rights, or not. Powers might include the right to amend, waive, or transfer my claim or privilege to the land. Having an immunity would mean that no one else has one of these powers. In addition, powers and immunities can be separate. Having the power to waive does not guarantee having the power to transfer.

Legal rights, or more properly, rights-bundles, are formed by combining sets of the four basic incidents and their variations. The right to vote, for instance, contains a claim and a specific type of power. I have a claim on the American government to vote, and a power to accept or waive that claim to vote. However, I do not have a privilege to exclude others from voting, nor do I have a power to transfer my vote to someone else. I also lack an immunity in some respects – the government can take away my right to vote if I commit a felony, renounce my citizenship, or otherwise fail to meet voting qualifications.

There are two important points to take away from Hohfeld's analysis. First, every duty will have some corresponding right, and vice versa. Hence, if there is a perceived obligation, there should also be a perceived right. But the analysis does not rule out that one should take priority over the other. That is, a right might be grounded in an obligation, an obligation might be grounded in a right, or both may

have the same priority. Second, not all rights have the same form. Rights can vary in how incidents are bundled. For instance, some rights might include privileges, some may not; some rights might include immunities against waiving without the privilege to waive, and so on. The exact bundle must be determined by whatever grants the right in the first place.

In his characterization of rights and rights-bundles, Hohfeld was only describing legal rights. However, I believe that his categorization can apply to any set of rules, whether the rights within those rules are explicit or not. I ground this claim in the work of H.L.A. Hart, especially his paper “Legal and Moral Obligation”. In this paper, Hart discusses the use of legal concepts, like obligation and rights, in the moral realm. He first explores the authority and grounding of legal obligations and rights within the law, distills the necessary criteria for legal rights and obligations, and then applies these criteria to the moral realm to determine what counts as a moral obligation or right. I will briefly outline his argument, then show how it applies to my own.

Legal rights and obligations, according to Hart, are, usually, deliberate human creations, which can be imposed, assumed, modified, destroyed, or released.⁶⁶ This conception maps on to Hohfeld’s concept of legal rights and obligations quite well -- within a system of law, a government or persons may have various powers and liabilities over duties and claims. Laws thus confer rights-bundles on governments

⁶⁶ Hart 1958, 84

and on individuals. The questions, then, is where these rights-bundles come from: how can legislative actions, mere words, have legal effects?

As we saw in the discussion of Hohfeldian legal concepts, in order for a legal claim or duty to be formed, someone must have a second-order right, a power, to create it. Thus, it must be that the legal system includes rules that state that particular rules create particular obligations and rights.⁶⁷ Powers are prior to claims and duties. But then how are these powers created? It seems that each power would require a higher-order power to create it, leading us to an infinite regress.

Hart rejects this regress by referencing the actual practice of the social group within which the laws are contained: these powers exist insofar as they “refer to the *actual practice* of the particular social group whose legal system is under consideration.”⁶⁸ In other words, there must be a norm of legal obedience, such that, in general among the social group, reference to the law is generally taken as a reason to follow the law, and disobedience is taken as a reason for criticism.⁶⁹ Common law gives us both legal obligations and obedience – reasons to obey the law – because it is grounded in our customary rules or practices.⁷⁰

⁶⁷ Hart 1958, 89

⁶⁸ Hart 1958, 88

⁶⁹ Hart 1958, 90

⁷⁰ Hart 1961, 103

Hence, if legal obligations and rights are to exist, they must be based on the actual social practice of the group within which they exist. That social practice must also contain the possibility of sanctioning violation, since part of the description of a social practice contains the requirement that deviation from the law is a reason for criticism. Now, the sanctions are certainly not formal sanctions, since they must exist before formal institutions in order to ground those institutions. They are informal sanctions, including even reactive attitudes like resentment and indignation or approval. These types of sanctions, according to Peter Strawson work to “inhibit or at least to limit our goodwill” toward offenders, while still recognizing that they are part of our moral community.⁷¹ Though mild, these sanctions, for Hart, help create the foundation for a working formal institution.

In addition, according to Hart, legal obligations and rights must have “possible independence of content.”⁷² Independence of content means that the right or obligation comes from the “use of the procedure by the appropriate person in the appropriate circumstances,” not from the specifics of the content of the case, such as the specific identities of the people involved.⁷³

If there are obligations and rights in the moral realm, according to Hart, they must fulfill all three of these criteria – they must be supported by social practice, allow for sanctioning, however mild, and allow for independence of content.⁷⁴ The

⁷¹ Strawson 1974, 21-22

⁷² Hart 1958, 100

⁷³ Hart 1958, 102

⁷⁴ Hart 1958, 100

same arguments that apply to legal rights and obligations apply to all rights and obligations. They must be based in a social practice so that they can be created. They must allow for sanctioning because the social practice allows for the possibility of sanctioning. And they must have independent content in order to be part of a system of rules. Hart then argues that the concepts of obligation and rights can apply to some, but not all, of the moral realm, though they apply to the entire legal realm.⁷⁵

Hart's concept of rights and obligations, as we have seen, concords with Hohfeld's. They agree that the basic units of law are (claim-)rights and corresponding obligations (or duties), with powers that allow individuals and governments to transfer, amend, accept, and waive those claims and duties.⁷⁶ Hence, Hohfeld's schema fits into Hart's concept of rights and obligations, and Hart's arguments should apply to Hohfeld's concept of law as well, at least with regard to claim-rights, duties, and powers. Thus, Hohfeld's categorization applies to any system that meets Hart's three criteria – support of the social practice, possibility of sanction, and possibility of independence of content.

Fortunately, any system of rules, whether formal or informal, can theoretically meet Hart's three criteria. According to Hart, obligation implies the existence of a rule, though not all rules imply obligation. Rules only properly imply

⁷⁵ Hart 1958, 83

⁷⁶ Hart 1961, 81

obligations – and thus corresponding rights – when “the general demand for conformity is insistent and the social pressure brought to bear upon those who deviate or threaten to deviate is great.”⁷⁷ This social pressure, or sanctioning, in turn exists because the rules are socially beneficial, allowing us to coordinate even when we desire to do otherwise. Rules supported by social pressure, especially when thought to be necessary to the social system or contradict the desires one might otherwise have, are properly duties,⁷⁸ and duties, as we have seen, have correlated claims.

This essentially sums up Hart’s first two criteria: the rules must be based in, or formed according to, a social practice, and must allow for sanctions. Hence, a system of rules contains obligations and rights if and only if there is social pressure to conform to those rules. Again, it may be mild pressure, like praise and blame or approval and disapproval. This social pressure may exist regardless of whether one experiences the pressure, or is motivated to comply with the rules because of the pressure. But as long as the pressure exists to back up the rules, the rules imply rights and obligations. And thus any system of rules could, in theory, contain rights and obligations. It simply depends on the social group in which they are present.

There is another sense in which rules, especially informal rules, could create informal rights. As we will see in the next section, on certain conceptions, the

⁷⁷ Hart 1961, 86

⁷⁸ Hart 1961, 87

existence of some informal rules is dependent on the existence of particular expectations with a given social group. According to Cristina Bicchieri, expectations within informal rule systems (in this case social norms) often *seem* to be legitimate, and hence seem to imply a right to some behavior, and a corresponding obligation on behalf of others. Even if there is no sanction or threat of sanction, a rule that an agent accepts as legitimate will be perceived as an obligation entailing a right.⁷⁹ In the case of certain social norms, and moral rules as well, this perception of legitimacy, or legitimate expectations, may be enough to make a rule into an obligation, or its fulfillment into a right.

This analysis of rights is important because, as we have seen in chapter one, promises are best explained as rights-transfers. But basing a theory of promising on a theory of rights has other advantages as well. Rights theory gives us several tools that make the analysis of promises – and predictions regarding promissory behavior – easier. First, Hohfeldian bundle theory allows rights to be constructed in many different ways. Hence, it gives us a tool to explain how promises can differ in different cultures and situations, or between different people. They are simply transferring different bundles of rights. This will be especially important when we explore the use of promises in social-change interventions in chapter four.

Second, as we have seen in chapter one, framing promises around rights can help us to explain several aspects of promising, including why individuals make and

⁷⁹ Bicchieri, forthcoming, 4

trust promises, not just why they keep promises. For instance, according to David Owens, the fact that promises can give someone else legitimate authority – or rights – over our actions often provides a reason for making promises in the first place.⁸⁰ The explanation of why we make promises can be cashed out in terms of duties, but it would be more cumbersome. It is simpler to think of ourselves wanting to give rights to others than wanting to give duties to ourselves with regard to others.

This point about simplicity leads to a final advantage of a rights-based theory of promises. A rights-based theory correctly emphasizes the promisee, and the relationship between the promisee and the promisor, in a way that a duty-based theory does not. Since promissory rights belong to the promisee, whereas duties belong to the promisor, a rights-based theory treats the promisee as fundamental. And a promisee-centric theory theory better describes the actual experience of promising: our promises are fundamentally about others, not about ourselves.

Further, the description of promises as transfers emphasizes the relational aspect of promising, another important part of the experience of promising. Giving a duty to oneself – even if that duty regards someone else – is at best a cumbersome way to describe creating a relationship with another person. Describing the relationship as a rights-transfer is simpler because it already includes the interpersonal, relational aspects of promises. However, it is not yet clear what kinds of rights are transferred when one promises. In the next section, I will explore the

⁸⁰ Owens 2006, 70

differences between rights grounded in social norms and rights grounded in moral norms.

III. SOCIAL NORMS AND SOCIAL RIGHTS

To determine whether promises transfer social rights or moral rights, we need a conception of how informal social rights differ from moral rights. Based on the idea of rights outlined above, social and moral rights will differ according to the rules upon which they are based. Thus, we need a conception of how social rules differ from moral rules – in other words, we need to contrast social norms with moral norms. We can then analyze the practice of promising to determine whether it is based upon moral norms or social norms, according to these definitions, and explore in what sense these norms imply moral or social rights.

I will begin this discussion of social and moral norms by defining social norms, using the theory of social norms set forth by Cristina Bicchieri. Using Bicchieri's arguments, I will show that this conception of social norms is the most appropriate for my purposes because it allows us to clearly distinguish behaviors that are in accordance with social norms from those that are not. I will argue that, using Bicchieri's model and experimental evidence, promises appear supported by social norms, at least in many circumstances. However, there are circumstances in which the behaviors surrounding promising do not adhere to the model of social norms. In the next section, I will explore the senses in which promising could be

supported by moral norm, and conclude that the data is best explained by a concept of promises as supported by moralized social norms.

Using this model, social behaviors can be classified into several types: customs or habits, descriptive norms and conventions, social norms, legal norms, and moral norms. Customs, or social regularities, are not (socially) normative, and are not properly social rules. They are simply behaviors that most people within a population do, regardless of what others do, because the benefit from it. Examples include carrying an umbrella, or wearing shoes. We all engage in these behaviors because we all want to stay dry or protect our feet, not because anyone else carries an umbrella or wears shoes. It is a collective behavior because we have, in these respects, similar interests.⁸¹ Since customs or habits are not social rules, they will not factor into any further discussion.

Legal norms are also not particularly relevant to our discussion of promising and social rules. Legal norms are formal social rules, characterized by formal sanctions and formal secondary rules for rule-creation.⁸² Though promises have a legal analogue -- namely, contracts -- they are not themselves legal norms. We do not keep our promises because they are imbedded into a formal legal system. Hence, I will leave the discussion of legal norms until I discuss contracts as a form of promising in the next section.

⁸¹ Bicchieri (forthcoming), 17.

⁸² Bicchieri (forthcoming).

A descriptive norm or convention is a behavioral regularity for which the preference to conform is conditional by social expectations. They are supported by what Bicchieri calls 'empirical expectations', or my beliefs about the actual behavior of other members of my reference network.⁸³ Unlike in the cases of customs and legal norms, in which behavior is motivated by one's own interests or by formal sanctions, respectively, regardless of what other members of the social group do, in the case of descriptive norms, one's behavior depends on what he or she expects others to do. In this sense, behavior is *conditional*.

Lastly, Bicchieri defines a social norm by referencing two types of social expectations: empirical and normative expectations. In her account, a behavioral rule R in situations of type S is a social norm for an individual member of population P iff:⁸⁴

1. He or she knows that R exists and applies in S.
2. He or she prefers to conform to R in S on the condition that
 - a. He or she believes enough members of P conform to R in S (empirical expectation)
 - b. He or she believes enough members of P believe he or she should conform to R in S (normative expectation) and

⁸³ Bicchieri 2006, 31-32.

⁸⁴ Bicchieri 2006, 11.

- i. either believes that other members of P will informally sanction behavior (social sanction), or
- ii. believes the normative expectations of others are legitimate (legitimacy).⁸⁵

This definition of social norms builds on the definition of descriptive norms and conventions, adding normative expectations, along with sanctions and legitimacy, to empirical expectations. Like conventions, then, social-norm behavior is conditional on social expectations; in this case, both normative and empirical expectations.

To illustrate the difference between a convention and a social norm, we can expand the language example. Suppose, for example, that I do not use language to communicate; instead, I use it to make a political statement. I believe that Esperanto is a superior language, as should be accepted as the standard language in all nations. Since I am no longer trying to communicate with you, the fact that you do not speak Esperanto is not enough to motivate me to speak our common language, English. Hence, English is no longer the language convention between us.

In this world of political-language, as opposed to communication-language, I have no motivation to speak your language, and hence we cannot coordinate. In order to coordinate on a language, we could use a social norm. In this case, even if empirical expectations are not enough to motivate me, normative expectations would be. Perhaps I believe that you believe I should speak English, and that you

⁸⁵ Bicchieri 2006, 15.

will sanction me if I do not speak it. Then I will be motivated to speak English, we will be able to coordinate. Or perhaps you believe that I believe you should speak Esperanto, and you think that this normative belief you have attributed to me is legitimate. You will then be motivated to speak Esperanto.

Of course, this is a relatively unbelievable case, but it does serve to illustrate the differences between a social norm and a convention. In the case of a social norm, empirical expectations are not enough to motivate conformity; behavior is also conditional on normative expectations supported by sanctions or a belief in legitimacy.

The differences between conventions and social norms also help to show why descriptive norms and conventions are not obligation- or right-granting. As we saw in the last section, for a rule to create rights and obligations, one of two conditions must be met. First, in Hart's account, rules create obligations only when there is sufficient social pressure. This social pressure corresponds to the possibility of sanctions within normative expectations. Second, in Bicchieri's account, rules can create rights and obligations when the rules are perceived to be legitimate. This corresponds to the legitimacy condition within normative expectations. Descriptive norms and conventions cannot by definition fulfill either of these conditions -- if they did, they would become social norms. Hence, descriptive norms and conventions cannot create rights or obligations, but social norms do create them.

Since descriptive norms, conventions, and customs cannot be right-granting, we can conclude that none of them are the type of rules that govern promising behaviors. In the next section, I will show that behaviors surrounding promising, in general, adhere to Bicchieri's model. But before I do so, I would like to explain why I am using Bicchieri's account of social norms and other social practices, and argue that it is the best account to use for the purposes of categorizing behavior. I use this account because it is a generally good model of socially normative behavior, and because it is testable.

First, Bicchieri's account allows us to model social behavior better than other models, such as social preference models, because it includes conditional preferences. The inclusion of conditional preferences allows for changes in behavior -- it is a dynamic model. If we assume, for instance, that individuals have fixed social preferences, we will not be able to account for conformity to particular rules in some situations but not in others. Bicchieri's model can explain this: since norms are based on individual expectations, individuals may or may not perceive a social norm in a particular situation.⁸⁶

Second, and more importantly, Bicchieri's model makes testable predictions that allow us to classify behavior.⁸⁷ Since preferences are conditional, we can test for the existence of a social norm by changing those conditions -- or more precisely,

⁸⁶ Bicchieri 2006, 128.

⁸⁷ Bicchieri 2006, 3, 102.

by manipulating expectations. In addition, we can categorize behavior into the various types of regularities by measuring these expectations and their influence on behavior. Since I am interested in categorizing promising as either moral or social, a model allows for such a categorization, and allows for testing of that categorization, is particularly useful.

IV. PROMISING IS SUPPORTED BY SOCIAL NORMS... MOSTLY

To test whether behaviors surrounding promising are social norms, we should be able to test whether they fulfill Bicchieri's conditions for the existence of social norms. Namely, promising would have to be conditional in the sense that individual preferences to conform are determined by the behaviors and perceived beliefs of their social group. We can test whether promising is supported by social norms by manipulating those conditions, and observing whether individuals continue to make, trust, and keep promises. In reviewing the experimental data, I will show that promising behaviors, and in particular promise-keeping, are conditional, but only somewhat conditional. That is, more individuals prefer to keep their promises even without empirical and normative expectations than we would expect. In the next section, I will explain this divergence from the definition of a social norm, while maintaining that promising is, fundamentally, based on social norms.

There are two major studies that help to establish promising-keeping behavior as a social norm. The first is Orbell, van de Kragt, and Dawes (1988),

which used promises to explain why communication increases cooperation in social dilemma games. Orbell and colleagues essentially set up a prisoner's dilemma between groups. Each group of 14 subjects was told that they would be split into two groups of seven, and each individual would be given a \$5 promissory note. Individuals could keep the \$5, give it to their own group, or give it to the other group. If given to their own group, each of the six other group members would receive \$2. If given to the other group, each of the seven members of the other group would receive \$3. Before being divided into groups, the 14 subjects were allowed ten minutes of discussion.⁸⁸ Half of the seven-person groups were allowed to discuss again before making their decisions. The game was set up as a dilemma between groups – each group was better off if no members gave to the other group, but both groups did best if they gave to each other.⁸⁹

Orbell and colleagues also kept track of how many, if any, of the group members made promises to give to the other group. In cases where all members promised, an average of 84% gave to the other group, compared to 58% in cases without universal promising. But the important finding was from groups in which some individuals promised, but others did not. In groups without universal promising, there was a correlation of only 0.09 between the number of subjects who made promises and the number of subjects who gave to the other group.⁹⁰ This

⁸⁸ Orbell et al. 1988, 815-816

⁸⁹ Orbell et al. 1988, 816

⁹⁰ Orbell et al. 1988, 817

suggests that only universal promising increases cooperation, which implies that subjects are willing to break promises when other members of their group do not promise.⁹¹

The willingness to break promises when promising is less than universal supports the idea that promising-keeping is a social norm. Recall that, in order to be a social norm, promising-keeping behavior must be conditional on empirical and normative expectations. The fact that subjects break promises because others did not promise shows that promise-keeping is somehow conditional on expectations about other people: subjects only take their promises seriously when others promise as well. When others do not promise to give to the other group, subjects expect that they will not give. The expectation that others will not give informs their choices not to give, and to break their promises.⁹²

These data show us that promising-keeping depends on empirical expectations: subjects do not think the rules of promising apply if they expect that others will not apply the rules either. It does not show, however, that subjects have normative expectations with regard to promising-keeping. Unfortunately, there is not much research available to establish that promising is conditional on normative

⁹¹ This study is somewhat complicated by the in-group-out-group dimension. However, given that the main conclusion I would like to draw is about the importance of universality, and not about overall giving to in-groups and out-groups, the in-group element is not relevant. The relevant data concerns promising and giving to the out-group.

⁹² For further discussion, see Bicchieri 2004, pages 111-112.

expectations. However, there is some research suggesting that keeping contracts is at least somewhat conditional on normative expectations. If we treat contracts as the legal analogue of promises, and stipulate that individuals view contracts as promises,⁹³ we can use this data to support the idea that promise-keeping is a social norm, and thus that promising in general is supported by social norms.

The support for the conditionality of contract-keeping on normative expectations comes from Eigen 2012. In this study, Eigen tests whether subjects view form-adhesive contracts as binding, and, if not, what sort of prompts would cause them to see these contracts as binding. A form-adhesive contract is a contract that does not allow for negotiation, only consent. To test whether subjects view these contracts as binding, Eigen had subjects agree to perform an undesirable on-line task under several conditions: a conventional 'terms and conditions' form, an agreement with a non-substantive choice, an agreement with a substantive choice, and a control with no agreement.⁹⁴ To test the prompts, when subjects tried to exit the task, one of four prompts would appear. Subjects would either be reminded that they signed a legal contract (legal), that they would not receive compensation without completing the task (instrumental), that they had made a binding promise

⁹³ I will discuss contracts and psychological promises more in a later chapter. For now, it should suffice to say that there are several ways in which individuals treat contracts as promises: they punish contracts-breakers more than stipulated by law, they justify punishment with promise-based reasons, and they differentiate between loss-avoidance and breach-to-gain in punishments. For more discussion, please see Wilkinson-Ryan and Baron 2009.

⁹⁴ Eigen 2012, 10-12.

