CHAPTER 2

Outline of a Theory of Prudence

There are many theories of normative rationality—that is, theories of how agents rationally ought to act—including Kantian theories of categorical rationality[1] and realist theories of rationality as responsivity to objective normative reasons.[2] However, the dominant conception of rationality in everyday life[3] and academic theorizing[4] appears to be instrumentalism, the 'means-end' view that rationality is a matter of agents adopting optimal means for achieving their ends (or goals). Notice further that this appears to be how we normally understand prudence. We say, for example, that prudent students study hard to maximize their chances of meeting their academic goals, that prudent people look both ways before crossing streets to maximize their chance of getting to the other side safely, and so on. Hence, although multiple theories of normative rationality exist, I will simply assume for theory-construction that normatively prudent behavior (henceforth 'prudence') is a matter of acting in ways that have the best-expected outcome for the agent acting, *vis-à-vis* satisfying their ends. Notice that prudence thus defined could involve what the ancient Greeks called *phronesis*, or practical wisdom and excellence of character for living an extraordinary life.[5] As we will see, this book's unified theory of prudence and morality does require forms of excellence: the development and exercise of virtues of fairness over a complete life (a kind of moral excellence), as well as excellence in the ability to judge which actions (constrained by principles of fairness) have the best-expected long-term aggregate outcomes. However, we will also see that there is a sense in which this book's theory does not demand 'excellence', at least as some laypeople and theorists understand it. For although some moral

theories (such as utilitarianism) and proposed moral exemplars (e.g., Jesus Christ, etc.) appear to espouse extreme forms of altruism as 'moral excellence', this book's theory holds that true moral excellence (fairness) does not require extreme sacrifice.

In any case, assuming that prudence involves acting in instrumentally-optimal ways that possess the best-expected outcomes for the acting agent, the next question is how to understand ends. Some theorists argue that ends and their value are subjective[6]. Others (hedonists) argue that ends and their value are reducible to pleasure and pain[7], others still that ends and their value reduce to preference-satisfaction[8], and so on.[9] This book cannot resolve these questions. Instead, I will simply contend that the theory of prudence that I outline below (and allow potential exceptions to in Chapter 5) is plausible given an intuitive, everyday understanding of ends and their value, with which any adequate theory of ends and their value should defeasibly cohere. I leave it to readers to judge whether this is the case.

1. Prudence as Maximizing Expected Aggregate Lifetime Utility

If prudence involves acting in ways that have the best-expected outcome for the acting agent, questions also arise about the nature and persistence of agents. Do agents like you and I literally persist over time, or are we mere 'person-stages' who exist only for a moment before new 'person-stages' come into existence?[10] As Michael Smith and I argue[11], we do not need to settle these matters to do normative theorizing. For regardless of these metaphysical issues, throughout 'our lives'—that is, as different 'person-stages'—we tend to care about 'our' past and future. For example, at present, I wish that certain things in 'my' past had gone better than they did. I also have all kinds of hopes for 'my' future—for example, that my 'future self' sees this book to be

well-received, that 'I' happily live to old age with my spouse, and so on. To be sure, empirical research shows (and commonsense recognizes) that we do not always care about our past or future 'selves'—that all too often we treat our past and future 'selves' akin to strangers.[12] Still, it is equally clear, especially in our 'cooler' moments—particularly when we reflect on the subjective experiences of 'our' past and future selves[13]—that we often care deeply about 'our' past and future. We ordinarily have standing hopes, dreams, desires, and worries about 'our' future, as well as retrospectives wishes and desires about 'our' past. Consequently, for the sake of theory-construction, we have sufficient grounds to treat agents as though they persist over time—based on our standing interests about 'our' past and future.[14]

We can use these facts to settle another important issue, at least well-enough for our purposes: namely, the relevant timeframe for evaluating outcomes. As Aristotle posited, our highest good (prudentially) seems to be a good life as a whole.[15] This claim not only appears widely accepted in the philosophical literature[16], but also a fair approximation of what laypeople normally care about: namely, for our *life* to go well. Consequently, for the purposes of theory-construction, it seems safe to suppose that prudence is a matter of acting in ways that have the best-expected aggregate outcomes across our lives as a whole (i.e. 'best-expected lifetime value').[17] Of course, there are many questions here, such as whether prudence permits or requires temporal favoritism, such as favoring the well-being of one's future selves over the past.[18] However, I want to set aside these matters as well. I will now argue that we can outline a compelling theory of prudence without presupposing any controversial answers to them.

2. Problems Maximizing Expected Aggregate Lifetime Utility in an Uncertain World

If we assume that prudence involves acting in whichever ways have the best-expected lifetime value, the next question concerns what it *is* for any given action to have this property. As many argue[19], the most natural way to understand to this is 'additively', identifying prudent actions as whichever series of individual actions that, when aggregated across a full life, generate the greatest-expected *sum* of well-being. However, while this seems plausible, we nevertheless face two sets of problems in attempting to determine which actions achieve this.

One problem concerns inter-world utility comparisons.[20] For an agent to determine which actions maximize their expected sum of lifetime value, it seems like the agent must be able to compare aggregate lifetime utilities of different possible lives they might have lived. However, in cases of 'big decisions'—such as whether to write a book, have a child, get married, or change careers—such comparisons seem infeasible.[21] For example, would my life be better as whole if had I never gone into academia? Might my life be better now if I abandoned my academic career in favor of another? These things seem impossible to estimate reliably: comparing alternative lives we have never lived appears to 'require a kind and amount of knowledge which seems far out of reach (at least currently) for mere mortals like us'.[22]

Some contend that inter-world comparisons are irrelevant to prudence, contending that the only thing that matters to prudence is how one feels about one's actual life.[23] Perhaps this is true, though I am skeptical. Because we cannot settle this, let us simply recognize that there seem to be *prima facie* reasons to think inter-world comparisons are relevant to prudence. First, given that prudence is widely understood to be acting in ways that have the best-expected aggregate

lifetime utility for the agent acting, and given that different life choices clearly can result in very different aggregate life-outcomes, alternative lives seem relevant to maximizing expected lifetime utility. Second, in everyday life we sometimes dwell on inter-world comparisons for these reasons. When our lives frustrate us, it appears common to wonder whether alternative choices might have been more prudent. We sometimes ask questions like, 'What if I had chosen a different career? Might my life have been happier? Indeed, might I be happier *now* if I changed careers?'. Inter-world comparisons thus seem relevant to prudence, at least in the counterfactual sense that if we *could* make such comparisons reliably, then we would be able make choices with better expected aggregate lifetime outcomes.

In addition to problems of inter-world comparisons, we also face profound difficulties evaluating the probabilities of different outcomes over life as a whole. First, the future—particularly the distant future—is formidably difficult to predict. Although short-term events often go how we expect, sometimes they turn out very differently, vastly transform the probabilities of future events. Consider a victim of a serious crime, such as a school shooting. Prior to the crime, the person may have regarded the likelihood of them dedicating themselves to gun-control activism to be near-zero. However, after the crime, the probability of them dedicating their life to activism may become extremely high. Although an extreme example, our lives routinely appear to transformed by the unexpected—by who we meet or fall in love with, unexpected job-opportunities, illnesses, and so on. How can we reliably estimate which life-choices have the best-expected aggregate lifetime outcomes (viz. prudence) when we do not know how the probabilities of outcomes might change over our lifetime?

Second, human beings are generally poor affective-forecasters—that is, at estimating how we are likely to experience the value of different outcomes.[24] For instance, people tend to think they will be profoundly happy winning the lottery. However, empirical studies indicate only short-term spikes in subjective well-being for lottery-winners, not long-term gains.[25] More generally, we often think that particular outcomes (getting a promotion at work, etc.) will bring us satisfaction, only to find the value of the outcome to be very different than expected.

Third, human beings often display preference-reversals[26], desiring one thing prospectively in the present (e.g., winning the lottery) only to retrospectively desire the opposite later (e.g., wishing we had never won). Similarly, one may want to tell a lie in the present—and then tell it because one expects to get away with it—only to wish one hadn't when the future comes (because of getting caught or feeling guilty).

Finally, there are transformative experiences: experiences that may transform us both epistemically—giving us salient information about the subjective value of outcomes only after a choice—as well as personally, altering our core preferences in unpredictable ways.[27] For example, a person at one point might think their life as a whole might be happier if they remain childless, yet they may find the experience of giving birth to a child so transformative that they now value having children tremendously. Over a complete life, we plausibly undergo a variety of transformative experiences ranging from having children, to falling in love, to taking a university course that fundamentally alters one's values and career goals, and so on.

Thus, although prudence requires acting in ways that have the best-expected aggregate lifetime outcomes, we appear to face an important set of problems. Life is so uncertain, and our capacities to reliably compare and predict the value and likelihood of different outcomes over the course of our lives so limited, that it is unclear how we can reliably determine which actions maximize aggregate expected lifetime personal utility. Various solutions to these problems have been proposed.[28] However, instead of evaluating them I want to explore a new solution: a novel normative and descriptive theory of prudence and prudential cognition.

3. Bruckner's Prudential Original Position and Minimax Regret

There is a long tradition of thought that (a) prudence is a matter of grappling with life's immense uncertainty, and (b) morality is only prudent solution to that uncertainty. For example, the Bible repeatedly emphasizes life's unpredictability[29], holding that moral behavior is a prudent response.[30] Similarly, the Stoics argued that life's uncertainty makes it rational to develop self-mastery over one's desires, emotions, and beliefs, holding in turn that such self-mastery constitutes morality.[31] I will now defend a new theory of prudence in this tradition.

Donald Bruckner has argued that because life is profoundly uncertain, prudence requires acting on principles that presuppose radical uncertainty.[32] Bruckner argues that this prudential requirement can be modeled by a Prudential Original Position, a hypothetical model closely akin to John Rawls' famous 'original position', but where an agent deliberates behind a veil of ignorance applied not to questions of social justice, but rather to their own life as a whole.[33] Rawls, of course, understands principles of domestic and international justice (respectively) as principles that instrumentally rational citizens in a domestic society[34] and peoples

internationally[35] would agree to from an extreme position of *inter*personal uncertainty—from

behind a veil of ignorance that prevents them from arbitrarily privileging themselves over any

other citizen or people. Bruckner's innovative idea is that principles of prudence should be

understood in a similar fashion—as principles that are instrumentally rational from a standpoint

of extreme *intra*personal uncertainty; that is, from behind a veil of ignorance that withholds from

an individual agent any knowledge of how their life as a whole is likely to go.

