University of Bayreuth

Faculty of Humanities and Social Sciences

Department of Philosophy

Prof. Dr. Julian Fink

# Bachelor Thesis

## Avoiding Moral Wrongs under Normative Uncertainty

## Biases Involved in Maximizing Expected Choice-Worthiness

A thesis submitted for the degree of Bachelor of Arts

Submitted by:

Silvana Hultsch
Albert-Preu-Str. 10
95444 Bayreuth
silvana.hultsch@gmx.de
Philosophy & Economics
Matriculation Number: 1530770
Date: 11/03/2020

ABSTRACT

This thesis addresses morally conscientious agents, who want to avoid committing moral wrongs under normative uncertainty. Besides theoretical discussions about what the correct metanormative theory is, practical problems have not been addressed sufficiently. I assume Maximizing Expected Choice-Worthiness to be the correct decision-making approach under normative uncertainty for interval-scale measurable and intertheoretically comparable first-order moral theories, asking the question how we can improve decision making under the possible influence of cognitive biases on metanormative judgments. A few pathways for the influence of biases on metanormative decisions are identified, however, this thesis solely addresses the distribution of subjective probabilities (credence) across first-order normative theories.

After an argument for the relevance of cognitive biases in metanormative thinking this thesis follows a three-step approach towards improving decision-making, as a necessary means to avoiding moral wrongs. First, a preliminary normative model of what good thinking under normative uncertainty requires is developed, consisting of a few rules. In the descriptive part it will be shown that availability and status quo thinking lead to biases of which we have reason to believe also exist in metanormative judgments. In the prescriptive part, the debiasing process is sketched and possible starting points are suggested, concluding that research needs to provide descriptive insights and highlighting the relevance of cognitive biases for practical philosophy.

# Content

# 1. Introduction

When facing decision theoretic problems, it is not only empirical uncertainty that complicates decisions. Even with full knowledge about the state of the world in any situation, we need to choose normative criteria by which we evaluate the rightness or wrongness of consequences and actions. Which criteria that should be is not a trivial question. Ethics is extremely difficult. To work out the correct criteria or moral theory, it is necessary to correctly balance intuitions and different theoretical considerations like plausibility, thought experiments or the elegance of a theory. Doing this, even the most intelligent and well-informed humans ever lived, still disagree. This may lead one to accept moral uncertainty as a relevant consideration, especially when one pursues the goal of avoiding moral wrongs (MacAskill, Bykvist, & Ord, 2020).

This paper addresses morally conscientious agents (MCA). MCAs are uncertain about which normative criteria are correct and have the general goal of avoiding committing moral wrongs.[1]

It has been argued that the default approach for rational decision-making under moral uncertainty should be analogous to the standard "maximizing expected utility" under empirical uncertainty. An account that currently receives a large part of the academic attention in this field is called "maximizing expected choice-worthiness" (MEC), where the expected choice-worthiness of an option A represents the strength of reasons one has to choose A, taking into account all views one has non-zero credence in (MacAskill, 2014). MEC applies in the following scenario as described by William MacAskill and Toby Ord in their 2018 paper *Why Maximize Expected Choice-Worthiness?*: Decision-makers choose from a set of exhaustive and mutually exclusive options (A, B, C, …). Being morally uncertain, the decision maker has credence in each of a set of first-order normative theories ($T_1$, $T_2$, $T_3$, …). Credence describes the subjective probability that one believes a theory to be true and can range between 0% and 100%. "We will normally talk about these theories as if they are complete stand-alone normative theories, such as a particular form of utilitarianism. However, they could often just as well represent partially specified theories, or particular normative considerations regarding

[1] I assume that cognitivism is true for the purpose of this paper, even though there may be ways of non-cognitivists to make sense of the concept of moral uncertainty as discussed by MacAskill, Bykvist & Ord (2020). This discussion will be put aside to dive deeper into ideas that grow on the pro cognitivism ground.

the options at hand, such as whether killing is equivalent to letting die." (MacAskill & Ord, 2018, p. 2)

Theories under consideration assess options in terms of choice-worthiness, which represents the strength of reasons for choosing a given option. This is not necessarily quantifiable, however, MEC is only straightforwardly applicable "when all moral views under consideration give an interval-scale measurable and interthereoretically comparable notion of choice-worthiness". (MacAskill & Ord, 2018, p. 2)[2]

Expected choice worthiness of an option A is then calculated by multiplying credence one assigns to a theory ("How likely is it that this theory is true?") with the choice-worthiness that that specific theory assigns to the option at hand ("The theory says this is 10 choice-worthy"). Then one would do this for all theories one has non-zero credence in and add it up. The result is some number of how overall choice-worthy that option is for the agent according to his beliefs. Doing this for all options allows one to compare and choose the option with the highest choice-worthiness (A specific example is given in section 2.3). More specifically speaking, the expected choice-worthiness of an option A is calculated according to the following formula:

$EC(A) = \sum_{i=1}^{n} C(T_i)CW_i(A)$.[3]

This is how MEC works. I assume for the purpose of this paper that MEC is the correct approach to decision making under normative uncertainty for interval-scale measurable and intertheoretically comparable theories. This research field has been increasingly growing only for the last two decades, even though the idea of moral uncertainty is not new. People have traced such ideas back to Aristotle. It is currently pioneered by William MacAskill, Krister Bykvist and Toby Ord, who have published a large share of relevant work. They recently

[2] Looking at all theories one could possibly consider, the following problems remain: (i) How should we incorporate theories that are incomparable to other theories? This is important, as we need to find out for which theories more is at stake. (ii) How should we incorporate merely ordinal theories that do not define how much greater the difference of choice-worthiness between two options is on theory A compared to theory B?
For work on an overarching decision-making approach that accounts for the varying information that theories give, see MacAskill et al. (2020, Ch. 3 and 4). They propose an overarching account where MEC is supplemented by the following procedures. For comparable but merely ordinal theories, a voting procedure is borrowed from social choice theory, where theories are treated like individuals who vote for options (It is argued that the so called *Borda Rule* is the best aggregation rule.) For interval-scale measurable but incomparable theories the so called *Variance Voting* is suggested, as it is argued to best capture the notion of giving theories "equal say". A multi-step procedure is proposed, starting with the informationally richest theories (where MEC can be applied) and step by step taking less informationally rich theories into account with the respective procedure to yield the most desirable option in a decision situation. This multi-step procedure is not yet fully developed. MEC plays an important role in it, which is why the focus in this thesis lies on MEC.

[3] Formula from MacAskill (2014, p. 14)

structured all that is to know about moral uncertainty today, in their book *Moral Uncertainty* (2020). This thesis departs from their progress rather than reproducing it. Extensive arguments and critical discussions on why MCAs should be morally uncertain or why MEC is promising are found in their book.

So, MCAs adopt MEC. What else do they need to consider in order to be practically successful?

From the large set of problems that could further lead MCAs to unintendedly commit moral wrongs one important consideration as I will argue are cognitive biases that may influence thinking under normative uncertainty. Cognitive biases are most commonly described as deviations from some standard of rationality, which is the definition I adopt here (Baron, 2009). They are caused by unwanted, sometimes subconscious, mental processing.

This paper will be guided by the following question: How can we improve decision making under the possible influence of cognitive biases on metanormative judgments? The expression *metanormative judgments* includes a wide range of decisions which are made on or about issues on a second- or higher-order normative level.[4] For instance, a decision about whether MEC is correct as a second-order decision making approach or a decision about how to apply second-order normative principles are metanormative judgments. In the following, I will be mostly concerned with a subset of metanormative judgments, which I call *credence judgments,* which are subjective judgments about the plausibility of first-order normative theories.

After arguing for the relevance of cognitive biases, discussing the value of debiasing and identifying pathways for the influence of cognitive biases in metanormative thinking, this thesis follows a classical three-step approach towards improving decision-making, starting with the normative model, to then evaluate the descriptive world and lastly develop prescriptive suggestions (Baron, 2009; Larrick, 2008).

The first step is developing a normative standard of how MCAs should think under uncertainty, more precisely as we shall see later, how they should distribute their credences over first-order normative theories. To be able to improve decision-making later on, we first need to figure out what we want to improve towards, in other words, what good thinking under uncertainty would look like. Since this is not the place to develop a full normative standard of decision making under normative uncertainty, I will rather address the question of whether we can define some

[4] MacAskill (2014) introduced the term *metanormativism* for the discussion of whether there are norms that govern decisions under normative uncertainty. “Meta” points to higher levels of decision-making.

rules that would ensure better thinking under normative uncertainty. This will lead to a preliminary set of rules, which is enough for my purposes.

The second step is to analyse what is known about descriptive behaviour in order to identify where we may make judgments differently from what the normative standard suggests. I consider availability and status quo thinking as causes of biases. Since there is little research specifically on the role of cognitive biases in people's judgments over metanormative issues under uncertainty, this part relies on careful analysis of existing studies with the aim to show that there are reasons to believe that those biases may exist in credence judgments, too. This section should be seen as first steps towards making sense of how our thinking about such matters may fail, not as actual knowledge about the descriptive world.

In the third step the challenge is to give prescriptive suggestions. If those deviations I analyse in the descriptive part should actually play a significant role: Is there something we can do to deviate less from how we should think? And what would a debiasing process look like?

# 2. Cognitive Biases in Metanormative Judgments

## 2.1 On the Relevance of Cognitive Biases

I believe these questions to be highly relevant for the following reasons.

Psychological research has shown that we tend to make decisions differently than posited in normative models. Rather, many factors play into the formation of our moral beliefs. The notions of cognitive biases was introduced by Tversky and Kahneman in the early 1970s and research has since then exposed a long list of errors that our brains tend to perform (Tversky & Kahneman, 1974).