(social-normative), that many others had complied (social-empirical), or they would simply be asked to finish (generic).⁹⁵

For my purposes, the crucial element of this study is how the different prompts affect compliance. If promises are social norm-based, the social-normative and the social-empirical should have the greatest effect on compliance. That is, if more subjects comply with the contractual promise when they know that others comply, or when they are reminded of a normative rule to comply, then compliance is conditional on empirical or normative expectations. If both the social-normative and social-empirical prompts have a large effect, it will show that contractual promise-compliance is a social norm.

Indeed, Eigen found a significant difference among the social-empirical prompt, the social-normative prompt, and the other prompts. Of the subjects who tried to exit the task – which was the vast majority of subjects – 70.5% returned after receiving the social-normative prompt and 60% returned after the social-empirical prompt. Only 43% returned after the instrumental and 46% after the legal prompt, compared to 50.5% of those who received the generic prompt.⁹⁶ The

⁹⁵ Eigen 2012, 14-15. It should be noted that Eigen refers to what I call the ‘social-normative’ prompt as the ‘moral’ prompt. I call it social-normative because it reminds the subject of the normative beliefs that others have by telling him he has a binding promise. In this way, it signals to the subject the experimenter’s normative beliefs, which he may or may not share. I also renamed Eigen’s ‘social’ prompt to ‘social-empirical’ to avoid confusion, and make it clear that Eigen is only testing empirical expectations with this prompt.

⁹⁶ Eigen 2012, 18-19

fact that 70.5% returned after receiving a social-normative prompt signals the presence of a normative expectation that affects behavior.

On the surface, then, promising-keeping appears to fulfill the conditions of a social norm: it is a behavior conditional on empirical and normative expectations. The problem, however, is that it does not appear to be conditional at all times, particularly on normative expectations. There are several studies that show that promises are still kept even when social norms cannot be enforced – that is, when there is no way to sanction, and normative expectations are not relevant – and that promises may be broken for seemingly irrelevant reasons. I will briefly outline two of these studies: Bicchieri, Lev-On, and Chavez 2010; and Ostrom, Walker, and Gardner 1992.

Ostrom and colleagues discuss the idea, dating back at least to Hobbes, that promises in prisoner's dilemma-type games need to be externally enforced to be kept. The basic point, which they argue against, is that “communication without a change in the payoff function does not eliminate a Nash equilibrium”, and that there must be an external enforcer to change the payoff function.⁹⁷ Ostrom and colleagues found that while sanctions do increase cooperation, and subjects will pay to sanction, cooperation occurs even in one-shot games, as long as there is

⁹⁷ Ostrom et al. 1992, 404.

communication and agreement.⁹⁸ This disproves the idea that an external enforcer is necessary, though it may be helpful.

Bicchieri and colleagues expand on this work regarding one-shot, anonymous promising and cooperation. They studied cooperation in one-shot trust games, varying the method of interaction. Subjects either communicated with each other face-to-face, or mediated by a computer; they were either allowed to discuss the game, have irrelevant discussions, or not communicate at all. Relevant communication increased cooperation in both the face-to-face condition and the computer-mediated condition. Trust did not differ by the method of communication, but reciprocity did – reciprocity was greater in the face-to-face condition.⁹⁹ The general finding is that while the content of the message greatly increases cooperation, media richness has less effect, if any at all.¹⁰⁰

This work with computer-mediated communication, in which expectations are more difficult to establish, and subjects are more anonymous, again casts doubt on the idea that promising is supported by social norms, as opposed to other rules. If subjects do not actually interact with their partners, and do not know who they are, it is unclear why social expectations about promises would matter to them. In that case, it seems that promising behavior, such as promise-making, -keeping, and –

⁹⁸ Ostrom et al. 1992, 413.

⁹⁹ Bicchieri et al. 2010, 140.

¹⁰⁰ Bicchieri et al. 2010, 139.

trusting, is, in some sense, unconditional. That is, these behaviors are not conditional on whether the social interactions matter for the individual.

Bicchieri and colleagues explain the apparent unconditionality by positing that “when a social norm becomes situationally salient, it causes a shift in the individual’s focus, provides the individual with the right kind of expectations, and thus directs her actions toward norm compliance.”¹⁰¹ In other words, once triggered, norm-obedience becomes somewhat automatic. Since agents are focused on the norm, they will expect others to comply, even without an enforcement mechanism. To show that the behavior is unconditional, we would have to show that individuals prefer to comply even when the lack of similarity to normal situations is made salient. That is, they would prefer to keep their promises even when they are reminded that the promisee is a computer. And this further assertion has not yet been shown.

Though we can explain seemingly unconditional promise-keeping within a framework of conditional preference, and have reason to believe that promising is supported by social norms, one still might think that promising has some moral characteristics. After all, we often categorize promise-keeping as more, and promise-breaking as immoral, and may want to do so without reference to social practices. For instance, Seana Shiffrin attempts to formulate “an account of a binding promise” that does not “depend on... the promisee’s belief in

¹⁰¹ Bicchieri et al. 2010, 127.

performance.”¹⁰² Even if promising appears to be supported by social norms, we may want to know if a theory like Shiffrin’s is behaviorally plausible.

Further, if we show that promising is not supported by moral norms, it will lend credence to the idea that it is supported by social norms by ruling out alternatives. If promising has a moral element, behaviorally speaking, there are two options: that promising is based on moral norms, and that promising is based on moralized social norms, and the moralization explains the divergence. I will eventually argue for the latter explanation.

V. IS PROMISING SUPPORTED BY MORAL NORMS?

In order to support the idea that promising is based upon social norms, I will reject an alternative approach: the idea that promising is based upon moral norms. I will explore two possible explanations. First, it may be that promising is not social norm-based at all, but rather based in moral norms. I reject this explanation by showing that promising behaviors do not fit any of the most common definitions of moral norms, in-principle unconditionality and universality. The second explanation is that promising behaviors are both moral norms and social norms – that is, a moralized social norms. Thus, we can keep the idea that promises are in some sense moral, behaviorally speaking, while still using a social norm-based account. In the final section, I will illustrate this view by using an analogy of

¹⁰² Shiffrin 2008, 486.

promises to manners and modesty. I will conclude with a final, full account of promises as the transfer of moralized social rights.

V.A. IN-PRINCIPLE UNCONDITIONALITY

To further argue that promises are supported by social norms, we can try to show the opposite: we may hypothesize that promising is not supported by social norms, but rather by moral norms. We then need an account of what constitutes a moral norm, and to determine whether promising behaviors fit into this account. In this sub-section and the next, I will explore two ways of distinguishing a moral norm from a social norm, and then apply these definitions to the behaviors surrounding promising.

First, Bicchieri distinguishes moral norms from social norms through the concept of unconditionality. In her view, whereas social norms depend on the normative and empirical expectations of others, and our behavior is conditional on these expectations, moral norms do not depend on these expectations -- our moral behavior is not conditional on the expectations of others.¹⁰³ One way to explain, and distinguish in practice, this unconditionality is by means of justification. In the case of conditional behavior, we often justify our actions by appealing to the actions of others. For instance, when I justify my table manners to people of different cultures, I might simply say, "this is how everyone eats in America." My table manners are conditional upon the actions of my reference group. Furthermore, for conventions

¹⁰³ Bicchieri and Muldoon 2011

like etiquette, this type of explanation is generally accepted. However, when I try to justify breaking moral norms like the norm against the physical abuse of innocents, I generally do not refer to the fact that others violate this norm as well, nor would that be an acceptable justification.

There are a few problems with this account of moral norms. First, a lot of behavior that we want to call moral does not appear to be unconditional; that is, we fail to act according to moral norms all the time. Bicchieri uses the example of giving to charity:

One may value giving to charity, but if one is reasonably sure very few others give to charity, then one's preference for charitable giving may be superseded by another preference, for example spending the money in a fancy restaurant.¹⁰⁴

The idea is that giving to charity is certainly a moral norm, but it is apparently not unconditional – the agent's preference here is conditional on his beliefs about others. Shaun Nichols argues for this view of moral norms in "Emotions, Norms, and the Genealogy of Fairness". According to Nichols, even if a norm is moral for an individual, there are several reasons why it might appear conditional. First, different expectations may lead me to weigh my competing desires differently. I might have an unconditional desire to give to charity, but it appears conditional because it is often *outweighed* by other desires, especially when others do not give

¹⁰⁴ Bicchieri 2010, 300.

to charity. In this case, the desire is unconditional, but its weight in decision-making is not.¹⁰⁵

This explanation of conditionality of moral norms appears to confuse desires with preferences, and beliefs with norms. According to Nichols, “what Bicchieri has in mind by ‘preference’ is something like desire or liking,” and so what is unconditional in her view is the *desire* to follow the norm, not the action itself.¹⁰⁶ And so Nichols substitutes the concept of desire for the concept of preference. A desire then supports a *belief* about which action one should choose. And so an unconditional desire supports a belief about correct action directly, whereas a conditional desire supports the belief only when the correct beliefs about others’ beliefs and desires are present.¹⁰⁷

But preferences, according to Bicchieri, are more akin to choices than they are to desires. Preferences are all-things-considered decisions, not the desires that are weighed when we make those decisions. Furthermore, norms are actions – they are collective behaviors – not beliefs. A moral norm of giving to charity is not a collective belief that we should all give to charity, but rather a collective behavior: we all *do* give to charity. Hence, Nichols’s suggestion is that moral norms are supported by unconditional *desires*, not unconditional preferences. The problem with this suggestion is that it cannot behaviorally differentiate between a moral

¹⁰⁵ Nichols 2010, 6

¹⁰⁶ Nichols 2010, endnote 8.

¹⁰⁷ Nichols 2010

norm and a social norm. It is not testable. I may have an unconditional desire, for example, to always give to charity, but my charity-giving behavior may be conditional on many other factors. However, the general objection may still apply to preferences and norms as Bicchieri understands them: there are norms that we consider moral, but sometimes are outweighed.

Second, Nichols argues that we may be sensitive to the expectations of others, and they may affect our behavior, without our behavior being conditional on them. Namely, “people might regard others’ behavior as providing information about whether the ... norm applies” in a particular case.¹⁰⁸ In circumstances in which we use behavior to signal the presence of the norm, the norm might appear conditional when it in fact is not. The norm is thus unconditional in situations in which it is triggered, but we may not always know when it applies.

Because of these criticisms, Bicchieri adds that while moral norms may not always be unconditional, they are always at least ‘in-principle’ unconditional.¹⁰⁹ Moral norms are unconditional when they are triggered, or when we realize they apply. In addition, in extreme cases, as in a Hobbesian state of nature or severe crisis, we may justify our deviation from a moral norm by referring to the behavior of others, or deviate from the norm because others are. But in these cases – cases of

¹⁰⁸ Nichols 2010, 6.

¹⁰⁹ Bicchieri 2006, 20.

survival – we still may feel guilty and recognize the immorality of our behavior. It is not fully justified by referring to the behavior of others.¹¹⁰

This version of unconditional preferences can also take Nichols's first worry into account. Having an unconditional preference when we *do* follow moral norms does not mean we always follow moral norms. We might have an unconditional preference to follow moral norms, provided that those norms are not overridden by other considerations, such as weakness of will, other moral norms, or other norms in general. In all of these cases, the moral norm is not conditional on social expectations -- we lack a preference for it for other reasons. For instance, our preference to give to charity may be unconditional with regard to social expectations, but still susceptible to weakness of will.

Second, it looks like unconditionality is not enough to distinguish moral norms from other sorts of behavioral regularities. While unconditionality can separate moral norms from social norms, it cannot distinguish between moral norms and customs. Recall that customs are also unconditional -- we behave in accordance with customs because it is in our personal best interest, regardless of what others do. We thus have no way, as of yet, to distinguish the moral from the prudential, though we often perceive these as different.

There are several ways to salvage unconditionality by adding extra criteria to distinguish moral rules from prudential rules. First, we can distinguish them with

¹¹⁰ Bicchieri (forthcoming), 28.

regard to subject. Prudential rules regulate my actions in accordance with my own interests; moral rules regulate my actions in accordance with the interests of some other or others. They are, in this sense, other-regarding. On this view, moral norms would be socially unconditional other-regarding rules. Of course, this view will only be acceptable if you accept that there are no moral rules that pertain to the self. Many moral theories, including utilitarianism and Kantianism, hold that there can be moral duties to the self.¹¹¹

We might also want to differentiate moral norms from prudential rules on the basis of acceptable justification for breaking these rules. Bicchieri alludes to this in her discussion of the contrast between moral norms and social norms. In the case of a social norm, we accept the behavior or beliefs of others as justification for our own actions. In the case of prudential norms, we accept any sort of justification. Generally, the justification for, say, not wearing a coat in the winter boils down to ‘I don’t want to’. Desire or inclination is an acceptable justification for breaking prudential norms. In the case of moral norms, neither inclination nor the actions of others is (generally) enough to justify violation. Usually, only other moral norms can justify violation, though changing factual beliefs can justify an assertion that the norm no longer applies.

¹¹¹ For instance, Bentham held that promoting utility would include one’s own utility. See Crimmins 2015. Kant also described duties to the self, such as the duty not to commit suicide, or the duty to improve one’s talents. See Kant 1997, pp. 38-40.

The idea of justification, however, seems to dissolve into self- and other-regardingness. Acceptable justification must be backed up by *reasons* why we accept a certain justification, but not another. We may accept the 'I don't want to' justification for prudential norms, but not for moral norms, because prudential norms are self-regarding. Norms that are other-regarding may simply require more justification because they involve the interests of other people.

Or, we might want to distinguish moral norms from prudential rules on the basis of universality: individual moral rules do not differ in form or content across cultures, persons, or situations. They are part of human nature. I will discuss universality in the next sub-section. Third, we can distinguish these rules according to their phenomenology – moral rules would be those we perceive as moral. This is the position I will take when I argue that promising behaviors are both moral and social. However, it cannot be used to distinguish moral norms from social norms because we may perceive norms that are structurally social as moral; that is, norms may be moralized. Lastly, we could distinguish moral norms from prudential rules using content. We would simply stipulate the kind of content that moral norms have; for instance, we might say that all moral norms relate to harm. However, we must have a principled way to decide on content, so I suspect that this method will boil down to some other method, whether other-regardingness, universality, or phenomenology.

Regardless of how we differentiate moral rules from prudential rules, it is clear that promising is not based upon unconditional behaviors, or even in-principle unconditional behaviors. We have seen that promising behavior, especially promise-keeping, often does depend on the existence of social expectations, like empirical beliefs about the behavior of others. We want to explain why promising behaviors appear to be unconditional some of the time, though they are often conditional on social expectations. A definition of moral norms as unconditional behaviors seems to explain the opposite. It would explain generally unconditional behavior that is rarely conditional on social expectations, but often conditional on other obligations or needs. Promising behaviors do not meet this description.

V.B. UNIVERSALITY: APPLICATION AND UBIQUITY

In addition to unconditionality, we may differentiate moral norms according to universality. Universality can be explained in two ways: in terms of application, and in terms of ubiquity. I will explore each of these proposals. First, one might claim that moral norms are simply a subset of norms that we believe applies to everyone. Put differently, and perhaps more canonically, moral norms are categorical imperatives. Moral norms apply to everyone, at all times, regardless of their specific ends or desires. This type of universality is what Kant describes in the *Groundwork to the Metaphysics of Morals*. Kant distinguishes moral principles, or rules, from other principles through the structure of the imperative. An imperative, according to Kant, is either hypothetical or categorical. A hypothetical imperative

commands only conditionally on a set end: if you want end x, you should do y. It is a means-ends imperative, and so only holds conditionally on set ends.¹¹²

Kant rejects the idea that moral principles are hypothetical imperatives because he assumes that morality has a law-like structure. If morality is law-like, it must be able to command unconditionally and objectively. That is, morality commands everyone at all times and in all circumstances, without reference to any subjective states or desires. It is in this sense, according to Kant, that morality is universal. It is a system of law, or perhaps a system of law-like rules, that applies to everyone. This universality in application, at least in part, is what separates morality from other forms of rules.

Kant's version of universality is closely tied to Bicchieri's idea of unconditionality. Kantian universality – or unconditionality – holds that moral duties are not conditional on empirical desires or preferences. This definition of unconditionality overlaps with Bicchieri, who holds that moral norms are not conditional on social expectations or preferences. Since social expectations, desires, or preferences are a subset of empirical desires and preferences more generally, we can view Bicchieri's unconditionality as a narrow slice of Kantian universality or unconditionality. Hence, moral norms cannot be universal in Kant's sense without being unconditional in Bicchieri's sense. However, they unconditional without

¹¹² Kant 1997, 30-31

being universal – Kant’s form of universality requires more than Bicchieri’s unconditionality.

Elliot Turiel also uses a similar idea of universality in order to distinguish moral from conventional reasoning, but replaces contingency on empirical desires with authority-dependence. That is, Turiel’s version of universality, like Kant’s, holds that moral norms should apply like law, to everyone at all times. But instead of construing this application as a categorical imperative that applies regardless of your particular desires or preferences, he sees it as an imperative that applies regardless of whether an authority commands or forbids it. Though these two forms of universality overlap in practice, they are conceptually distinct. However, they share the common element that morality applies at all times, irrespective of various contingent desires or situations.

In his experiments, Turiel distinguishes moral reasoning in children by asking them whether an action is still wrong even if it is allowed by an authority. For example, a child may be classified as thinking that waiting in line is conventional because she believes it is wrong to violate it only if an authority says so. That is, if the teacher says to wait in line, one should wait in line, but if a teacher at another school does not demand queuing, then the students in that school do nothing wrong if they refuse to wait in line. On the other hand, a child views hitting as moral, on Turiel’s view, if she believes that hitting is always wrong, no matter what the

teacher says.¹¹³ In Turiel's words, "Morality, too, applies to social systems, but contrasts with convention in that it is not determined by existing uniformities."¹¹⁴ Morality always applies, no matter what the existing rules or behaviors are.

Turiel's idea of using universality to distinguish between moral and conventional is questioned by Royzman, Leeman, and Baron in their article "Unsentimental ethics: Towards a content-specific account of the moral-conventional distinction". They use Blair's work on psychopathy to show that, while children distinguish moral from conventional on the basis of universality, psychopaths do not.¹¹⁵ Hence, there must be something that children have, and that psychopaths lack, that allow them to make this universality distinction. In other words, universality of application is not enough to explain the distinction between the moral and conventional. We also need a deeper explanation of the reasons we classify some actions as universal in application, while others are not. Though universality of application may be necessary for distinguishing the moral domain, it is not sufficient.

There is a second version of universality with which we can supplement or replace universality of application, namely, ubiquity. Ubiquity is the type of universality John Mikhail uses in his theory of Universal Moral Grammar (UMG). In developing UMG, Mikhail focuses, in part, on the question of what constitutes moral

¹¹³ See Weston and Turiel 1980

¹¹⁴ Turiel 2002, 110

¹¹⁵ Royzman et al 2009, 160. See also Blair 1995.

knowledge.¹¹⁶ He claims that moral knowledge is fairly constant across cultures. For example, all languages have words for deontic concepts like permission, obligation, and forbidden. In addition, there are many moral prohibitions that appear to be ubiquitous, or exist in all cultures, including judgments in trolley cases.¹¹⁷ Mikhail uses this data to support his claim that we have a universal, or ubiquitous, moral grammar: there is a basic, underlying structure to our judgment and knowledge that we all share. Our moral competence is shared across humanity.

The idea that underlies Mikhail's theory is that we can differentiate what counts as moral based on its connection to universal moral grammar. That is, the moral domain is constituted by whatever rules are structured by our universal moral grammar. In that sense, that moral is delineated by ubiquity – morality is constituted by the rules, principles, or reasoning processes that we all share.

Ubiquity is also a way to check if a given rule is objectively moral. If a rule is objectively moral, the reasoning goes, any culture that does not have it, or that distorts it, will be immoral, or amoral. As a corollary, if we do not want to label any culture as immoral or amoral, the rule must be shared by all cultures. If we find that a culture does not follow this rule, or distorts it, and we do not believe that this culture is immoral or amoral, then we must conclude that the rule is not actually moral. This is the sense in which I will be using the idea of ubiquity. I will check

¹¹⁶ Mikhail 2007, 143.

¹¹⁷ Mikhail 2007, 144. See also Thomson 1976.

whether promising practices are moral practices by exploring whether they are followed by all cultures that we would consider to be moral.

Using these two versions of universality, we can disprove the existence of a moral right in several ways. First, we can show that the right in question is not applied universally – that is, it applies in some situations, or to some people, but not others. Second, we can show that the right in question does not exist in all moral cultures, or that it varies across cultures that we would consider moral. Using Hohfeldian incidents, we have seen that rights can differ in a two ways. First, rights-bundles are made up of various sub-rights – they can differ in their specific combination of incidents. Second, rights may vary in content, or what sorts of things we can have particular rights-bundles to. If promissory rights fail the second test – that is, if they do not exist or varies across cultures – then they will automatically fail the test of application. If a culture has divergent practices of promising, or does not promise at all, and the divergence or non-existence is morally acceptable, then our rules of promising do not apply to them.