Let us assume for the sake of argument that Bruckner is correct. Which principle(s) of prudence

are rational from the Prudential Original Position? Bruckner rejects several principles:

intrapersonal utilitarianism (simple utility-maximization), maximin (choosing whichever action

has the best worst outcome) and minimax loss (choosing whichever action minimizes the

maximum shortfall experienced by any particular 'person stage'), arguing that all of these

principles are implausible in test-cases.[36] There are also well-known problems with another

principle for choice under extreme uncertainty: the principle of insufficient reason (treating all

outcomes as equally likely)—as it seems arbitrary to treat all outcomes as equally probable when

one has no evidence of this.[37] Ultimately, Bruckner argues that only minimax regret, the

principle of acting in ways whose maximum possible amount of regret is lowest, gives plausible

results.[38] Finally, Bruckner derives the further result that the more an agent cares about their

future and past, 'the more [they] will be acting as if following the utilitarian principle of

prudence [viz. utility-maximization] when actually following minimax regret.'[39]

It is not my aim to evaluate Bruckner's argument. Instead, I will now argue that two parts of his

solution—minimax regret, and its convergence with maximum expected aggregate value the

more one cares about the past and future—cohere with a kind of 'moral risk-aversion' that prudent children, adolescents, and adults typically progressively internalize for grappling with life's uncertainty.

4. 'Moral Risk-Aversion' as Maximizing Expected Aggregate Lifetime Utility

Prudence, again, is intuitively a matter of acting in whichever ways have the best-expected aggregate outcomes for one's life as a whole. Consequently, when it comes to first-order choices—that is, to deciding how to act in any given situation—prudent agents will have to weigh and aggregate different possible outcomes, in turn estimating the relative probabilities of different weighed aggregate outcomes given the ways they might change over time.[40] However, or so I will now argue, when life's immense uncertainty is combined with several empirical regularities, we can see that prudent agents should also typically internalize higher-order dispositions (beliefs and preferences) as constraints on first-order prudential deliberation. As we will now see, a set of moral dispositions constituting a kind of 'moral risk-aversion'—that is, aversion to risking the violation of moral norms—appear to be higher-order prudential dispositions of this sort: dispositions that prudent individuals typically learn are likely to make their lives as a whole go better in an otherwise uncertain world.

4.1 Depictions of the Prudence of 'Moral Risk-Aversion' in Fiction and Religion

Consider so-called 'morality plays': fictional works conveying moral and prudential lessons. One common type of morality play is 'tragedy' or 'tragic drama'. A central lesson of tragedies is this: although violating (plausible[41]) moral norms may appear to the acting agent to have greater expected personal benefits than moral behavior, it is nevertheless imprudent to risk violating

9

such norms, given how immoral actions can lead to maximally regrettable outcomes that moral actions typically do not have.

Consider the standard plot of tragedies. Typically, the protagonist risks violating moral norms because it appears to them to have the best-expected outcome. However, their risk goes horribly wrong. Unexpected events cause their moral transgressions to result in prudential disaster—disaster that they regret far more than they likely would have regretted had they obeyed moral norms. The lesson implicit in tragedies, as such, is that prudent people should not risk violating moral norms because of just how much they might regret it. Let us call this first trope of morality plays *immorality is not worth the risk of maximal regret*.

This trope is widely embedded in tragedies across history and cultures. Consider *Hamlet*, considered one of the most powerful tragedies in world literature. Why is *Hamlet* so tragic? Upon close examination, we see that not just one but three main characters suffer maximally regrettable personal misfortune for risking immoral behavior. At the outset, we are introduced to King Hamlet, the recently deceased king of Denmark, who we learn slew King Fortinbras of Norway many years ago (a 'moral risk' insofar as killing people is generally immoral). Shortly thereafter, Prince Hamlet encounters his father's ghost, who tells him that his brother Claudius (Prince Hamlet's uncle) murdered him to seize the throne and marry his widow, Queen Gertrude. After encouraging Prince Hamlet to seek revenge on Claudius (another 'moral risk' by King Hamlet, since 'taking the law into your own hands' is plausibly immoral), Prince Hamlet hatches a plan to kill Claudius and avenge his father's death (Prince Hamlet's first 'moral risk'). Alas, the plan goes horribly wrong. When Hamlet goes to kill Claudius, he finds Claudius praying

and—because Hamlet reasons that Claudius would go to heaven if killed while praying—Hamlet relents. This causes Claudius to fear Hamlet, leading Claudius to hatch a plan to kill Hamlet in return (Claudius' 'moral risk'). Finally, after Hamlet tragically kills another innocent character (Polonius) as a result of mistaken identity (another 'moral risk' on Hamlet's part), Hamlet and Claudius end up killing each other, Queen Gertrude is killed, Hamlet's entire royal family is killed, and the new King Fortinbras—the son of the former king killed by the original King Hamlet—seizes the Danish throne.

Hamlet is such a powerful story—a moral and prudential tragedy *par excellence*—because each character's willingness to risk violating moral norms unexpectedly leads to profoundly regrettable outcomes for them and others they care about. King Hamlet's 'moral risks' (invading Norway, killing another king, and encouraging his son to seek revenge) unexpectedly lead to the loss of everything he valued during his lifetime: his kingdom and family. Similarly, Prince Hamlet's 'moral risks' (plotting to kill Claudius and killing Polonius by mistake) unexpectedly result in him losing everything he values: his life, the throne, and his family's lives. Finally, Claudius' 'moral risk' (plotting to kill Hamlet) results in *him* losing his life and the throne. The moral and prudential lesson of Hamlet thus seems to be this: 'Do not be like King Hamlet, Prince Hamlet, or Claudius. Although violating moral norms may appear to have the best-expected outcomes, it is not worth the risk. Life is profoundly uncertain and the potential disasters that immoral actions can cause are maximally regrettable, far more regrettable that doing the morally safe thing'.

Hamlet is far from alone in presenting this lesson. It is a common storytelling trope across history and cultures. Many modern works of tragic-fiction illustrate the same lesson.[42] We also see it in ancient works, such as Old Testament stories ranging from Adam and Eve's expulsion from the Garden of Eden for disobeying God's commands (which Christians consider to be moral requirements), to God punishing King David for infidelity and the treachery of sending Bathsheba's husband off to be killed in battle. We also see the trope in non-Western heroic fiction. For example, in a recent 'Bollywood' film[43], a corrupt police officer cooperates with two drug dealers, accepting bribes in return for letting them continue their illegal activities. Later, through chance events, the drug dealers brutally rape a young woman the police offer unexpectedly falls in love with. This devastates the corrupt officer, leading him to regret his past corruption, change his ways, and become a hero. Similarly, in Asian martial arts films, the following is a common type of plot: a powerful mob boss immorally kills the protagonist's family or loved-ones. Then, months or years later, the protagonist uses their martial-arts mastery to avenge their deaths, killing the boss' henchmen, and ultimately, the boss themselves.[44] In every case, the implicit moral-prudential lesson seems to be this: 'Do not be like Adam and Eve, King David, the corrupt police officer, evil henchmen, or the mob boss. Do not risk violating moral norms because it seems to have likely positive benefits: you might profoundly regret it'.

Now, in tragedies, the tragic characters are typically never able to learn from their moral-prudential mistake(s) until it is too late—that is, until after their death or death of their loved ones. This is what makes such stories so powerful. However, another genre of morality play just mentioned—heroic fiction—typically combines this first moral-prudential lesson with another: the *morality has better likely personal outcomes in the long-term* trope. Indeed, both tropes

together constitute a distinct subtype of morality play: *redemption stories*. In redemption stories, the protagonist typically begins as a 'rogue'—a thief, smuggler, or criminal who risks violating moral norms in seeking personal gain.[45] However, after their actions unexpectedly result in prudentially disastrous outcomes they regret, they get a 'second chance.' Instead of their story ending in tragedy, the hero learns their lesson: they become no longer willing to risk violating moral norms, and instead come to treat the expected long-term aggregate benefit of moral behavior as greater than the expected benefits of immoral behavior.

For example, in a recent popular film, a young warrior prince disobeys his father, the King, thus destroying a fragile truce between two factions—a 'moral risk' in the sense that the prince's presumptive obligation is to serve the King.[46] The King banishes the prince from the kingdom, punishing the prince for his transgression. For a time, the prince is dejected and hopeless, regretting his transgression. His regrets, however, lead him to change his ways, culminating in him choosing moral behavior even when it appears imprudent: he chooses to sacrifice himself to save the lives of his friends. Fortunately, to the hero's surprise, the events that led him to sacrifice himself turn out to be a test his father created to see whether he learned his lesson. In the end, the prince is welcomed back into the kingdom as a hero. The implicit lesson here, and in many redemption-stories like it, is: 'Learn the moral-prudential lesson the prince learned. Even when moral behavior does not seem like it will benefit you, if you are patient and stick to it, then it probably will benefit you in the long run, far more than immoral behavior is likely to'.

Morality plays thus tend to contain two related moral-prudential lessons: a negative lesson and a positive one. The negative lesson of tragedies is that, given life's great uncertainty, violating

moral norms is not worth the risk because of the immensely regrettable outcomes immoral actions can cause. The negative lesson, in short, appears to be that life's uncertainties should lead one to (A) act on minimax regret, and by extension (B) avoid immoral behavior, because (C) one can regret immoral behavior far more than moral behavior. The positive lesson of redemption-stories then is that one should internalize the negative attitudes just described—avoiding immoral behavior on minimax regret grounds—precisely because this strategy ('moral risk-aversion') has greater expected aggregate personal lifetime benefits than immorality.