Research so far has (more or less arbitrarily) focused on empirical uncertainty. However, biases are not a result of some specific decision situation in a specific field, but rather patterns of thinking which we apply to any kind of decision making under uncertainty often independent of the specifics of the decision situation.

For example, in 2002, Kahneman and Frederick let subjects perform different tasks including "(1) categorical prediction (e.g., 'In a set of 30 lawyers and 70 engineers, what is the probability that someone described as 'charming, talkative, clever, and cynical' is one of the lawyers? '); (2)

summary evaluations of past events (e.g., ‘Overall, how aversive was it to be exposed for 30 minutes to your neighbor’s car alarm?’); and (3) economic valuations of public goods (e.g.,’What is the most you would be willing to pay to prevent 200,000 migrating birds from drowning in uncovered oil ponds?’).” (Kahneman & Frederick, 2002, p. 17) They proposed “that a generalization of the representativeness heuristic accounts for the remarkably similar biases that are observed in these diverse tasks.“ (Kahneman & Frederick, 2002, p. 17)

Therefore, it would be a much stronger claim to propose that we are unbiased in our metanormative judgments under normative uncertainty. The burden of proof rather lies with those that make this claim. This is the first motivation for analysing biases in metanormative decisions. In short: they are probably present.

The second motivation results from the fact that the outcome of a decision situation under normative uncertainty is highly sensitive to the distribution of credences over theories. Being only slightly affected by irrelevant factors, could change which action would be the adequate one in a given situation. Then, counteracting such biases seems like a necessary step towards avoiding moral wrongs.

A third motivation is that any deviation from the best possible decision-making procedure are especially harmful in this realm, as there is often much at stake in moral decisions. Additionally, a bias in metanormative thinking would affect a range of decisions on a first-order level negatively. The impact of steps toward debiasing seems highly valuable due to such butterfly effects.

## 2.2 The Value of Debiasing

The motivations mentioned above rest on an implicit assumption: that biases are always harmful to metanormative judgments and do not work in our favour. However, the following could be the case:

> *The Lucky One*
>
> Suppose, Utilitarianism is the correct first-order moral theory. In any decision situation under normative uncertainty, it is true that the option that Utilitarianism regards as most choice-worthy is the option that should be chosen. Leon is a MCA, who has 99% credence in Utilitarianism, and 1% credence in some other form of consequentialism, which due to similarities of the theories yields the same options choice-worthy as having

> 100% credence in Utilitarianism. His credence distribution will lead him to make the best possible decisions under normative uncertainty. However, he only does so because he has recently taking a convincing seminar on Utilitarianism. Besides Utilitarianism, he only remembers some other form of Consequentialism, which he has learned about years ago and which he judges very unlikely, because he cannot remember it completely.

Clearly, the formation of his credence distribution is heavily influenced by morally irrelevant factors (the point at and intensity with which he has learned about the theories). Let's assume that this is what we would call "biased judgments". Then *The Lucky One* shows that biases can work in one's favour by making one choose the correct option. In some cases, we come to judgments through faulty thinking and get lucky. Should Leon inform himself and engage in critical thinking about his credence distribution in this case, if doing so would lead him to believe in Utilitarianism (the correct theory) less?

The answer seems to be no, because the outcome is not improvable. Leon already chooses the correct option. However, can he *know* that he is in such a situation? I argue that he cannot know and that this dooms particular cases irrelevant. If he knew that he just got it right, there would be no normative uncertainty left and no biases can result. Leon would know that Utilitarianism is correct and would followingly know exactly which options to choose. Objective recommendations like "debias only if it *actually* improves judgments", even if theoretically true, cannot guide actions. Practically speaking, it seems not useful to give different recommendations for cases where people get lucky and other cases, where they do not, as it is practically impossible to separate them.

We are left with a discussion about what is the better *general strategy* for MCAs under normative uncertainty. The process of reaching outcomes in *The Lucky One* does not seem like something we would generally recommend. It introduces randomness and allows judgments to be influenced by irrelevant factors like the time at which someone learned about a theory.

If it is not the case, that most people actually get lucky most of the time, we should promote to improve decision making towards whatever rationality requires, i.e. debiasing. It seems like we must accept that there will be cases like *The Lucky One*, where debiasing will not only not improve judgments but may even make them worse. This thesis rests on the assumption that debiasing makes MCAs believe moral truths and fewer moral falsities on average.

Can biases make one more rational?

The definition of biases as deviations from a standard of rationality excludes the possibility of biases making one more rational. A similar argument as above applies:

The kind of rationality that is relevant for MCAs is instrumental rationality. "Someone displays instrumental rationality insofar as she adopts suitable means to her ends." (Kolodny & Brunero, 2013) MCAs have a well-defined moral goal: to avoid moral wrongs. Theoretically, if a necessary means is to believe in a contradiction, then that is what rationality requires MCAs to do. Rationality, as I apply it here, is therefore a context dependent set of means that are suitable to reach the given goal.

In particular cases it may be true, that biases lead MCAs to choose means, i.e. form judgments or believes or behave in a way which is closer to *what a perfectly rational person would do,* as it happens in *The Lucky One*. However, I would not call Leon in *The Lucky One* more rational. Leon got it right for the wrong reasons. He was not guided by what rationality requires. Faulty thinking lead him to form believes that by accident happened to be a good set of beliefs, which makes him choose correct options.

However, this discussion becomes more complex when one considers different parts of rationality. By solely addressing MCAs, who have a specific goal, I am, again, only concerned with the part of practical rationality referred to as instrumental rationality. There is much more to consider, when discussing biases and rationality in general. For example, the use of heuristics, which often leads to what is called "biased judgments", is in many cases rational. It reduces the complexity of option sets and allows agents to make fairly good decisions without high investments of time or cognitive effort, which are of scarce resources that need to be distributed across the agent's goals. What we call biases, ultimately depends on how we understand rationality in the given context.

I next turn to the question, where biases may interfere in metanormative judgments.

## 2.3 Pathways for Biases in Metanormative Judgments

The influence of biases on metanormative judgments may be manifold. A few possible pathways will be identified in this section. First, consider MEC again, to better understand the following pathways. The following is a classic example for how MEC works, borrowed from MacAskill (2019, p. 3).

*Vegetarianism*

> Harry is considering whether to eat meat or a vegetarian option for dinner. He thinks it's pretty unlikely animals matter morally, but he's not sure. If he eats meat and animals do matter morally, then he commits a grave wrong. If he eats the vegetarian option, he will certainly not commit a grave wrong, though he will enjoy the meal less than he would have done had he eaten meat. (MacAskill, 2019, p. 3)

Suppose his credence in the "Animals matter" view is 90% and his credence in the alternative view is 10%. His decision situation looks the following:

| | Animals matter – 10% | Animals do not matter – 90% |
|---|---|---|
| Eat meat | -1000 | 10 |
| Eat vegetarian | 5 | 5 |

Suppose the numbers in the table represent how choice-worthy the respective theory judges the option at hand. For instance, on the "Animals matter" view eating the meat is choice-worthy -1000 while eating vegetarian is choice-worthy of 5 to both views.

Now, the expected choice-worthiness of eating meat is calculated by the multiplication of credence and choice-worthiness. Eating meat has an expected choice-worthiness of $-1000 \times 10\% + 10 \times 90\% = -9100$ [units of choice-worthiness], whereas eating vegetarian has an expected choice-worthiness of $5 \times 10\% + 5 \times 90\% = 500$ [units of choice-worthiness]. In this simplified example, eating vegetarian would be the option with the maximum expected choice-worthiness.

Where in this process and beyond could biases possibly interfere? In general, there are at least 4 possible pathways through which cognitive biases could affect our metanormative judgments.[5]

(1) Through the variable *credence*.

Biases may play into metanormative judgments by affecting how we distribute credences over first-order normative theories. Making judgments about the plausibility of normative theories is a subjective evaluation process. Subjective evaluations always open the floor for biases. Awareness and a deep understanding of this influence will be

[5] Thank you to Christian Tarsney for his thoughts on the possible influence of cognitive biases on metanormative reasoning beyond MEC, which inspired this section.

crucial for a successful use of MEC. This is the pathway that will be addressed in this paper.

(2) Through the variable *choice-worthiness*.

In this paper I assume that the choice-worthiness rankings of theories are given and that there are no controversies about what theories recommend. Obviously, this is one of the biggest remaining problems for the MEC approach in general. Does anyone really know what the choice-worthiness rankings of all relevant theories look like? When it comes to Kant for example, people disagree greatly about the implications of his theory. Even when a sensible approach to finding choice-worthiness rankings is found, it may open several pathways for biases to influence MCAs in doing that work.

(3) By affecting how we distribute credences over *higher order normative theories*.

I assume the second-order theory MEC to be true for interval-scale measurable and intertheoretically comparable theories, however, in practice we need to decide which decision-calculus is appropriate, also for theories that do not fulfil these requirements, and not only on the second level. Any level would be relevant. The scope of this is practically limited by cognitive capacity. Until then however, biases can interfere on every level. Once we are biased on a higher level, all lower-level judgments will suffer. Suppose I choose MEC as the correct normative theory on a second-order level due to a bias towards quantifiable theories. Suppose also that MEC is incorrect. Now, I would need to be really lucky to still be able to consistently make correct judgments, as the effects of the bias on the second-order normative level distort judgments on a first-order normative level. This paper deals only with (1), however this consideration shows that research on the influence of biases on higher levels might be just as valuable.