There is evidence to suggest that promising does not exist in some cultures, cultures that we do not find morally reprehensible, and that, as a set of behaviors, it diverges across cultures. The best example of divergence and non-existence comes from the island of Tonga in the south Pacific. In their article “Where People Don’t Promise”, anthropologist Shulamit Dector Korn and philosopher Fred Korn investigate promise-like behavior in Tonga. They first observe that there is no

corollary to the term “promise” in the Tongan language, where “promise” is “speech act ... whereby the speaker undertakes an obligation to perform some future act.”¹¹⁸ They have words for giving assurance without creating an obligation (‘fakapapau’), taking a legal oath (‘faukava’), and pledging in order to mark subordinate status (‘tukupa’). The word ‘tala’ofa’ is most commonly translated as ‘promise’, but actually means “to tell of a loving gift.”¹¹⁹

Second, Korn and Dector Korn observe that there is no way for Tongans to create obligations through speech acts. One can, of course, enter into obligation-generating relationships or otherwise choose to take on obligations, like parenthood, marriage, and legal contracts, but he cannot do so via speech. Instead, Tongans predict future actions as a way to signal solidarity, an intention to act, or the desire to act. But these predictions are seen as hypothetical or conditional. Korn and Dector Korn explain this lack of promising with the idea that, for Tongans, the future is too uncertain for them to be able to obligate themselves by predicating a future action. They conclude that “promising is incompatible with the Tongan view of the future and Tongans’ conduct of social relationships.”¹²⁰

Whether or not this study proves that Tongans do not have a version of promising, it does show that their version of promising is, at the very least, drastically different from those of other, Western cultures. Even though Tongans

¹¹⁸ Korn and Dector Korn, 447

¹¹⁹ Korn and Dector Korn 1983, 447

¹²⁰ Korn and Dector Korn 1983, 450

can obligate themselves – or transfer rights – through legal contracts and social arrangements like marriage, and these are arguably forms of promises, they are limited in scope, especially in terms of sanctioning. In terms of Hohfeldian incidents, whereas an American can transfer a claim to a future action and the power to accept or waive that claim, a Tongan cannot, except perhaps in a legal context. Furthermore, an American can transfer a right-bundle to a variety of actions, using only a speech act, whereas a Tongan can only do so in legal or marital matters, and cannot use merely a speech act. Hence, there is certainly variation in the practices surrounding promising, and in particular promise-making and promise-trusting.

In addition, it is not clear that we should morally disapprove of the Tongans for not having a way to make, trust, or keep promises. Taking Korn and Dector Korn's study at face-value, it does not seem that the Tongans are morally lacking without promissory behaviors; rather, the authors conclude that Tongans have genuinely ethical social interactions, but that their society is incompatible with promising. The basic idea is that in situations of great uncertainty, like those that exist in Tongan daily life, the promising practices may not emerge. There would simply be no use for it. Further, they use predictions and intentions to build relationships – to show a desire to help – instead of obligating themselves. This seems to be a morally permissible, if not admirable, form of interaction. We can conclude that we have what seems to a morally acceptable culture that has a different promissory practices, if they have them at all.

Even within American culture, we may notice some varieties of promissory rules that apply at different circumstances and to different people. These more intuitive examples include promises to oneself and promises to God. When one promises herself, she gives herself a claim to her own future action, and the related duty to perform, without giving herself any powers to waive, revise, or transfer the claim. This differs from most promises to others, which include the power to waive the claim, but not the powers to revise or transfer. Promises to God, if made in earnest, include the claim, along with all of the powers – who would argue God’s request to waive, revise, or transfer the claim?

The reply to the use of these examples is that they are not promises, or that they are deficient versions of promises, because they are not made to others. It may, on the other hand, be a benefit to this theory of promises that it can explain promises to the self and to God. However, even if these examples are not accepted, there is still intuitive divergence in different cases of promising. Imagine, for instance, promising to buy your friend a beer at a bar. She then informs you that she has transferred her claim to the beer to her sister. You may, in some circumstances, view the transfer as completely reasonable. In that case, your promise included the power to transfer – when you transferred the right, it included the right to transfer. You may, however, instead insist to your friend that you promised her a beer, not her sister. In that case, the right you transferred included only the claim and the power to waive. This example illustrates the variation that promises can have, even in structurally identical circumstances.

The variety of promises allows us to reject the idea that a single set of promissory practices is ubiquitous. And, since there is no ubiquity, and the lack of ubiquity is not morally reprehensible, we can conclude that promising is not universal in application, either. For example, I cannot make a promise to a Tongan, and he cannot make a promise to me. Hence, if moral norms are differentiated on the basis of universality, of either form, promising is not based on moral norms.

VI. MORALIZATION

If behaviors surrounding promising do not fit any of our definitions of moral norms that exclude social norms, we have more reason to believe that promises are supported by social norms. But if we do not want to give up the belief that promising is, in some behavioral sense, moral, we may be able posit that promising is both moral and social. In this section, I will explore and adopt this explanation, concluding that promising is based on moralized social norms that imply moralized social rights.

The essential idea is that we can differentiate the moral domain on the basis of phenomenology. In other words, we can distinguish moral norms based on our affective reactions to them, or simply by considering them to be moral. One way to explain this is through moralization: some norms become deeply entrenched within

our practices and psychology, such that they are strengthened (more likely to be followed) and have strong emotions associated with them or their violations.¹²¹

Moralization is the sort of process that Hume describes when he explains the moral justification of property and other artificial virtues in the *Treatise* and in the second *Enquiry*. I will further examine an example of moralized artificial virtues in the next sub-section. In brief, Hume argues that there are two types of moral norms, natural and artificial. Natural moral norms are those norms that have a natural sentiments associated with them, like norms that require us to care for our children. There are natural moral sentiments that support parental duties. Artificial moral norms are moralized social practices: they depend on the existence of a social practice, and the moral sentiments that support them are developed over time.¹²² Hume's examples of artificial virtues include justice, or the regulation of property, honesty, modesty, and manners, to name a few. We do not have a natural sentiment to be honest, for example, because honesty can only exist as a part of convention-dictated social interactions – namely, linguistic interactions. Moralization, then, occurs when we develop moral sentiments to support these practices.

The advantage of this method of separating moral norms from other norms is that it does not fully separate moral norms from other norms – moral norms will overlap with other norms. For example, a social norm may be moralized and

¹²¹ Nichols 2010, 14.

¹²² Hume 2000, 320-321.

become a moral norm, while still retaining the structure of a social norm. A norm, on this view, can be both moral and social. Additionally, we can still separate natural from artificial moral norms, or pre-social moral norms from *moralized* norms. Thus, we can still say a norm is essentially social when it is moralized, and essentially moral when it is not.

Moralization can also allow us to further explain why a social norm, like promise-keeping, may appear unconditional in some settings. Moralization occurs when a norm becomes deeply entrenched; that is, it becomes a default rule, followed in any situation where that rule is triggered, or in any situation similar enough to its usual application. We can describe moralization in terms of scripts. The moralized social norm is part of a script that is automatically activated in particular social interactions, like Bicchieri and colleagues' case of subjects keeping promises to computers.¹²³ In the case of promise-keeping and -trusting, when the norm is made salient – that is, when promises are made – it “places agents in a familiar context... where people who make promises tend to keep them.”¹²⁴ What appears to be unconditionality is actually script-following; subjects keep and believe promises because promising practices are so deeply entrenched.

Another aspect of moralization is the one that Hume focuses on: moral sentiments. When a social norm is moralized, it takes on various moral emotions.

¹²³ Bicchieri et al. 2010, 127.

¹²⁴ Bicchieri et al. 2010, 128.

That is, agents will feel positive, internally-based emotions when adhering to the norm, and negative emotions, like guilt, when violating it, regardless of the social situation they are in.¹²⁵ The notion that agents will feel guilty when violating a norm, even when it is anonymous and not sanctioned, can explain why subjects keep their promises in anonymous, one-shot, computer-mediated games. In fact, Charness and Dufwenburg use guilt to explain why subjects prefer to keep their promises in one-shot trust games. The basic idea is that guilt alters a player's preferences; players are guilt-adverse. In their view, there is a norm that shapes expectations, and subjects live up to these expectations because they would feel guilty if they did not.¹²⁶

In addition, moralization explains why sanctioning is not necessary for promise-keeping. Moralization causes us to view a norm as legitimate, and thus view the normative expectations of others as legitimate. The fact that we would feel guilty if we broke our promises makes us see the rules governing promising, which tell us to keep our promises, as legitimate, fulfilling a criterion for normative

¹²⁵ For a more detailed discussion of moralization, see Nichols 2002. For a critique of Nichols's account of moralization, see Royzman et al. 2009. Both agree that moralization includes entrenching rules and moral emotions; they simply disagree about which rules are moralized, how they are moralized, and the moral implications. These distinctions are not relevant to my present discussion.

¹²⁶ Charness and Dufwenberg 2006, 19. Charness and Dufwenberg believe that guilt aversion shapes promise-keeping behavior because subjects would feel guilty if they let someone else down, not if they violated a norm. However, their data only support the idea that subjects adhere to the norm because they feel guilty, and do not allow us to conclude why they feel guilty. Vanberg 2008 has shown, on the contrary, that subjects feel guilt for norm violation, not for disappointing expectations. I will discuss this debate in a later chapter.

expectations. Hence, since moralization entrenches norms and adds motivating moral sentiments, it can explain any divergence between the practice of promising and the definition of social norms. We can conclude that there are moralized social norms that support promising.

VI.A. TWO ANALOGIES: ALLEGIANCE AND MODESTY

We can now translate this discussion of rules into a discussion of rights. As long as there is a system of rules, and either sanctions or a perception of legitimacy, there will be rights. Since promising is based on moralized social norms, it is based on a system of rules that is, because of moralization, considered to be legitimate. Once a norm becomes moralized, it gains, by the definition of moralization, a perceived legitimacy. Therefore, promissory rights can be based on these moralized social rules. They become what I call “social rights”.

One way to explain what I mean by moralized social rights is through two examples, taken from Hume, of allegiance and modesty. I use these examples because they are relatively uncontentious – most of us would likely agree that they are both based on social rules – but we might think they differ in terms of their legitimacy. In this section, I will show how manners and modesty are analogous to promising under my Humean view of promising as the transfer of moralized social rights. I will begin with manners.

Hume’s distinction between artificial and natural virtues tracks the distinction between moral norms and moralized social norms. Whereas

benevolence, parental love, and other such virtues are natural, virtues like fidelity (or promising), allegiance, chastity and modesty, and justice are artificial, and thus rely on social practices. However, in order for these artificial virtues to actually be virtues, the practices upon which they are based must be moralized. According to Hume, allegiance, the virtue that concerns our interaction with the state, begins as a social practice, but is then moralized:

By means of these two advantages, in the execution and decision of justice, men acquire a security against each other's weakness and passion, as well as against their own, and under the shelter of governors, begin to taste at ease the sweets of society and mutual assistance.¹²⁷

As we see how well our government meets our needs, and the needs of others, we develop moral sentiments toward it – we internalize the rules and give them a moral character, namely, patriotism or allegiance, as well as the duties of magistrates.

But, naturally there can be many varieties of government that meet our needs, and so there may be many moralized varieties of allegiance. The fact that governments, and allegiance to them, are useful, though, can explain their ubiquity – almost all societies have some government and a moralized sort of patriotism. However, there are circumstances under which governments are likely to be

¹²⁷ Hume 2000, 345

useless, or perhaps harmful. According to Hume, allegiance would not develop, or develop very differently, under these conditions.¹²⁸

Lastly, allegiance is primarily a virtue regulating how we interact with our government, and how our government interacts with us. These rules give us expectations about the behavior of others, expectations that are reinforced by the rules and observed behavior. We come to feel we have a right to expect certain behaviors, and may be offended if, say, someone burns a flag, or morally outraged if someone breaks a law, even if it has no negative effect. We feel we have a claim to correct magisterial interactions, and so government allegiance forms a system that regulates and gives us claims to things. It is a system of rights.

Many of us may agree that allegiance is socially-based, and perhaps that they are morally legitimate. Many may agree that allegiance or patriotism *should* take on this moral character. But there are other artificial virtues that we may not think are morally justified, but still take on a moralized character. Regardless of whether we believe that the rights based on these virtues are actually or 'objectively' moral, they can qualify as moralized. An example of such a virtue is modesty.

Hume describes modesty and chastity as virtues that spring from a need to organize families and establish paternity for children.¹²⁹ Rules of modesty differ across cultures and depend on the needs and circumstances of that culture.

¹²⁸ Hume 2000, 345-346

¹²⁹ Hume 2000, 365

Modesty regulates how we dress and act in potentially sexual situations and may take on a moral character. Individuals within a social system that has modesty rules might view these rules as legitimate, and that they have a right to proper sexual behavior, and a right to punish deviance from that behavior. Like promising, modesty might require several different behaviors and rights, and thus may be based on several distinct, though related, norms. But, as long as those rights are *perceived* to be legitimate – as they often are – we can say the modesty system is supported by moralized social norms.

Bicchieri concurs with this interpretation of artificial virtues. She claims that Hume's artificial virtues rely on conditional preferences and social expectations, and meet her definition of a social norm. What distinguishes Hume's artificial virtues like promising, justice, allegiance, and modesty from other, non-moralized social norms "is that many of us would have a conditional preference for abiding by such norms because we acknowledge that the normative expectations ... are *legitimate*."¹³⁰ This belief in legitimacy signals a moralized social norm.

In addition, artificial virtues are the only type of virtue that can be rights-based because they are the only type of virtue that is rule-based. According to Hume, natural virtues admit of indetermination, indistinguishability, and gradation in a way that the virtues of allegiance and modesty, and other artificial virtues, do

¹³⁰ Bicchieri 2006, 21.

not.¹³¹ Since artificial virtues are so determinate, distinguishable, and law-like, they meet Hart's standards for admitting of a moral obligation, and thus moral rights.

The Humean models of allegiance and modesty that I have just laid out has three crucial aspects: it is a moralized social practice, it is variable, and it is rights-based. I have also made these three claims about promising. Hence, we can use Hume's account of allegiance and modesty as models for promising behaviors. First, we have seen that promising is supported by moralized social practices, and that these practices are variable under certain conditions, though generally ubiquitous. For example, Korn and Dector-Korn argued that promising does not exist in Tongan culture because of their uncertainty about the future. This situation is analogous to allegiance and modesty. If we have no need to regulate sexual interaction to establish paternity, as in a commune in which children are raised by the entire community, we may not have modesty norms. Similarly, if future action is uncertain, it is not in one's self-interest to promise. However, in most circumstances, promising will be useful, and societies will develop the practice.

The main advantage of using Humean artificial virtues as a model for promising is that it shows us how rights are embedded in a social practice. Just as the social rules that govern manners give us claims to actions, the social rules that govern promising can give us claims over particular actions as well. The rules that govern promising concern the performance of a specific action by and for specific

¹³¹ Hume 2000, 339

individuals. The rules allow the action to be claimed by particular individuals.

Rules of promise-making and –trusting govern the transfer of these actions, whereas rules of promise-keeping govern the fulfilment of the claim.

We now have a full picture of the practice of promising. It is a moralized, rights-based social practice, analogous to Humean allegiance and modesty – an artificial virtue. Whereas the conditions of modesty, for instance, might include the need to establish paternity, the conditions of promising include future-certainty and –predictability. Like manners, promising begins as a way to coordinate action, and eventually develops moral characteristics, judgments, and sentiments.

CHAPTER 3: RIGHTS-BASED MOTIVATION FOR PROMISE-KEEPING

I. INTRODUCTION

In the previous chapters, I showed that the behaviors surrounding promising are best understood as the transfer of a moralized social right. From this account we can draw two related conclusions: promises transfer rights, and the normative basis of promise-keeping lies is the fact that the promisee has acquired a right. Since promising is supported by social norms, most notably a norm of promise-keeping, we can also conclude that promise-keepers are motivated by a conditional preference to follow the social rules that surround promising. I will further explore this last conclusion in this section.

A social norms approach, by itself, explains motivation to adhere to a social norm, and this motivation is testable. In this chapter, I will test my account of promising by showing that it can account for the experimental data on promise-keeping. Further, I will argue that the experimental data support my view of promising: not only can my view account for all of the data, but it is the best explanation of it. Though I believe promising may be made up of several norm-governed behaviors, I will focus primarily on promise-keeping. I will begin this chapter by reviewing several experiments about promise-keeping that suggest that

when promises are made, they tend to be kept. I will then review several experiments that aim to test various explanations of why these promises are kept.

There are three general explanations of why promises tend to be kept. First, subjects may be guilt-averse, and thus want to avoid disappointing expectations. Second, subjects may be harm-averse, and thus want to avoid harming promisees. Third, subjects may have a preference for keeping promises, which could be conditional or unconditional on social expectations. This last explanation includes my social-rights-based explanation. I will argue that, taking all the experiments together, my explanation can account for the results better than any of the other proposed explanations.

I begin with prisoners' dilemma-games and trust-games that establish that promises are generally kept and promise-breaking is both condemned and punished. These experiments come from Dawes et al., Bicchieri et al., and Wilkinson-Ryan & Baron. I will then analyze several experiments and their proposed explanations for promise-keeping, including the expectation-based accounts from Charness & Dufwenberg and Ederer & Stremitzer, the commitment-based accounts from Vanberg and Ellingsen & Johannesson, and my social-rights-based account. I conclude that my social-rights explanation of promising best accounts for all of the data.

II. PROMISES ARE KEPT AND BREACHES ARE PUNISHED

II.A. KEEPING PROMISES IN COMMONS DILEMMAS AND TRUST GAMES

In this section, I will review some experiments that show that cooperation increases in commons dilemmas and trust games when there is communication that includes promises. Since cooperation increases with promises, we can conclude that these promises are kept, even though the games are set up such that individuals have higher payoffs when they defect, or, in these cases, break their promises. I will begin with prisoners' dilemmas, and then move onto trust games.

II.a.1. Commons Dilemmas (Dawes et al. and Orbell et al.)

In their 1977 article "Behavior, Communication, and Assumptions About Other People's Behavior in A Commons Dilemma Situation", Robyn Dawes and colleagues attempt to distinguish which types of communication lead to increased cooperation in commons dilemma games. They hypothesized that communication could increase cooperation for three reasons: it could raise interest in others' welfare, provide information about others' beliefs and expectations, or provide assurance that a particular decision will be made.

Dawes and colleagues attempt to separate these explanations with four different communication treatments in their commons dilemma game, including a no-communication condition, an irrelevant communication condition, a relevant

communication condition, and a condition that included both relevant communication and a vote to determine the group's decision.¹³²

If communication increases cooperation because it raises interest in the welfare of others, and thus motivates subjects to make decisions that benefit others, irrelevant communication should increase cooperation as much as relevant communication. However, if cooperation increases because it gives players information about the beliefs, preferences, and expectations of others, only relevant communication – communication that reveals these beliefs, preferences, and expectations – would increase cooperation. Lastly, if cooperation increases because it provides assurance regarding others' choices, voting should increase cooperation more than mere relevant communication.

In this commons dilemma experiment, cooperative responses earned \$2.50, defecting responses earned \$12 with a fine of \$1.50 to everyone. Thus, if everyone defected, no player received any money, and there was a possible negative payoff for cooperators.¹³³ Subjects were recruited in groups of four. These foursomes were then split into different groups, each consisting of eight participants. After the game, the groups of four pooled divided their earnings.¹³⁴

¹³² Dawes et al. 1977, 3.

¹³³ Dawes et al. 1977, 2.

¹³⁴ Dawes et al. 1977, 4

Dawes and colleagues found that relevant communication significantly increased cooperation, but that voting did not.¹³⁵ They concluded that communication increases cooperation because it gives players information about the beliefs and preferences of other players, and not because it gives players assurance regarding the actions of others. But this conclusion seems hasty; Dawes and colleagues did not analyze the content of the relevant discussion. If the discussion involved promises, relevant discussion itself would provide assurance, not just information. A vote would not increase cooperation because it would be redundant – promises provide as much assurance as voting.

To separate these variables, Orbell, van de Kragt, and Dawes ran a social dilemma experiment in which they coded the discussions according to its content – that is, whether or not they contained promises. The game was set up as a dilemma between and within groups. Each player was given a \$5 promissory note, which she could keep, give to her own group, or give to the other group. If she gave to her own group, each of the other members received \$2. If she gave to the other group, each member of that group received \$3. Hence, keeping the money was dominant for each player, but each member would do best if every member of each group gave to the other group.

