Univerzita Karlova

Pedagogická fakulta

Katedra anglického jazyka a literatury

BAKALÁŘSKÁ PRÁCE

Téma racionality v díle Harry Potter a metody racionality

The Theme of Rationality as Depicted by Yudkowsky in Harry Potter and the Methods of Rationality

Kateřina Macků

Vedoucí práce:	PhDr. Tereza Topolovská, Ph.D.
Studijní program:	Specializace v pedagogice
Studijní obor:	AJ-IT

2023

Declaration

I hereby declare that this bachelor thesis, titled „The Theme of Rationality as Depicted by Yudkowsky in Harry Potter and the Methods of Rationality“, supervised by PhDr. Tereza Topolovská, Ph.D. is a result of my own work and research cited in the „Sources used“ section. I further declare that this bachelor's thesis has not been used to attain any other degree.

Prague, July 8 2023

I would like to extend my gratitude to PhDr. Tereza Topolovská, Ph.D. for her invaluable assistance in helping me narrow down the topic of my thesis and unwavering support throughout the writing process. Her expertise and guidance have been invaluable in shaping the direction of my work and her commitment to providing an environment for me to explore my own ideas is deeply appreciated. I would also like to thank my family and friends for their emotional support and encouragement they selflessly provided throughout the writing process and in general as well.

ABSTRAKT

Cílem této bakalářské práce je prozkoumat téma racionality v fanfikci Eliezera Yudkowského *Harry Potter a metody racionality* (2010) se zaměřením na autorovo ztvárnění tohoto tématu a jeho potenciálního dopadu na čtenáře. Teoretická část poskytuje úvod do témata fanfikce a jejího přínosu pro literaturu, zkoumá autorův vztah k racionalismu a dále se zaměřuje na témata Bayesovy věty, bayesovského usuzování a kognitivních zkreslení. Praktická část pak staví na konceptech v teoretické části a demonstruje, jak Yudkowsky vyobrazuje racionalitu jako hlavní téma fanfikce a jak tyto koncepty ovlivňují osobnost a činy Harryho Pottera a dalších postav. Praktická část též názorně ukazuje výskyt racionalistických konceptů napříč dějem a hodnotí, jak téma racionality přidává ději na zajímavosti a vede k jeho rozvoji. Na základě analýzy důležitých momentů a charakterizace postav pak práce posuzuje účinnost Yudkowského propagace racionality a způsob, jakým je racionalita ztvárněna.

KLÍČOVÁ SLOVA

Eliezer Yudkowsky, racionalita, fanfikce, kognitivní zkreslení, Harry Potter a metody racionality

ABSTRACT

This thesis aims to explore the theme of rationality as it is developed in Eliezer Yudkowsky's fan fiction series *Harry Potter and the Methods of Rationality* (2010). The theoretical part provides an introduction to the concept of fan fiction and studies its position within the contemporary literary context, introduces Eliezer Yudkowsky and his work, explores his background as a rationalist, and focuses on the topics of Bayes' Theorem, Bayesian reasoning and cognitive biases, which are incorporated into *Harry Potter and the Methods of Rationality*. The practical part then draws upon the concepts established in the theoretical part and exemplifies Yudkowsky's use of rationalist concepts, specifically as a central theme of the narrative and how it shapes the character and actions of Harry Potter and other characters. It demonstrates Yudkowsky's use of rational concepts and also evaluates how the theme of rationality adds intrigue to the plot and drives it forward. By examining key moments and characterisation throughout the narrative, the thesis then assesses the effectiveness of Yudkowsky's advocacy for rationality and the way in which rationality is portrayed.