(4) By affecting how we *apply* second-/first-order normative theories.

For the application of MEC, it is far from being clear what the appropriate choice in a given situation will be. Most papers on moral uncertainty rest on many heavy assumptions. The application of MEC will probably not be as straight-forward in practice, as examples like *Vegetarianism* suggest. A lot can go wrong by trying to translate appropriate credences and best-guesses on choice-worthiness rankings into appropriate actions. As MacAskill notes, “invoking moral uncertainty alone is not

> sufficient to conclude that vegetarianism is right or that abortion is wrong. Instead, one must also invoke substantive and potentially controversial assumptions about what credences one ought to have across a wide array of moral views and across different choices of intertheoretic comparisons." (MacAskill, 2019)
>
> MacAskill's paper discusses how various interactions between implications of normative uncertainty and the possibility of different intertheoretic comparisons complicates practical application. So, it is not only relevant that we are unbiased in our credence formation. We also need to figure out at least what the implications of moral views on a certain issue are from their choice-worthiness ranking, and what the most plausible intertheoretic comparison is, and then work out the highest expected-choice-worthiness (MacAskill, 2019). All of which may be subject to bias.

These and possibly other pathways of influence need to be considered when talking about debiasing on a metanormative level. As said, I will exclusively address pathway (1) in this paper: forming (un)biased credences over first-order normative theories, however, insights here can be useful for the understanding of other pathways, too.

The next section addresses the first step of the three-step model: the normative question. The definition of biases as deviations from a standard of rationality requires a definition of what rationality requires MCAs to do. More specifically, how should MCAs ideally distribute their credences over first-order normative theories?

# 3. The Normative Step: How Should We Think?

## 3.1 Metanormative Biases as Deviations from Good Thinking

This part contains the first step towards improving credence judgments. I develop a preliminary normative framework for thinking under normative uncertainty that best helps MCAs achieve their moral goal. It will then allow us to identify deviations descriptively (step 2) as a prerequisite for step 3, which is concerned with possible improvements.

This normative section is strongly inspired by Jonathan Baron's book *Thinking and Deciding* (2009), which is of high relevance in psychology. His definition of rationality and his conception of thinking has been frequently adopted in psychological discussions.

As said, biases can be described as deviations from some standard of rationality (Baron, 2009). Biased judgments violate a standard of rationality. Since we have already accepted MEC as the correct normative theory for interval-scale measurable and intertheoretically comparable first-order moral theories under normative uncertainty, classical instrumental rationality requires MCAs to maximize expected choice-worthiness, based on their credences in first-order normative theories. But that is not enough to avoid moral wrongs. MCAs need to form the best possible credence-distribution, those that are closest to the true moral landscape, in order to reach their goal.

I suggest that the formation of adequate credences requires what Baron calls *good thinking* and that good thinking comes with a normative force for MCAs. It would be irrational for a MCA to refuse to do good thinking, if it is true that this would best reach her goal of avoiding moral wrongs. This leaves us with having to figure out what good thinking about the plausibility of normative theories is.

Based on this, biases in metanormative decisions, at least in this particular pathway, are defined as deviations from good thinking or as I will put it: violations of good thinking rules.

## 3.2 What Does Good Thinking Require?

The precise question that motivate this part is the following. What would good thinking under normative uncertainty require and can we formulate any rules for MCAs that would lead to the best possible credence-distribution?

The following comes with no claim to completeness. What I am proposing are what seems to me to be the minimal necessary conditions that must be met by MCAs thinking styles to end up with appropriate, (as far as possible) unbiased, judgments about the plausibility of first-order normative theories. Those are enough for showing in step 2 how the analysed thinking patterns violate even those basic requirements. However, a complete normative model of what good thinking requires, would obviously be necessary to pursue this further.

Generally, it seems like two major things can go wrong when we think about our credence-distributions. We may not have looked enough for new moral information or we may have sufficient information, but we draw inappropriate inferences from it.

Avoiding this leads to a standard of rationality, characterized by a few rules that require MCAs to draw inferences without violating the basic rules of coherence, as well as to perform fair and sufficient search for evidence.

Let's look more closely at where these components come from and what they contain.

## 3.3 Coherence

Classical coherence requires one not to make logical errors in the formation of probability beliefs. It has been criticised for being unattainable. However, if coherent moral beliefs are necessary for avoiding moral wrongs, then we should *ceteris paribus* try to be as coherent as possible. Scarce resources such as time or cognitive capacity may lead to compromises between perfect coherence and other requirements. Suppose it would take up all available time to reach perfect coherence so that there is nothing left to for example look for new evidence. In such cases there may be an optimum degree of coherence below perfect coherence. I assume that the closer we get to the ideal state, the better the decisions are.

In the theory of epistemic justification coherentism "implies that for a belief to be justified it must belong to a coherent system of beliefs." (Murphy, 2020, para.1) It usually involves the following components: logical consistency (no contradictions involved), explanatory relations, and various inductive (non-explanatory) relations (Murphy, 2020). A coherent credence distribution is an important prerequisite for MEC to produce the correct expected choice-worthiness of options, given the agents credence distribution. This is because the MEC approach cannot generate the overall best choice if the inputs are incorrect and being incoherent means that at least one of the held moral beliefs must be incorrect.[6] However, as noted above, we may need to accept some incoherence and thereby some inferior MEC outputs as compared to perfectly coherent credence inputs, in order to meet other requirements.

Imagine someone having a 50% credence in Kantianism, 50% credence in consequentialism and if she is asked how her credence is distributed among the sub-forms of consequentialism, she says 40% utilitarianism and 40% rule consequentialism. This adds up to 80% credence in consequentialist theories which together with the 50% in Kantianism is a total of 130%. Clearly,

[6] I shall note that Coherentism is often contrasted with epistemic foundationalism which addresses the structure of an agent's beliefs. In foundationalism some beliefs are only justified because of other beliefs that are justified. In my approach however, the combination of coherence and updating incorporates the basic idea of foundationalism and I see no need for discussing this in length. It is not the standard approach in expected utility theory either. It may, however, be possible to construct a normative model using foundationalism as a basis.

we would say that this agent's probability beliefs could be improved, as they violate a basic probability rule:

(1) Probabilistic consistency: Credences in first-order normative theories should add up to one. This is what John Broome calls *Bayesian requirement*: "When *p, q* and *r* are mutually contrary propositions such that (*p* or *q* or *r*) is necessarily true, rationality requires of *N* that *N*'s degrees of belief in *p, q* and *r* respectively add up to one." (Broome, 2013, p. 198) Where N represents a moral agent.

Baron composed a list of rules that fall within the concept of coherence. "Other constraints on probability judgments have to do with the need for coherence among judgments concerning various propositions. This means that our probability judgments must obey the rules of probability as a mathematical system. Related judgments must 'cohere' according to these rules." (Baron, 2009, p. 114) Without going into detail, his list contains rules about additivity, conditional probability, independence and the multiplication rule. These could be used as a starting point for a full normative model but are not necessary for the purposes of this paper. Rather, what is important here is the following second rule, which captures a basic idea of coherence:

(2) Logical consistency: Be logically consistent in your probability judgments, i.e. avoid contradictions in your moral beliefs. In Broome's words: "Rationality requires of *N* that *N*'s degrees of belief at t that p and also believe at t that not p." (Broome, 2013, p. 155)

However, as Broome notes, it does not follow that if rationality requires (1) and (2) that it also requires us to identify violations of (1) and (2) and change our beliefs accordingly. This indicates that I am merely scratching the surface of the discussion of what rationality generally requires one to do or believe under normative uncertainty.

Coherence requires more than (1) and (2). MCAs cannot sensibly hold any credence distribution as long as it is probabilistically and logically consistent.

> Besides being probabilistically consistent with one another, coherent beliefs gain in justification from being inferred from one another in conformity with the canons of cogent inductive reasoning. (Murphy, 2020, bii para. 2)

Kahneman and Tversky also note:

> For judged probabilities to be considered adequate, or rational, internal consistency is not enough. The judgments must be compatible with the entire

> web of beliefs held by the individual. Unfortunately, there can be no simple formal procedure for assessing the compatibility of a set of probability judgments with the judge's total system of beliefs. The rational judge will nevertheless strive for compatibility, even though internal consistency is more easily achieved and assessed. (Tversky & Kahneman, 1974, p. 1130)

What rules can we construct from compatibility? One intuitive rule for drawing from evidence is the following.

(3) "If you have stronger evidence for theory A then you should assign higher probability to A." (Baron, 2009) Irrelevant factors should not change the probability distribution.

Another rule that has been raised in the context of coherence is:

(4) The psychological realization condition, which requires that one's beliefs must stand in some relation in the persons mind. It must be inferred from another or at least depend on other belief such as if the agent's credence in consequentialist theories would change to be close to zero that would change the agent's credence distribution over the other theories in some way.

These rules put limits on which credence judgments count as appropriate, i.e. suitable to reach the given moral goal. Improving on this part alone may improve overall decision making. Striving for coherent beliefs is necessary, but not sufficient for making the overall best decisions under moral uncertainty. There is more that MCAs need to do, to avoid to perfectly coherently commit moral wrongs. They need to be open to and search for new evidence.

## 3.4 Search for Evidence

In their book *Moral Uncertainty* (2020), MacAskill, Bykvist and Ord write a whole chapter on the value of moral information and the difficulties involved in calculating how much we should be willing to "pay" for new moral information, especially imperfect information. They argue that the value of moral information can be calculated in the same way as we would calculate the value of empirical calculation. For details see chapter 9 in their book.

Moral information is what I call evidence here.