Participants were allowed to discuss in 14-person groups, then separated into two groups of seven each before making their decisions. Half of these 7-person

¹³⁵ Dawes et al. 1977, 5.

groups were then allowed another discussion period before deciding.¹³⁶ Orbell and colleagues coded all of the discussions according to their content. They found that, of the groups allowed subsequent discussion, seven of thirteen had universal promising, and in the other six groups, an average of 7.8 people promised. Without subsequent discussion, six of eleven groups had universal promising, and the other five groups averaged four promisors. There were only four groups that lacked promises entirely.¹³⁷

On average, 84% of participants contributed when promising was universal, and 58% when only some promised. When promising was not universal, the correlation between the number who promised and the number who contributed was 0.09.¹³⁸ This low correlation indicates that promises are kept if and only if promising is universal. Universality is important, according to a social norms approach, because only universal promising would signal that others expect the rules of promising – including promise-keeping – apply. If promising is not universal, individuals do not have normative expectations that come with promise-making. If others do not promise at all, promising behaviors, and the norms that support them, may not be triggered. Universal promising allows individuals to trust that rules regarding promising apply in this situation, and thus trust that others will keep their promises, and they should do the same. Similarly, if promising is not

¹³⁶ Orbell et al. 1988, 815.

¹³⁷ Orbell et al. 1988, 816.

¹³⁸ Orbell et al. 1988, 817.

universal, and individuals do not think that others expect rules of promise-keeping to apply, they will not think that they would be punished for breaking a promise. After all, perhaps it was not really a promise. Universal promising signals that others think the rules of promising apply, and hence the promise (in this case, cooperation) should be kept.

From this study, we can conclude that communication in social dilemma games will often include promises, but that these promises are only taken to be meaningful, as true promises, when promising is universal. Further, when they are meaningful, or when they fully trigger social norms surrounding promising, promises increase cooperation significantly. Hence, communication increases cooperation in social dilemma games often because subjects make and keep their promises.

II.a.2. Trust Games and Computers (Bicchieri et al.)

Bicchieri, Lev-On, and Chavez add to these data on promise-keeping in social dilemmas by expanding it in two ways. First, they use trust games instead of social dilemmas. Trust games differ from social dilemmas in that they are asynchronous. Instead of each party deciding whether to cooperate simultaneously, in a trust game, an investor is given a certain amount of money, and must decide whether to invest it with a trustee. If she invests, the amount of money is increased, and the trustee must decide whether to keep the entire sum or split it. Hence, instead of each party promising to cooperate at the same time, the trustee might promise to split the sum

if the investor agrees to invest. In many ways, this type of game mirrors real-life promising to a greater degree.

Trust games also limit the degree to which promises serve to give players information rather than provide assurance. When the trustee promises, she is not communicating any expectations or beliefs to the investor. The trustee's expectations are not relevant. She is then merely providing assurance to the investor that she will divide the money if the investor decides to invest it. And, if she promises, she is not merely expressing an intention; rather, she is making the declaration of intention trustworthy. If promises are used in trust games, they are not merely information-gathering techniques.

Second, Bicchieri and colleagues vary the method of communication. While researchers like Dawes and Orbell established that promises, once made, tend to be kept in face-to-face discussions, they did not investigate whether promises are made or kept in written or computer-mediated interactions. If promises are made, believed, and kept only in face-to-face interactions, it may be that the face-to-face interactions themselves are changing expectations and preferences, not the promises. That is, the interpersonal interaction might explain the cooperation, not the promise. If promises are made, believed, and kept even in written and computer-mediated interactions, it would testify to the relevance and the motivating force of the promise itself.

In their trust game, Bicchieri and colleagues had five conditions: no communication, irrelevant face-to-face communication, relevant face-to-face, irrelevant computer-mediated communication, and relevant computer-mediated communication. They measured three separate dependent variables. They measured trust by the amount the investor sent to the trustee, reciprocity by the amount the trustee sent back, and expected reciprocity by the amount the investor expected to return.¹³⁹

Bicchieri and colleagues found cooperation increased with relevant communication, regardless of the medium, though cooperation in the face-to-face treatment with relevant communication was higher than in the computer-mediated treatment.¹⁴⁰ In addition, though the communication medium did affect reciprocity and expected reciprocity – reciprocity was higher in the face-to-face relevant communication condition – it did not affect trust. That is, investments were higher after relevant communication, no matter the medium, though returns did differ: “Media richness, instead, had no effect on trust.”¹⁴¹

Lastly, they also found an abundance of promise-making, promise-trusting, and promise-keeping in both relevant communication conditions. According to Bicchieri and colleagues, “Especially in face-to-face conditions, when participants

¹³⁹ Bicchieri et al. 2010, 133.

¹⁴⁰ Bicchieri et al. 2010, 134.

¹⁴¹ Bicchieri et al. 2010, 139.

could discuss the game they always made promises.”¹⁴² If promises were generally made, trust was high when communication was allowed, and media richness did not affect trust, then these promises were generally believed, regardless of the medium.

However, more promises were made and kept in the face-to-face relevant communication condition. Hence, though communication media do not affect whether promises are believed, they do affect whether they are made and kept, at least to some degree. From this, we can conclude that face-to-face interaction does change preferences to some extent, but that the change in preferences resulting from promising is also significant. Thus, it is not just interpersonal interaction that is relevant, but promises themselves.

From these three experiments, we can draw several conclusions. First, communication increases cooperation, even when players have an incentive to defect, and even when decision-making is asynchronous. Second, when players are allowed to communicate, they tend to promise, and these promises tend to be believed, regardless of the means of communication. Third, players tend to keep their promises, especially when they communicate face-to-face. However, these studies do not address moral judgments of promise-breaking – that is, they do not address punishment of breach.

II.B. PUNISHMENT OF BREACH

¹⁴² Bicchieri et al. 2010, 135.

Punishment of promise-breaking is important in that it reveals another element of promising: third-party reactions. In these cases, promise-breaking is punished by an uninterested, third-party observer. Showing that promise-breaking – and promise-breakers – are generally punished, even by uninterested parties, allows us to draw several conclusions about promising in general. First, it shows that promise-breaking is generally disapproved to the extent that third parties take an interest. That is, third parties have normative expectations for upholding promises. This conclusion supports the idea that promising is supported by social norms. Second, if third-party punishment comes with a moral judgment against the promise-breaker, it can indicate that a norm is moralized. Lastly, third-party punishment gives us reason to doubt the guilt-aversion explanation of promise-keeping. Even if guilt-aversion can explain why individuals keep their promises, it could not explain why they would be willing to punish others for breaking them because they could not feel guilty about someone else's perceived wrongdoing. I will expand on this idea in section III.

Some of the best work on punishment of promise-breaking is in the study of contract breach. For these studies to apply to promising, we must first accept that contracts are the legal analogue of promises, and hence that studies of contracts will allow us to make conclusions about promises. However, there are reasons to think that promises differ from contracts in important ways. For instance, contracts are enforced by a different mechanism than promises, they are part of a formal system

of rules, and they are made in more formal contexts between parties with fewer emotional ties.¹⁴³

If we are to use work on contracts to draw conclusions about promises, it will be useful to distinguish between two aspects of contracts: the enforceable legal contract and the psychological contract. This distinction comes from the work of Robinson and Rousseau (1994), and was adopted by Tess Wilkinson-Ryan (2012), among others. In brief, the psychological contract is how involved parties actually understand the contract.¹⁴⁴ It includes perceived legal and social norms, expectations about the other party, beliefs about the legal system, and so on. Since parties do not tend to know background contract law, they assume that the law aligns with their moral intuitions, and they include moral and social norms within their psychological contracts.¹⁴⁵ Extending the idea of a contract as the legal analogue to promises, Wilkinson-Ryan writes, “A contract is a promise that the legal system recognizes; the psychological contract is the promise that the parties themselves recognize.”¹⁴⁶

In their paper “Moral judgment and moral heuristics in breach of contract”, Wilkinson-Ryan and Jonathan Baron study the divergence between the

¹⁴³ For a lengthier discussion of the differences between contracts and promises, see Seana Shiffrin (2007), “The Divergence of Contract and Promise”.

¹⁴⁴ Again, for a lengthier discussion of psychological contracts, see Robinson and Rousseau (1994), “Violating the psychological contract: Not the exception but the norm”

¹⁴⁵ Wilkinson-Ryan 2012, 845.

¹⁴⁶ Wilkinson-Ryan 2012, 846.

psychological contract and the legal contract, as well as the tendency to punish efficient breach of contract. Efficient breach occurs when one of the contracting parties – the one analogous to the promisor – breaches the contract only for his or her own economic gain. That is, the breaching party (respondent) finds a new opportunity which would benefit him more than the existing contract.

Generally, in the legal context, when damages are assessed in cases of efficient breach, they are measured by the expected benefit to the claimant, advancing her interests to the point they would be had the contract been fulfilled. Economically this makes sense: the costs and benefits of breach are partially determined by benefit to the other party.¹⁴⁷ That is, the respondent is forced to take the claimant's economic interests into account when calculating the cost of breach. In addition, motivation is considered irrelevant, and neither punitive damages nor specific performance are awarded. The respondent need only consider the exact cost of breach to the claimant, and then reimburse her to that extent.

Wilkinson-Ryan and Baron hypothesized that the psychological contract would differ from the enforceable legal contract on the issues of motivation, punitive damages, and specific performance. To test this idea, they conducted two experiments in which they asked subjects to assign damages to claimants in various hypothetical cases of breach. In the first experiment, they compared assigned damages for efficient breach in order to avoid loss with damages for efficient breach

¹⁴⁷ Wilkinson-Ryan and Baron 2009, 4-5

in order to gain. The two cases were otherwise identical; they were simply framed differently. Thus, in a legal setting, the awarded damages would (likely) be the same in both cases.

The results, however, indicated that subjects do not see the difference between the cases as a mere framing issue; rather, it is an issue of motivation. And subjects find the difference between the motivation to avoid loss and the motivation to gain relevant. Setting the expected loss to the claimant at 3, subjects set damages at a mean of 4.08 for breach-to-gain and 3.14 for breach-to-avoid-loss.¹⁴⁸ This result was replicated in within-subject and between-subject trials.¹⁴⁹

In short, subjects assigned an average of around 30% more damages to parties who suffered a breach to gain rather than to avoid loss. Further, in cases of breach-to-avoid-loss, subjects assigned nearly the same amount in damages as would have been assigned in a legal setting. They simply repaid the claimant for her loss. However, in cases of breach-to-gain, subjects awarded punitive damages – damages that exceed the cost to the claimant and are thus intended either to punish the respondent or to compensate the claimant for something beside her economic loss. Hence, the psychological contract, in this case, differs from the legal contract, and it differs based on motivation. In the psychological contract, breach motivated by gain deserves either punishment or extra compensation to the wronged party.

¹⁴⁸ Wilkinson-Ryan and Baron 2009, 11.

¹⁴⁹ Wilkinson-Ryan and Baron 2009, 12.

In the second experiment, Wilkinson-Ryan and Baron compare damages awarded for negligent breach to damages awarded for efficient breach. In addition to asking subjects to assign damages in hypothetical cases, in this experiment, they asked subjects specify the form the damages should take, and to rank the immorality of the breach. Again, they found that motivation was important to subjects: they assigned higher damages to efficient breach than to negligence.¹⁵⁰ And, based on the immorality ranking, the differences in motivation were morally relevant to subjects. On a 7-point scale, where 1 was not immoral and 7 was most immoral, intentional breach was 5.1 and negligent harm was 3.3. From this, Wilkinson-Ryan and Baron conclude that subjects assign punitive damages based on their moral assessment of the motivation to breach.

Further, Wilkinson-Ryan and Baron found that subjects were likely to assign damages including specific performance, though specific performance is rarely awarded to claimants in a legal setting. Subjects preferred that contractors actually perform – that is, “honor the contract rather than pay damages and breach,” – and 66.7% of subjects thought that the law should generally force the contractor to perform, not just pay damages.¹⁵¹ Thus, the form of damages differs between the legal and the psychological contract. The legal contract awards monetary damages; the psychological contract favors specific performance.

¹⁵⁰ Wilkinson-Ryan and Baron 2009, 19.

¹⁵¹ Wilkinson-Ryan and Baron 2009, 19.

So far, we can conclude from these experiments that:

1. Motivation matters for subjects assigning damages to contract breach.
2. Subjects, acting as third parties, are willing to punish breach based on the motivation to breach.
3. Punishment is dependent on motivation because subjects find motivation – specifically, intention to breach and breaching to gain – morally relevant.
4. Subjects prefer specific performance to mere monetary damages when assigning damages.

If we allow contracts to be analogous to promises, we can make several conclusions about promises as well. First, we can conclude from the fact that subjects assign punitive damages that they disapprove of breach – in certain cases – to the extent that they are willing to sanction. The presence of a sanction, even when the law does not require (or even encourage) it, indicates the existence of normative expectations. Subjects expect general disapproval of breach in certain cases, and thus find punishment appropriate. The normative expectations, in turn, indicate the presence of a social norm against breach.

Second, since subjects morally disapprove of efficient breach, and they distinguish between cases of breach based on features the situations that they find morally relevant – that is, motivation – we can conclude that the social norm of keeping one’s contract is moralized. Subjects believe that breach is immoral. So far,

the analysis of third-party punishment of breach of contract supports the idea that promises are moralized social norms.

Lastly, the fact that subjects prefer specific performance to monetary damages supports – and is best explained by – the theory that promises are moralized social *rights*. Specifically, a rights-based explanation of the data fits better than a harm-based explanation. Recall that the right a promise transfers is the right of the promisee to a specific action of the promisor. The preference for specific performance reflects the idea that the claimant is entitled to the respondent's action. And this entitlement cannot be reduced to correcting a harm.

Suppose the subjects punished because they believed that the respondent harmed the claimant by breaching, above and beyond the cost of the breach. That is, breaching a contract – or breaking a promise – would cause two distinct harms. It would harm the promisee (or claimant) by not performing, and by inflicting a psychological harm, or damaging the relationship between the promisor and promisee.

According to this theory, subjects assign specific performance to rectify the first harm – the harm of non-performance – and assign punitive damages because of the second, psychological harm. However, it does not explain why subjects would prefer specific performance to monetary damages. On the harm-based explanation, the promisor must rectify the breach, but it does not matter how, or by whom, the harm is rectified. If the morally relevant feature is harm, then as long as the

promisee is compensated, either by performance or monetary damages, the harm is rectified.

In other words, if harm is the important part, and the promisee is harmed by the lack of performance, then she is sufficiently compensated as long as there is performance, or if she benefits to the same degree but in a different way. It does not matter *who* fulfills the contract, and it may not matter if anyone fulfills the contract at all. But the harm explanation then cannot account for the fact that subjects prefer specific performance (by the contractor) to mere monetary damages.

The social rights-based account explains punitive damages using the concept of normative expectations, and explains the preference for specific performance using the concept of a right. If promise-keeping is a social norm, we would predict that a broken promise would be punished. If the promises transfer rights, and thus entitlements to action, we would predict specific performance to be preferable to compensation, if not required. Favoring, or requiring, specific performance reflects an entitlement to an action.

II.C. CONCLUSIONS FROM EXPERIMENTS

From the experiments discussed, we can discern several features of promising. First, promises tend to be made, believed, and kept in trust games and social dilemma games that allow communication, regardless of the method of communication and time of decision, even though promise-breaking is incentivized. We can thus conclude that promisors keep their promises against their own

apparent interests, and that promisees believe that promisors will keep their promises against their own apparent interests.

Second, in cases of promise-breaking, third parties are willing to punish promise-breaking above and beyond the harm inflicted by non-performance, and they punish according to their own moral distinctions. In addition, they prefer specific performance to other forms of compensation. We can conclude that third parties morally condemn promise-breaking, and it is best to make up for the breach by performing the promised action. That is, the promisee is still entitled to the promisor's action.

In the next section, I will discuss several theories that purport to explain these features of promising, using further experiments to support their conclusions. I will begin by describing two expectation-based explanations. These are essentially harm-based approaches: subjects keep their promises because they are averse to harming others by disappointing their expectations. I will then reject these accounts using a commitment-based explanation. Lastly, I will argue for my social-rights account as a middle ground between these two theories, and that it explains the data better than each theory individually. Hence, I argue, we should adopt the social-rights account to explain the motivation behind promising, as well as the normative and descriptive basis of it.

III. EXPLANATIONS OF PROMISING

III.A. EXPECTATIONS AND GUILT-AVERSION

There is a group of explanations of promise-keeping behavior that center around the idea that breaking a promise harms the promisee, and the promisor wants, for whatever reason, to avoid this harm. Specifically, the promise creates expectations of future behavior in the promisee, and the promisee is somehow harmed when those expectations are disappointed. The harm of disappointed expectations goes above and beyond the content of the promise. That is, there is an extra harm in having one's expectations left unmet, and it is added to any harm resulting from the particular promised action not occurring.

In the first chapter, I discussed a similar account of promising from a theoretical point of view. I looked at accounts of what we do when we promise, and the normative basis of promising, including theories that hold that, when we promise, we are fundamentally creating expectations. The wrongness of breaking a promise, then, would lie in the disutility (or some other wrong) of disappointing expectations.

In the discussion in the first chapter, I reduced these expectationalist accounts of the foundations of promising to a rights-based account. In the second chapter, I completely rejected the expectationalist view in favor of a conventionalist, or social norms, view of promising. However, neither the philosophical reduction of expectationalism to entitlements and rights, nor the rejection of expectationalism as

the normative or theoretical basis of promising implies that individuals do not *act* because of expectations.

In other words, I have already reduced and eliminated expectationism as a theoretical explanation of promising. We must now explore its uses as a practical explanation of promise-keeping behavior. I will ultimately reject this account in the practical realm as well, but first I will discuss the data and arguments that support it. I will begin with an experiment from Charness and Dufwenberg, and then look at an experiment defending their conclusions from Ederer and Stremitzer.

III.a.1. Charness and Dufwenberg

In their 2006 paper, Charness and Dufwenberg hypothesize that players in trust games allowing pre-play communication tend to keep their promises because they are generally guilt-averse. A guilt-averse player feels guilty whenever she disappoints another's expectations, and prefers to avoid this guilt to some extent – she takes guilt into account when calculating her expected utility. That is, “A guilt-averse player suffers from guilt to the extent he believes he hurts others relative to what they believe they will get.”¹⁵² According to this explanation, players are not only avoiding guilt that comes from breaking promises, but they feel guilty because promise-breaking disappoints expectations.

To test this hypothesis, Charness and Dufwenberg ran an experiment using a trust game with a probabilistic twist: the principal may get a payoff of zero without

¹⁵² Charness and Dufwenberg 2006, 5.

knowing whether it was luck or the agent's decision.¹⁵³ The game began with each of the two players, the principal and the agent, receiving \$5. The principal could then choose whether to invest his share or not. If he chose not to invest, both parties would keep their \$5. Once the principal invested, the agent would decide whether or not to roll a die. If she chose not to roll, she would keep the entire \$10 – the \$5 from the principal, along with her own \$5. If she chose to roll, there was a multiplier: there was a possibility that the investment could be multiplied and become a larger amount. There was a 5/6 chance that the principal would receive \$12 and a 1/6 chance he would receive nothing. In addition, neither knew the other's choices (in/out or roll/don't roll).¹⁵⁴

There were three variations of the game, varying the communication. One had pre-play free-form written communication from principal to agent, allowing for requests. The second had pre-play free-form written communication from agent to principal, allowing for promises. The final variation did not allow communication.¹⁵⁵ All variations were between-subject. Further, to test their theory of guilt-aversion, the experimenters measured the principals' expected payout, as well as the agents' beliefs about the principals' expectations. They measured these latter beliefs using incentivized questions: the agents were asked

¹⁵³ Charness and Dufwenberg 2006, 3.

¹⁵⁴ Charness and Dufwenberg 2006, 7.

¹⁵⁵ Charness and Dufwenberg 2006, 4.

what they thought the principals expected their payout to be, and were given \$5 if they were within 5% of the principals' actual expected payout.

Given their theory of guilt-aversion, Charness and Dufwenberg hypothesized, first, that *Roll* choices would increase if agents believed principals had higher expectations. Second, they predicted that *In* and *Roll* would be more common with communication, especially when communication included promising. The idea is that communication, and promises in particular, are coincident with higher expectations and beliefs about expectations, and thus would increase *In* and *Roll* choices.¹⁵⁶

The results seem to support the predictions quite well. Without communication, 56% of principals chose *In* and 44% of agents chose *Roll*. With agent messages, it was 74% and 67% respectively. The combination of *In* and *Roll* occurred 20% of the time without communication and 50% of the time with communication.¹⁵⁷ Messages from the principal to the agent had no effect.¹⁵⁸ In addition, expectations and second-order beliefs predicted action. Principals who chose *In* had higher expectations of *Roll*, and agents who chose *Roll* had higher guesses of the principals' expectations.¹⁵⁹ Messages increased the probability of *Roll*, the principal's expectations, and the agent's second-order beliefs.¹⁶⁰ That is,

¹⁵⁶ Charness and Dufwenberg 2006, 8.