KEYWORDS

Eliezer Yudkowsky, rationality, fan fiction, cognitive biases, Harry Potter and the Methods of Rationality

Table of Contents

1 Introduction

The *Harry Potter* books and movies are an undeniably iconic set of stories originally written by J. K. Rowling - from the year 1997, many people across the world have been taking interest in the story of young Harry Potter, an orphan living with his troubled relatives, who one day finds out he is a wizard and is swept up in a world full of magic and mystery. This story is very popular to this day, as many people can relate to the desire of one day finding out they harbour hidden powers that enrich their everyday lives and allow them to embark on exciting adventures. The *Harry Potter* series has shaped my childhood[1], as I strongly believe it has done to many others. Consequently, many fans of Wizarding Britain have since wondered how the story would change if some of its aspects were to be altered, or even erased entirely. This is where fan fiction finds its footing within the Harry Potter fandom, fan fiction authors trying to answer many "what if" questions and reimagining the stories to explore unexplored aspects or even insert their own. Fan fiction unfortunately faces a bad reputation due to its perception as disrespectful or derivative. However, having been a fan of the *Harry Potter* world as well as a part of the fan fiction community for many years, I have always had interest in exploring both phenomena.

After having considered many different examples of fan fiction as the subject of my thesis, browsing Internet forums, and consulting other fan fiction enthusiasts, I ultimately settled on the 2010 fan fiction *Harry Potter and the Methods of Rationality* by Eliezer Yudkowsky. This work stands out even amongst the *Harry Potter* fandom, as it is an enthralling work which delves into the topic of rationalism and rationality within the *Harry Potter* universe. It changes the character of Harry Potter and others around him, equipping them with an insatiable want for wisdom and the knowledge of rationality. Furthermore, Eliezer Yudkowsky uses the story of *Harry Potter* to advocate for rationality in an unconventional way, by discussing aspects of rationality within the story and showcasing the power it holds when one knows how to wield it.

This thesis attempts to analyse the intricacies of *Harry Potter and the Methods of Rationality*, investigating its portrayal of rationality and its influence on the plot and the

[1] I also feel it pertinent to mention that I, the author of this thesis, denounce the words and opinions of J. K. Rowling, specifically those regarding the LGBTQIA+ community.

readers. I wanted to evaluate how effectively Yudkowsky advocates for rationality and how approachable his explanations of rational aspects were within the context of fan fiction. In order to do so, I have studied Yudkowsky's rationalist forum *LessWrong* to gain an insight into the topic, the results of which I then compared academic monographs such as Russel Hardin's *Rationalism* (1998) and William Bristow's *Enlightenment* (2017) to understand the appeal of rationalism. Furthermore, I delved into the fan fiction itself, searching for instances of the use of rationality as a motif or a theme in order to see how the significant changes to characterization (and various plot points) affected the story overall and how straightforward the explanations or uses of specific rational concepts are.

The theoretical part of the thesis introduces the concept of fan fiction, while also characterising the type of fan fiction, *Harry Potter and the Methods of Rationality*, the subject of this thesis, is. Furthermore, the author of *Harry Potter and the Methods of Rationality* is introduced, along with his popular rationalist forum *LessWrong* and other works, to showcase the author's interest in rationality and rationalism. Consequently, I focused on Bayes' theorem and Bayesian reasoning, a concept which Yudkowsky often likens to the basis of rationality. Additionally, cognitive biases were introduced as a topic of interest, as they fundamentally hinder our perception of the world around us and many types of cognitive biases are repeatedly mentioned within the story. There were two biases, which are described in detail – confirmation bias and anchoring bias, as the first is mentioned explicitly and the other implicitly. These two specific biases are of interest due to their nature as sources of conflict and hardship in Harry James Potter-Evans-Verres' aspiration to rationalism.

The practical part seeks to establish rationality as a central theme within *Harry Potter and the Methods of Rationality*, at first as a theme and leitmotif and then as exemplified through the thoughts and actions of the main character of *Harry Potter and the Methods of Rationality*, Harry James Potter-Evans-Verres. Furthermore, I strive to showcase specific instances of the use of Bayes' theorem, Bayesian reasoning, confirmation bias and anchoring bias, explaining the context and demonstrate Yudkowsky's choice of language to be in favour of rationality and of an approbatory nature. The aim of the practical part is to analyse the different ways in which these concepts operate in the story, the ways they are portrayed and the ways they influence the plot and the reader’s experience.