> [S]omething is a piece of moral information iff coming to possess it should, epistemically, make one alter one's beliefs or one's degrees of belief about at least some fundamental moral proposition. (MacAskill et al., 2020, p. 230)

They explain that "the term 'moral information' could apply to experiences, arguments, intuitions, or knowledge of moral facts themselves", which is the definition I adopt here. In the following I will often merely speak of arguments as examples for moral information/evidence for reasons of simplicity. Moral information will most likely be imperfect information: it will improve our epistemic state rather than giving us certainty.  (MacAskill et al., 2020, p. 242)

Being aware that new evidence may change one's credence distribution and the fact that it is ceteris paribus better to have more information when making judgments, produces the following rules.

(5) If you reasonably believe that it is likely that there is evidence that may change your credence distribution, you should look for that evidence.

MacAskill et al give an argument for what counts as reasonable: One should calculate the expected choice-worthiness of gaining new moral information. Without going into detail, it ultimately depends on how reliable one takes the information to be. "If one believes that one will not learn very much from doing study, research or reflection on ethical matters, then the expected choice-worthiness of gaining that moral information will be low." (MacAskill et al., 2020, p. 245)

How do we gain moral information? Examples which MacAskill et al name include studying or researching moral philosophy, engaging in ethical reflection. The classical sources of knowledge and justification are perception, introspection, memory, reason and testimony (Steup & Neta, 2005). Perception cannot be a source of knowledge for metanormative questions, as we cannot perceive it. Introspection may be possible, however, it is not clear to me what metanormative introspection would look like. Reasoning can be simply translated to thinking. Thinking is arguably the biggest source of knowledge here as the examples cited above show, which all relate somehow to thinking. I understand testimony in the sense that someone presents a theory, normative criteria or an argument of theirs, as it happens when one listens to a lecture on moral philosophy, which inspires the listener to think in new directions or reflect on their current credence distribution. Memory as a source of moral information would be retrieving a piece of information that has been forgotten. Arguably, this information is not new, however, has the capacity to change one's moral beliefs about at least some fundamental moral proposition.

Another question remains: How long should a MCA search for new moral information? The answer is captured by the following rule:

(6) Search must be sufficient in the sense that "it best serves the thinker's personal goals, including the goal of minimizing the cost of thinking." (Baron, 2009, p. 63)

Meaning that at some point the cost of further search for moral information may pose such high opportunity costs that it is not appropriate for an agent to pursue the search further. Since the personal morally relevant goal of MCAs is to avoid moral wrongs search must reach a level where appropriate fulfilment of other requirements of rationality in order to reach that goal are ensured.

There is one more crucial requirement for both good search and inferences:

(7) "Search and inference are 'fair' when they are not influenced by factors other than the goals of the thinking itself." (Baron, 2009, p. 63)

This means giving for example each argument the chance to change one's mind, regardless of how plausible it seems intuitively and incorporating it neutrally as far as possible.

So, after having checked that one is coherent, (5) - (7) cut out at least some ignorant or lazy thinking, which means being one step closer to good thinking.

All of this together is what Baron would call a step towards actively open-minded thinking (good thinking). This is what I believe MCAs wanting to avoid moral wrongs should strive for.

> In general, then, actively open-minded thinking is most likely to lead to true beliefs. In addition, when we cannot be sure that a belief is true, good thinking will ensure that our confidence in the belief is in proportion to the evidence available. Appropriate confidence is, in most cases, a more realistic goal than certainty. (Baron, 2009, p. 70)

One last note at this point: Two people with exactly the same evidence might still arrive at different conclusions even when considering requirements for both inferences and evidence search as much as possible. This is precisely because credence remains a subjective variable. We can point to a few constraints and show that the thinking that lead to some credence judgments is not appropriate. We *cannot* point to a unique way of forming credences under every circumstance for every agent just like this is not possible for empirical uncertainty problems.

This account is a first attempt to point to at least a few promising components of the best possible style of thinking about the plausibility of first-order normative theories in order to form appropriate credences. It remains rather vague and is surely in need for completion. For practical and theoretical purposes, it would be great to develop more precise rules. However, as they may become more complex and detailed, the developed standard of good thinking may not be too bad for practical purposes after all.

Now that we have worked out how rationality may require MCAs to form their credence distribution, we can better understand what it means to be biased. It would simply mean to systematically violate at least one of the 7 given rules. Now, we can look for deviations (step 2) from that standard in order to find out where we need to improve.

# 4. The Descriptive Step: How Do We Think?

## 4.1 The Methodology of Analysing Biases

There are around 200 cognitive biases to be found in the scientific literature. All of which could potentially be relevant to decisions under normative uncertainty.

I focus on availability and status quo thinking as causes of biases. We seem to have a fairly good picture of their nature and sufficient empirical evidence of when and how they influence decision making. This gives us the necessary basis to theoretically analyse them in the context of credence judgments. All other biases are not necessarily less relevant. Again, this descriptive part should be seen as an example rather than a complete analysis. It is a first attempt to unravel the influence of some thinking patterns on our judgments under normative uncertainty.

The procedure will be the following: Since there is little empirical information on biases in credence judgments, we must consult the general literature on cognitive biases.

For both of the following cognitive biases I will first sketch what we know from psychological research (4.2.1 and 4.3.1). Building on that I show that there are many reasons in favour of those thinking patterns being present in our thinking about the plausibility of normative theories, too (4.2.2 and 4.3.2). Looking back at the normative standard, I will further show for both of them, where they deviate from it (4.2.3 and 4.3.3).

## 4.2 Availability Thinking

### 4.2.1 What We Know from Research

The availability heuristic was discovered in the early 1970s by Kahneman and Tversky as part of their investigation into mental shortcuts (heuristics) that people subconsciously use to solve complex tasks under uncertainty (Tversky & Kahneman, 1973).

Our brains frequently replace rather complex questions with similar easier questions. “A person who is asked ‘What proportion of long-distance relationships break up within a year?’ may answer as if she had been asked ‘Do instances of swift breakups of long-distance relationships come readily to mind?’” (Kahneman & Frederick, 2012, p. 4). This is an example of the availability heuristic.

Doing this, we are not aware of the substitution of questions and since “the target attribute and the heuristic attribute are different, the substitution of one for the other inevitably introduces systematic biases.” (Kahneman & Frederick, 2012, p. 5) We are biased towards answers that are easily available to the mind in a given situation.

Tversky and Kahneman first wrote about the availability heuristic in the context of evaluating frequencies or probabilities of events. They suspected that the judgment process is mediated by availability considerations. Their research suggests that the number of relevant examples that subject can recall, influences their frequency and probability judgments (Tversky & Kahneman, 1973).

Later research has confirmed that availability is a relevant mediator in probability evaluations and that such effects remain even when accessibility of positive/negative memories are actively manipulated by priming subjects or inducing certain moods (Gabrielcik & Fazio, 1984; MacLeod & Campbell, 1992). For frequency estimations, a series of studies contrarily indicates that availability is not a relevant factor when judging frequencies (Manis, Shedler, Jonides, & Nelson, 1993).

Besides this, the effect has been shown to exist in set size judgments (Manis et al., 1993) and even in social judgments, where people demonstrated an egocentric bias when assessing their own and their partners’ contributions to certain activities. The authors ascribe the bias to the increased availability of examples that have been experienced by the self (Fiedler, 1983).

Other research has shown that the mere ease of recall serves as a distinct source of information next to the content that is recalled. In this study subjects recalled of examples for their own assertive behaviour. When this is easy, the subject’s self-assessment showed higher assertiveness. If it is hard to recall examples of assertive behaviour, they reported lower assertiveness (Schwarz, Bless, Strack, Klumpp, & al, 1991; Schwarz & Vaughn, 2012).

Availability has also been used to partly explain biases in general scientific hypothesis generation, where specified and unspecified hypotheses are used, and subjects consistently showed overconfidence in specified hypotheses because difficulty retrieving unspecified hypotheses. This shows how some categories of instances that are by nature harder to remember, can be systematically neglected in judgments due to their complexity (Mehle, Gettys, Manning, Baca, & Fisher, 1981).

A similar effect exists with vivid information: In two studies it was found that concrete and colourful language influences judgments about a woman’s fitness as a mother and that the presence of photographs affected judgments about the proportion of male and female students at two universities. Vivid presentation of information affected its ready accessibility in memory (Shedler & Manis, 1986).

From this overview, we can see that with the availability heuristic, Kahnemann and Tversky seem to have uncovered one of the predominant mental shortcuts. Over various decision situations with diverging tasks, subjects’ judgments are at least partly influenced by which relevant instances are most available to the mind. Now, how could availability mediate the process of finding a subjective credence distribution?

### 4.2.2 Application to Credence Judgments

The robustness of these findings gives reason to believe that judgments under normative uncertainty are mediated to some extent by availability, too, if the decision-situations are similar in relevant aspects.

We know that heuristics are applied when uncertainty is involved. Thinking about our credence distribution, we are clearly dealing with high levels of uncertainty. In empirical estimations, we have some objective measure to calibrate to. In metanormative thinking, we are left with subjective judgments alone, which demands increased cognitive effort and opens the floor to the use of heuristics instead of appropriate complex thinking.

We also know that we make use of the availability heuristic when judgment tasks are complex. Judgments about credences are complex. Arguably even more complex than the rather simple estimation tasks that are used in most studies on the availability heuristic. If the question whether the letter "K" is more often the first or the third letter in English words gives rise to an availability caused bias (Tversky & Kahneman, 1973), what do questions like "How plausible do you judge Kantianism?" produce? Thinking about such questions requires evaluating a wide range of possible theories, their up- and downsides, their implications and much more. It seems to be a more multi-dimensional task than all the experiments that have been run on the availability heuristic so far, giving reason to believe that at least some points, availability considerations may play in.