¹⁵⁷ Charness and Dufwenberg 2006, 9.

¹⁵⁸ Charness and Dufwenberg 2006, 10.

¹⁵⁹ Charness and Dufwenberg 2006, 11.

¹⁶⁰ Charness and Dufwenberg 2006, 12.

higher expectations and second-order beliefs were correlated with a higher probability of choosing *Roll*.

Charness and Dufwenberg also coded the agent-to-principal messages according to whether they contained promises. They found that “In all cases but one, the *In* rate, the *Roll* rate, and the ex post (*In*, *Roll*) realizations were much higher following a promise than otherwise.”¹⁶¹ That is, promising increased *In* and *Roll* choices more than communication without promises. However, it must be noted what Charness and Dufwenberg counted as a promise. They included explicit promises, such as ‘I promise to choose *Roll* if you choose *In*’, as well as mere statements of intent, such as ‘I will choose *Roll*’ or ‘I intend to choose *Roll*’. They did not include vague statements of fact, such as ‘If you choose *In* and I choose *Roll*, we will both benefit’.

From these results, Charness and Dufwenberg conclude that individuals are guilt-averse when they make promises. Guilt-aversion predicts that second-order beliefs about expectations will increase cooperation (and promise-keeping), and that promising itself will increase cooperation because promising increases second-order beliefs about expectations. In their words, “words may affect the agent's beliefs (about what the principal expects), and so may change the degree of guilt he experiences.”¹⁶²

¹⁶¹ Charness and Dufwenberg 2006, 13.

¹⁶² Charness and Dufwenberg 2006, 18.

However, guilt-aversion cannot account for one of their results: principal-to-agent messages had no effect on choices. If communication increases cooperation, especially when agents make promises, because it increases agents' second-order beliefs, then it should not matter who communicates. If promising increases the agent's second-order beliefs about expectations, then a statement of expectations from the principal should certainly increase them as well. How better to assess the expectations of the principal than by a statement by the principal himself?

Charness and Dufwenberg explain this divergence by claiming that promises affect beliefs and motivation, but statements of expectations do not.¹⁶³ But they give no explanation of why this would be the case. There is no reason to think that a statement of expectations would have no effect on second-order beliefs about expectations. It is counter-intuitive. It is possible – even probable – that promises affect motivation in a way that statements of expectations do not, but guilt-aversion does not give us the tools to explain why. Guilt-aversion reduces motivation to guilt about disappointing expectations. If promising and statements of expectations have the same effect on second-order beliefs, they should have the same effect on motivation. The theory cannot explain the divergence in motivation.

Further, guilt-aversion also cannot explain the punishment of breach discussed in the last section. If subjects keep their promises because they feel guilty about disappointing expectations, then it is still unclear why any third party would

¹⁶³ Charness and Dufwenberg 2006, 18.

punish promise-breaking. Essentially, they would be punishing the lack of guilt: the promisor would be punished because their guilt was insufficient to motivate them. But, generally, it would be strange to think that a third party would fine someone for not feeling guilty.

It seems possible, however, that the third party would not punish the lack of guilt, but rather punish the promisor for harming the promisee by disappointing their expectations. But we have already discarded this explanation because it does not account for the preference for awarding specific performance over monetary damages. Monetary damages and specific performance can make up for a harm equally well; only specific performance can make up for the wrong of breach. Hence, a guilt-based account cannot explain the actions of third parties awarding damages in efficient breach.

But even if the guilt-aversion explanation could account for all of the results, the results do not eliminate other explanations. For instance, all of the predictions are consistent with – and perhaps better explained by – a social norms approach. A social norms approach also predicts that cooperation will increase with increasing expectations and second-order expectations, but it requires us to infer a causal link. The guilt-aversion explanation posits that second-order beliefs *cause* agents to choose to cooperate, whereas a social norms approach would posit only correlation. The action and the expectations would have a common cause: the activation of a social norm. Once a social norm is activated, agents have the correct second-order

beliefs and the motivation to act from them. Charness and Dufwenberg do not give us a way to choose between these two competing explanations.

II.a.2. Ederer and Stremitzer

In their 2013 paper, Ederer and Stremitzer attempt to defend a guilt-aversion account similar to that of Charness and Dufwenberg. To avoid the objections that Charness and Dufwenberg face, Ederer and Stremitzer amend the guilt-aversion account to include only promises. Their main thesis is that “promisors (senders of promises) are motivated to keep their promises in order to avoid disappointing the expectations created in the promisees (receivers of their promises).”¹⁶⁴ The amendment concerning the idea that promises create expectations allows them to avoid the problem of the principal-to-agent communication.

To test their thesis, Ederer and Stremitzer ran a trust-game experiment with varying probabilities. First, agents were allowed to promise, and principals could then choose whether to opt in or out. After the initial choice, principals learned that there was some probability that the agent would not be able to keep her promise. This probability would affect the expected payout of the principal: the lower the probability that the agent could keep her promise, the lower the expected payout to the principal. The agent then learned whether she would be able to fulfill

¹⁶⁴ Ederer and Stremitzer 2013, 2.

her promise, and then decided whether or not to do so.¹⁶⁵ The agent acted as a dictator: once she made her decision, the principal could not reject the outcome.

The hypothesis was that the agent would be more likely to keep her promise if she believed the principal expected her to do so. According to Ederer and Stremitzer, “If higher second-order expectations lead to higher contribution rates by the dictators [agents] who promised to perform, this would constitute evidence for the expectation-based explanation for promise keeping.”¹⁶⁶ That is, if promises only increase cooperation when they are accompanied by expectations – cooperation decreases along with expectations, even when agents promise – then expectations are providing the motivation for cooperation. And that is exactly what guilt-aversion predicts.

Like Charness and Dufwenberg, Ederer and Stremitzer coded communication as promises, empty talk, and no message. Promises included both explicit promises and predictions of actions, and empty talk include everything else. They ran two versions of the experiment, one with a low probability (unreliable) that the agent could keep her promise, and one with a high probability (reliable). They found that 40% of agents chose not to perform under the unreliable device, while less than 5% chose not to perform when it was reliable.¹⁶⁷

¹⁶⁵ Ederer and Stremitzer 2013, 4.

¹⁶⁶ Ederer and Stremitzer 2013, 8.

¹⁶⁷ Ederer and Stremitzer 2013, 16.

In addition, both promises and reliability increased the principals' expectations when calculated in terms of expected payout.¹⁶⁸ More importantly, the agents' second-order beliefs about the principals' expectations increased significantly with increased reliability. "When a dictator made a promise to share money with the recipient, her belief about the amount the recipient expected she would contribute was significantly higher when the random device was reliable than when it was unreliable."¹⁶⁹ Lastly, promises decreased opt-out rates of principals. Opt-out rates went down from 21.6% and 12.8% with no message and empty talk, respectively, to 7.3% when agents promised.¹⁷⁰

Ederer and Stremitzer conclude that, since promise-keeping varies with reliability, and expectations depend on reliability, promise-keeping depends on reliability. This conclusion supports the guilt-aversion view, which holds that agents try to avoid the guilt that comes along with disappointing expectations – guilt that increases as expectations increase.

However, this conclusion runs into the same problems that Charness and Dufwenberg faced. First, the experiment allows us to conclude that expectations are important, but not why they are important. The guilt-aversion account tells us that expectations directly impact the agent's motivation, but this is not the only possible explanation. For instance, in this case, the expectations may signal the existence of a

¹⁶⁸ Ederer and Stremitzer 2013, 18.

¹⁶⁹ Ederer and Stremitzer 2013, 19.

¹⁷⁰ Ederer and Stremitzer 2013, 20.

social norm. A social norm of promising would trigger certain expectations, but once they ceased to exist, the social norm would cease to exist as well.

In other words, the guilt-aversion account tells us that agents cooperate because of expectations, not because of promises. When these two come apart, agents will act based on expectations. The social norms account tells us that the expectations determine whether a promise exists. If agents do not expect a promise to be kept, then it is not really a promise. Agents cease to cooperate not because there are low expectations, but because they are no longer obligated to – the promise is no longer a promise.

Ederer and Stremitzer claim that there are two ways to explain the moral force of promising: unconditional preference to keep a promise once made, or a preference conditional on the expectations of others.¹⁷¹ They argue that their experiment supports the latter view. But the social norms view essentially reduces the first explanation to the second. Expectations determine whether a promise exists. Hence, once again, the experimental conclusions do not allow us to differentiate between these two explanations.

Second, Ederer and Stremitzer claim that “A guilt-averse player suffers from guilt to the extent she believes she hurts others relative to what they believe they will receive... Guilt aversion only plays a role if the dictator sent a promise ... but not

¹⁷¹ Ederer and Stremitzer 2013, 3.

if she did not send a promise.”¹⁷² The idea is that we feel especially guilty when we break our promises, and so we are motivated not to do so.

Even without its other defects, the guilt-aversion explanation is at best incomplete in at least two ways. First, there is no explanation as to why promises affect expectations and guilt to a greater extent than other expectation-creation or – communication methods. Second, as we have seen, not only do individuals tend to keep their own promises, but they also punish promise-breaking as third parties. Guilt-aversion cannot explain why individuals would be averse to someone *else* feeling guilty, nor does it explain why the harm of disappointing expectations is punished with specific performance, and so it cannot account for the sort of third-party punishment we see in cases of efficient breach. Several experimenters have picked up on these defects of the guilt-aversion accounts, and have attempted to replace them with commitment-based accounts.

III.B. COMMITMENTS

Commitment-based approaches represent the other horn in Ederer and Stremitzer’s dichotomy. Whereas expectation-based accounts see promise-keeping as conditional on the perceived expectations of the promisee, commitment-based accounts view promises as socially unconditional commitment-producing devices.¹⁷³ That is, individuals have preferences for keeping their promises that do

¹⁷² Ederer and Stremitzer 2013, 9.

¹⁷³ That is, neither promising nor promise-keeping is conditional on social expectations, or beliefs about the beliefs and actions of others. However, they might

not depend on normative or empirical expectations. My social rights approach combines these two: individuals have an unconditional preference for keeping promises once the promise exists, but the existence of a promise depends on expectations. In other words, if you classify an action as a promise, meaning that the normative and empirical expectations that allow you to transfer your right to an action exist, your decision to keep the promise will not be contingent on social expectations. The rules apply because of social expectations; once the rules apply, the social expectations become unnecessary.

In this section, I will describe Christoph Vanberg's commitment-based approach, and then explore its weaknesses. In his 2008 paper, Vanberg defends a commitment-based approach by attempting to refute expectation-based accounts. In other words, he accepts the dichotomy between commitment- and expectation-based explanations, and then designs an experiment to eliminate the expectation-based option, thereby supporting the commitment view. He picks up on the problem of competing causal explanations, and then uses an experiment to test the causal connection between expectations and promise-keeping.

Vanberg agrees that second-order beliefs predict behavior, communication affects second-order beliefs, and trustworthiness increases with communication. The problem is that, in their experiments, Charness and

still be conditional on other factors, such as when one promises in conditions of uncertainty.

Dufwenberg, among others, do not show a causal link. That is, do promises change behavior because second-order beliefs change, or do promises change second-order beliefs because they change behavior in a predictable way?¹⁷⁴ The guilt-aversion theory relies on the idea that the motivation to keep promises can be reduced to beliefs about expectations. Hence, promises change behavior only insofar as they change beliefs about expectations. The experiments, however, only show correlation, not causation.

To show causation, one must show necessity: all other things equal, and eliminating multiple causes, the effect would not have occurred without the cause. For instance, Ederer and Stremitzer tried to show that second-order beliefs cause cooperation by showing that cooperation decreases as second-order beliefs decrease, regardless of promises. That is, promises correlate with expectations, and expectations cause cooperation. Hence, when promises and expectations diverge, cooperation is predicted by expectations. However, they did not eliminate other possible causes, nor did they show that the two events were not jointly caused by something else, namely, a triggered social norm. Hence, they did not establish causation.

Vanberg, then, attempts to show that the relationship between second-order beliefs and cooperation is not causal in the way that Charness and Dufwenberg or Ederer and Stremitzer think it is. He claims that promises cause both cooperation

¹⁷⁴ Vanberg 2008, 1469.

and changes in expectations – they both have a common cause. To prove this relationship, Vanberg rejects the opposite; he proves that expectations cannot cause full cooperation without promises. Thus, it seems either that promises and expectations jointly cause cooperation, or that promises cause expectations. Either way, promising is not explicable in terms of guilt-aversion – avoiding disappointing others’ expectations – alone.

To prove his theory, Vanberg conducted an experiment similar to that of Charness and Dufwenberg, but with an added step. Vanberg’s trust game not only included a random element in the game, but also with the players: half of the players were re-matched with new partners after communication, and before playing. And, while the trustees knew if they were assigned new partners, the principals did not. Thus, their expectations and actions would be based on the initial trustee’s communication or promise.¹⁷⁵

The expectation-based model suggests that behavior will be constant because expectations are constant.¹⁷⁶ If subjects keep promises only because of their second-order beliefs about the promisees’ expectations, then it should not matter who made the promise, as long as the promisees are unaware of the switch. That is, a promisee has expectations of a particular agent performing a particular action. If another agent fails to perform that action, the promisee’s expectations

¹⁷⁵ Vanberg 2008, 1470

¹⁷⁶ Vanberg 2008, 1470

cannot be said to be disappointed. That is why, for instance, a successful political candidate does not feel the need to make good on his opponent's campaign promises, nor would we blame him for not doing so. However, if the promisee does not know that the agents have been switched, her expectations may be disappointed by the new agent. Hence, according to the guilt-aversion model, the new agent should still feel guilty for thwarting the promisee's expectations, even if she did not create them. After all, on this model, expectations are the only explanans.

Vanberg's result show that promise-keeping after switching agents is not nearly as constant as the guilt-aversion model predicts. Though trustees' second-order beliefs did not change between the switch and no-switch conditions, their behavior did – they were more likely to choose *Roll* in the no-switch condition. 73% of trustees who promised to roll in the no-switch condition did so.¹⁷⁷ 54% of trustees chose *Roll* when rematched with a partner who had received a promise, and 52% when rematched with a partner who did not receive a promise. For trustees who did not promise, the percentages who chose *Roll* in the three conditions were 52%, 54%, and 56% respectively.¹⁷⁸ Hence, once partners were switched, neither making nor receiving a promise changed behavior – it was as if no promise had been made.

¹⁷⁷ Vanberg 2008, 1474

¹⁷⁸ Vanberg 2008, 1475

According to Vanberg, these results are inconsistent with the expectation-based approach: it matters who makes a promise, and to whom. Hence, he tries to substitute the guilt-aversion theory with a theory that can account for the fact that promisors are only motivated to keep their own promises, regardless of the knowledge and expectations of promisees. He proposes a commitment theory, which holds that a promise creates a particular relationship between the promisee and promisor. Expectations may matter, but they are secondary to this commitment relation:

A dictator's [trustee's] own promise has an effect on behavior toward a particular person lends support to the commitment-based explanation for promise keeping, according to which a promise creates a contractual obligation toward the person to whom it is made.¹⁷⁹

This commitment theory seems to support my theory of rights-based promising. A promisor's perception of the promisee's right to her action, and only her action, can explain Vanberg's idea of a contractual obligation. In other words, the agent-specific commitment relation may be easily explained by the perception that a promisee holds a right to an agent-specific action.

Vanberg's model, then, involves unconditional preferences. Since promises involve forming a commitment relationship, agents always prefer to keep their promises, given that they are in this relationship. Hence, Vanberg does not, and

¹⁷⁹ Vanberg 2008, 1476

cannot, account for the importance of expectations, and the conditionality of preferences based on expectations, from Ederer and Stremitzer's experiment.¹⁸⁰ Recall that Ederer and Stremitzer found that promise-keeping was conditional on the degree to which principals expected the agents to be able to keep their promises. Agents were not as likely to keep their promises – even if they could – when they believed the principals thought it unlikely. If there were an unconditional preference for following the rules of promising, or classifying a commitment as a promise and fulfilling it, agents would keep their promises when possible, given that they committed, regardless of the principals' expectations.

Furthermore, like the expectation-based accounts, the commitment-based account does not fully eliminate competing theories. Though it refutes certain aspects of the guilt-aversion model, it does not explain why the commitment device (in this case, a promise) works, and so leaves room for other theories. In other words, the commitment explanation passes the buck: it proposes an unconditional preference for keeping one's commitments, but is completely neutral as to what constitutes a commitment or why it would give us unconditional preferences.

III.C. SOCIAL NORMS AND RIGHTS

Hence, both the commitment-based account and the expectation-based accounts have gaps. The expectation-based accounts cannot explain why promises

¹⁸⁰ This is probably because Ederer and Stremitzer published their experiment five years after Vanberg published his.

increase cooperation to a great extent than mere prediction or statement of expectations. They also cannot explain why promise-keeping is agent-specific even when the promisee does not know which agent can fulfill the promise. Lastly, the theory itself gives us no reason to choose it over other theories, such a social-norms account.

The commitment-based account fills in these gaps, but has problems of its own. It cannot explain why promising would be conditional on the expectations of the promisee in cases in which the probability of performance is variable. It also does not give an explanation of the unconditional preference promises are supposed to give us. That is, it posits a commitment device without explaining what that device is.

Given the flaws with these two types of theories – in short, they cannot explain each other's data – I propose to amend the commitment account to include the expectationalists' experimental results. My social rights theory is essentially a hybrid between these two accounts. It explains the agent-specific commitment data from Vanberg while allowing conditionality based on expectations.

The reason that my social rights explanation can account for the data used to support both theories, though they diverge in their views on conditionality, is that it changes the level of explanation. Recall that in previous chapters, I have been attempting to explain the suite of practices surrounding promising in general, not just promise-keeping. Promise-keeping is a part of promising, but it likely does not

constitute the entire set of behaviors related to promising. If we analyze promising as a set of behaviors, and not just promise-keeping, we will be able to account for all sets of data.

On this theory, promising may include multiple behaviors that are conditional on social expectations – that is, beliefs about what others do and what they believe one should do. When those expectations change, the rules no longer apply. Let us examine, for instance, promise-making expectations. Expectations about promise-making do not tell us whether or not we should keep our promise, or change our motivation to keep our promise, but rather they signal whether or not the behavior counts as a promise, or falls under the rules of promising-making. When expectations about promise-making change, cooperation may change because the promise ceases to be a promise. The rules do not apply.

The social norms theory can thus explain the sense in which expectations matter for promising as a set of norm-governed behaviors, and so it can explain Ederer and Stremitzer's data better than Vanberg's commitment theory can. The rights-based aspect of the theory allows us to use, and further explain, Vanberg's idea of commitment to fill in the gaps in Ederer and Stremitzer's account. If a promise transfers a right over a particular action of the promisor, then, as long as the promise counts as promise, the action will belong to the promisee. Hence, in this case, promising-making is conditional on expectations, but promise-keeping may

not be. If the promisor views the promise as a transfer of rights, she will be unconditionally motivated to fulfill the right.

Further, since the right transferred is agent-specific, no other agent would be able to fulfill it without the promisee allowing the promise to be modified. Hence, no other agent would be obligated to fulfill the right, even if – or perhaps especially if – the promisee did not know who would be able to fulfill the right. Hence, the theory also explains decreased cooperation when partners are switched. Given the right you have transferred, I can have no obligation to fulfill your promises.

Thus, the social rights explanation best explains the data on promising by adding conditionality on expectations to the commitment-based account. In other words, it takes the best parts of both theories so it can explain all of the data. It uses rights to explain the idea of unconditional, agent-specific promise-keeping while using social norms to allow for conditionality in promising as an overarching behavior. These advantages give us good reason to adopt the social rights approach not only as the metaphysical and metaethical basis of promising, but also as the motivational basis.

CHAPTER 4: APPLYING A THEORY OF PROMISES TO SOCIAL NORM CHANGE

INTRODUCTION

In the previous several chapters, I have outlined and defended a theory of promising, a theory that takes promising to be supported by social norms that transfer rights to particular actions. I argued that classifying promising as social norm-supported – and particularly as a social right-transfer – allows us to explain our motivation to keep promises. In this chapter, I apply this motivational account to the process of norm change. Not only will the applicability of the theory give us further reason to accept it, but this particular application will also provide insight into the process of norm change.

Analysis of interventions aimed at altering harmful social norms shows that the more successful norm-change programs often contain a public commitment. A public commitment can take many forms, but, whatever form it takes, it always involves a public space in which individuals within a community declare their intention to adopt or abandon a norm to each other. Public commitments have been used to change harmful practices ranging from open defecation to female genital cutting (FGC). In this chapter, I will use my theory of promising to show why public commitments are effective in norm change, when they are effective, and provide tools for analysis of less effective public commitments. I will argue that public

commitments ideally have all of the essential features of promising, and will work best when they conform to strong existing norms of promising within the community.