These same lessons are also implicitly embedded in doctrines of divine or cosmic justice across a wide variety of world religions. Many religions envision heaven as a long-term reward for moral behavior, and hell as punishment for evil. Similarly, in Hinduism, moral behavior is seen as rewarded and bad behavior punished in reincarnation via *karma*. Taken literally, these doctrines teach that it is certain that one will be eternally rewarded for good and punished for evil. However, given our epistemic situation—that is, given that the existence of heaven, hell, reincarnation, and so on, can seem remote and uncertain—these doctrines implicitly convey the same two moral-prudential lessons as tragedies and redemption-stories. The doctrine of hell conveys that immorality is not worth the maximally-regrettable risk of divine punishment. Conversely, the doctrine of heaven conveys that morality is likely to result in greater long-term benefits. These are the same moral-prudential lessons as in morality plays: they aim to convince the hearer that immorality is not worth the risk of greater maximal regret, and that morality has greater expected long-term value.

Notice, finally, how closely these lessons hew to Bruckner's theory of prudence. Bruckner holds that prudence requires minimax regret given life's uncertainty, and that the more we care about the past and future, the more this rule should converge with maximization of expected long-term personal benefit. This is what morality plays teach, with the added substantive claims that (A) given certain empirical regularities (viz. observed consequences of moral and immoral behavior), (B) prudent agents *should* care a great deal about the past and future, (C) learning from the maximally regrettable consequences of moral mistakes (by ourselves and others), such that (D) we should treat immoral behavior as not worth the risk (viz. minimax regret), and instead (E) internalize attitudes that moral behavior has greater expected long-term aggregate value.

Finally, morality plays also convey a further lesson that will prove to be important in Chapter 3: specifically, that how one sees the likelihoods--of whether moral or immoral action is likely to have better personal outcomes—depends on whether one has internalized the above moral-prudential lessons in one's own attitudes (i.e., one's beliefs and desires). Prior to 'learning their lesson'—that is, prior to internalizing 'moral risk-aversion'—characters in redemption-stories perceive the likely benefits of immoral behavior as outweighing the risk. However, after learning their lesson, they see things the opposite way, judging the benefits of immorality to be a mirage, believing instead that the long-term benefits of moral behavior are greater.

The next question is whether these moral-prudential lessons are actually true for real human agents, morality plays and religious myths aside. We will now see that there are ample reasons to think that prudent people do indeed tend to internalize the above lessons in their standing psychological attitudes, precisely because of life's great uncertainty. Chapter 3 will then outline

how a normative moral and social-political theory, Rightness as Fairness, is derivable from this prudential foundation. Chapter 4 will then detail how this unified theory of prudence and morality explains a variety of philosophical and empirical phenomena. Finally, Chapter 5 will consider potential counterexamples.

4.2 The Prudence of Learned 'Moral Risk-Aversion' in Childhood

Consider childhood and adolescent television programming: so-called 'after-school specials' that present moral and prudential lessons to children. A typical storyline here involves a child or adolescent choosing to risk violating moral norms for some expected benefit, such as lying to their parents to attend a music concert. The child usually hatches a 'perfect plan' for getting away with the risk. Alas, as in tragic dramas, unexpected events lead the child to suffer negative consequences they highly regret, such as punishment by parents, school officials, or law enforcement. Finally, however, like redemption-stories, such programs usually end on a happy note, with the child or adolescent coming to believe that violating moral norms is not worth the risk, and that moral behavior has better long-term benefits than immorality.

We can see how ubiquitous these moral-prudential lessons are by considering a few examples. In one television episode in a series targeting children[47], a father allows his daughter to use his credit-card as a reward for academic success.[48] The daughter then succumbs to temptation, going overboard with her spending, in turn hatching several different plans with her classmates—all of which violate moral norms—to earn back the money back. However, every plan backfires, leading her to eventually regret her behavior and 'come clean' to her father, who then rewards her for telling the truth. Similarly, in a more fanciful episode, a teenager is struck by lightning,

16

enabling him to see the future.[49] The character then uses this power to help a friend cheat on an exam. The power wears off, however, and his friend fails the exam due to lack of preparation—leading them both to regret their plan. Finally, in a third episode, a teenager goes to dangerous lengths to get a female student to go to a school dance, bugging the girl's bedroom with microphones to overhear her conversations with female friends.[50] Unfortunately for the boy, the young women find out and take revenge on him, leading him to regret his actions.

As we see in these stories and many others, the same two moral-prudential lessons depicted in tragic fiction and redemption-stories are common in youth television programming. Such programs give children many examples, in vivid contexts they understand, of how violating moral norms is not worth the risk of immense regret, and how moral behavior has greater expected long-term benefits.

These lessons are also typically socially-reinforced in children on an everyday basis. In early childhood, young children often misbehave, risking violating moral norms as a result of impulsivity, such as playing in the house when their parents tell them not to, getting into fights on the playground, or engaging in minor forms of theft. Here, though, is what often happens. Although the child commits these acts impulsively, children who are well-supervised and well-raised are regularly caught and punished for such behavior. Typically (though not always), such forms of punishment lead to regret—if only for 'getting caught'. Finally, children are also socialized to believe there are greater longer-term benefits—such as good performance at school, as well as parental and other social rewards—if they *do* obey moral norms. Consequently, unless they have psychopathic tendencies[51], the typical child will learn to internalize the above

17

empirical regularities in their attitudes. They will come to believe, on the basis of empirical regularities they observe around them, that (A) immoral behavior is not worth the risk of immense regret, and (B) moral behavior has greater expected long-term personal benefits, in turn developing (C) standing negative desires not to behave immorally, and (D) standing positive desires to act morally as a general 'life policy'.

In other words, childhood moral-prudential learning appears to occur in three stages (Figure 1). In Stage 1, children risk violating moral norms but are (at least occasionally) punished or see other children punished for similar behavior. This tends to lead them to learn, in Stage 2, that it is imprudent to risk violating moral norms because their parents or other authority figures may punish them and instead reward moral behavior in the longer run. Finally, in Stage 3, the child internalizes this reward-punishment structure into their own beliefs and preferences by believing they should behave morally, wanting to behave morally, punishing themselves (viz. guilt) if they are tempted to violate moral norms (and even more severely if they do violate moral norms), and rewarding themselves (viz. inner self-esteem) if they 'do the right thing'.

FIGURE 1.
Childhood Moral-Prudential Learning

Stage 1: socially-punished moral-norm violations

Impulsivity/present focus	Violation of moral norm	External Punishment	Regret
Temptation to steal candy. Desire to hit another child. Cheating on school test.	*First-personal*: Stealing candy. *Second/third-personal*: observing schoolyard fight or cheating.	Parental discipline. School suspension. Legal punishment.	'I wish I hadn't stolen the candy.' 'It was stupid for Jones to get in a fight. They got in trouble'.

Stage 2: externally-incentivized 'moral risk-aversion'

Temptation/present focus	Memory of punishment & regret	Minimax regret / moral-risk aversion	Delayed social reward
Temptation to steal candy. Desire to hit another child. Cheating on school test.	*First-personal*: 'I don't want to get in trouble for that again'. *Second/third-personal*: 'Jones got suspended from school for that'.	'I shouldn't risk stealing, fighting, or cheating. I might regret it. It's not worth the risk'.	Parent rewards good behavior. Better friendships at school. Good academic performance.

Stage 3: internalized 'moral risk-aversion'

Internalized moral-risk-aversion	Temptation to immorality	Internalized moral-risk aversion	Delayed internal reward
'I shouldn't get into fights, steal candy, or cheat on tests. I might regret it. And I see that people tend to benefit more later when they follow the rules'.	Impulse to steal candy or hit another child Etc.	'I don't want to risk it. If I do, I might feel guilty and regret it'.	'I'm proud of myself, and my parents and teachers would be proud that I followed the rules'.

4.3 The Prudence of Consolidating 'Moral Risk-Aversion' in Adolescence

This schema is not only typically repeated throughout childhood. It ordinarily continues well into adolescence and adulthood. Sometimes adolescents simply extend the moral-prudential lessons they learned as children to new situations—for instance, avoiding cheating on exams in high school and beyond because of the standing negative attitudes (against risking violating norms) and positive attitudes (in favor of obeying moral norms) they began internalizing as children. Other times, however, adolescents can fail to generalize the moral-prudential lessons from childhood to new cases. Consider the kinds of risky violations of moral norms that adolescents may engage in: cheating on school exams, cutting class to do drugs with friends, sneaking out of the home to go to a party or concert, and so on. Here again, as in childhood, the same pattern of moral-prudential learning typically appears to play out (Figure 2). The adolescent who fails to study for an exam may be tempted to cheat, figuring they are likely to succeed. Perhaps they even cheat successfully a few times. However, what often happens is that they are eventually— and, from their standpoint, unexpectedly—caught. The end-result is punishment, which can range in severity: they may end up failing the exam, get sent to the principal's office, receive detention or a school suspension, or face parental punishment at home (Stage 4). In more severe cases, such as drug use or petty crime, adolescents may be subject to serious legal ramifications. These are negative outcomes which, if the adolescent is not a criminal delinquent or a psychopath, they will typically find highly regrettable: far more regrettable than outcomes they experience when they behave morally (where, e.g., the worst they can do is perform poorly on an exam). Next, just as in childhood, these lessons are often also reinforced by second and third-hand experience: by the adolescent seeing what happens to others around them who violate moral norms (other children being expelled from school, arrested by police, and so on)—

presenting them, just as in childhood, with the following empirical regularities: adolescents like them tending to regret their choice to risk violating moral norms, and benefitting more over the long-term by behaving morally.

Consequently, the prudent adolescent (at least typically) comes to progressively believe that violating moral norms is not worth the risk of immense regret (viz. minimax regret), and that moral behavior has better expected aggregate value over the long run given life's many uncertainties (Stage 5). Further, if they are prudent, they consolidate these beliefs into standing desires: negative desires not to risk violating moral norms *in general*, and positive desires to obey moral norms for longer-term benefit. These attitudes in turn constitute and give rise to a progressive number of other attitudes and dispositions readily identifiable (in commonsense) as 'having a conscience': the adolescent increasingly believes they shouldn't violate moral norms, desires not to violate them, worries about what might happen if they do violate them; feels guilt and regret when they do; and so on.