Credence judgments require lots of thinking, retrieving and remembering. There is nothing observable to be found when it comes to metanormative questions. Precisely the kind of cognitive tasks (remembering, accessing information, retrieving, …) that we seem to frequently escape using shortcuts, are especially important for metanormative decision making. If theories which I have learned about/talked about last, are those that are more likely on the top of my mind and retrieved more easily and then followingly likely to be judged with higher credences, the availability heuristic could largely distort credence judgments. This may also mean that theories which one learned about a while ago might not play a role in the calculus at all. And theories that we do not know about in general can naturally not be available.

We may here take the experienced ease of recall as a source of information, too. If I am thinking about whether Rule Utilitarianism is more plausible than Act Utilitarianism, I might come to compare the pros and cons I learned about for both of them. Listing instances of two categories like pros and cons, has been the setup of many of the above cited studies and a classic set up in which availability has been proven to distort judgments. If the pros for Hedonism come more readily to my mind, this may bias me towards hedonism. This effect could happen for various reasons: For instance, I may have for example just taken a seminar on Hedonism or I had a really good book on hedonism which explained the theory vividly and with colourful language, which, as we know, increases availability.

Also, we know that own experiences are more readily available than reports from others. It therefore may make a difference where I got information about moral theories from: did I reason myself or did I merely listen to another's reasoning. The intensity of active involvement in the thinking process may lead to higher availability of the information at some later point. Having reconstructed an argument and understood it thoroughly seems to be a relevant factor here. Understanding an argument usually involves active cognitive effort and thereby increased availability. So, we may end up placing too much weight on our own arguments, at least if we do not make an active effort to understand other arguments we hear or read about.

What we also learned from research is that some instances may be underestimated because they are for some reason systematically harder to recall (unspecified theories compared to specified theories). Is there something we may systematically underestimate because we have a harder time remembering it? Probably. We can think of mathematically complex theories or long, counterintuitive arguments as possible candidates. Obviously, a theory does not become less plausible due to its property of being complex (unless we argue that good theories ought to be simple). If that property leads to systematically lower credences in such theories via availability effects, or we judge such theories less plausible, because we do not fully understand them, we end up being biased towards simple and easily understandable theories for no good reason.

Summing up, there are many reasons to believe that at least some credence judgments are mediated by availability.

### 4.2.3 The Deviation

Comparing the descriptive insights into human decision making to the normative standard, it can be shown why the availability heuristic leads to biases. The use of the availability heuristic violates two basic requirements:

(3) "If you have stronger evidence for theory A then you should assign higher probability to A." (Baron, 2009) Irrelevant factors should not change the probability distribution.

(7) "Search and inference are 'fair' when they are not influenced by factors other than the goals of the thinking itself." (Baron, 2009, p. 63)

Giving the most available information disproportionate weight in the decision process hinders us from drawing fair and evidence-based inferences from the information we have, creating suboptimal metanormative decision-outcomes.

In sum, these possible deviations are what can be called availability bias in credence judgments.

## 4.3 Status Quo Thinking

### 4.3.1 What We Know from Research

Let's turn to the second thinking pattern: status quo thinking. First, I sketch what research has shown about this particular kind of thinking (4.3.1), then I apply this knowledge to the context of credence judgments (4.3.2) and again make the bias, i.e. the deviation from the standard of rationality developed in the normative section explicit (4.3.3). So, what do we know from research?

In 1988 Samuelson and Zeckhauser first published a series of decision-making experiments which demonstrated that individuals disproportionally stick to the status quo. Their results implied that an alternative becomes significantly more popular when it is designated as the status quo (Samuelson & Zeckhauser, 1988).

Today, it is well-known that we show a context-independent tendency towards the status-quo option in decision-situations, which cannot be explained by the actual superiority of the status-quo option (Kahneman, Knetsch, & Thaler, 1991).

Since the discovery of the tendency towards the status quo, research has focused on unravelling the underlying factors that *cause* this tendency. Simple preference for the status quo option is not enough to explain the full scope of the observed bias in judgments. The following list contains explanations that seem to play a role in biasing people's judgments toward the status quo: Loss aversion, regret avoidance, preference for the status quo, bias toward omission, repeated exposure, rationalization, existence, longevity and limited attention.

*Loss aversion, regret-avoidance & preference for the status quo*

A first explanation came from Daniel Kahnemann and colleagues in 1991. Loss aversion and a preference for keeping things as they are were shown to account for most of the status quo bias in their experiments. Loss aversion describes behaviour where subjects prefer avoiding losses to having an equivalent gain (Kahneman & Tversky, 1984). Loss aversion is tightly connected to regret avoidance. Kahneman and Amos Tversky observed in 1982 that people's regret is greater for bad outcomes that they actively bring about than for bad outcomes that result from inaction.

Later research, which aimed at clarifying the underlying neural mechanisms, also suggests that status quo bias is regret-induced. It confirmed that subjects regret (measured by neural processes in the brain) was higher after an erroneous status quo rejection compared with acceptance (Nicolle, Fleming, Bach, Driver, & Dolan, 2011).

Overall, the tendency towards the status quo is therefore at least partly an implication of some loss-aversive thinking, our wish not to feel regret and a preference for the status quo. At least for decisions under empirical uncertainty.

*Bias toward omission*

Baron and Ilana Ritov (1994) emphasized that doing nothing as opposed to something or maintaining one's previous belief as opposed to change are two distinct factors that need to be separated. In a series of three experiments, they showed that much of the tendency towards the status quo can be explained by a preference for inaction instead of a preference for keeping things as they are. They tested scenarios in which change occurs unless action is taken, where subjects preferred inaction over action independent of whether inaction meant change. They concluded that the tendency towards the status quo is at least partly caused by a bias toward omission. This bias stems from the common preference for harmful omissions over equally harmful actions (Baron & Ritov, 1994; Spranca, Minsk, & Baron, 1991). And of course, if we show a tendency towards inaction, then that will favour retaining the status quo.

*Repeated exposure*

When subjects are exposed to something repeatedly, they tend to start liking it more than they would otherwise do (Bornstein, 1989; Harrison, 1977; Zajonc, 1968). This simple effect also explains the tendency towards the status quo because we are naturally exposed to what exists all the time as opposed to non-existing alternatives. Thereby, the status quo will be evaluated more favourably and perceived as more true (the truth effect) (Eidelman & Crandall, 2012). The authors note: "Like mere exposure, the truth effect is not specific to any one domain but instead occurs for a broad array of topics, including people, politics, history, art, geography, religion, science, and marketing." (Eidelman & Crandall, 2012, p. 272)

*Rationalization*

Especially, when existing states are the result of our own choices, we are motivated to justify those decisions, upgrading what was chosen and downgrading what was not (BREHM, 1956). It was demonstrated that people like the world to be a just, manageable and predictable place and tend to see the world that way (Lerner, 1980). Similar motivations to rationalize have been shown to exist with extant social systems (Jost, Banaji, & Nosek, 2004) and even members of underprivileged groups rationalize their own disadvantage (Jost, Pelham, Sheldon, & Ni Sullivan, 2003).

*Existence & Longevity*

We treat existence as a prima facie case for goodness. Without giving it much thought, we simply assume the goodness of existing states (Eidelman, Crandall, & Pattershall, 2009). Taking what is as what ought to be is a classical naturalistic fallacy but by definition favours the persistence of the status quo (Eidelman & Crandall, 2012).

People value the status quo regardless of the costs associated with change. This has been shown for system changes and aesthetic qualities of galaxies. It holds across different contexts (Eidelman & Crandall, 2012).

Besides existence, the longer something exists, the better justified it is perceived (Eidelman, Pattershall, & Crandall, 2010). Even negative stimuli benefit from longevity. The use of torture was judged as more justified when it was around for long then when it was newly introduced (Crandall, Eidelman, Skitka, & Morgan, 2009).

*Limited attention*

The status quo bias increases with the number of alternatives. This has been shown in an experimental (Samuelson & Zeckhauser, 1988) and later in an empirical setting, where individuals in the US equity mutual fund market demonstrated more severe status quo bias in segments, where there are more funds to choose from (Kempf & Ruenzi, 2006).

In sum, this sketches what is known about the status quo bias and its causes. Many factors seem to contribute to the frequently observed tendency towards the status quo in choices. Now, how could the status quo mediate the process of finding a subjective credence distribution?

### 4.3.2 Application to Credence Judgments

In the context of credence judgments, “status quo” could refer to at least two things.

(i) the credence distribution we currently hold

(ii) the set of most believed theories in our environment (friend group, some institution, society, …)

We could be biased towards our current credence distribution, or we could be biased towards the most popular theories in for example our society, for the similar reasons. Both fall within the scope of the status quo bias because they are characterized by a tendency to remain in the current state as opposed to change.

Note that we may have good reasons to prefer popular theories or to rather stick towards the credence distribution in which we have invested much thought and consideration. What can be called status quo bias are merely those extra tendencies towards the status quo that cannot be reasonably explained by the superiority of the status quo.

Do we have any reason to believe that the above listed causes of the status quo bias influence credence judgments? If so, we have reason to believe that a status quo bias exists.

*Loss aversion, regret-avoidance & preference for status quo*

We may simply show a preference for the status quo, of any kind, too, since this is an observable phenomenon which occurs whenever there is a status quo default option.