2 Theoretical part

2.1 Fan Fiction and its Contributions to Literature

Fan fiction, often abbreviated to *fanfic*, is the literary genre comprising stories produced by fans of specific pieces of work using the pre-established plotlines of the source text, which is referred to as "canon" (Thomas 1) among the fan fiction community. This phenomenon has taken the world by storm and has gained popularity, mostly due to the connective nature of the digital age. For the sake of establishing the significance of this literary form, this thesis will work with Rebecca Tushnet's definition of fan fiction, which establishes fan fiction to be "any kind of written creativity that is based on an identifiable segment of popular culture, such as a television show and is not produced as 'professional' writing" (qtd. in Lipton 435). This would include any work that is written by a fan based on a "canon", and which expands upon the original (Dawson 13).

Fan fiction is often published on a chapter-by-chapter basis, traditionally on online websites such as *FanFiction.Net*, where readers have the option to easily track whenever new chapters of their favourite fan fiction get published (Thomas 9). As such, specific instances of fan fiction are often labelled as *Work in Progress*, which is a descriptor within the fan fiction sphere that denotes that such a piece of literature is currently being worked on and not yet completed (Busse, Hellekson). This is particularly attractive to readers, since it opens the communication channel between the writer and the readers, making the experience overall transformative for all parties involved, as the writer is usually an active participant within the community around their own work and has the option to be receptive to their reader's feedback.

Consequently, many "fanfics", although they are published under a single name, could not exist without the support and feedback of the community (Busse, Hellekson). For example, such is the case for a type of fan fiction generally tagged as "prompt", which, as the name suggests, is a type of fan fiction written based on a short prompt. These prompts range in theme as well as source, as they can be either provided by the author or the community. A prompt in the context of fan fiction is a scenario, or theme that serves as inspiration to the writer to develop fan fiction stories. A prompt can be sentence, a paragraph or even a simple collocation. Prompt fan fiction underlines the collaborative nature of fan fiction, as authors often come together to share and respond to prompts, often even creating community-wide events, such as the June's Prompts Challenge. This challenge is as of writing this thesis being held on the *Reddit* website, specifically on the *r/FanFiction* subreddit, and is loosely themed around Pride month. The writers are asked to publish their works in shared spaces, so that others have the opportunity to witness their interpretations. This collaborative aspect cultivates a sense of community and encourages writers to explore all kinds of different perspectives within the established universes of the original works.

Another notable example of direct communication between the writer and their audience is the case of *The Life and Times* (2009), a well-known Marauders Era[2] *Harry Potter* fan fiction published on the website *FanFiction.Net* by Jewels5. Author's notes (A/N) are included at the beginning and end of each chapter, in which Jewels5 directly replies to reviewers, answering their questions and expressing gratitude for the reader feedback. This proves that fan fiction has facilitated a community of readers and writers that offers a supportive environment for writers to share their work, receive feedback, and develop their skills as well as for readers to share their opinions with like-minded fans and the author.

Yet another merit of fan fiction in general is the expansion of universes and characters of existing works, allowing writers to explore new possibilities and expand on the source material in ways they wish to see themselves. One such facet would be the option for the representation of minorities. According to Abigail Derecho, fan fiction is "a genre that has a long history of appealing to women and minorities, individuals on the cultural margins who used archontic[3] writing as a means to express not only their narrative but their criticism of social and political inequities as well" (*Archontic Literature* 876). Fan fiction offers endless possibilities for exploring "what if" scenarios, including the representation of diverse characters and perspectives. Such scenarios include the popular theme of alternate timelines and romantic relationships between characters, but it also provides opportunities to challenge and expand the representation of diverse identities, including characters of different

[2] Marauder-era refers to an era within the *Harry Potter* universe during the school years of Harry's father James Potter. The main characters are traditionally James, Sirius Black, Remus Lupin, Peter Pettigrew, and Lily Evans (later known as Lily Potter).