FIGURE 2.
Adolescent Moral-Prudential Consolidation

Stage 4: negative consequences of failures to extend 'moral risk-aversion' to adolescent choices

Temptation to immorality
'I learned as a child I shouldn't get in fights on the schoolyard or steal candy. But I think I can cheat on my high-school exam and sneak out to a concert, get away with it, and not feel guilty'.

Immoral behavior (first-personal)
Cheating on exam.
Sneaking out of house.

Observed immorality (third-personal)
Hearing friend cheated on exam.
Hearing friend snuck out of home.

External and/or Internal Punishment
Parental, school, legal punishment
'I got away with it, but I feel guilty for cheating and lying to everyone'.

Observed External / Internal Punishment
Others face parental, school, legal sanctions.
Seeing friends feel guilty for their actions.

Consolidation of 'moral risk-aversion'
'It is dumb to cheat on exams, sneak out of the house, or do drugs. I've either gotten punished myself, felt guilty for it, or seen others get punished or feel guilty. Violating social moral norms is not worth the risk: I might regret it far worse than behaving morally, and I have seen the lives of people my age go better on the whole when they behave morally. They have better friends, go to college, and so on'.

Stage 5: consolidated 'moral risk-aversion' (conscience)

Consolidated 'moral risk-aversion'
'I don't want to cheat, sneak out, or use drugs. I know that it's a bad risk I might regret, both because of potential social punishment but also because I have learned to feel guilty for bad behavior'.

Temptation to immorality
'Blanking out' at an exam, and being tempted to look at another student's exam.

Prospective guilt/regret
'I shouldn't risk cheating. I might regret it. I might end up expelled, unable to go to college, or just feel guilty and have to confess to get the guilt off my chest'.

Delayed internal and/or external rewards
'I might not get the best grade on the exam. But I am proud that I didn't cheat. I might have regretted it far worse than a bad grade. I can always study harder and do better next time'.

4.4 The Prudence of Consolidating Categorical 'Moral Risk-Aversion' in Adulthood
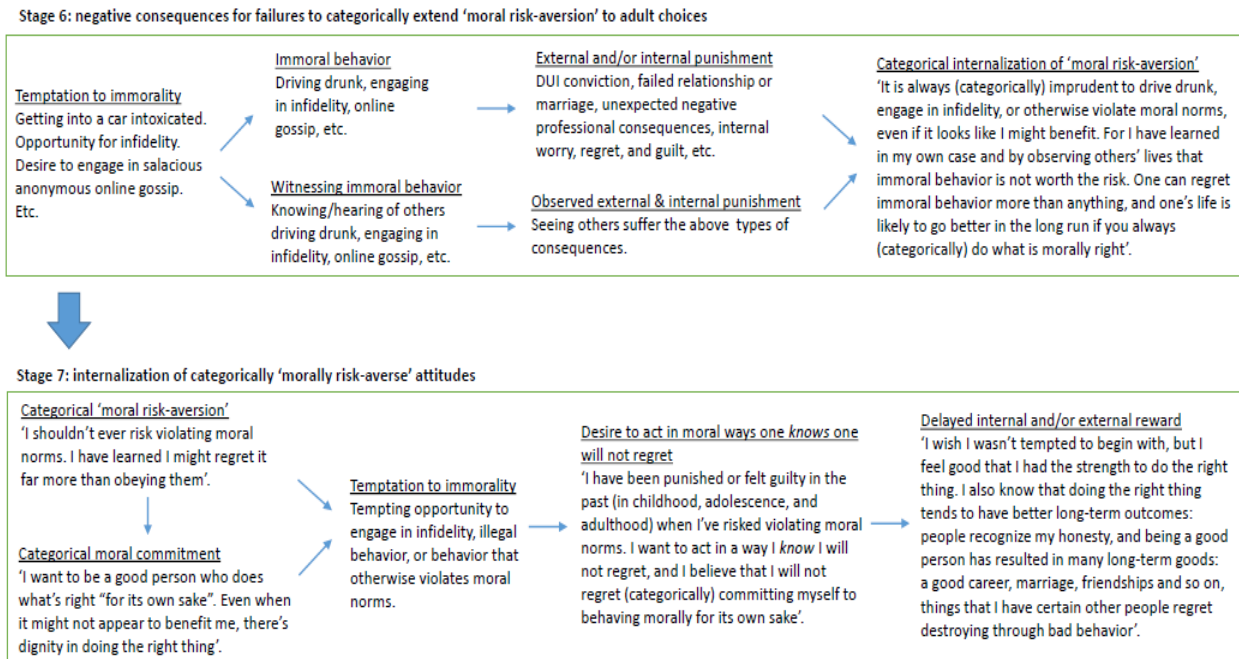
Finally, the same learning process often continues to play itself out—with increasingly high stakes—in adulthood. Once in adulthood, people either make moral and prudential mistakes of their own (often with very serious consequences), or observe others doing so. For example, some people fall prey to the temptation to drive under the influence of drugs or alcohol—often suffering extreme consequences, such as a serious car crash or jail time. Alternatively, they may witness dire consequences for others who engage in these behaviors—for instance, friends or associates dying in a vehicular accident or suffering jail time for driving under the influence. Similarly, some adults succumb to temptation of sexual infidelity, cheating on a romantic partner or spouse. Although romantic or sexual affairs can be tempting, one common result is that the affair is found out and the person's primary romantic relationship irreparably damaged, something that can lead to great, long-lasting regret ('I wish I had never cheated. I ruined everything with someone I love.'). Further, even if an adult does not make these mistakes themselves, chances are they will often see others suffer long-term from such decisions.

For example, consider recent sexual misconduct scandals. Although many alleged perpetrators got away with such behavior for many years, many are now facing serious consequences.[52] Second, consider pseudonymous philosophy blogs.[53] Many such forums flourished for a time, hosting content alleged by some to involve unethical forms of abuse or defamation.[54] In each case, the forum eventually closed down because the blog administrators were either unexpectedly found out or found themselves in danger of having their identity revealed.[55] These cases both fit with the theory of prudence proposed above. For although sexual harassers and

owners of pseudonymous blogs might have thought they could benefit from their actions, the

seemingly-unlikely happened: their actions and identities were either revealed or threatened to be

revealed, putting their future in jeopardy in ways they could (and perhaps did) end up deeply

regretting.

What prudent adults typically do, then, is observe these kinds of empirical regularities—the same

kinds of empirical regularities they observed in childhood and adolescents—but now with much

greater stakes. We often see just how profoundly, and in many cases unexpectedly, immoral

behavior can negatively impact perpetrators' lives, ruining their careers, marriages, public

reputation, well-being, and autonomy—up to and including legal ramifications. Conversely, we

also normally see how moral behavior appears beneficial in the long-term, enabling people to

cultivate stable careers, families, and long-term well-being by 'doing things the right way'.

Consequently, assuming they have already internalized moral risk-averse attitudes in childhood

and adolescence—becoming prudent people 'of conscience'—prudent adults typically

consolidate these morally risk-averse attitudes in the strongest way possible, internalizing them

as a set of *categorical* attitudes (Figure 3): (i) negative attitudes (beliefs and desires) that

immoral behaviors are never worth the risk of immense regret, and (ii) positive attitudes that they

should always behave morally regardless of what else they might desire. This is the final stage of

moral-prudential learning (Stage 7).

FIGURE 3.
Adult Categorical Moral-Prudential Consolidation

Stage 6: negative consequences for failures to categorically extend 'moral risk-aversion' to adult choices

Temptation to immorality
Getting into a car intoxicated. Opportunity for infidelity. Desire to engage in salacious anonymous online gossip. Etc.

Immoral behavior
Driving drunk, engaging in infidelity, online gossip, etc.

Witnessing immoral behavior
Knowing/hearing of others driving drunk, engaging in infidelity, online gossip, etc.

External and/or internal punishment
DUI conviction, failed relationship or marriage, unexpected negative professional consequences, internal worry, regret, and guilt, etc.

Observed external & internal punishment
Seeing others suffer the above types of consequences.

Categorical internalization of 'moral risk-aversion'
'It is always (categorically) imprudent to drive drunk, engage in infidelity, or otherwise violate moral norms, even if it looks like I might benefit. For I have learned in my own case and by observing others' lives that immoral behavior is not worth the risk. One can regret immoral behavior more than anything, and one's life is likely to go better in the long run if you always (categorically) do what is morally right'.

Stage 7: internalization of categorically 'morally risk-averse' attitudes

Categorical 'moral risk-aversion'
'I shouldn't ever risk violating moral norms. I have learned I might regret it far more than obeying them'.

Categorical moral commitment
'I want to be a good person who does what's right "for its own sake". Even when it might not appear to benefit me, there's dignity in doing the right thing'.

Temptation to immorality
Tempting opportunity to engage in infidelity, illegal behavior, or behavior that otherwise violates moral norms.

Desire to act in moral ways one _knows_ one will not regret
'I have been punished or felt guilty in the past (in childhood, adolescence, and adulthood) when I've risked violating moral norms. I want to act in a way I _know_ I will not regret, and I believe that I will not regret (categorically) committing myself to behaving morally for its own sake'.