We know that as soon as some option is designated as the status quo, it is favoured more often.
(i) If we are aware of "our credence" distribution, it has become a status quo. Simply by designating it as such, we may evaluate it more favourably.
(ii) The most popular theories in our society, form an implicit status quo. For example, Kantianism, Utilitarianism and Virtue Ethics form the basics of ethics taught in many schools and universities. The set of those theories enjoys much more attention than the set of all other theories, which may lead to a bias towards well-known, widely accepted theories.

Regret seems to play no major role in mediating credence judgments, as we most likely not know whether our credences lead to wrong or right actions. There can simply be no feeling of regret. A similar argument holds for loss aversion. There is nothing decision-makers "possess" that they could be reasonably afraid to lose through changes in their credence distribution. We may, however, be averse to taking moral risks. The loss would be lives lost or harm caused. However, whether this effect weighs equally strong as classical loss-aversion towards possessing goods, cannot be said from mere analysis.

This class of processes, which served as the youngest explanation for the status quo bias, seem to be rather irrelevant in the context of credence judgments.

*Bias toward omission*

There is more that may cause a status quo bias. The demonstrated preference for omission may lead people to believe that falsely sticking to the popular theories is better than actively seeking alternatives and being wrong about them. If inaction is preferred here, for reasons of omission or mere cognitive laziness, this may lead decision-makers (i) to not change their credence distribution enough or (ii) to not deviate enough from popular views in the sense that they would deviate more if it was not for the status quo effect.

*Repeated exposure*

People are by definition more frequently exposed to (ii) the popular theories in society. If people pursue ethics in some institutional setting, they will probably first be exposed to Kantianism, Utilitarianism and Virtue Ethics. Over the course of their studies, those are talked about, studied and discussed more frequently. At the same time, (i) engaging with our own credence distribution, we evaluate the status quo distribution much more often than alternatives. We are exposed to the status quo repeatedly, which may lead us to believe it is truer.

*Rationalization*

Credence distributions are the result of deliberate choices. We know that people like their choices to be justified. We are motivated to come up with good reasons for why we hold those beliefs, which gives them disproportionate weight. This is the classical process of rationalization. In addition, we like to see that we are making good choices, so we are inclined to believe that our credence distribution is a good choice. The increased uncertainty of metanormative issues may even increase such effects. The less we know what is right and wrong, the more we need to justify. Rationalization may therefore play a major role in credence thinking.

*Existence & Longevity*

The existence and longevity effect may lead to (ii) overconfidence in existing and long-standing theories. We may think that the theories that exist, exist for a reason. Those that do not exist, are hard to imagine, but there may be alternatives that we would judge plausible if we knew about them. We may also mistake theories that have been popular for a long time or (i) our own credence distribution that has not been changed in a long time as good. Whereas the majority could be mistaken and not having changes in the own belief set could also be due to a lack of search for alternatives.

*Limited attention*

The status quo effect increases with the number of alternatives. This is a problem for credence judgments, as there is an infinite number of alternatives. Think of all the possible combinations of theories that one could have non-zero credence in. Having to evaluate and compare all relevant scenarios requires immense cognitive effort and probably exceeds out attention limit.

All these effects could lead to a bias towards the status quo (i) credence distribution or the status quo (ii) theories in one's environment in credence judgments.

### 4.3.3 The Deviation

Having disproportionate credence in at least one theory because of the above mentioned effects as opposed to careful argumentation for why one should take the fact that the status quo is the status quo as reason to increase credence in this view, violates the basic requirements for drawing inferences:

(3) "If you have stronger evidence for theory A then you should assign higher probability to A." (Baron, 2009) Irrelevant factors should not change the probability distribution.

(7) "Search and inference are 'fair' when they are not influenced by factors other than the goals of the thinking itself." (Baron, 2009, p. 63)

The status quo credence distribution or popular theories should not be overvalued because they are perceived as the status quo. This is an irrelevant factor.

Such behaviour may also violate the following search rules from the normative part:

(5) If you reasonably believe that it is likely that there is evidence that may change your credence distribution, you should look for that evidence.

(6) Search must be sufficient in the sense that "it best serves the thinker's personal goals, including the goal of minimizing the cost of thinking." (Baron, 2009, p. 63) Meaning that at some point the cost of further search for arguments may pose such high opportunity costs that it is not appropriate for an agent to pursue the search further.

If we prefer inaction over action or prefer the status quo for some other goal-unrelated reason, we may not be actively looking for evidence enough, even if it is likely that it may change our credence distribution. Thereby, sufficient search may not be reached.

In sum, these possible deviations are what can be called status quo bias in credence judgments.

# 5. The Prescriptive Step: How Can We Improve Thinking?

## 5.1 The Debiasing Process

The use of the availability heuristic has benefits in many everyday life situations, where we want to decide quickly in reoccurring decision situations. What comes to mind quickly and easily is usually what we have used in recent decisions and therefore the most important ideas to think about now. And there are usually good reasons for why the status quo is the status quo. Not every sticking to the status quo or using available information is necessarily a sign of poor reasoning. However, the forgone section has shown, that we may apply availability and status quo thinking in situations where it violates some basic normative thinking rules which leads to suboptimal judgments. MCAs should therefore strive to counteract such availability and status quo effects.

We want to close the gap between the normative model and descriptive behaviour. What should such a debiasing process look like?

In 1994 Brekke and Wilson defined mental contamination as the process in which a decision-maker ends with an unwanted judgment because of mental processing that is unconscious or uncontrollable. "Unwanted" means the decision-maker would prefer not to be influenced by those processes. They presented a model which breaks the debiasing process down into a few steps, which are needed to close the gap between normative standards and descriptive behaviour if a bias is present. The process indicated by the blue arrows leads to perfect mental correction, i.e. successful debiasing. If the answer to any of the blocks is "no", a bias is created.

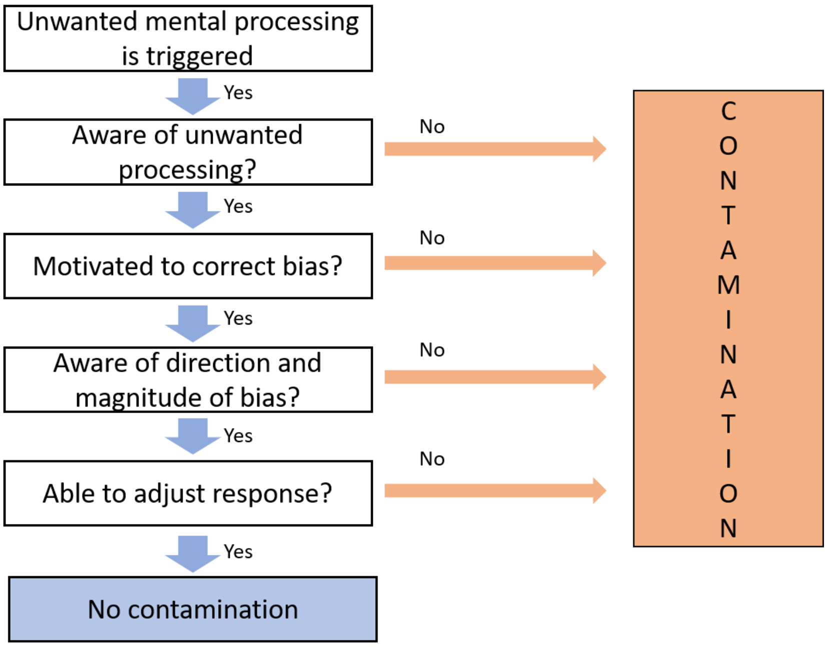
Unwanted mental processing is triggered
Yes
Aware of unwanted processing?
No
Yes
Motivated to correct bias?
No
Yes
Aware of direction and magnitude of bias?
No
Yes
Able to adjust response?
No
Yes
No contamination
CONTAMINATION

*Figure 1*. The process of mental contamination and mental correction from Wilson and Brekke (1994)

It starts with the decision-maker being *aware* of the bias, when performing it. Unfortunately, we seem to often not realize that our judgments are biased due to for example missing feedback or ambiguous causal determinants of the outcome (Hogarth, 2001). It is therefore crucial to spread knowledge about biases. Up to this point, this whole thesis aimed at introducing metanormative biases as a real issue and raising awareness for the possible presence of such cognitive biases in our metanormative reasoning. So, we have already taken a first step towards successful debiasing. But mere awareness is not enough. Several studies show that simply telling subjects about the bias and asking them to avoid it does not work (Pohl & Hell, 1996; Quattrone, Lawrence, Finkel, & Andrus, 1984). Therefore, the model contains more.

The second block is one on which much of the debiasing literature focuses. How can we *motivate* people to want to correct their biases? Debiasing revolves much around finding the best strategies to motivate people to make better decisions. Such approaches include giving incentives, nudging or making people accountable for the outcome (Larrick, 2008). None of this is relevant for our purposes, as we assume MCAs to be naturally motivated to debias, as this is the only way towards reaching their goal.

What is more interesting is the following block. *Awareness for the direction and magnitude* of the bias is necessary for the process to continue. This addresses the lack of information when it comes to biases in normative and metanormative reasoning. In the previous section I have shown that we have reason to believe that biases are present in credence judgments, but this does not compensate missing research. For successful debiasing, studies and experiments are needed that specifically tackle normative and metanormative thinking. Additionally, the general studies on biases and heuristics are often ambiguous and merely demonstrate that a bias exists. What is needed to develop fitting debiasing strategies in later steps, is knowledge about the actual underlying processes that cause the observed bias. Only then can we successfully counteract such effects.