[3] *Archontic* relates to the word *archive*, and I take it from Jacques Derrida's 1995 work *Archive Fever*, in which Derrida claims that any and every archive remains forever open to new entries, new artifacts, new contents (Derecho 890).

ethnicities and genders. An example of this is a fan fiction series based on *Harry Potter* (1997) called "Through the Quiet Emerald Eyes" by authors alwayslily22 and Des98 on Archive Of Our Own (AO3), an ongoing series of works that retells the story of Harry Potter with significant changes to the main characters. Hermione is a young black girl while Harry is of Iranian descent and deaf as a result of the abuse he faced (and continues to face) from the Dursleys. These characteristics are intrinsically important to the story and, as such, the fan fiction does not stray from discussing bigotry, racism, sexuality, and other contemporary topics. In this regard, fan fiction contributes to literature by promoting diversity and inclusivity by means of providing representation for marginalized communities and offering a space for them to express and celebrate their identity free from discrimination, while also making the canon and the mainstream their own. Although fan fiction does not have the same reach in society as traditional mainstream media, it still shapes the literary landscape and potentially influences the reader's view of different identities and life experiences.

2.1.1 Rationalist Fiction

The terms Rational Fic, Rationalist Fic and Rationalist Fiction are often used interchangeably, but there is a distinct difference as established by their largely overlapping communities (*TV Tropes)*. For example, Rational fan fiction is a recently developed subtype of fan fiction, which applies aspects of rational fiction to the literary genre of fan fiction. Rational fiction usually demonstrates at least some level of rationality, such as refusing to create specific situations or characters simply because "the plot requires it" (*Goodreads*). All elements must have a plausible explanation for their existence. More characteristics of rational fiction include emphasis on intelligent characters using their knowledge and resources in ingenious ways to solve problems, the story's climax features a satisfying and intelligent solution to its problems, and others. Although the terms *rational fiction* and *rationalist fiction* are often used in place of one another, it is generally understood that rationalist fan fiction is a subtype of rational fan fiction. Rationalist fan fiction exhibits many characteristics identical to rational fan fiction, but it also puts significant emphasis on the characters aspiring to rationalism by attempting to improve their reasoning abilities and trying to teach the reader certain aspects of rationality (*Reddit*). In the words of Eliezer

Yudkowsky, the originator of rationalist fan fiction, “a rationalist!hero[4] should excel by *thinking* – moreover, thinking in understandable patterns that readers can, in principle, adopt for themselves”. Such rationalist!heroes traditionally adhere to rational beliefs, through which they often question the logical inconsistencies and plot devices of the story they are a part of, frequently conducting experiments to apply science to non-scientific phenomena. As rationalist fan fiction traditionally deals with complex concepts and reasoning, it is generally more appealing to older readers who can grasp and appreciate such nuances. The perceived initiator of rational and rationalist fan fiction is Eliezer Yudkowsky’s *Harry Potter and the Methods of Rationality* (2010), which is the subject of this thesis. This fan fiction is a lengthy retelling of J. K. Rowling’s *Harry Potter* series where Bayes’ theorem, Bayesian reasoning, and cognitive biases (all important aspects of Yudkowsky’s rationality) play a central role. All these aspects will be further expanded upon in this thesis, as they shape the story and add intrigue to the plot, while also engaging the reader in an intellectually challenging exploration of familiar themes and characters.