Before applying my theory of promises to social norm change, I have two caveats. First, in the past several chapters, I have argued that promising is essentially a transfer of social rights to a future action of the promisor. This conclusion does not imply that all transfers of social rights to future action are promises. It is possible that there are other behaviors that transfer similar rights in a similar way. In order to classify a behavior as a promise – the aim of many parts of this chapter – we would seemingly need this further claim that all such rights-transfers are promises. The rights-transfer is a necessary but not sufficient condition. This, however, implies that we do not have a full theoretical account. There may be other practices, like consent, for instance, that transfer rights but are not promises. I am inclined to say that we can refine the theoretical account by restricting what is included in the rights-bundle. For example, perhaps consent never transfers an immunity, but promising always does.

However, although the lack of sufficient conditions specifying the set of rights-bundles is a theoretical gap, it is not practically as pressing because the particular rights-bundle will have to be specified in each case anyway, as we shall see. Hence, for the purposes of this chapter, it is not important that the public commitments used in norm change are promises, but rather that they have the

essential features of promises. The essential features of promising, according to the social rights theory, are a triggering act or speech act and a transfer of some sort of right to future action to another individual through a social rule. As long as the public commitment has these essential features, we can analyze it using the social rights view of promising, whether or not it is a promise as opposed to another form of commitment. I then reserve the specification of promising rights-bundles, as opposed to other rights-transfer-bundles, for future work. With this caveat in mind, however, I will refer to behaviors as promises when they contain these essential features of promising.

Second, I want to emphasize a point about applicability. I do not want to argue that my theory of promising is the only theory that can explain why certain public commitment-based interventions work and others do not. I also do not want to argue that my theory is the only theory that can produce the programmatic lessons that I list at the end of this chapter. However, I do want to argue that the fact that the social-rights theory of promising can explain success and failure in interventions and provide other programmatic insights is a reason to accept the theory. Hence I will not make the broader claim that one should accept my programmatic implications regardless of the theory of promising they hold. Rather, I claim that one reason in favor of my theory is its useful programmatic implications.

I will begin this chapter with a short review of my theory of promising. I will also outline some of the practical implications of this theory. In the second section, I

will give a more detailed account of public commitments and their role in the norm change process. I will include examples of programs that successfully use public commitments, as well as an overview of the analysis of public commitments in the norm-change literature. Finally, I will argue that public commitments have the essential features of promising, and are effective in changing norms precisely because they draw on strong existing social norms. I will then briefly examine programmatic implications of this theory.

WHEN IS A PROMISE NOT A PROMISE? A SOCIAL-RIGHTS VIEW OF PROMISING

My social-rights theory of promising aims to answer two questions. First, it aims to explain what a promise is. Second, it aims to explain our motivation to keep our promises. That is, given a rather large body of evidence that shows that we often keep our promises even when we could easily benefit ourselves by breaking them without consequence, my theory should tell us why we are so inclined to keep them. In this section, I will explain what my social-rights view is, and its theoretical and practical implications for actual promises.

In previous chapters, I argued for a theory of promising called the “authority-transfer view”. Essentially, this is the view that promises are fundamentally rights held by the promisee to a particular action of the promisor. I call it the “authority-transfer view” because the promisor transfers the right he has to his own action, as an agent, to the promisee.

Fundamentally, promises are rights transferred from the promisor to the promisee. But, as we saw in chapter two, not all rights have the same form – that is, there are different forms of rights that create different relationships between the right-holder and the duty-bearer. A bundle-theory of rights states that ‘rights’ are made up of one or more sub-rights, called ‘incidents’. A right could include a claim to an action or thing, a privilege to exclude others from it, a power to waive the claim or privilege, a power to transfer the claim or privilege, a power to create a claim or privilege, or an immunity from others interfering with the claim or privilege. And the list goes on – the right might include a power to transfer a power, or an immunity from others waiving a power.¹⁸¹ For example, my right to clean water does not exclude others from accessing clean water, but my right to my water bottle does exclude others. Similarly, a student can transfer her right to her class notes, but she cannot transfer her right to the grade she receives in a class.

If promises are fundamentally rights-transfers, then we can use this taxonomy of rights to analyze them. Since rights can vary in their components, it is likely that promises can as well. For instance, usually the promisee has the authority to waive the claim to the action, but not to transfer the claim at her will. However, there may be a context in which the promisee is allowed to transfer her claim. We can conclude that promissory rights are bundles, but we cannot

¹⁸¹ Hohfeld 1964, p. 36

universally classify them as particular bundles. Thus, at least theoretically, promissory rights can differ according to the incidents that comprise them.

Hence, a promise is the transfer of the right to one's action, and as such gives the promisor an obligation to perform. I further claim that this right is essentially social – that is, grounded in a social norm. This claim needs a bit more explication. I begin with the idea that rights and obligations are grounded in, or created by, systems of rules or principles. For instance, a moral obligation is an obligation that stems from moral rules or principles; a legal right is conferred by a law. Any right or obligation can then be classified according to the distinctions we make between systems of rules or principles.

Cristina Bicchieri (2006) divides systems of rules according to the characteristic motivation and expectations of those to whom the rules apply. Her system is particularly useful for two reasons. First, it is testable and can explain behavioral motivation. Given that I aim at a practical theory of promising, a classification that allows us to measure behavior and the motivations behind it is useful because it provides tools for application. Second, it is theoretically useful because it is tied to expectations, just like theories of promises that I have been using.

Bicchieri distinguishes between two types of social expectations in her system of social practices. First, empirical expectations are beliefs about how others within the relevant reference group act with regard to the rules or practice. They

are expectations about what others do.¹⁸² Second, normative expectations are second-order beliefs about what others in one's group think one should do, again with regard to the rules or practice. They are expectations about the normative beliefs of others.¹⁸³ Bicchieri distinguishes among different types of rules and practices according to which social expectations motivate the actor.

By combining these types of motivation, we get three levels of social rules and practices. First, what Bicchieri calls "customs" are socially unconditional – the actor's behavior is not conditional upon, or motivated by, either empirical or normative expectations.¹⁸⁴ Actors might be motivated by their own desires or self-interest, or by a set of non-social rules, like pragmatic or moral rules. Second, descriptive norms are conditional on only social expectations.¹⁸⁵ Actors follow these rules because others do so, and would not if others did not. Examples of descriptive norms include fashions and fads, as well as conventions like speaking a particular language.

Lastly, social norms are conditional on both normative and empirical expectations. The actor is motivated by both her beliefs that others follow the rules, and that they believe she should follow the rules.¹⁸⁶ The latter, normative expectations can be cashed out in two ways. First, the actor might be motivated by

¹⁸² Bicchieri 2006, 13

¹⁸³ Bicchieri 2006, 14

¹⁸⁴ Bicchieri 2006, 20

¹⁸⁵ Bicchieri 2006, 31-32

¹⁸⁶ Bicchieri 2006, 11

the belief that the perceived beliefs of others is legitimate – that is, she agrees that she should follow the rule. Second, she might be motivated by the belief that others will sanction her for non-compliance.¹⁸⁷

When the actor is motivated by the belief that the perceived normative beliefs of others are legitimate, the social norm is, in a sense, moralized. Instead of being motivated by sanctions, she is, at least partially, motivated by the notion that it is right or good that others think she should conform. She is thus motivated by a moral belief, and connects moral sentiments to the social norm. Moralization, however, does not imply unconditionality. If others stopped following to the norm, or if the actor stopped believing others think she ought to conform, she would no longer follow the norm.

I have argued in previous chapters that promises are supported by the last category of social rules, social norms. Thus, we can now specify what type of rights is transferred – namely, rights based on social norms. Hence, when we promise, we transfer socially-based rights to our actions, rights based on social expectations. The existence of the rights, as well as the corresponding obligations, depends on the existence of the norm, and thus the social expectations.

From the analysis of rights, we concluded that promissory rights might vary in their structure. That is, promissory rights might vary in the bundle of incidents that constitutes them. From the analysis of social norms, we can find several other

¹⁸⁷ Bicchieri 2006, 25

sources of possible variation in promissory rights: content, application, mechanism, triggers, strength, and the very existence of a right. I will discuss each of these in turn.

CONTENT: WHAT DO PROMISSORY RIGHTS APPLY TO?

Rights supported by social norms will differ just as social norms may differ from each other. Social norms specify their application for us – they apply in particular contexts or situations, within particular groups, and govern particular actions. Which actions can be transferred, or the content of the promise, might be limited by the norms supporting it, or by other conflicting social norms. However, there are constraints on how much the content of a promise can differ and still be considered a promise. We have seen that a promise transfers a right to a future action via social rules. Hence, the content cannot differ to the extent that the promise predicates a future action, the right to the future action is has meaning in that the promisor would not perform it anyway, and the future action is something one could have an obligation or right to.

These constraints are essentially John Searle's felicity conditions for promises, and come from the definition of a promise. According to Searle, a promise is made correctly if it predicates a future action of the promisor; the promisee prefers the action over its negation, and the promisor believes this; it is not obvious that the promisor would normally perform the action; the promisor intends to do

the action; and it counts as creating an obligation to do the action.¹⁸⁸ These conditions restrict how much the content of a promise can differ.

How the felicity conditions are fulfilled, however, can vary. For example, the norms supporting promising – promise-making, promise-keeping – might cover promises to sell oneself into slavery, or they might not. Suppose Eric tries to sell himself as a slave to Kyle. The promise might be overruled by a conflicting norm against slavery, or it might not be considered a promise in the first place, given the forbidden content, since it violates the condition that the promised act can be an obligation. In the former case, Eric has broken his promise and violated Kyle's right, but the breach is required and thus not blameworthy because Kyle is not allowed to hold the right. In the latter case, Eric never transferred a right to Kyle. Other examples might include promising an action to which the promisor does not have a right (promising for someone else), or promising something impossible. These do not count as promises because the promisor cannot transfer the right.

APPLICATION: WHEN DO PROMISSORY RIGHTS APPLY, AND TO WHOM?

Similarly, promises can differ according to the situations in which they apply, and the people to whom they apply. For example, a promissory obligation will not apply when another norm outweighs it. That is, if there is a more powerful conflicting norm, the promise will exist, but breach would be required, and thus blameless, as in the previous case. However, there may also be cases in which the

¹⁸⁸ Searle 1969, 63

promise does not actually exist simply because of the context in which it was made. For instance, a coerced promise may not be binding because it was never a promise in the first place. The social norm governing promise-making might exclude coercion.¹⁸⁹

The context of promising also includes the individuals involved in the promise. Social norms can differ in who they apply to, both generally and specifically. First, promises can differ in their subjects (who can make them) and their objects (who can hold them). For instance, we might discount promises made to children, God, or the deceased because it is plausible that these individuals cannot claim the promised action. Likewise, we might view promises made by children as less- or non-binding, or non-promises, because they do not fully understand the rules of promising. Second, on the group level, we might apply our norms supporting promising universally, or only to our own group. That is, our promissory norms might apply to only the in-group. A promise from the member of an out-group might not be considered a promise, and thus might not be trusted, because the norms governing promise-keeping and promise-making do not apply to her. Similarly, a promise to an out-group member might not be considered binding

¹⁸⁹ An example of promisors not keeping coerced promises, and not feeling guilty, might be virginity pledges. Bearman and Brueckner (2000) found that pledgers who broke their pledges and engaged in sexual intercourse did not necessarily feel guiltier than those who did not pledge. The lack of normative pull of the pledgers may have come from the socially-coercive nature of the promise.

either, and so the out-group member would have no claim.¹⁹⁰ Hence, there may be variation in who can transfer rights, and who they can transfer rights to.

MECHANISM AND STRENGTH: HOW DO PROMISSORY RIGHTS APPLY?

Social norms not only tell us the conditions under which to apply them, but also how to use and invoke them. That is, the norm will specify how to make a promise, and (ideally) tell us if a promise was made in a given instance. In practice, this leads to some ambiguity. For example, saying “I promise to X”, under normal conditions, might be an acceptable way to make a promise, as dictated by the particular social rules. But saying “I intend to X”, or “I will X”, might be less clear. The speaker might classify it as a promise, while the recipient might classify it as an intention. Diverse mechanisms can thus lead to variation not only among different promissory rights, but also within the same social norm. The mechanism can also vary the strength of the norm. Written promises might be considered stronger under some norms, while oral promises might be considered stronger under others.

In addition to the various mechanisms by which we can make promises, promissory norms can also differ in the situations that trigger the behavior of promise-making and promise-keeping. Social norms are embedded into behavioral scripts that are automatically activated in particular situations, and guide our behavior in those situations. We classify new situations according to our familiar

¹⁹⁰ For example, Morell (2014) finds that the influence of second-order expectations in promise-keeping situations greatly decreases when the promise is to an out-group member as opposed to an in-group member.

scripts. According to Bicchieri, “Once a script has been activated, the corresponding beliefs, preferences, and behavioral rules (norms) are prompted. I have argued that activating a norm informs the subject about what to expect others will do and what he himself is expected to do.”¹⁹¹ Different promissory rights may be embedded in different scripts, and those scripts may be activated in different ways. We will usually only engage in promising when certain scripts are triggered – social requests, business deals, and so on – and these triggering conditions will differ based on the script and its usual application.

Promissory rights can also differ in their strength. Promissory rights, in any social context, are part of a web of other norms, beliefs, and scripts. These webs will differ based on cultural context, but will likely contain norms and beliefs that can conflict with each other, or are ambiguous in their application. Within this web, promissory rights can be stronger or weaker than other norms – they might outweigh some, but not others. A relatively strong promissory norm will tend to be obeyed even when there are conflicting rights and obligations, whereas a relatively weak one may be easily set aside. Other factors, like content, context of application, and promissory mechanisms, can also affect the strength of the norm.

Another factor that might affect the strength of norms supporting promising, like promise-making or promise-keeping, is whether or not individuals have moral motivations to conform to the norms. Recall that normative expectations can

¹⁹¹ Bicchieri 2006, 57

include either sanctions or a belief in the legitimacy of the expectations, or both. If the individual believes that the perceived normative beliefs of others are legitimate, it may indicate that the norm has been moralized – that is, the norm gains “affective backing” in the form of moral sentiments, and the individual begins to view it as legitimate.¹⁹² According to Nichols, moralized norms are stronger and more deeply entrenched.¹⁹³ The idea that moralized norms are stronger is also supported by experimental evidence from Baron and Spranca, who find that individuals are less likely to make trade-offs with regard to their “protected values,” or values supported by moral sentiments.¹⁹⁴ Hence, we might expect norms supporting promising to be stronger when they are moralized.

However, there are also factors that make norms weaker. One that deserves mention is self-serving bias in ambiguous situations. Individuals, we assume, strive to satisfy their preferences, broadly construed. In pursuing their preferences, they often face constraints which, naturally, they try to avoid. Social norms are one such constraint. One way to avoid constraints is to make excuses, or interpret the norm in a self-serving way if it is ambiguous. Rabin describes the tendency to make excuses and to search for self-serving interpretations of a norm in order to circumvent it as a type of self-serving bias in the moral domain. It is “strategic belief manipulation”, in which an individual convinces herself that she is not constrained

¹⁹² Royzman et al. 2009, 11

¹⁹³ Nichols 2010, 14. See also Rozin 1999 for a more detailed account of moralization.

¹⁹⁴ Baron and Spranca 1997

by the norm by interpreting it in a lenient way.¹⁹⁵ Since social norms are generally side constraints, not goals in themselves, individuals are likely to interpret norms to serve their actual goals whenever they can, in any sufficiently ambiguous situation.¹⁹⁶ In terms of promises, this conclusion implies that norms like promise-keeping will be weaker in any ambiguous situation because the promisor will look for a reason why the promise what not really a promise, or an excuse to override the promise.

EXISTENCE: DOES PROMISING APPLY?

At this point, it is important to note that promissory rights can take on three forms: multi-lateral, unilateral, and promises to oneself. A multi-lateral promise is like a contract: each party transfers a right on the condition that the other party, or parties, transfer a right as well. Alternatively, a promise can be unilateral. A unilateral promise, like a vow or oath, is when only one party makes a promise, and it is not conditional on the other parties promising. Unlike a unilateral promise, a promise to oneself – a true vow or oath – is not the transfer of a right, but rather the waiver of a right to one’s own action, and hence is not a true promise. These distinctions affect the content of the norm in that they affect how many promises exist, and the conditions of those promises. For instance, a multi-lateral promise is no longer a promise if the promisor does not receive rights-transfers from the other parties to the promise.

¹⁹⁵ Rabin 1995, 2

¹⁹⁶ Rabin 1995, 1

In addition to the multi- or unilateral nature of promises, particular bundle of incidents that comprises them, the situations that trigger them, and the factors that strength or weaken them, promises can differ along many lines, include when and how they apply, what they apply to, and to whom they apply. Lastly, promises can also differ in their existence – that is, whether it is possible, within a society, to make promises at all. It is plausible that some cultures do not have promising at all; it would thus be impossible for them to promise. As we look at promising cross-culturally, and try to use norms that support promising to change other norms, we must first determine whether promising is possible within a society.

Some of the axes of variation, such as strength, will tell us whether a promise is binding, or whether individuals within a group are likely to follow the norm. Others, like mechanism, triggers, and situational application, tell us whether or not a promise exists. That is, if a promise was not made correctly, or according to the rules of promising, it was never a promise to begin with; no right was transferred. Other axes of variation, like content and application to individuals, could fall into either category. If I promise something impossible, it seems I have never transferred a right because I had no right to the action in the first place. If I transfer something forbidden, it might be that the right I transferred was merely outweighed by the rule forbidding it.

Take, for instance, a promise that transfer a right to an illegal substance, like heroin. Suppose I promise to deliver a kilo of heroin to you. In some sense, you

have a right to that delivery of a kilo of heroin, and you are likely to sanction me if I do not deliver it. However, there is also a conflicting sense in which you cannot have a right to the delivery, namely that heroin is illegal. But the conflicting legal prohibition does not imply that I do not have some sort of right (though not a legal one) to the heroin delivery from your perspective, or that I cannot transfer that right to you. From the government's point of view, I may have never made a promise because I have no right to the heroin, but it is unlikely that we would care about the government's perspective regarding promises in this situation anyway. We likely have our own black-market rules instead. Hence, whether or not a promise exists or applies might depend on the perspective one takes.¹⁹⁷

If we want to apply this theory of promising, and use it to change other, harmful social norms in various cultures, we must consider all of the axes of variation. Different cultures will have different norms that support promising, if they have these norms at all. In later sections, when I analyze situations in which promises tend to work and situations in which they do not, I will explain the difference using these axes of cross-cultural variation. In some cases, promise-based interventions are less successful because the promise was not binding

¹⁹⁷ We might also relate this conclusion to self-serving biases. I am likely to take the perspective on application and existence that serves my own interest. This can lead to disagreement on whether or not a promise exists even when the norms supporting the promise are identical between the two parties.

enough, and in other cases, because the promise never really existed in the first place.

In the next section, I will briefly explore the process of norm-change, as used by organizations such as UNICEF and Tostan, to end harmful practices. I will emphasize the use of public commitments in these interventions. Then, I will use the social-rights theory of promises to analyze the commitment aspect of such interventions, showing why they may work, and why they may not, in various circumstances.

THE NORM CHANGE PROCESS AND THE IMPORTANCE OF PUBLIC COMMITMENTS

Before discussing norm-change, I want to briefly note that I am specifically discussing *changing* a norm, and not *creating* a new norm where one did not exist to govern the specific behavior. The processes for norm-creation and norm-change are not necessarily the same¹⁹⁸, and though promises may be important for norm-creation as well, my discussion here will be limited to changing norms that already exist. In discussing the norm-change process, I will focus primarily on examples of interventions aimed at abandoning female genital cutting (FGC). The main examples of successful interventions are the Tostan program, which originated in Senegal, KMG in Ethiopia, and the FGM-Free Village model and Deir al Barsha in Egypt.

¹⁹⁸ Bicchieri holds that the norm-change process is distinct from the norm-creation process (See, for instance, Bicchieri and Mercier 2014). However, Mackie has articulated in lectures and conversations that these processes differ practically only because of context, and do not differ in form.

Analysis of the Tostan program in particular, as well as other programs, has been adopted by UNICEF's Innocenti Research Center, and the program is often used as a model for other programs.¹⁹⁹

In this section, I will use these examples, along with others, to illustrate a basic formula for norm-change. I will focus on the importance of a particular step in the process, public commitments. I will then discuss previous analysis of public commitments. In the final section, I will show how public commitments, and the norm-change process in general, connect to my theory of promises.

THE NORM-CHANGE PROCESS

From the definition of social norms, we can conclude that norms are supported by social expectations; the expectations we have are self-fulfilling in that they give us reasons to conform to the norm.²⁰⁰ To change a norm, then, we must change the relevant expectations. However, norms are embedded into scripts. Scripts, once activated by situational cues, tell us how to behave in particular circumstances, usually without conscious decision-making. According to Bicchieri, "Interpreting and understanding a social situation involves comparing it to similar ones we have encountered in the past"²⁰¹; once we interpret the situation, our action is guided by whatever script applies.