Delayed internal and/or external reward
'I wish I wasn't tempted to begin with, but I feel good that I had the strength to do the right thing. I also know that doing the right thing tends to have better long-term outcomes: people recognize my honesty, and being a good person has resulted in many long-term goods: a good career, marriage, friendships and so on, things that I have certain other people regret destroying through bad behavior'.
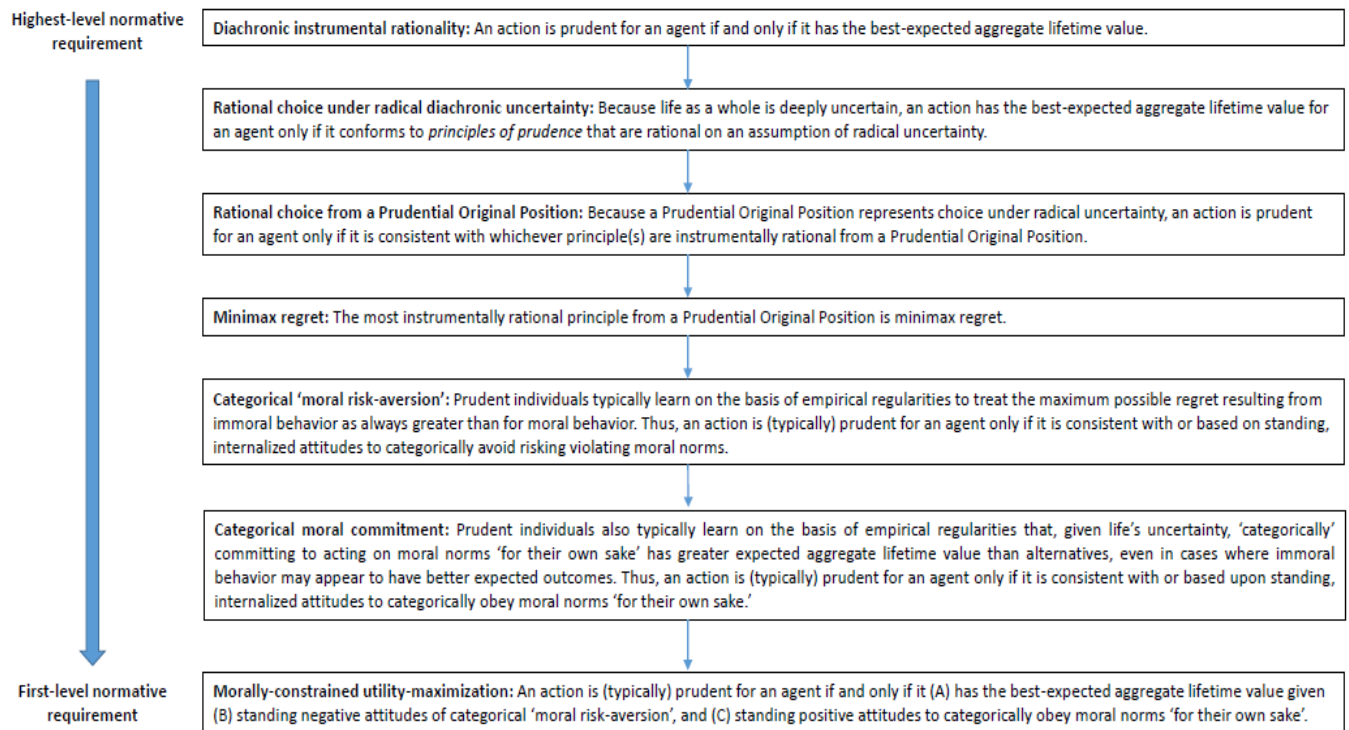
In sum, throughout childhood, adolescence, and adulthood, the very same prudential lessons that we observe in fictional tragedies, redemption-stories, and world religions appear to be both externally incentivized by empirical social regularities, and then—at least typically—progressively internalized by prudent individuals in their own attitudes. This process culminates in the internalization of 'categorical' attitudes that violating moral norms is never worth the risk of immense regret (viz. minimax regret), and morality always has better expected long-term value (viz. lifetime aggregate utility maximization), given life's immense uncertainties. Importantly, this account coheres with well-established facts of moral development: namely, that children first obey moral norms on largely instrumental grounds,[56] but over time progressively internalize a commitment to act on moral principles for their own sake[57]—though even then, elements of instrumental reasoning still remain prominent.[58]

23

5. Outline of a Normative Theory of Prudence

This chapter's theory of prudence is as follows (Figure 4). First, diachronic instrumental rationality is prudence's highest-level normative requirement: prudence is a matter of acting in ways that have the best-expected aggregate outcomes for an agent's life as a whole. Second, because life as a whole is radically uncertain, a prudent agent should treat an action as rational only if it conforms to principles of prudence that are instrumentally rational from a standpoint of radical uncertainty: a Prudential Original Position. Third, the principle of prudence most rational from that standpoint appears to be minimax regret. Fourth, due to observed empirical regularities, minimax regret in real life typically leads prudent agents to become deeply averse to risking violating moral norms (viz. 'moral risk-aversion'), and—because morally-risk averse individuals learn to care a great deal about the past and future—to regard this strategy as maximizing expected aggregate utility over the course of their life as a whole (viz. moral behavior). Consequently, what appears to be a form of prudential risk-aversion (unwillingness to risk immoral behavior) just is a form of long-term self-control (delaying immediate gratification viz. temptation to immorality, for the sake of greater long-term expected benefits of moral behavior). Thus, descending from the highest-level requirement of prudence—that prudence involves maximizing aggregate expected utility over life as a whole—prudence at a first-order level is (at least typically) a matter of 'morally-constrained aggregate-lifetime-utility maximization'. An action is prudent for an agent if and only if it has the best-expected aggregate lifetime personal utility (viz. rationally weighing and aggregating costs and benefits[59]) constrained by standing negative categorical attitudes (beliefs and desires) against ever violating moral norms and positive categorical attitudes in favor of 'behaving morally for its own sake'.

FIGURE 4
Outline of a Normative Theory of Prudence

Highest-level normative requirement

Diachronic instrumental rationality: An action is prudent for an agent if and only if it has the best-expected aggregate lifetime value.

Rational choice under radical diachronic uncertainty: Because life as a whole is deeply uncertain, an action has the best-expected aggregate lifetime value for an agent only if it conforms to *principles of prudence* that are rational on an assumption of radical uncertainty.

Rational choice from a Prudential Original Position: Because a Prudential Original Position represents choice under radical uncertainty, an action is prudent for an agent only if it is consistent with whichever principle(s) are instrumentally rational from a Prudential Original Position.

Minimax regret: The most instrumentally rational principle from a Prudential Original Position is minimax regret.

Categorical 'moral risk-aversion': Prudent individuals typically learn on the basis of empirical regularities to treat the maximum possible regret resulting from immoral behavior as always greater than for moral behavior. Thus, an action is (typically) prudent for an agent only if it is consistent with or based on standing, internalized attitudes to categorically avoid risking violating moral norms.

Categorical moral commitment: Prudent individuals also typically learn on the basis of empirical regularities that, given life's uncertainty, 'categorically' committing to acting on moral norms 'for their own sake' has greater expected aggregate lifetime value than alternatives, even in cases where immoral behavior may appear to have better expected outcomes. Thus, an action is (typically) prudent for an agent only if it is consistent with or based upon standing, internalized attitudes to categorically obey moral norms 'for their own sake.'

First-level normative requirement

Morally-constrained utility-maximization: An action is (typically) prudent for an agent if and only if it (A) has the best-expected aggregate lifetime value given (B) standing negative attitudes of categorical 'moral risk-aversion', and (C) standing positive attitudes to categorically obey moral norms 'for their own sake'.

This book does not aim to prove this theory. As we will see in Chapter 5, there may be plausible counterexamples (which, hopefully, are outliers). The point for now is simply that there is much to be said for this account when considering moral-prudential development, the empirical regularities that underlie it, and the ubiquity of moral-prudential lessons in tragic fiction, redemption-stories, and other cultural artifacts across cultures and different eras.

Now, a theory may be reflected in common social practices and yet still be false. So let us look a bit deeper. Why do prudent people appear to focus so much on minimizing maximum possible regret, at least when it comes to avoiding violating moral norms? Second, why do we tend to think that immorality is liable to generate greater possible regrets than moral behavior? Isn't it possible to regret moral behavior just as immensely as immoral behavior? Third, why do we tend to think morality is more beneficial over the long-term than immorality? I believe there is a more

general psychological story we can tell here to buttress the normative plausibility of this chapter's theory in response to these kinds of questions.

First, typical human adults display a pronounced negativity bias, such that when it comes to outcomes of equal positive and negative intensity, negative outcomes weigh far more heavily in determining our thoughts and actions.[60] Human beings also display proclivities to loss-aversion, with some studies indicating that perceived losses are twice as powerful, psychologically, as prospective gains.[61] In concrete terms, people disprefer a loss of $100 about twice as much as they prefer a gain of $100. People also tend to display status quo bias, preferring things to remain as they are rather than change[62], as well as dispositions to keep things they already have (the endowment effect).[63] Finally, as noted in Chapter 1, risk-aversion is known to be implicated in prudential cognition.[64] People tend to weigh improbable events more heavily than moderate or high probability ones, be highly averse to even a small chance of severe loss, and desire 'sure things' over moderate or even high-probability gains.[65] People also appear derive more utility from guaranteed good outcomes than they do from risky good outcomes of equal expected value.[66]

Are there any good normative grounds for these psychological phenomena? One plausible answer is that they are rational responses to immense lifetime uncertainty: if one does not know how the future is likely to go, keeping what you already have is a sure thing—it protects your future from unexpected jeopardy. The problem with immoral actions, or so we learn from experience, is that although they can produce immense gains (e.g., riches, power), they are risks that can produce immense losses, both externally *and* internally. Consider social ostracism (e.g.

job and reputation loss) for violating moral norms, or going to jail for violating moral and legal norms. Both can systematically undermine a person's ability to pursue their goals for many years to come. Further, for a person who has internalized minimax regret (given human dispositions to negativity bias, etc.), the internal worry or guilt they may experience from violating moral norms may be long-lasting. For example, an individual who cheats on their spouse may not only end up divorced and alone, but also regretting their actions 'for the rest of their life'. Conversely, moral actions rarely appear to have these kinds of intensely negative consequences. First, the maximum negative consequences of moral behavior tend to be not losses from the status quo, but rather foregone gains: wealth or power with which one may have no prior experience. Second, for a person who has internalized the desire to behave morally based on the belief that morality has greater additive value over their life, even when things go externally badly (e.g., one could have become rich by stealing) one can at least take solace in having done the 'right thing'. Moral actions, in other words, tend to be *regarded* by prudent people as far less potentially regrettable than immoral actions. And to the person who has internalized this kind of moral risk-aversion— the person who wants to behave morally because they believe it has better long-term outcomes and is less-regrettable—moral actions contain very little risk, and the prospect of at least some sure reward: a sense of internal pride for having 'done the right thing'.