The block "*Able to adjust response?*" is ultimately what Baruch Fischhoff has defined as debiasing: the application of methods to reduce and remove unwanted mental processing (he calls this "biases" in his paper) that introduce biases (he calls this "errors") into judgments and decisions (Fischhoff, 1982). The mental contamination process suggests that debiasing cannot be reduced to the mere application of correct methods, yet it is the core part of the debiasing process.

It contains two distinct challenges. We need to develop effective debiasing strategies and we need to be able to apply them correctly in relevant situations. Fischhoff has already extracted 25 possible different strategies for debiasing, depending on the kind of bias (Fischhoff, 1982). His and other results have been recently evaluated by Richard Larrick (2008). There remains much disagreement about which kinds of strategies work best for which kinds of biases. Since biases in metanormative judgments will differ to some degree from other biases, we may need to adapt existing strategies or come up with new, fitting ones. This is a project that cannot be addressed yet, as the knowledge about the concrete influence of biases is missing.

However, it would be wrong to conclude that there is nothing we can say at this point about counteracting biases on in credence judgments. The following are some starting points that can give practical guidance even at this early stage in the process.

## 5.2 Towards Successful Debiasing

One thing we can learn from three decades of research on debiasing strategies is that successful biasing strategies share a common feature: Moving away from intuitive thinking towards deliberate reflective thinking (Croskerry, Singhal, & Mamede, 2013).

This idea rests on the *dual process theory* which suggests that our brains generate judgments through two different kinds of processes (Wason & Evans, 1974). One being a fast, automatic and unconscious process (System 1), the other being a slower, conscious process of active thinking (System 2). The names “System 1“ and “System 2” are commonly used to distinguish between the two thinking processes (Kahneman, 2011). Biases are mostly present in our System 1 thinking, however, persist in System 2 judgments if not actively tackled. Debiasing can only work in System 2 thinking, where intuitive judgments are evaluated and updated. Being able to perform this switch in thinking may be the critical feature of cognitive debiasing (Croskerry et al., 2013; Kahneman & Frederick, 2002).

So, be it availability, status quo or any other cognitive bias that we want to avoid, “applying cognitive effort, rationalizing, slowing down, using tools and aids, and bringing more information and facts to the decision-making process” (Kahneman, 2011), in other words, System 2 thinking, is a good way to start.

Beyond that, a general good cognitive strategy seems to be “*consider the opposite*”. It is simple and has been effective at reducing different kinds of biases (Arkes, 1991; Mussweiler, Strack, & Pfeiffer, 2000). “’Consider the opposite’ works because it directs attention to contrary evidence that would not otherwise be considered” (Larrick, 2008, p. 8) However, thinking in the opposite direction can backfire as shown in two studies that attempted to debias hindsight by thinking about alternative outcomes. Counterfactual thoughts, which was perceived difficult, consistently increased the targeted hindsight bias instead of reducing it (Sanna, Schwarz, & Stocker, 2002). What we can learn from this is that considering the opposite is a good idea, but too much of is counterproductive.

A second general strategy could be called “*name your reasons*”. There is evidence that simply asking subjects for the reasons behind their judgments can help with debiasing hindsight and some type of framing effects (Arkes, Faust, Guilmette, & Hart, 1988; Miller & Fagley, 1991). Having to justify our decisions may therefore help us identify faulty thoughts.

Useful rules are what we ultimately need for debiasing to be practically successful. We want to find optimal debiasing strategies based on the biases present and break them down to a set of useful heuristics that help MCAs in decision situations. As Baron describes:

> Such heuristics may take the form of ‘words to the wise’ that we try to follow, such as ‘Make sure each paragraph has a topic sentence’ or (in algebra) ‘Make sure you know what is ‘given’ and what is ‘unknown’ before you try to solve a problem.’ In studying

> probability, one might learn the general rule 'All sequences of equally likely events are equally likely to occur.' (Baron, 2009)

Training in rules has been proven to be successful at least for simple rules, that subjects should remember. In the field of economics, Larrick and colleagues (1990) showed that besides other, students could be trained to ignore sunk costs in financial domains. Darrin Lehman & Richard Nisbett (1990) demonstrated that students of social sciences and psychology show improved reasoning about statistical problems after three years of coursework in this area, while showing no improvement in other domains. So, training may be an effective tool in future debiasing processes, because such simple recommendations, if derived correctly, can help MCA's become better decision makers.

The effort needed to overcome major cognitive biases may be larger than thought. Eric Schwitzgebel and Fiery Cushman (2015) examined the effects of framing and order of presentation on professional philosophers' judgements about two simple moral puzzles. Neither framing effects nor order effects appear to be reduced even by high levels of academic expertise, they conclude.

In sum, these are some relevant considerations when entering the debiasing process.

## 5.3 Example: The Reversal Test

For the status quo bias, a heuristic exists that seems to be transferable to metanormative decision-making. Nick Bostrom and Toby Ord (2006) proposed a simple self-check to eliminate status quo bias in applied ethics: the reversal test.

> When a proposal to change a certain parameter is thought to have bad overall consequences, consider a change to the same parameter in the opposite direction. If this is also thought to have bad overall consequences, then the onus is on those who reach these conclusions to explain why our position cannot be improved through changes to this parameter. If they are unable to do so, then we have reason to suspect that they suffer from status quo bias. (Bostrom & Ord, 2006, p. 10)

It stems from the idea that it is very unlikely that the parameter in question is in fact at its optimum. So, if one opposes to change in both directions, they might do so unreasonably.

That way, credence judgments could be checked. Imagine changing your credence distribution in direction x. If you are hesitant, imagine changing it in the opposite direction. If you judge this to be a bad adjustment, too, you need to explain why the current credence distribution is the best distribution you can reach. If this cannot be reasonably done, some adjustment would be good, after all, and the original views should be rethought.

The reversal test is one example of how practical advice for MCAs and all other decision-makers could look like. It rests on the assumption that a status quo bias exists in credence judgments, that debiasing is the overall best strategy as opposed to not debias and that the reversal test is a sufficient strategy to debias the error in its magnitude and direction.

# 6. Conclusion

This thesis was meant to support morally conscientious agents (MCA) in their project of avoiding moral wrongs under normative uncertainty. It was guided by the following question. How can we improve decision making under the possible influence of cognitive biases on metanormative judgments?

Currently, the most promising metanormative approach to minimize risks of wrongdoings requires MCAs to maximize expected choice-worthiness across options. In order to do that, decision-makers need to make judgments about the plausibility of certain first-order moral theories (credence judgments). I have made a simple but relevant point here. Cognitive biases exist in all classes of judgments. We should assume that they also distort our metanormative judgments. I identified a few pathways through which this could happen and focused on the influence of cognitive biases on credence judgments.

If I am right, MCAs need to limit biases in their thinking in order to reach their goal. I introduced a three-step approach towards improving decision-making, containing a normative, descriptive and prescriptive part. As a first step, I developed the basic framework of a normative model which serves to evaluate metanormative plausibility judgments (credence judgments). In the second step, I analysed descriptive studies for two biases in detail, the availability caused bias and the bias towards the status quo, showing for each that we have reasons to believe that they exist in plausibility judgments about first-order moral theories and if this is so, where they deviate from the normative model. In the third step, I suggested a framework for the future debiasing process as well as some relevant considerations which are useful even at this early stage in the debiasing process.

This thesis has most generally been an introduction to metanormative debiasing. To improve decision-making in the future, we need to be aware of the existence and robustness of cognitive biases across contexts and tasks. We need a more thorough understanding of how and where they influence judgments. Insights from moral psychology, especially the studies of moral reasoning and moral development can be helpful here. Only then can we develop and train fitting counteracting strategies.

Looking beyond the scope of this thesis, there remain additional hurdles which need to be overcome for decision-making under normative uncertainty to be practically successful. How do we translate choice-worthiness into judgments? How can humans with limited cognitive capacity make good evaluations about infinite possible credence distributions? How do we know what to do in a specific situation, even if we successfully figured out our credence distribution?

Even if the assumptions about normative uncertainty being relevant and MEC being correct changed, it would not make cognitive biases irrelevant. Whenever we talk about avoiding moral wrongs and we enter the practical landscape, we need to have a discussion about cognitive biases for this project to be successful. Therefore, this branch of interdisciplinary research deserves more attention. Whatever the path from here will look like, there is a lot to be learned from psychology when it comes to the practical application of meta-/normative principles.