¹⁹⁹ See, for example, Mackie 2009.

²⁰⁰ Bicchieri 2006, 23.

²⁰¹ Bicchieri 2006, 98

Norms are constructed through expectations and desires, but they may be supported by subconscious scripted behaviors and categorizations. Hence, changing a norm will not only require changing expectations, but also disrupting the categorizations and scripts that support them and allow the norm to be followed subconsciously. In his description of the process of norm change, Gerry Mackie emphasizes both making people aware of the alternatives to their norm-governed actions – in Bicchieri’s theory, bringing scripts to the conscious level – as well as shifting expectations. He outlines six steps in the norm-change process: creating a core group, community discussion and deliberation, conditional commitment, public commitment, organized diffusion, and sanctions and monitoring.

The first step in changing a harmful norm is to form a core group. A core group is a relatively small group of first movers who can subsequently recruit others to the process of abandonment.²⁰² The reason for starting with a small group is both practical and theoretical. Practically, it is easier to convince and deliberate with a smaller group of individuals. Theoretically, a group of people within the community may be a more credible source of information, and has the incentives and resources to reach others. This core group does not initially change their behavior – they are not necessarily trendsetters – but rather they recruit others into the process of change.

²⁰² Mackie and LeJeune 2009, 11

The group may be composed of those most likely to accept new norms or ideas, those who are most influential to others, those who have the most opportunities to talk to others, or a combination. The core group should ideally be composed of a relatively representative sample of the community, including all sectors of the community. If the members of this group were expected to shift their behavior before the rest of the community – that is, if they were trendsetters – it would be difficult to recruit a representative sample. Only those most accepting of new practices would participate.²⁰³ But because the core group does not change their behavior first, and instead simply spreads information to create reasons for change, it is possible to recruit a more diverse group. And, since norms are sustained by expectations about the beliefs and behaviors of the entire reference group, the entire reference group, and not just the target population, needs to be represented. Otherwise, the intervention would be less able to alter the social expectations of the population, and perhaps less credible.²⁰⁴

Once the core group is recruited, they can begin the second step of the process, discussions and deliberation. Discussions may be about the practice in question, or about values more generally, or both. The discussions serve to change the “attitudes”, or personal and factual beliefs, surrounding the norm, and to bring those beliefs to the conscious level. According to Mackie and LeJeune, “Community

²⁰³ Bicchieri, forthcoming

²⁰⁴ Mackie 2009, 7. The example Mackie uses is of a less successful intervention against FGC in Kenya that targeted only young girls at risk for FGC.

discussion changes attitudes. People must gain awareness of the existence of an alternative, and the alternative must become valued more highly than the practice.”²⁰⁵

Hence, community discussion has three functions. First, it makes the practice salient to those engaged in it by bringing the subconscious scripts to the surface, allowing a community to find alternatives to the practice that fits within the script. Second, it can change any self-enforcing or supporting beliefs about the practice within the script. Self-enforcing beliefs are beliefs about the practice that keep it from changing. They might include beliefs that the practice is universal and that there is no alternative, or that speaking about the practice might bring bad omens.²⁰⁶ There may be other supporting beliefs as well, like beliefs that the practice is required by religion, or that it is necessary to preserve honor.²⁰⁷ By discussing the practice, individuals can learn about alternatives and evidence to amend their beliefs.

Lastly, community discussion can change how individuals view or value the practice. That is, after the practice becomes salient and individuals gain knowledge of alternatives or change other beliefs that support the practice, they may change their attitude toward it and no longer support it. In addition, discussion allows for a process of reflective equilibrium, in which community members can evaluate their

²⁰⁵ Mackie and LeJeune 2009, 14

²⁰⁶ Mackie 2009, 9; Mackie and LeJeune 2009, 15-16

²⁰⁷ Mackie 2009, 13

beliefs and values to make sure they are coherent.²⁰⁸ Given new knowledge, individuals may realize that one of their social practices contradicts their values. For instance, they may realize that the practice of FGC is inconsistent with the value of protecting one's children. According to Mackie, "Enunciation of more fundamental moral norms allows for culturally sensitive revision of more derived social norms."²⁰⁹

Additionally, it is important that these discussions are positive and non-directive. Since contradicting entrenched beliefs seems ridiculous at first, both the source and the message must be credible.²¹⁰ Holistic, non-directive, and positive programs demonstrate the good will and competence that makes the source and message credible.²¹¹

The third step of the process is conditional abandonment and diffusion. Once the core group has discussed the practice and changed their beliefs and attitudes, they may decide to conditionally abandon the practice. Since the decision to conform to the practice is dependent on the perceived behaviors and beliefs of others, the core group cannot decide to abandon the practice themselves. This is because, even if they develop reasons or incentives to abandon the practice, the norm-based motivation to conform will usually override these other incentives.

²⁰⁸ Bicchieri and Mercier 2014, 12; Rawls 1971, 20.

²⁰⁹ Mackie 2009, 15.

²¹⁰ Mackie and LeJeune 2009, 18.

²¹¹ Mackie 2009, 9-10.

However, they can decide to abandon conditionally – that is, they will abandon if enough others in their reference network do so as well.

Given the conditional abandonment and revised beliefs and attitudes, members of the core group now have reasons to recruit others in their community to conditionally abandon the practice as well. This is the process of diffusion. Since the message comes from community members, it is more likely to be credible. According to Paul Thagard, when individuals come across discrepant information, “such claims should not be automatically rejected either, because there may be more evidence for the claim than for what you believe: the fact that a credible source states the claim is one such piece of evidence.”²¹² Credibility gives us reason to accept a new claim, and community members are more likely to be credible than unknown outsiders. According to Mackie and LeJeune, the core group has “an incentive to recruit remaining members of the community to conditionally join in the effort, until a large enough portion, called the *tipping point*, is willing to coordinate on stable abandonment.”²¹³

The last steps of the norm change process are public commitment, organized diffusion, and community monitoring and sanctioning. Organized diffusion allows norm change to spread to other communities through various networks of communication.²¹⁴ Community monitoring and sanctioning makes the change more

²¹² Thagard 2005, 298.

²¹³ Mackie and LeJeune 2009, 11.

²¹⁴ Mackie and LeJeune 2009, 12.

stable by providing further incentives against reverting back to the old norm.²¹⁵

Since my focus is on promising and commitment, I will focus on the fourth step of the process, public commitment, for the remainder of this section, using examples to illustrate its importance.

EXAMPLES OF PUBLIC COMMITMENTS: THE NEED FOR TRUST

In this section, I will show how public commitments are helpful for successful norm-change interventions. I will illustrate this point using several examples, all of which are interventions targeting FGC. FGC is useful as an example of a harmful norm simply because of the quantity of research on norm change surrounding it. However, I will show that public commitment can be used to change other norms as well, depending on whether there are social norms that support promising.

Public commitment is what allows conditional abandonment to become actual abandonment. After community deliberation, and organized diffusion, individuals have incentives to abandon the social norm as long as others do as well – their abandonment is conditional on everyone else's. Public commitment allows individuals to see and believe that others will abandon the practice so that they can do so as well. According to Mackie, "there must be a moment of coordinated actual abandonment, so that each individual sees that most others do abandon... it

²¹⁵ Mackie and LeJeune 2009, 15.

requires the commitment of the greater part of the... community to abandon the practice together.”²¹⁶

In terms of social norms theory, public commitment changes empirical expectations by transmitting trustworthy information that others actually will abandon the norm. Once empirical expectations shift, normative expectations change as well – if individuals believe that enough others have abandoned, they will no longer fear social sanctions for abandoning themselves. That is, “The act of public commitment shifts reciprocal expectations, including each family’s [or individual’s] expectations.”²¹⁷

Hence, since public commitment relies on a background of trust, it is important that the commitment is credible. In the next section, I will connect the need for trust to the need for a social norm of promising or commitment-making. But there are other requirements for establishing trust that may be important as well. Recall that credibility depends on the reliability of the source,²¹⁸ and so a message that comes from an outside source may not be considered reliable. For instance, a public commitment that is imposed from the outside – or ‘top-down’ – would not be as effective because community members would not find it credible. Similarly, a public commitment without discussion and deliberation would not be as

²¹⁶ Mackie and LeJeune 2009, 14-15.

²¹⁷ Mackie and LeJeune 2009, 11.

²¹⁸ Thagard 2005

effective because individuals would not believe others have an incentive to change, and might still fear sanctions.²¹⁹

There are several examples of public commitments that were less successful because they did not have this requisite background of trust. For instance, in Ethiopia, the Amhara and Wolayta regions each had discussions at the subdistrict level, followed by one or two days in villages, and the communication of new laws. Commitment was downward, from subdistricts and regional government to villages. Afterward, participants were asked whether FGC was still practiced in their community; a third of survey respondents believed the practice to continue in secret.²²⁰ Because of the lack of community involvement, individuals did not trust others to abandon, even though there was public commitment.

In Kenya, alternative rite of passage (ARP) programs target only girls – that is, those at risk of being cut – instead of the entire community. Even though these girls may commit to not being cut, since only they commit to ending the practice, they may still be stigmatized by the broader community. Without full community commitment, individual actors cannot trust each other to abandon the practice, and not to sanction their abandonment.²²¹ In both of the cases, commitment did not work because of lack of trust – it was done in a way that could not guarantee to individuals that others in their relevant reference networks would abandon as well.

²¹⁹ Mackie and LeJeune 2009, 12.

²²⁰ Mackie 2009, 6

²²¹ Mackie 2009, 7.

The public commitments that worked did so because they could build this background of trust. As we will see, in the case of commitments, trust depends on the existence of social norms that support promising. In the next section, I will show how all of the successful programs I discuss relied on background social norms surrounding promising. For now, I will describe several forms of public commitment and their use in various programs.

In Senegal, the Tostan program uses public declarations for commitment and coordinated abandonment. Initially, these declarations were not part of the Tostan program; rather, they arose organically from communities who decided that continuing to practice FGC conflicted with their values and goals. Soon, public declarations were adopted into the Tostan program, but only after communities complete three years of learning and deliberation, and after organized diffusion.²²² Public declarations bring together multiple intermarrying communities to abandon together, and generally include speeches, dancing, music, and other festivities.

In Ethiopia, the KMG programs used *edir* commitments to coordinate abandonment of FGC. An *edir* is “an insurance and self-defense organization for all whose decisions are binding on members.” *Edir* commitments can be scaled-up to subdistricts and districts, and can be made publicly as well. Public commitments

²²² Mackie 2009, 8.

through the *edir*, as well as weddings of uncut women, were celebrated and publicized.²²³

In Egypt, the FGM-Free Village program used public declarations – including celebrations – with a written component. Community leaders publicly signed a written commitment at the end of the celebrations, which then became a source of pride for the communities.²²⁴ Similarly, in Deir al Barsha, also in Egypt, town leaders announced that they would not cut their daughters, and publicly signed a document to that effect, along with cutters and other families.²²⁵

Though the form of the public commitment varied in different programs and cultural contexts, all of these successful programs used some form of commitment to ensure trust. To see why they were successful, we can apply the theory of promising developed in earlier sections to the public commitments.

PUBLIC COMMITMENTS AS PROMISES: PUTTING THE THEORIES TOGETHER

In this section, I will show that the public commitments used in the examples of successful interventions have all of the essential features of promises under my definition. In addition, I will argue that these public commitments were successful *because* they had these features, and could thus ensure trust. In essence, then, we can use the social norms that surround promising to change other, more harmful

²²³ Mackie 2009, 5.

²²⁴ Mackie 2009, 4-5

²²⁵ Mackie 2009, 13.

social norms. In the last part of this section, I will offer some practical suggestions and caveats, given the types of commitments that work and the types that do not.

PUBLIC COMMITMENTS ARE PROMISES, AND THAT IS WHY THEY WORK

Recall that our theory of promising states that the essential features of promises are the transfer of a social-norm-based right to a particular action. We can now analyze whether the four examples of successful public commitments – Tostan, KMG, FGM-Free Village, and Deir al Barsha – have these features. We can also explore whether the commitments used in the less successful programs – ARP in Kenya and regional commitments in Ethiopia – had the essential features of promising, and how the mechanism of commitment effected their success.

Tostan

In Senegal, Tostan uses public declarations to demonstrate the change in the practice so that individual empirical expectations shift. But how do public declarations achieve this shift? According to Mackie, public declarations are connected to honor, a social sanction that incentivizes individuals to keep their word. In Senegal, “It is a matter of sacred honor for individuals and communities to keep their word.”²²⁶ Public declarations are explicitly linked to word-giving, or promising, and to the sanctions that come with being honorable or dishonorable. Hence, not only are these public declarations instances of word-giving, but they are social-norm-based: keeping one’s word is connected to normative expectations.

²²⁶ Mackie 2009, 8

When a public declaration is made, the word-giving transfers the right to cut – or not to cut – to all of the other individuals within the community.

Since the whole community commits, this is a multi-lateral promise: each member of the community is transferring the right to cut, and is in turn receiving a right from every other pledger. The right each member transfer is the privilege to cut, and the rights he receives are each other community members' privileges to cut (or to marry a cut girl). Since each member has given up his privilege to cut, he now has a duty not to cut (or not to marry an uncut girl). The social sanctions of honor support this transfer by adding a social norm of promise-keeping.

Further, the fact that the public declarations transferred social rights was important to the success of the program. In their research comparing Tostan to other FGC intervention programs, Diop and Askew found that it was crucial that the declaration is a commitment, as opposed an expression of approval or disapproval: "in these village-based societies, as is the case across rural Africa, a public declaration is more powerful than individual expressions of opinion about a socially approved practice."²²⁷ The key part of the declaration is that it is binding, not that it expresses attitudes. Bindingness creates trust. And this bindingness comes from the transfer of rights, which are in turn supported by social expectations and sanctions. The bindingness is also the reason why it is important that the public declarations are promises, and not just statements of intention. A statement of

²²⁷ Diop and Askew 2006, 139

intention does not bind because there are no sanctions or legitimate normative expectations to create the motivation to follow through, and without the knowledge of the added motivation, individuals might not believe the declarations of others.

KMG

In Ethiopia, the KMG program used an already-established social practice to produce public commitment: the *edir*. An *edir* is a social association made up of individuals connected by work, family, friendship, district, ethnic group, or gender. Originally, *edirs* dealt with life-cycle events; families could pay into the *edir* to receive financial assistance for marriage, death, and birth. However, *edirs* have expanded to include general financial assistance, as well as other types of assistance in emergency situations.²²⁸ Belonging to an *edir* compels members to assist other members, and to follow the by-laws of the institution. Members are incentive to comply through social sanctions and, in some cases, monetary sanctions: members may be fined for disobeying by-laws.²²⁹

Given that it is supported by sanctions, and expectations to comply with particular rules, following the dictates of the *edir* falls under the definition of a social norm. But is it, in this case at least, an instance of promising? A promise is a binding agreement that gives others the right to one's action. An *edir* is just that. By joining, members agree to follow the by-laws, and in doing so, give other members

²²⁸ Aredo 1993, 79

²²⁹ Aredo, 1993, 80.

the right to their compliance. Though *edirs* are not explicitly forms of word-giving, they are, in a sense, agreements. And since they do transfer rights, supported by social sanctions, they can be considered a form of promising. Like the public declarations in the Tostan program, *edirs* produce multi-lateral promises. Upon joining an *edir*, each member pledges to every other member, and hence gives up the privilege to disobey the *edir's* dictates, while in turn gaining the privilege of each other member. Since claims correspond to the loss of a privilege, each member thus has a claim over the actions of other members.

Edir agreements are also effective in promoting and sustaining FGC abandonment. Between 2008 and 2009, the number of *edir* commitments – that is, adding sanctions against FGC to by-laws – increased from 11 to 292.²³⁰ The position of the *edir* in the community, and the credibility it adds to FGC abandonment, allow the abandonment to be credible to members. According to KMG, *edir* involvement “has secured buy-in and ownership, which sustain the dialogue mechanism” and reinforce change.²³¹ *Edirs* allow the community to invest in change and in commitment, so the commitment becomes reliable. And the reliable commitment allows individuals to trust that others will abandon the practice so they can as well.

FGM-Free Village Model and Deir al Barsha

²³⁰ KMG Ethiopia 2015

²³¹ KMG Ethiopia 2015

In Egypt, the FGM-Free Village program uses a similar commitment mechanism to Tostan, though they add written commitment as well. Like Tostan, they use public declarations to allow individuals to “declare publicly their commitment to abandon FGC” in a credible way, so abandonment can be coordinated.²³² In order to achieve this credibility in commitment, the FGM-Free Village program focuses on commitment community leaders: “Declarations make villagers aware that opinion makers in their communities... are opposed to FGM.”²³³ At the end of the celebratory and festive declaration ceremony, participants, particularly opinion leaders, sign a document pledging to abandon FGC, which becomes a source of pride for the village.²³⁴ This pride acts as a positive social sanction to keeping one’s commitment.

The village of Deir al Barsha, also in Egypt, used a similar written commitment, and similarly focused on village leaders. There, too, using a written pledge made the promissory aspect of abandonment explicit, and provided extra social sanctions for non-compliance. In both of these cases, the signed declaration is a formal multi-lateral commitment – a contractual agreement – and includes social sanctions, we can conclude that it, too, is a form of a social norm-based promise.

In addition to the verbal commitment in public declarations, the written declarations used by the FGM-Free Village program and Deir al Barsha made the

²³² Barsoum et al. 2011, 18

²³³ Barsoum et al. 2011, 17

²³⁴ Barsoum et al. 2011, 18

contractual nature of the declaration explicit, and added sanctions – for instance, pride – to make it binding. If abandonment can only be achieved through altering empirical expectations, the bindingness that comes along with pride in the community is crucial because it shows change in behavior, not just attitudes. Though it is important for leaders to express their disapproval, the pride created by the written commitment binds them to abandonment, and thus allows empirical expectations to be created. In this case, then, the promise within the public declaration is causally efficacious.

Amhara and Wolayta

In the previous four cases, public commitments, including declarations and *edir* commitments, worked well to secure stable abandonment of FGC. However, there are other cases – such as the Amhara and Wolayta districts in Ethiopia – that include public commitments, but are less successful. In these cases, we can use my theory of promises to explain why the public commitments did not work as well.

In Amhara and Wolayta, individuals did commit, but they committed either at the subdistrict level, or only after new laws and higher-level commitment was communicated to the villages. It is possible that these commitments were actual promises; that is, it is possible that they triggered a social norm that transferred rights to the rest of the community. However, since the promises were made at a high level, in subdistricts instead of villages, individuals could not be sure that they would be held accountable for their commitment. They were committing to others

in their districts, not their friends and neighbors. In other words, they were making multi-lateral commitments, but did not necessarily know to whom they were giving their privilege to cut, or whose privileges they were receiving. Sanctioning others would thus be difficult, and one may not believe that he would be sanctioned.

Without sanctions, or commitment within the relevant network, the norm of promise-keeping may not have been strong enough to outweigh the incentives to keep practicing FGC. When the two conflicted, individuals would have believed others would choose FGC over promising, and would thus choose the same. Hence, this program likely did not fully consider the application of the norm of promise-keeping, or to whom it applies.

It is also possible that the commitment, in this case, did not even qualify as a promise. A promise transfers a right from one individual to another. If the promisee does not believe that the promisor is actually transferring a right, and the promisor knows this, it is doubtful that the promisee accepts the right that is transferred. She would not believe she has a claim, and the promise would not exist. In the case of commitment at a subdistrict level, as opposed to the village level, given that the norm of FGC is so strong, it is possible that individuals do not believe that others intend to transfer a right, especially since they might not know who was transferring and receiving the right, and hence do not believe they have a claim. In that case, commitment is not only ineffective, but it also not normative – it does not count as a promise.

From the case, it is not clear which explanation best accounts for the failure of the program – whether the promise was weak because it lacked sanctions, or whether it did not exist because it was not made under the right conditions. More information would be needed from population involved to be able to determine this. However, both explanations rely on a social-rights-based account. The first explanation admits that commitments may require independent social sanctions to be kept, and hence that keeping the commitment may itself be a social norm. The second explanation also relies on a social-norms account: commitments must be accompanied by the correct second-order beliefs in order to even count as promises. Whichever explanation is correct, the social-rights theory is informative.

ARP

In Kenya, Alternative Rite of Passage programs used commitment, but like the cases of Amhara and Wolayta, the commitments were less successful because they targeted the wrong groups. ARP programs target the at-risk population: young girls who have not yet been cut. These girls commit to each other that they will not be cut. Like the other programs, it is a multi-lateral promise, like a contract. However, often it is not the girls alone who make the decision to cut; rather, the parents or the family make the decision. If the girls do not believe that they have control over the action, then they cannot transfer the right to the action via a promise. This program likely did not fully account for the interaction of the content of the promise and to whom it applies. Hence, the commitment was likely not a promise, and it was likely not normative.