2.2 Rationality as a philosophy

Rationality has lengthy roots in philosophy dating back to the 17th century, during which the rationalism philosophical movement gathered significant momentum due to the work of many rationalist figures, such as Descartes, Spinoza, or Leibniz (Vanzo 253). Rationalism has many subtypes depending on how they perceive reason, however within this thesis, rationalism will be understood as employing reason (a.k.a. our intellectual abilities) in order to search for evidence both in accordance and also contrary to potential beliefs (Hardin 75). When mentioning this version of rationalism, it is pertinent to mention 18th century’s Enlightenment, as the beliefs of this movement directly draw from rationalism. The Enlightenment also plays a role within the *Harry Potter and the Methods of Rationality* fan fiction, as Harry is explicitly introduced as “armed with Enlightenment ideals and the experimental spirit” (*LessWrong*) in the introduction to the work. The Enlightenment was a philosophical movement most prominent in France during the 18th century, but to Enlightenment thinkers, The Enlightenment cannot be confined to a specific historical era,

[4] This is an example of a descriptor of a fan fiction, specifically about the characters therein. This descriptor often takes the form „characteristic!character“.

to them, as it is considered a continuous process of social, psychological, or spiritual development that transcends time and space (Bristow). Enlightenment does not suggest *what* to think, but rather *how* to think, which can be demonstrated by the different beliefs held by Enlightenment thinkers. Enlightened thinking highlights mental autonomy and using one's mental capabilities to determine beliefs and actions. Along with this confidence in people's cognitive abilities, Enlightened thinkers also profess deep scepticism toward other sources of authority, such as tradition, superstition etc., as those are deemed in opposition to the authority of individual reason and experience (Bristow).

Rationality and rationalism are inherently related concepts; however, they do have different meanings. Rationalism, as mentioned above, is a philosophical position, which claims that certain truths can be discovered through reason along with our sensory experience. Rationality, to Yudkowsky, is a two-part statement; Epistemic rationality is the process of "systematically improving the accuracy of your beliefs" and instrumental rationality helps one to systematically achieve one's values. With these two definitions, Yudkowsky fittingly summarizes rationality as "forming true beliefs and making decisions that help you win" ("What Do We Mean By "Rationality"?"). Yudkowsky further refers to rationality as "the martial art of mind". He claims that as long as one has a hand with correctly positioned tendons and muscles, one can "learn to make a fist". This is to say, that if one has a correctly formed brain, one can learn rationality ("The Martial Art of Rationality").

2.2.1 Bayes' theorem and Bayesian reasoning

As Bayes' theorem and Bayesian reasoning play a vital role within the subject of this thesis and within rationality in general, the author of this thesis finds it suitable to establish these concepts to an appropriate level. Bayes' theorem is a mathematical formula often used in calculating conditional probabilities (Joyce). James Joyce defines Bayes' theorem as "The probability of H condition on E is defined as $\mathbf{P}_E(H) = \mathbf{P}(H \;\&\; E)/\mathbf{P}(E)$, provided that both terms of this ratio exist and $\mathbf{P}(E) > 0.1$". Although this formula might seem impenetrable at first glance, it is referred to by Efron as "an algorithm combining prior evidence with current evidence". It is no wonder this algorithm is often used in the field of medicine, which is where the real-life example which will be used to help understand this theorem comes from. In "Bayes' Theorem in the 21st Century," Bradley Efron writes:

> A physicist couple I know learned, from sonograms, that they were due to be parents of twin boys. They wondered what the probability was that their twins would be identical rather than fraternal. There are two pieces of relevant evidence. One-third of twins are identical; on the other hand, identical twins are twice as likely to yield twin boy sonograms, because they are always same-sex, whereas the likelihood of fraternal twins being same-sex is 50:50. Putting this together, Bayes' rule correctly concludes that the two pieces balance out, and that the odds of the twins being identical are even.

In this instance, the prior experience is the fact that one-third of twins are identical, while the current evidence is the sonogram. Referring to the aforementioned formula, $\mathbf{P}_E(H)$ is the final probability that the odds of the twins being identical are even, P(H) is the probability that the twins are identical, P(E) is the probability of observing twin boy sonograms, and $\mathbf{P}(H \,\&\, E)$ is the probability that the twins are identical AND the sonogram shows twin boys.