# References

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, *110*(3), 486–498. https://doi.org/10.1037/0033-2909.110.3.486

Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology*, *73*(2), 305–307. https://doi.org/10.1037/0021-9010.73.2.305

Baron, J. (2009). *Thinking and Deciding* (4. ed., reprinted.). Cambridge: Cambridge Univ. Press. Retrieved from http://www.loc.gov/catdir/enhancements/fy0729/2007020449-b.html

Baron, J., & Ritov, I. (1994). Reference Points and Omission Bias. *Organizational Behavior and Human Decision Processes*, *59*(3), 475–498. https://doi.org/10.1006/obhd.1994.1070

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, *106*(2), 265–289. https://doi.org/10.1037/0033-2909.106.2.265

Bostrom, N., & Ord, T. (2006). The reversal test: Eliminating status quo bias in applied ethics. *Ethics*, *116*(4), 656–679. https://doi.org/10.1086/505233

Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *Journal of Abnormal Psychology*, *52*(3), 384–389. https://doi.org/10.1037/h0041006

Broome, J. (2013). *Rationality Through Reasoning* (1. Aufl.). *The Blackwell / Brown Lectures in Philosophy*. s.l.: Wiley-Blackwell. Retrieved from http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10719160 https://doi.org/10.1002/9781118609088

Crandall, C. S., Eidelman, S., Skitka, L. J., & Morgan, G. S. (2009). Status quo framing increases support for torture. *Social Influence*, *4*(1), 1–10. https://doi.org/10.1080/15534510802124397

Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 1: Origins of bias and theory of debiasing. *BMJ Quality & Safety*, *22 Suppl 2*, ii58-ii64. https://doi.org/10.1136/bmjqs-2012-001712

Eidelman, S., & Crandall, C. S. (2012). Bias in Favor of the Status Quo. *Social and Personality Psychology Compass*, *6*(3), 270–281. https://doi.org/10.1111/j.1751-9004.2012.00427.x

Eidelman, S., Crandall, C. S., & Pattershall, J. (2009). The existence bias. *Journal of Personality and Social Psychology*, *97*(5), 765–775. https://doi.org/10.1037/a0017058

Eidelman, S., Pattershall, J., & Crandall, C. S. (2010). Longer is better. *Journal of Experimental Social Psychology*, *46*(6), 993–998. https://doi.org/10.1016/j.jesp.2010.07.008

Fiedler, K. (1983). On the Testability of the Availability Heuristic. In R. W. Scholz (Ed.), *Advances in Psychology: Vol. 16. Decision making under uncertainty: Cognitive decision research, social interaction, development and epistemology* (2nd ed., Vol. 16, pp. 109–119). Amsterdam: North Holland. https://doi.org/10.1016/S0166-4115(08)62196-2

Fischhoff, B. (1982). *Debiasing/Kahneman, D., Slovic, P. and Tversky, A*. Retrieved from https://philpapers.org/rec/FISDDS

Gabrielcik, A., & Fazio, R. H. (1984). Priming and Frequency Estimation. *Personality and Social Psychology Bulletin*, *10*(1), 85–89. https://doi.org/10.1177/0146167284101009

Harrison, A. A. (1977). Mere Exposure. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology. Advances in experimental social psychology* (Vol. 10, pp. 39–83). New York: Academic Press. https://doi.org/10.1016/S0065-2601(08)60354-8

Hogarth, R. M. (2001). *Educating intuition*. Chicago: Univ. of Chicago Press. Retrieved from http://www.loc.gov/catdir/description/uchi052/2001027562.html

Jost, J., Banaji, M., & Nosek, B. (2004). A Decade of System Justification Theory: Accumulated Evidence of Conscious and Unconscious Bolstering of the Status Quo. *Political Psychology*, *25*(6), 881–919. https://doi.org/10.1111/j.1467-9221.2004.00402.x

Jost, J., Pelham, B., Sheldon, O., & Ni Sullivan, B. (2003). Social inequality and the reduction of ideological dissonance on behalf of the system: evidence of enhanced system justification among the disadvantaged. *European Journal of Social Psychology*, *33*(1), 13–36. https://doi.org/10.1002/ejsp.127

Kahneman, D. (2011). *Thinking, fast and slow* (1st ed.). New York: Farrar Straus and Giroux.

Kahneman, D., & Frederick, S. Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In *Gilovich, Griffin et al. (Hg.) 2002 – Heuristics and Biases* (Vol. 13, pp. 49–81). https://doi.org/10.1017/CBO9780511808098.004* (Original work published 2002).

Kahneman, D., & Frederick, S. (2012). Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases* (Vol. 13, pp. 49–81). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives, 5*(1), 193–206. https://doi.org/10.1257/jep.5.1.193

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty* (pp. 201–208). Cambridge University Press. https://doi.org/10.1017/CBO9780511809477.015

Kempf, A., & Ruenzi, S. (2006). Status Quo Bias and the Number of Alternatives: An Empirical Illustration from the Mutual Fund Industry. *Journal of Behavioral Finance*, *7*(4), 204–213. https://doi.org/10.1207/s15427579jpfm0704_3

Kolodny, N., & Brunero, J. (2013). *Instrumental Rationality*. Retrieved from https://plato.stanford.edu/entries/rationality-instrumental/

Larrick, R. P. (2008). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 316–338). Malden, MA: Wiley Interscience. https://doi.org/10.1002/9780470752937.ch16

Larrick, R. P., Morgan, J. N., & Nisbett, R. E. (1990). Teaching the Use of Cost-Benefit Reasoning in Everyday Life. *Psychological Science*, *1*(6), 362–370. https://doi.org/10.1111/j.1467-9280.1990.tb00243.x

Lehman, D. R., & Nisbett, R. E. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology*, *26*(6), 952–960. https://doi.org/10.1037/0012-1649.26.6.952

Lerner, M. J. (1980). The Belief in a Just World. In M. J. Lerner (Ed.), *The Belief in a Just World: A Fundamental Delusion* (pp. 9–30). Boston, MA: Springer US; Imprint: Springer. https://doi.org/10.1007/978-1-4899-0448-5_2

MacAskill, W. (2014). Normative Uncertainty.

MacAskill, W. (2019). Practical Ethics Given Moral Uncertainty. *Utilitas*, *31*(3), 231–245. https://doi.org/10.1017/S0953820819000013

MacAskill, W., Bykvist, K., & Ord, T. (2020). *Moral Uncertainty*. [S.l.]: Oxford University Press.

MacAskill, W., & Ord, T. (2018). Why Maximize Expected Choice-Worthiness? *Noûs*, *46.6*(1), 898. https://doi.org/10.1111/nous.12264

MacLeod, C., & Campbell, L. (1992). Memory accessibility and probability judgments: An experimental evaluation of the availability heuristic. *Journal of Personality and Social Psychology*, *63*(6), 890–902. https://doi.org/10.1037/0022-3514.63.6.890

Manis, M., Shedler, J., Jonides, J., & Nelson, T. E. (1993). Availability heuristic in judgments of set size and frequency of occurrence. *Journal of Personality and Social Psychology*, *65*(3), 448–457. https://doi.org/10.1037/0022-3514.65.3.448

Mehle, T., Gettys, C. F., Manning, C., Baca, S., & Fisher, S. (1981). The availability explanation of excessive plausibility assessments. *Acta Psychologica*, *49*(2), 127–140. https://doi.org/10.1016/0001-6918(81)90024-X

Miller, P. M., & Fagley, N. S. (1991). The Effects of Framing, Problem Variations, and Providing Rationale on Choice. *Personality and Social Psychology Bulletin.* Advance online publication. https://doi.org/10.1177/0146167291175006

Murphy, P. (2020, February 27). Coherentism in Epistemology. Retrieved from https://www.iep.utm.edu/coherent/

Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the Inevitable Anchoring Effect: Considering the Opposite Compensates for Selective Accessibility. *Personality and*

*Social Psychology Bulletin.* Advance online publication. https://doi.org/10.1177/01461672002611010

Nicolle, A., Fleming, S. M., Bach, D. R., Driver, J., & Dolan, R. J. (2011). A regret-induced status quo bias. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 31*(9), 3320–3327. https://doi.org/10.1523/JNEUROSCI.5615-10.2011

Pohl, R. F., & Hell, W. (1996). No Reduction in Hindsight Bias after Complete Information and Repeated Testing. *Organizational Behavior and Human Decision Processes*, *67*(1), 49–58. https://doi.org/10.1006/obhd.1996.0064

Quattrone, G. A., Lawrence, C. P., Finkel, S. E., & Andrus, D. C. (1984). *Explorations in anchoring: The effects of prior range, anchor extremity, and suggestive hints*. Manuscript. Retrieved from https://scholar.google.com/citations?user=qsgf4qpk9vic&hl=de&oi=sra

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*(1), 7–59. https://doi.org/10.1007/BF00055564

Sanna, L. J., Schwarz, N., & Stocker, S. L. (2002). When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 497–502. https://doi.org/10.1037/0278-7393.28.3.497

Schwarz, N., Bless, H., Strack, F., Klumpp, G., & al, e. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, *61*(2), 195–202. https://doi.org/10.1037/0022-3514.61.2.195

Schwarz, N., & Vaughn, L. A. (2012). The Availability Heuristic Revisited: Ease of Recall and Content of Recall as Distinct Sources of Information. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases* (Vol. 13, pp. 103–119). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.007

Schwitzgebel, E., & Cushman, F. (2015). Philosophers' biased judgments persist despite training, expertise and reflection. *Cognition*, *141*, 127–137. https://doi.org/10.1016/j.cognition.2015.04.015

Shedler, J., & Manis, M. (1986). Can the availability heuristic explain vividness effects? *Journal of Personality and Social Psychology*, *51*(1), 26–36. https://doi.org/10.1037/0022-3514.51.1.26

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, *27*(1), 76–105. https://doi.org/10.1016/0022-1031(91)90011-t

Steup, M., & Neta, R. (2005). *Epistemology*. Retrieved from https://plato.stanford.edu/entries/epistemology/

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Wason, P. C., & Evans, J.S.B.T. (1974). Dual processes in reasoning? *Cognition*, *3*(2), 141–154. https://doi.org/10.1016/0010-0277(74)90017-1

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

# Eidesstattliche Erklärung zur Bachelorarbeit

*Silvana Hultsch, Matrikelnummer: 1530770*

Ich erkläre ausdrücklich, dass ich die von mir eingereichte Bachelorarbeit mit dem Titel „Avoiding Moral Wrongs under Normative Uncertainty – Biases Involved in Maximizing Expected Choice-Worthiness“, selbständig angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle wörtlichen oder sinngemäß übernommenen Textstellen als solche kenntlich gemacht habe. Diese Arbeit wurde nicht bereits zur Erreichung eines anderen akademischen Grades eingereicht oder veröffentlicht.

Datum, Ort

Unterschrift