Consequently, this chapter's theory of prudence appears to have good normative foundations. 'Moral risk-aversion'—aversion to risking violating moral norms, and instead categorically preferring to behave morally for its own sake—appears to be an effective means to preserving one's social and economic status, personal autonomy, and internal well-being. Conversely, willingness to risk immoral behavior places all of these features of our lives in serious jeopardy,
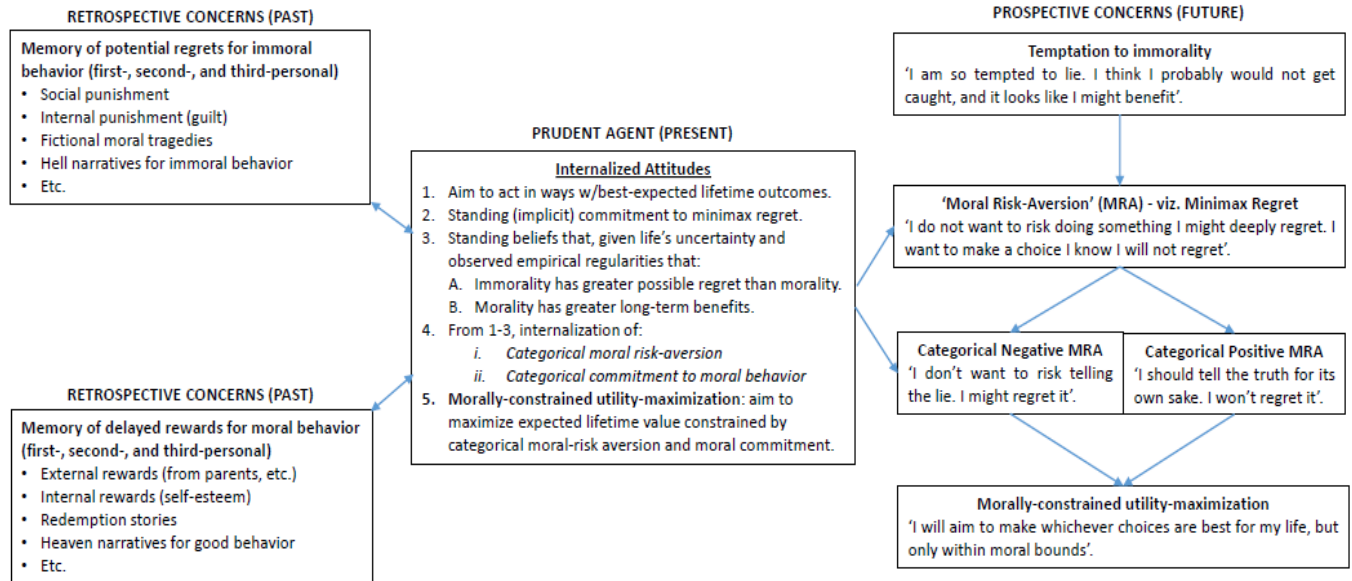
as illustrated by examples of criminals, psychopaths, and ordinary people (including, often enough, ourselves) who routinely suffer serious consequences for immoral behavior.

None of this is to say that everyone internalizes 'moral risk-aversion', or that it is always prudent. Some people become criminals; others become habitual 'cheaters' in romantic relationships; and so on—and some people appear to fare well in life through immoral means. It remains an open question whether such behaviors can be prudent. We will return to these potential counterexamples in Chapter 5.

6. Outline of Descriptive Psychological Theory of Prudence

Finally, let us reflect on the psychology of a person who has internalized this chapter's account of prudential 'moral risk-aversion'. How will such a person approach first-order life-decisions? Consider first negative categorical attitudes against ever risking violating moral norms on minimax regret grounds. A person who has internalized this standing attitude will approach morally salient cases by wanting to avoid even the smallest likelihood of acting in an immoral way they might deeply regret. Second, if they have developed the standing positive categorical attitudes that I have argued are (typically) prudent, they will have positive attitudes of the following sort: 'Whatever happens, if I do the right thing for its own sake, I will *regret it less* than if I do the wrong thing.' In sum, Figure 5 represents the psychology of a person who conforms to this chapter's theory of prudence.

FIGURE 5.
A Descriptive Model of Prudential Psychology

RETROSPECTIVE CONCERNS (PAST)

Memory of potential regrets for immoral behavior (first-, second-, and third-personal)
- Social punishment
- Internal punishment (guilt)
- Fictional moral tragedies
- Hell narratives for immoral behavior
- Etc.

RETROSPECTIVE CONCERNS (PAST)

Memory of delayed rewards for moral behavior (first-, second-, and third-personal)
- External rewards (from parents, etc.)
- Internal rewards (self-esteem)
- Redemption stories
- Heaven narratives for good behavior
- Etc.

PRUDENT AGENT (PRESENT)

Internalized Attitudes
1. Aim to act in ways w/best-expected lifetime outcomes.
2. Standing (implicit) commitment to minimax regret.
3. Standing beliefs that, given life's uncertainty and observed empirical regularities that:
   A. Immorality has greater possible regret than morality.
   B. Morality has greater long-term benefits.
4. From 1-3, internalization of:
   i. Categorical moral risk-aversion
   ii. Categorical commitment to moral behavior
5. Morally-constrained utility-maximization: aim to maximize expected lifetime value constrained by categorical moral-risk aversion and moral commitment.

PROSPECTIVE CONCERNS (FUTURE)

Temptation to immorality
'I am so tempted to lie. I think I probably would not get caught, and it looks like I might benefit'.

'Moral Risk-Aversion' (MRA) - viz. Minimax Regret
'I do not want to risk doing something I might deeply regret. I want to make a choice I know I will not regret'.

Categorical Negative MRA
'I don't want to risk telling the lie. I might regret it'.

Categorical Positive MRA
'I should tell the truth for its own sake. I won't regret it'.

Morally-constrained utility-maximization
'I will aim to make whichever choices are best for my life, but only within moral bounds'.

Is this the moral-prudential psychology that prudent people exhibit in everyday life? It indeed appears to be. For consider again an opportunity to cheat on an exam or a romantic partner. Everyone knows it is possible to get away with these kinds of actions. In some cases, it may even seem likely. However, prudent people typically become unwilling to take these risks. Why? *We worry we might regret it*. We also tend to believe that moral behavior is advantageous in the long-run (viz. sayings like 'honesty is the best policy'). We also tend to *want* to behave morally ('I want to do the right thing'), wanting to know 'the right thing to do' before making a choice so that we can derive inner-satisfaction from doing the right thing 'for its own sake'. Finally, we even tend to counsel others in these terms when they struggle with temptation, impressing the rationality of moral risk-aversion upon them in minimax regret terms (e.g. 'Don't do go through with cheating on your partner. You might regret it more than anything!'), the long-term benefits of moral behavior ('Cheaters never prosper'), and morality's internal benefits ('If you tell the truth, you will at least have a clear conscience'), and so on.

7. Three Clarifications

Three points are worth clarifying. First, this theory does not hold that prudent agents should directly use minimax regret to make first-order life decisions, such as what to eat, which career to choose, and so on. As we have seen, Bruckner argues that prudence requires minimax regret as a high-level principle, but that the more an agent cares about the past or future, the more an agent following minimax regret will *appear* to be acting on a simple principle of expected-personal-utility maximization.[67] This chapter's theory coheres with this result: it holds that normatively prudent agents typically internalize morally risk-averse attitudes on minimax regret grounds, but that these attitudes should lead prudent individuals to act as though they are engaging in 'morally-constrained utility-maximization', making everyday choices that appear to them to *maximize their own long-term well-being* within moral constraints. I believe this result accurately represents how prudent people typically make decisions. Prudent people like you and I typically seek to make choices that maximize our own long-term benefit constrained by our 'moral conscience'.

Second, the theory does not hold that most people actually tend to care about their past or future when making decisions. As we have seen, empirical science has found that people are often present-focused, treating their past and future selves akin to strangers.[68] The theory defended here merely holds that *normatively prudent* people care about their past and future in the broad sense of aiming to live happy lives as a whole, as well as in specifically morally-salient cases, recalling and regretting past punishment for immoral behavior, and not wanting to risk potential future regret when temptations to immorality arise. I argued this is also intuitive, as prudent people appear to respond to moral temptations this way: namely, by recalling past regrets and

worrying about potential future regret. Further, empirical science suggests these are indeed

central features distinguishing prudent, moral people from criminals and psychopaths.

Psychopaths lack retrospective regret for their past immoral behavior[69], and criminals and

psychopaths appear to engage in imprudent and immoral behavior because of not worrying about

the future[70] or being able to use experiences of prospective regret to inhibit immoral behavior.[71]

Finally—and this is important—this chapter has not argued that 'moral risk-aversion' always has

greater expected aggregate long-term personal utility than immoral behavior. As Chapter 5

discusses, there may be cases where immoral actions have better-expected long-term outcomes

than moral behavior. My argument in this chapter has merely been that across childhood,

adolescence, and adulthood, prudent individuals typically learn to internalize attitudes that *treat*

moral actions as always having greater expected long-term personal outcomes than immoral

ones. I argued that this is because prudent individuals appreciate how life as a whole is

profoundly uncertain, while internalizing two broad empirical regularities: (1) people often

paying dearly for and regret immoral behavior, and (2) moral behavior tending to pay off in the

long-run (viz. stable and productive lives, careers, relationships, and so on). Consequently, my

theory has not been that moral actions always in fact have greater long-term expected value than

immoral ones. Rather, it is that prudent agents typically internalize attitudes that *treat* moral

actions as having greater long-term expected value—as a kind of 'safe' life policy for grappling

with the otherwise-uncertain nature of life as a whole. We have seen that this is amply attested to

in moral-prudential development, as well as in common adages such as 'Honesty is the best

policy'. Although most of us recognize (at least implicitly) that it may be possible for immoral

actions like dishonesty to pay off, prudent people typically appear to internalize the opposite

attitude: that moral actions such as honesty are better 'life-policies' likely to have better long-term outcomes than immoral behavior.