Bayes' theorem is the basis for Bayesian reasoning and serves as a useful framework for understanding rational decision-making, which is a fundamental aspect of rationality.

2.2.2 Cognitive Biases and Heuristics

Cognitive biases and heuristics are key topics when studying human rationality and rational decision-making in the realm of cognitive science[5], specifically cognitive psychology. These biases and heuristics often affect the ability to make rational decisions, which leads to errors in judgment that are not in accordance with available evidence. In the realm of rational fan fiction, biases and heuristics play a vital role in how characters and events are portrayed, as well as how readers respond to these portrayals. The subject of this thesis has made cognitive biases and heuristics one of the central themes in trying to teach

[5] Cognitive science is an interdisciplinary scientific investigation of the mind and intelligence. It integrates principles from psychology, linguistics, philosophy, computer science, artificial intelligence, neuroscience, and anthropology. (*Encyclopaedia Britannica)*

the readers how to recognize these phenomena and efficiently mitigate them when aiming for rational decision-making.

Heuristics, also known as judgmental rules, are utilized to simplify difficult mental tasks in order to speed up decision-making (H. A. Simon qtd. in Barnes 129) They can help people come up with reasonably accurate judgments with limited time and effort. Unfortunately, while useful in certain situations, heuristics can lead to persistent cognitive biases, and although they tend to be accurate most of the time, awareness of their existence is pertinent in terms of rational decision-making. Heuristics can be further divided by their type, for example the availability heuristic (the more is a specific piece of information available, the more it is perceived as likely to be true) or the representativeness heuristic (the assumption that something or someone belongs to a specific group due to one's idea of what characteristics this group usually consists of) (Ehlinger et al.).

Cognitive biases in decision-making encompass a broad range of deviations from what is commonly considered purely rational judgment and decisions (Ehrlinger et al.). An important aspect of cognitive bias is the reality that it is not a random error caused by our ignorance, but a *systematic error in how we think.* This means that, in a way, cognitive bias causes us to perceive reality inaccurately. Cognitive bias is a prevalent issue in scenarios ranging from everyday tasks to events of great significance. An example of such an occurrence is the general wariness that takes hold of society after a major disastrous event, such as the crash of an airplane, or a large-scale car collision. This societal behaviour is to an extent the case of hindsight bias. Not only does the incident become a vivid example of the potential dangers of plane or car travel, but hindsight bias directly contributes to this as many people retroactively view the incident as preventable or assume that there were obvious signs pertaining to the impeding accident. This is suggested to cause issues in terms of analysing incidents, crashes, collisions, or disasters (Murata et al. 45).

Confirmation Bias

According to Barbara Koslowski and Mariano Maqueda, psychologists generally agree that people do not explicitly seek to have their minds changed or to be wrong. One of the only times someone might be willing to consider evidence disproving their stance is when it is presented to them explicitly (Koslowski, Maqueda 104). This is an instance of confirmation bias, a universally observed bias which is prevalent across all social circles and as a result has been well documented. Confirmation bias is also among the hurdles presented to humans when intending to make rational decisions, as individuals tend to seek out, gather, or recall information which is in accordance with their pre-existing beliefs while often disregarding evidence of the contrary, which is an on-going theme within the subject of this thesis (Oswald, Grosjean 79). This pattern of behaviour leads to distorted perception of reality which people draw evidence from, thus causing further problems with rational thinking.

Confirmation bias is a commonly occurring cognitive bias. A situation which many are familiar with will be used to exemplify it. A manager of a coffee shop strongly believes in the correlation of "hard work" and prosperity. As a result of this, they attribute the problem of gradually declining sales to their staff's perceived lack of effort – particularly the fact that the staff takes lunch breaks which are longer than intended. Therefore, the manager employs threats of disciplinary action and extended work hours to attempt to mitigate this issue. This leads to an unfavourable outcome where sales do not increase, but labour costs do.