But even if the commitment were a promise, it might again be less effective because the norm would be too weak – given the sanctions involved, the FGC norm would outweigh the norm of promise-keeping. Again, the program did not adequately account for the relative strength of the norm of promise-keeping, compared to the FGC norm. To illustrate, suppose a girl promises to her schoolmates that she will not be cut. If she were cut, she would face sanctions from her schoolmates for breaking her promise. If she remained uncut, she would face sanctions from the rest of the community, including the possibility of never being able to marry. Even if she does meaningfully promise not to be cut, the incentives to cut outweigh the incentives to keep her promise.

Neither of these two explanations requires a social-rights theory of promises – any theory of promises will say that an agent cannot promise something she has no control over, or that a promise can be outweighed – but the theory can easily account for them. Though this case does not necessarily prove the explanatory power of the theory, it at least shows that the theory can work in the analysis of a variety of situations of failed commitment. Hence, it adds to the thoroughness of the theory's application.

Commitments may fail to be effective for several reasons. If the wrong group is targeted – the at-risk group instead of the decision-makers – the commitment may not count as a promise. But even if the commitment is a promise, it may not be

effective because the promise is too weak, and the existing social norm outweighs it. I will discuss these issues further in the next section.

CONCLUSIONS: USING PROMISES IN PRACTICE

We have seen that, in cases of norm interventions, public commitments ideally have all of the essential features of promises, and are causally efficacious to the extent that they function as promises. That is, in these cases, public commitments work best when they take on the features of promises – when they draw on existing social sanctions, trigger expectations of compliance, and do so by transferring rights to actions. When they do not take on the features of promises, or do not account for the specific form of the culture’s norms that support promising, they are less likely to work. Promises work because they create the trust necessary for empirical expectations to shift.

This conclusion does not imply that public commitments – or promises – are either necessary or sufficient for norm change, or even for the narrower goal of ending FGC. They are not sufficient because they are merely one step in a norm change process. And they may not be necessary because there may be other ways to build up enough trust to shift empirical expectations. But what we can show is that in some of the instances where promises were used to change norms, they were effective, given that certain conditions were in place, though we still need to further specify those conditions with continued research. Promises can be a powerful tool for norm change.

Public commitments have the essential features of promises, and promises are themselves supported by social norms. So how can we use the norms surrounding promising to change other, more harmful social norms? The key is to be able to trigger the norms that surround promising, if they exist, and to make sure they provide strong enough incentives. The promise must actually exist, and it must not be outweighed. Given that promises are supported by social norms, we can expect that they will vary by culture or community, just like other social norms. We cannot assume that the same actions will count as promise-making across a variety of cultures. For instance, in the United States, individuals often interpret legal contracts as promises, and so they have expectations of punishment and particular performance in the case of breach.²³⁵ In Tonga, on the other hand, it is doubtful that promises exist at all.²³⁶ Instead of triggering a norm of promise-keeping, Tongan contracts will be interpreted simply as legal documents that include only the provisions and sanctions included within them.

Hence, the first step in using promises to end other harmful social norms is to recognize cultural variability in promissory norms, and to assess what will trigger a promise – and what will count as a promise – in a particular cultural context, if it has norms that support promising at all. I have already outlined how promises can vary cross-culturally: when and how they apply, to what they apply, and to whom they apply. We can see this cultural variability even in the FGC abandonment cases

²³⁵ Wilkinson-Ryan and Baron 2009

²³⁶ Korn and Dektor-Korn 1983

previously outlined. The declarations in Senegal relied on the connection between oral word-giving and honor, while the declarations in Egypt used written commitment and pride. These commitments may have been successful partly because they took the culture-specific aspects of promising into account.

Second, once we recognize that promises supported by social norms, we can understand why they are effective. Social norms give individuals incentives to comply through normative expectations, either in the form of sanctions or in the form of belief in the legitimacy of others' beliefs. However, to use a norm effectively, the incentives for compliance created by normative expectations needs to outweigh any contrary incentives. If the promise-keeping norm is too weak compared to the harmful norm – for instance, if sanctioning is unreliable or not severe enough – the incentives will not outweigh the incentives to keep following the stronger harmful norm. It is likely that ARP programs and the programs in Amhara and Wolayta had this problem.

Given the weakness of the promise-keeping norm, individuals will not believe the promises of others, and will not shift their empirical expectations. In this case, either the promise-keeping norm must be strengthened, or the incentives for complying with the harmful norm must be diminished. Otherwise, the promise cannot be effective in creating the necessary trust. Alternative Rite of Passage (ARP) programs are a good example of reducing incentives to comply with a harmful norm. By taking away the incentive to perform FGC as a rite of passage, they strength the

incentives to keep one's promise relative to the incentives to perform FGC. But since they only reduce certain incentives, these programs are ineffective in isolation.

The second step, then, is evaluating the relative strength of each norm in terms of their incentives. It may then be necessary to bolster the strength of the promise-keeping norm or ameliorate the incentives to comply with the harmful norm. Third, interventions must target the decision-making group, as opposed to the at-risk group, if they are different. If the target group does not have control over the relevant action, they cannot meaningfully promise to stop it. They cannot transfer rights if they do not have the rights in the first place. This is likely the problem with using ARP programs in isolation. The girls cannot transfer the right to cut (or not) if they do not have it. This step takes variation in both content and application of promissory rights into account.

Hence, the third step is to target decision-makers, the group that has the ability to transfer a right to the action. If interventions are sensitive to cultural differences in the norms supporting promising, and to individuals' incentive structures, and target the right-holding population, promises can be an effective tool in changing harmful social norms.

CONCLUSION

Throughout this dissertation, I have described a social-rights theory of promising, and provided reasons to accept such a theory. I have argued that promises transfer rights, and that these rights are essentially social – they are supported by social norms. I have also argued that this theory of promising explains behavioral data on promising and cooperation better than any of the competing theories. Lastly, this theory can inform programming for behavioral change, including behavior change in development contexts. The usefulness of the theory provides another reason to accept it.

In this final section, I will explore some implications and areas for further study using the social-rights theory of promising. The first broad area for further study concerns further work on promising as a tool for cooperation and social change. The second area concerns implications for political philosophy, specifically the social contract position, and suggests using social science in these philosophical inquiries. The third area concerns investigation in other related promise-like or rights-transfer phenomena, like consent, apology, and debt. I will discuss each of these in turn.

PROMISING FOR SOCIAL CHANGE

Given a social-rights theory of promising, there is more work to do on promises as a means for behavior change. There are several remaining questions that I have not addressed. First, in describing promises as social rights-transfers, I

have only dealt with ideal or perfect promises, promises that actually transfer rights. But as I discussed in Chapter Four, not everything that looks like a promise is actually a promise, and not all promises actually succeed in motivating promise-keeping. Though I did discuss reasons why a promise might fail, a more thorough study of imperfect or failed promises would be helpful for predicting and motivating behavior.

Relatedly, I only discussed one application of the social-rights theory of promising: public commitments in FGC abandonment. There could be many other applications, but in the development context and outside of it. For example, public commitments, when they are actually strong enough promises, are effective in FGC abandonment. There are other harmful practices, like child marriage or gender-based violence, which might be changed with a similar approach, given other relevant similarities. This approach to promising might be helpful in assessing intervention strategies for these practices as well. In addition, promises have been shown to be effective in solving common-pool resource dilemmas, most notably by Elinor Ostrom.²³⁷ My analysis of promising might aid our understanding of how, and under what circumstances, communities can solve this type of dilemma.

PROMISING AND POLITICAL PHILOSOPHY

If the social-rights theory of promising is correct, it also has implications for political philosophy, and in particular, social contract theory. This point relates to

²³⁷ Ostrom 1990, 43-45

Hume's critique of the social contract tradition: political and social institutions arise from conventions, but "this convention is not of the nature of a *promise*: For even promises themselves ... arise from human conventions."²³⁸ Modern iterations of the social contract position, including contractualism and contractarianism, in many forms, require promises to be prior to moral principles or social practices. In particular, we need to be able to agree, and believe that others will follow through on their agreement. For instance, T.M. Scanlon states that, when forming institutions, we are motivated by "the reason we have to live with others on terms that they could not reasonably reject insofar as they are also motivated by this ideal."²³⁹ This motivation requires the belief that others are motivated as well, or that others agree.

This objection does not necessarily apply to all of those working in the social contract tradition. For example, most contractualists do not assume that contracts are pre-social. They are thought experiments done from within an existing society, with given institutions and values, and may not apply outside of that society and value system. Rawls, for instance, argues that his contractualist approach may not apply to all peoples: "The Law of Peoples holds that decent nonliberal points of view exist, and that the question of how far nonliberal peoples are to be tolerated is an essential question of liberal foreign policy."²⁴⁰ The objection may not even hold for

²³⁸ Hume 2000, 314

²³⁹ Scanlon 1998, 153-154

²⁴⁰ Rawls 2001, 10

most contractarians, but, given the social-rights theory of promising, it is an objection that needs to be reconsidered. And, given a plethora of social science data on promises, we have added tools to reconsider it.

PROMISING AND RELATED PHENOMENA

The social-rights theory of promising concerns only one specific phenomenon, namely, promising. But there are other phenomena that are clearly related to or similar to promising that could benefit from a similar analysis, or could use this analysis of promising. For example, the social-rights theory predicts that promises are likely to be kept, given certain circumstances, but it might also contribute to inquiries into broken promises. For example, Benjamin Ho has studied the use of apologies to repair trust in various kinds of relationships.²⁴¹ A theory of promising might aid our understanding of why apologies work, or if they work, to ameliorate broken promises.

There are also other phenomena that are rather similar to promises in that they seem to transfer rights in a similar way. Consent, for example, seems to transfer a right, or perhaps a liberty, but that right may be revoked at any time. Debt seems to be the result of a right-transfer as well, but one in which the party transferring still retains some rights as well. Not only can a social-rights theory of promising help to analyzing these phenomena and predict behavior, but the analysis might inform exactly what bundle of rights can be transferred through a promise.

²⁴¹ Ho 2012

That is, the analysis might help us use rights theory to draw distinctions among these various phenomena. And these distinctions might help us predict behavior.

CLOSING THOUGHTS

I have hopefully provided sufficient reasons to accept my social-rights account of promising and apply it in intervention programming. I hope the applicability of the project provides motivation to others to continue this line of research as well. Most importantly, I hope I have provided some kind of a model for interdisciplinary work in philosophy, using resources from various disciplines to aid philosophical inquiry. If anything, this dissertation should show that philosophers can and should work across disciplines, and can and should apply their work to the real problems that exist around us. As Marx states in his *Theses on Feuerbach*, "The philosophers have only interpreted the world, in various ways; the point is to change it."²⁴²

²⁴² Marx 2002

REFERENCES

Aredo, D. 1993. "The Iddir: A study of an indigenous informal financial institution in Ethiopia." *Savings and Development*, 77-90.

Arrow, K. J. 1994. "Methodological individualism and social knowledge." *The American Economic Review*, 1-9.

Baron, J., & Spranca, M. 1997. "Protected values." *Organizational behavior and human decision processes*, 70(1), 1-16.

Barsoum, G., et al. 2011. "National efforts toward FGM-free villages in Egypt: The evidence of impact. A summary of the evaluation of The FGM-Free Village Project implemented by Egypt's National Council of Childhood and Motherhood." *Poverty, Gender, and Youth Working Paper No. 22*, Population Council.

Bearman, P. S., & Brückner, H. 2000. "Promising the Future: Virginitiy Pledges as They Affect Transition to First Intercourse." *ISERP Working Papers*.

Bicchieri, C. 2004, "Cooperation and Communication: Group Identity or Social Norms?" In: *Teamwork: a Multi-professional Workshop*, Palgrave Macmillan, 103-136.

Bicchieri, C. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.

- Bicchieri, C. Forthcoming. *Norms in the Wild*. Cambridge: Cambridge University Press.
- Bicchieri, C., Lev-On, A., and Chavez, A. 2010. "The medium or the message? Communication relevance and richness in trust games." In: *Synthese* 176, 125-147.
- Bicchieri, C., & Mercier, H. 2014. "Norms and beliefs: how change occurs." In: *The Complexity of Social Norms*, 37-54. Springer International Publishing.
- Bicchieri, C., & Muldoon, R. 2011. "Social norms". *Stanford Encyclopedia of Philosophy*, plato.stanford.edu.
- Blair, R. J. R. 1995. "A cognitive developmental approach to morality: Investigating the psychopath." *Cognition*, 57(1), 1-29.
- Charness, G., & Dufwenberg, M. 2006. "Promises and partnership". *Econometrica*, 74(6), 1579-1601.
- Crimmins, J.E. 2015. "Jeremy Bentham", *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, <<http://plato.stanford.edu/archives/spr2015/entries/bentham/>>.
- Dawes, R.M. et al. 1977. "Behavior, Communication, and Assumptions About Other People's Behavior in A Commons Dilemma Situation", *Journal of Personality and Social Psychology*, Vol. 35, No. 1, pp. 1-11.
- Diop, N.J. & Askew, I. 2006. "Strategies for Encouraging the Abandonment of Female Genital Cutting: Experiences from Senegal, Burkina Faso, and Mali". In: *Female*

Circumcision: Multicultural Perspectives. Ed. Rogaia Mustafa Abusharaf. Philadelphia: University of Pennsylvania Press. 125-141.

Ederer, F., & Stremitzer, A. 2013. *Promises and Expectations* (No. 1931). Cowles Foundation for Research in Economics, Yale University.

Eigen, Z. J. 2012. "When and why individuals obey contracts: experimental evidence of consent, compliance, promise, and performance." *The Journal of Legal Studies*, 41(1), 67-93.

Hart, H.L.A. 1958. "Legal and Moral Obligation." *Essays in Moral Philosophy*, ed. A.I. Melden, Seattle: University of Washington Press, pp. 82-107.

Hart, H.L.A. 1961. *The Concept of Law*, Oxford: Oxford University Press.

Hayek, F. A. 1948. *Individualism and economic order*. Chicago: University of Chicago Press.

Ho, B. 2012. Apologies as signals: with evidence from a trust game. *Management Science*, 58(1), 141-158.

Hohfeld, W. N. 1964. *Fundamental Legal Conceptions as Applied in Judicial Reasoning*. Ed. Walter Wheeler Cook. New Haven: Yale University Press.

Hooker, B. 2011. "Promises and Rule-consequentialism". *Promises and Agreements: Philosophical Essays*. Oxford: Oxford University Press, 237-254.

- Hume, D. 2000, *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton, Oxford: Oxford University Press.
- Johnson, M.C. 2000. "Becoming a Muslim, Becoming a Person: Female 'Circumcision,' Religious Identity, and Personhood in Guinea-Bissau". In: *Female "Circumcision" in Africa: Culture, Controversy, and Change*. Ed. Bettina Shell-Duncan and Ylva Hernlund. Boulder: Lynne Rienner Publishers. 215-233.
- KMG Ethiopia. 2015. "Comprehensive Sexual and Reproductive Health." Retrieved April 20, 2015, from <http://www.kmg-ethiopia.org/index.php/what-we-do/our-programs/comprehensive-sexual-and-reproductive-health>
- Kant, I. 1997. *Foundations of the metaphysics of morals; and What is enlightenment?* Trans. Lewis White Beck. Upper Saddle River: Prentice-Hall.
- Korn, F., & Korn, S. R. D. 1983. "Where people don't promise". *Ethics*, 445-450.
- Lewis, D. 1969. *Convention: A Philosophical Study*, Cambridge: Harvard University Press.
- Mackie, G. (2009). More effective and less effective programs to abandon harmful practices in five countries. *Florence, Italy: UNICEF Innocenti Research Centre*.
- Mackie, G., & LeJeune, J. (2009). Social dynamics of abandonment of harmful practices: a new look at the theory. *Special Series on Social Norms and Harmful Practices, Innocenti Working Paper, 6*.

- Marx, K. 2002, "Theses on Feuerbach", trans. W. Lough, *Marx/Engels Internet Archive*, www.marxists.org.
- Melenovsky, C. Forthcoming. " Promises, Practices, and Reciprocity"
- Menger, C. 1985. *Investigations into the Method of the Social Sciences with Special Reference to Economics*, trans. Francis J. Nock. New York: New York University Press.
- Mikhail, J. 2007. "Universal moral grammar: Theory, evidence and the future". *Trends in cognitive sciences*, 11(4), 143-152.
- Morell, A. 2014. "The Short Arm of Guilt: Does it only hit who is close?" *MPI Collective Goods Preprint*, (2014/19).
- Narveson, J. 1971, "Promising, Expecting, and Utility," *Canadian Journal of Philosophy*, Vol. 1, No. 2, pp. 207-233
- Nichols, S. 2010. "Emotions, norms, and the genealogy of fairness". *Politics, philosophy & economics*, 9(3), 275-296.
- Orbell, J. M., Van de Kragt, A. J., & Dawes, R. M. 1988. "Explaining discussion-induced cooperation". *Journal of Personality and social Psychology*, 54(5), 811.
- Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Ostrom, E., Walker, J., & Gardner, R. 1992. "Covenants with and without a sword: Self-governance is possible". *American Political Science Review*, 86(02), 404-417.

- Owens, D. 2006. "A Simple Theory of Promising," *Philosophical Review*, vol. 15, no. 1, pp. 51-77.
- Rabin, M. 1995. "Moral preferences, moral constraints, and self-serving biases." *Department of Economics, UCB*.
- Rawls, J. 1955. "Two Concepts of Rules." *The Philosophical Review*, Vol. 64, No. 1, pp. 3-32
- Rawls, J. 1971. *A theory of justice*. Harvard university press.
- Rawls, J. 2001. *The law of peoples: with, the idea of public reason revisited*. Harvard University Press.
- Robins, M.H. 1976. "The Primacy of Promising." *Mind*, vol. 85, no. 339, pp. 321-340.
- Robinson, S. L., & Rousseau, D. M. 1994. "Violating the psychological contract: Not the exception but the norm". *Journal of organizational behavior*, 15(3), 245-259.
- Ross, W.D. 1930. *The Right and the Good*. Indianapolis: Hackett.
- Royzman, E. B., Leeman, R. F., & Baron, J. 2009. "Unsentimental ethics: Towards a content-specific account of the moral-conventional distinction". *Cognition*, 112(1), 159-174.
- Rozin, P. 1999. "The process of moralization." *Psychological Science*, 10(3), 218-221.
- Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge, MA: Belknap Press.

- Schapiro, T. 2003. "Compliance, Complicity, and the Nature of Nonideal Conditions." *The Journal of Philosophy*, Vol. 100, No. 7, pp. 329-355.
- Searle, J.R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Sellars, W. 1963. "Philosophy and the Scientific Image of Man", in *Science, Perception, and Reality*, New York: Routledge & Kegan Paul, pp. 1-40.
- Sheinman, H. 2008. "Promise as practice reason". *Acta Analytica*, 23(4), 287-318.
- Shiffrin, S. V. 2007. "The divergence of contract and promise." *Harvard Law Review*, 708-753.
- Shiffrin, S. V. 2008. "Promising, intimate relationships, and conventionalism." *Philosophical Review*, 117(4), 481-524.
- Strawson, P. 1974. *"Freedom and Resentment" and Other Essays*. London: Methuen.
- Thagard, P. 2005. "Testimony, credibility, and explanatory coherence." *Erkenntnis*, 63(3), 295-316.
- Thomson, J.J. 1976. "Killing, Letting Die, and the Trolley Problem." *The Monist*, 59, 204-17.
- Thomson, J.J. 1990. "Giving One's Word." *The Realm of Rights*. Cambridge, MA: Harvard University Press, 294-321.

Turiel, E. 2002. *The culture of morality: Social development, context, and conflict*. Cambridge University Press.

Vanberg, C. 2008, "Why Do People Keep Their Promises? An Experimental Test of Two Explanations", *Econometrica*, Vol. 76, No. 6, pp. 1467-1480.

Weber, M. 1949. *The Methodology of the Social Sciences*, E. A. Shils and H. A. Finch (ed. and trans.), New York: Free Press.

Weber, M. 1968. *Economy and Society*, ed. Guenther Roth and Claus Wittich, Berkeley: University of California Press.

Weston, D. R., & Turiel, E. 1980. "Act-rule relations: Children's concepts of social rules". *Developmental Psychology*, 16(5), 417.

Wilkinson-Ryan, T., & Baron, J. 2009. "Moral judgment and moral heuristics in breach of contract". *Journal of Empirical Legal Studies*, 6(2), 405-423.

Wilkinson-Ryan, T. 2012. "Legal Promise and Psychological Contract," *Wake Forest Law Review*, Vol. 47, 843-873.