Conclusion

Chapter 3 will now argue that this theory of prudence and prudential cognition entails a novel normative theory of morality and empirical theory of moral cognition. Chapter 4 will then argue that this unified theory of prudence and morality is the best normative and descriptive explanation currently available of the behavioral neuroscience summarized in Chapter 1.

References

Anand, P. (1995). *Foundations of Rational Choice Under Risk*. Oxford: Oxford University Press.

Argyle, M. (1986). *The Psychology of Happiness*. London: Methuen.

Aristotle [1984]. *The Complete Works of Aristotle: The Revised Oxford Translation*. J. Barnes (Ed.). Princeton: Princeton University Press.

Arvan, M. (2016). *Rightness as Fairness: A Moral and Political Theory*. New York: Palgrave MacMillan.

Aurelius, M. [1862]. *Meditations*. G. Long (trans.). Logos Publishing, 2018.

Ayton, P., Pott, A., & Elwakili, N. (2007). Affective Forecasting: Why Can't People Predict Their Emotions?. *Thinking & Reasoning*, *13*(1), 62-80.

Baltzly, D (2019). Stoicism. *The Stanford Encyclopedia of Philosophy,* E.N. Zalta (ed.), https://plato.stanford.edu/archives/spr2019/entries/stoicism/.

Baron, M. (1995). *Kantian Ethics Almost Without Apology*. Ithaca: Cornell University Press.

Baskin-Sommers, A., Stuppy-Sullivan, A.M., & Buckholtz, J.W. (2016). Psychopathic

    Individuals Exhibit but do not Avoid Regret During Counterfactual Decision

    Making. *Proceedings of the National Academy of Sciences*, *113*(50), 14438-43.

Bazerman, M.H., Loewenstein, G.F., & White, S.B. (1992). Reversals of Preference in

    Allocation Decisions: Judging Alternatives Versus Judging Among Alternatives.

    Administrative Science Quarterly, *37*, 220-40.

Baumann, P. (2018). What Will Be Best for Me? Big Decisions and the Problem of Inter-World

    Comparisons. *Dialectica*, *72*(2), 253-73.

Baumeister, R.F., Bratslavsky, E., Finkenauer, C., & Vohs, K.D. (2001). Bad is Stronger than

    Good. *Review of General Psychology*, *5*(4), 323-70.

Blow (2001). https://www.imdb.com/title/tt0221027/, retrieved 29 July 2019.

Bricker, P. (1980). Prudence. *The Journal of Philosophy*, *77*(7), 381-401.

Brickman, P., Coates, D., & Janoff-Bulman, R. (1978). Lottery Winners and Accident Victims:

    Is Happiness Relative?, *Journal of Personality and Social Psychology*, *36*(8), 917-27.

Broome, J. (2007). Does Rationality Consist in Responding Correctly to Reasons? *Journal of*

    Moral Philosophy, *4*(3), 349-74.

Briggs, R. (2015). Transformative Experience and Interpersonal Utility Comparisons. *Res*

    Philosophica, *92*, 189-216.

Bruckner, D.W. (2013) Present Desire Satisfaction and Past Well-Being. *Australasian Journal of*

    Philosophy, *91*(1), 15-29.

----- (2003). A Contractarian Account of (Part of) Prudence. *American Philosophical*

    Quarterly, *40*(1), 33-46.

Bykvist, K. (2006). Prudence for Changing Selves. *Utilitas*, *18*(3), 264-83.

Carpendale, J. I. (2000). Kohlberg and Piaget on Stages and Moral Reasoning. *Developmental Review*, *20*(2), 181-205.

Crisp, R (2017.). Well-Being. *The Stanford Encyclopedia of Philosophy.* E.N. Zalta (ed.), https://plato.stanford.edu/archives/fall2017/entries/well-being/.

Dorsey, D. (2018). Prudence and past selves. *Philosophical Studies*, *175*(8), 1901-25.

----- (2017). Future-Bias: A Defense. *Pacific Philosophical Quarterly*, *98*(S1), 351-73.

----- (2012). Subjectivism without Desire. *Philosophical Review*, *121*(3), 407-42.

Dougherty, T. (2015). Future-Bias and Practical Reason. *Philosophers' Imprint, 15*(30), 1-16.

Epictetus [1865]. *The Complete Works of Epictetus*. E. Carter (Ed.). Cambridge: Little, Brown, and Company.

Gilbert, D.T. & Wilson, T.D. (2009). Why the Brain Talks to Itself: Sources of Error in Emotional Prediction. *Philosophical Transactions of the Royal Society B*, *364*, 1335-41.

Grether, D.M. & Plott, C.R. (1979). Economic Theory of Choice and the Preference Reversal Phenomenon. *The American Economic Review*, *69*(4), 623-38.

Hansson, S.O. (2005). *Decision Theory: A Brief Introduction*, https://people.kth.se/~soh/decisiontheory.pdf, retrieved 25 June 2018.

Haybron, D. (2011). Happiness. *The Stanford Encyclopedia of Philosophy.* E.N. Zalta (ed.), https://plato.stanford.edu/archives/fall2011/entries/happiness/.

Herpertz, S.C., & Sass, H. (2000). Emotional Deficiency and Psychopathy. *Behavioral Sciences & the Law*, *18*(5), 567-80.

Isaacs, Y. (2019). The Problems of Transformative Experience. *Philosophical Studies* https://doi.org/10.1007/s11098-018-01235-3, 1-21.

Ito, T.A., Larsen, J.T., Smith, N.K., & Cacioppo, J.T. (1998). Negative Information Weighs More Heavily on the Brain: The Negativity Bias in Evaluative Categorizations. *Journal of Personality and Social Psychology*, *75*(4), 887-900.

Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus, and Giroux.

----- & Tverksy, A. (1984). Choices, Values, and Frames. *American Psychologist*, *39*(4), 341–50.

----- & Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, *4*, 263–91.

Kant, I. [1797]. *The Metaphysics of Morals*. In M.J. Gregor (ed.), *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*. Cambridge: Cambridge University Press, 1996, 353-604.

----- [1788]. *Critique of Practical Reason*, in Ibid., 133-271.

----- [1785]. *Groundwork of the Metaphysics of Morals*, in Ibid., 38-108.

Kill Bill: Vol. 1 (2003). https://www.imdb.com/title/tt0266697/, retrieved 29 July 2019.

Kochanska, G., Gross, J.N., Lin, M.H., & Nichols, K.E. (2002). Guilt in Young Children: Development, Determinants, and Relations with a Broader System of Standards. Child Development, *73*, 461–82.

Kohlberg, L. (1973). The Claim to Moral Adequacy of a Highest Stage of Moral Judgment. *The Journal of Philosophy*, *70*(18), 630-46.

Korsgaard, C.M. (1996). *The Sources of Normativity*. New York: Cambridge University Press.

Lang, G. (2012). What's the Matter? Review of Derek Parfit, On What Matters. *Utilitas*, *24*(2), 300-12.

Langone, A. (2018). #MeToo and Time's Up Founders Explain the Difference Between the 2 Movements—and How They Are Alike. *Time,* http://time.com/5189945/whats-the-difference-between-the-metoo-and-times-up-movements/ , retrieved 25 June 2018.

Leiter, B. (2018). The Disappearing Anonymous and Pseudonymous Blogs. *Leiter Reports*. http://leiterreports.typepad.com/blog/2018/05/the-disappearing-anonymous-and-pseudonymous-blogs.html, retrieved 25 June 2018.

----- (2015). So What Happened to the 'Philosophy MetaMeta Blog', the blog that…?'. *Leiter Reports*, http://leiterreports.typepad.com/blog/2015/01/so-what-happened-to-the-philosophy-metameta-blog-the-unmoderate-forum-that.html, retrieved 25 June 2018.

Lewicka, M., Czapinski, J., & Peeters, G. (1992). Positive-Negative Asymmetry or 'When the Heart Needs a reason'. *European Journal of Social Psychology*, *22*(5), 425-34.

Lichtenstein, S. & Slovic, P. (1971). Reversals of Preference Between Bids and Choices in Gambling Decisions. *Journal of Experimental Psychology*, *89*, 46-55.

Lin, E. (2016). The Subjective List Theory of Well-Being. *Australasian Journal of Philosophy, 94*(1), 99-114.

Lord, E. (2018). *The Importance of Being Rational*. Oxford: Oxford University Press.

McNeil, B.J., Pauker, S.G., & Tversky, A. (1988). On the Framing of Medical Decisions. In D.E. Bell, H. Raiffa, & A. Tversky (Eds), *Decision Making: Descriptive, Normative, and Prescriptive Interactions,* Cambridge: Cambridge University Press, 562-68.

Menegatti, M. (2014). New Results on the Relationship Among Risk Aversion, Prudence and Temperance. *European Journal of Operational Research*, *232*(3), 613-17.

Moore, A. (2018). Hedonism. *The Stanford Encyclopedia of Philosophy*. E.N. Zalta (ed.), https://plato.stanford.edu/archives/win2018/entries/hedonism/.

Morewedge, C.K. & Giblin, C.E. (2015). Explanations of the Endowment Effect: An Integrative Review. *Trends in Cognitive Sciences*, *19*(6), 339-48.

Morton, J.M. (2013). Deliberating for Our Far Future Selves. *Ethical Theory and Moral Practice*, *16*(4), 809-28.

Murphy, M.C. (1999). The Simple Desire-Fulfillment Theory. *Noûs*, *33*(2), 247-72.

Myers, D.G. & Diener, E. (1995). Who is Happy?, *Psychological Science*, *6*(1), 10-19.

Olson, E.T. (2017). Personal Identity. *The Stanford Encyclopedia of Philosophy.* E.N. Zalta (ed.), https://plato.stanford.edu/archives/sum2017/entries/identity-personal/.

Parfit, D. (2011). *On What Matters*, *Vols. 1&2*. Oxford: Oxford University Press.