Seeking guidance among other coffee shop managers, the manager is told that the decline in sales is most likely due to the coffee shop's new, less visible location. Despite this, the manager's confirmation bias – preferring information which aligns with their pre-existing beliefs – had caused them to falsely attribute the problem to the staff of their establishment while subconsciously dismissing evidence in favour of the highly probable issue: the coffee shop's unfavourable location (*The Decision Lab*).

Anchoring Bias

"When people make judgments or estimates about an uncertain situation they tend to rely on initial, salient values, impressions, or pieces of information – often called "anchors"" (de Wilde et al.) Anchoring bias is a judgmental bias which occasionally occurs as a result of the anchoring heuristic. To reword the previous quote, individuals tend to rely heavily on the first piece of information they encountered (the "anchor") when making judgments or estimates. The initial piece of information serves as a reference point which influences any following decisions, even if it is unrelated to the specific context.

A significant exemplification of the anchoring bias by an experiment was realized by Tversky & Kahneman in 1974. This experiment consisted of two groups of high-school students, which were both given two different mathematical problems they were to approximate a result to within 5 seconds. The first group was asked to estimate the answer to 8 x 7 x 6 x 4 x 5 x 3 x 2 x 1, while the second group was instructed to do the same for 1 x 2 x 3 x 4 x 5 x 6 x 7 x 8. These equations mathematically result in the same number, but the first group of participants reportedly had a median result of 2 250, while the group given the latter equation reported a median result of 512.

These results were in accordance with Tversky & Kahneman's predictions, as they foresaw that the participants would likely multiply the first few numbers in the equation and then estimate a result based on this initial number (1128). The results of this experiment revealed a considerable disparity between the two groups, which further underscored the existence and importance of the anchoring bias.

2.3 *Harry Potter and the Methods of Rationality*

Harry Potter and the Methods of Rationality is a completed fan fiction written by Eliezer Yudkowsky and originally published on *FanFiction.Net*. During the time of its gradual chapter-based publishing, which began on the 28th of February 2010 and finished on the 14th of March 2015, it grew immensely popular and garnered many reviews due to its interesting and subversive premise. It offers an alternative narrative to J. K. Rowling's original *Harry Potter* series, inserting elements of rationality, critical thinking, and the scientific method into the world of magic portrayed in the world of *Harry Potter.* Its final version, containing just around 660,000 words and 122 chapters and also offering a glimpse into the author's mind through author's notes, has garnered attention and acclaim for its thought-provoking integration of rationality concepts, engaging readers in an intellectual exploration of familiar characters and themes within a context filled with scientific inquiry and rational analysis.

Nevertheless, the author of this fan fiction acknowledges the potentially polarizing nature of the work, as it explicitly aims to teach its readers about intellectually challenging concepts. Unlike conventional fan fiction, which primarily intends to entertain, *Harry Potter and the Methods of Rationality* can be perceived as work akin to a humorous textbook of rationalist principles. To illustrate this point, the second line of the first chapter contains a fitting piece of advice from the author: “This fic is widely considered to have really hit its stride starting around Chapter 5. If you still don’t like it after chapter 10, give up.” (Yudkowsky 1)

The main premise of *HPMOR*[6] revolves around a significant alteration in the life of Harry Potter. In this AU[7] Harry’s aunt marries an Oxford professor and biochemist Michael Verres, who goes on to become Harry’s stepfather. Due to this change in the family dynamic, he is raised in a severely scientific and academically challenging environment, and as a result acquires a level of rationality equivalent to that of Yudkowsky’s at 18 years old (*LessWrong*). The story then further delves into Harry’s first amusing encounter with magic, where Professor Minerva McGonagall transforms into a cat (breaking the principle of conservation of energy), continuing with his studies at Hogwarts and his constant attempts at applying science and rationality to magic.