----- (2001). Rationality and Reasons. In D. Egonsson, J. Josefsson, B. Petersson, T. Ronnow-Rasmussen, & I. Persson (eds.), *Exploring Practical Philosophy: From Actions to Values*, Farmham: Ashgate, 17-39.

Paul, L.A. (2015a). *Transformative Experience*. Oxford: Oxford University Press.

----- (2015b). What You Can't Expect When You're Expecting. *Res Philosophica, 92*(2), 1-23.

Peterson, M. (2017). *An Introduction to Decision Theory*, 2nd edition. Cambridge: Cambridge University Press.

Pettigrew, R. (forthcoming). *Choosing for Changing Selves.* Oxford University Press. https://philpapers.org/archive/PETCFC.pdf, retrieved 4 February 2019.

----- (2015). Transformative Experience and Decision Theory. *Philosophy and Phenomenological Research*, *91*(3), 766-74.

Price, B.W. (2002). The Worthwhileness Theory of the Prudentially Rational Life. *Journal of Philosophical Research*, *27*, 619-39.

Pronin, E., Olivola, C.Y., & Kennedy, K.A. (2008). Doing Unto Future Selves as You Would Do
  Unto Others: Psychological Distance and Decision Making. *Personality and Social
  Psychology Bulletin*, *34*(2), 224-36.

Pronin, E. & Ross, L. (2006). Temporal Differences in Trait Self-Ascription: When the Self is
  Seen as an Other. *Journal of Personality and Social Psychology*, *90*(2), 197-209.

Rawls, J. (1999a). *A Theory of Justice: Revised Edition*. Cambridge, MA: The Belknap Press of
  Harvard University Press.

----- (1999b). *The Law of Peoples, with 'The Idea of Public Reason Revisited'*. Cambridge, MA:
  Harvard University Press.

Rozin, P. & Royzman, E.B. (2001). Negativity Bias, Negativity Dominance, and
  Contagion. *Personality and Social Psychology Review*, *5*(4), 296-320.

Samuelson, W. & Zeckhauser, R. (1988). Status Quo Bias in Decision Making. *Journal of Risk
  and Uncertainty*, *1*(1), 7-59.

Saved by the Bell (1989a). The Lisa Card. https://www.imdb.com/title/tt0695233/, retrieved 5
  June 2019.

----- (1989b). The Gift. https://www.imdb.com/title/tt0695229/, retrieved 5 June 2019.

----- (1989c). Fatal Distraction. https://www.imdb.com/title/tt0695193/, retrieved 5 June 2019.

Scanlon, T. M. (2014). *Being Realistic About Reasons*. Oxford: Oxford University Press.

Simmba (2018). https://www.imdb.com/title/tt7212726/, retrieved 5 June 2019.

Smith, M (2017). Parfit's Mistaken Meta-ethics. In P. Singer (ed.), *Does Anything Really
  Matter?: Essays on Parfit and Objectivity.* Oxford: Oxford University Press, 99-120.

----- (2013). A Constitutivist Theory of Reasons: Its Promise and Parts. *LEAP: Law, Ethics, and
  Philosophy*, *1*, 9-30.

Star Wars (1977). Episode IV: A New Hope. https://www.imdb.com/title/tt0076759/, retrieved 5 June 2019.

Steele, K. & Stefánsson, H.O. (2016). Decision Theory. *The Stanford Encyclopedia of Philosophy.* E.N. Zalta (ed.), https://plato.stanford.edu/archives/win2016/entries/decision-theory/.

The Big Boss (1971). https://www.imdb.com/title/tt0067824/, retrieved 29 July 2019.

The Godfather Part II (1974). https://www.imdb.com/title/tt0071562/, retrieved 29 July 2019.

The New American Bible [2011]. Revised Edition. New Jersey: World Catholic Press.

Thor (2011). https://www.imdb.com/title/tt0800369/, retrieved 5 June 2019.

Tom, S.M. *et al.* (2007). The Neural Basis of Loss Aversion in Decision-making Under Risk. *Science*, *315*(5811), 515–18.

Tucker, K. (1991). Saved by the Bell. *Entertainment Weekly*, https://ew.com/article/1991/02/08/saved-bell/, retrieved 24 May 2019.

Tversky, A. & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, *5*, 297-323.

Ullmann-Margalit, E. (2006). Big Decisions: Opting, Converting, Drifting. *Royal Institute of Philosophy Supplements, 58*, 157-72.

Van Gelder, J.L., Hershfield, H.E., & Nordgren, L.F. (2013). Vividness of the Future Self Predicts Delinquency. *Psychological Science*, *24*(6), 974-80.

Wilson, T.D. & Gilbert, D.T. (2003). Affective Forecasting. *Advances in Experimental Social Psychology*, *35*, 345-411.

Wood, A. (2008). *Kantian Ethics*. Cambridge: Cambridge University Press.

Notes

---

[1] Kant [1785, 1788, 1797]. Cf. Baron (2005), Korsgaard (1996), Wood (2008).

[2] Lord (2018), Parfit (2001, 2011), Scanlon (2014). Cf. Broome (2007), Lang (2012), Smith (2017).

[3] Paul (2015a,b), Arvan (2016): ch 1, §3.

[4] Anand (1995), Hansson (2005), Peterson (2017), Steele & Stefánsson (2016).

[5] Aristotle [1984]. *Nicomachean Ethics*: Book VI, esp. 1140a, 1141b, 1142b.

[6] Dorsey (2012), Lin (2016).

[7] Moore (2018).

[8] Murphy (1999).

[9] See Crisp (2017) for an overview.

[10] Olson (2017).

[11] Arvan (2016): 45-6, Smith (2013).

[12] Pronin *et al.* (2008).

[13] Pronin & Ross (2006): 204-5.

[14] See also Morton (2013).

[15] Aristotle [1984]: *Nichomachean Ethics,* Book I, 1098^{a}15-20.

[16] Haybron (2011): §1, Bruckner (2003): 34-5, Bricker (1980), Kant [1797]: 223, Price (2002), Pettigrew (forthcoming).

[17] Bricker (1980), Bruckner (2003), Pettigrew (forthcoming).

[18] Bruckner (2013), Dougherty (2015), Dorsey (2017, 2018).

[19] Bruckner (2003): 34-5, Bricker (1980), Pettigrew (forthcoming), Price (2002).

[20] Baumann (2018) argues this is broader than Paul's (2015a,b) problem of transformative experience, which I explore shortly.

[21] Ullman-Margalit (2006), Briggs (2015). Cf. Paul (2015a,b).

[22] Baumann (2018): 264, Ullman-Margalit (2006).

[23] Bykvist (2006).

[24] Kahneman (2011): ch. 38, Gilbert & Wilson (2009), Wilson & Gilbert (2003), Ayton *et al.* (2007).

[25] Brickman *et al.* (1978), Myers & Diener (1995): 13, Argyle (1986).

[26] Kahneman (2011), Lichtenstein & Slovic (1971), Grether & Plott (1979), Bazerman, *et al.* (1992).

[27] Paul (2015a,b).

[28] Bykvist (2006), Paul (2015a), Pettigrew (2015, unpublished manuscript).

[29] *The New American Bible* [2011]: Ecclesiastes, Chapter 9.

[30] Ibid.: Wisdom of Ben Sira, especially Chapter 40. Also see Chapter 5:11-15, Chapters 8-9, Chapter 11, Chapter 14, esp. 12-14, Chapter 18 esp. 22-26. Chapter 20.

[31] Aurelius [1862]: especially Book IV, §8, Baltzly (2019): §5, Epictetus [1865]: *Enchiridion*.

[32] Bruckner (2003).

[33] Ibid.: §3. Cf. Rawls (1999a).

[34] Rawls (1999a): §4.

[35] Rawls (1999b): Part 1, §3.

[36] Bruckner (2003): §§4-7.

[37] Peterson (2017): §3.5.

[38] Bruckner (2003): §8.

[39] Bruckner (2003): 45.

[40] Pettigrew (forthcoming), though see Isaacs (2019) for a critique.

[41] In this and following sections where I argue prudence requires internalizing attitudes to 'moral risk-aversion', I mean merely that agents learn to avoid risking violating norms that *plausibly* (either from our perspective or theirs) appear to be 'moral', not that those norms *are* in fact moral in a normatively binding sense (viz. 'moral truths'). Instead, my argument will be that in internalizing attitudes against violating social norms *plausibly* treated as 'moral' (Chapter 2), the prudent agent internalizes attitudes that can be used to derive what is *in fact* moral in a normatively binding sense (Chapter 3).

[42] See e.g., *Blow* (2001), *The Godfather Part II* (1974).

[43] *Simmba* (2018).

[44] See e.g., *The Big Boss* (1971), *Kill Bill: Vol. 1* (2003).

[45] *Star Wars'* (1977) character Han Solo is a paradigm case, along with many others.

[46] *Thor* (2011).

[47] Tucker (1991).

[48] *Saved by the Bell* (1989a).

[49] *Saved by the Bell* (1989b).

[50] *Saved by the Bell* (1989c).

[51] Baskin-Sommers *et al.* (2016).

[52] Langone (2018).

[53] Leiter (2015, 2018).

[54] Ibid.

[55] Leiter (2018).

[56] Kochanska *et al.* (2002).

[57] Kohlberg (1973).

[58] Carpendale (2000).

[59] Pettigrew (forthcoming).

[60] Lewicka *et al.* (1992), Rozin & Royzman (2001), Ito *et al.* (1998), Baumeister *et al.* (2001). See also Kahneman & Tversky (1979).

[61] Tom *et al.* (2007), Tversky & Kahneman (1992).

[62] Samuelson & Zeckhauser (1988)

[63] Morewedge & Giblin (2015)

[64] Menegatti (2014).

[65] Kahneman & Tversky (1984).

[66] McNeil *et al.* (1988).

[67] Bruckner (2003).

[68] Pronin *et al.* (2008).

[69] Herpertz & Sass (2000).

[70] Van Gelder *et al.* (2013).

[71] Baskin-Sommers *et al.* (2016).