During Harry’s first year at Hogwarts School of Witchcraft and Wizardry (which spans the entirety of the fan fiction), he, along with his allies Hermione Granger and Draco Malfoy, faces numerous challenges. They partake in simulated battles in Defence Against the Dark Arts while Harry creates a new form of Transfiguration, devises an improved version of the Patronus charm capable of destroying Dementors, frees a criminal from Azkaban with the help of his mentor Quirinius Quirrell, and navigates the complexities of relationships with his peers and professors. All is seemingly lost when Hermione Granger, Harry’s best friend, gets murdered, but Harry prevails, trying to come up with a way to resurrect her. In a climactic moment, Harry’s mentor Professor Quirinius Quirrell is revealed to be possessed by Lord Voldemort and has been the mastermind behind much of the misfortune which befell Harry. Lord Voldemort is seeking to take the Resurrection stone the nature of which is revealed to be its capability of making any Transfiguration permanent. Furthermore, the secret behind Harry’s “dark side” comes to light, as on the night that Lord Voldemort killed Harry’s biological parents, he attempted to shape young Harry’s soul after his own, consequently imprinting his own thoughts onto the child. Ultimately, Harry

[6] *HPMOR* – an acronym for *Harry Potter and the Methods of Rationality*

[7] Alternate Universe – a descriptor in the realm of fanworks which denotes work which modifies one or more aspects of the established canon

succeeds in capturing Lord Voldemort, resurrecting Hermione Granger, and returning to school victorious and determined to further test the boundaries of magic and eventually conquer death itself. In essence, the storyline loosely follows the story of *Harry Potter and the Sorcerer's Stone* (1997), often adopting key shifts in narrative but coming to a much different but still rewarding payoff.

The central theme of *HPMOR* is rationality and critical thinking. Harry's high intellect, Enlightenment ideals, and rationalist state of mind offer intriguing approaches to certain challenges within the magical world and provide a unique perspective. The pursuit of rationality affects not only Harry's personal growth and that of the other main characters (namely Hermione Granger and Draco Malfoy), but also exposes and questions fixed ideologies, such as the wizarding justice system or the perceived limitations of magic. Scientific inquiry too occurs quite often in the fan fiction in order to gather evidence by posing a question, forming a hypothesis and conducting experiments. One such explicit example of scientific inquiry will be mentioned in the following chapters, as Harry uses the scientific method to awaken Draco Malfoy's scientific potential. To summarize, *Harry Potter and the Methods of Rationality* presents a compelling alternative universe narrative for mature readers who wish to be captivated by the unique combination of magic, rationality, and intellectual exploration.

2.3.1 Eliezer Yudkowsky

Eliezer Yudkowsky is a decision theorist born in the United States of America in 1979. Although he is a self-educated scholar with no formal secondary education, Yudkowsky has published many academic articles and has numerous friends in the academic field. He also co-founded and assumes a leadership role in research at the Machine Intelligence Research Institute, a non-profit organization which conducts research to safeguard artificial intelligence and to ensure it will be beneficial to humanity. Upon visiting his personal website (www.yudkowsky.net), one is greeted with a fitting title: "Welcome to the website of a man who wears more than one hat". As can be discerned from his rationalist writings and involvement with the Machine Intelligence Research Institute, Yudkowsky's main areas of interest lie within two distinct realms: rationality in the realm of cognitive science and artificial intelligence. These areas have as many similarities as they have differences, however, due to the topic of this thesis, the primary focus will be kept on Yudkowsky's work in connection to rationality. Yudkowsky began publishing his rational writings in the early 2000s, his most influential work being *Rationality: From AI to Zombies* (2015) which greatly contributed to the popularisation of rationality and sparked interest in rationalist concepts in various communities. Therefore, it can be said that while he draws upon scientific research and concepts, Yudkowsky's contribution in the realm of rationality and science is mostly the popularisation of the topics through his writing and online presence. Yudkowsky is the author of the subject of this thesis, which speaks to his rationalist beliefs and even briefly mentions issues regarding artificial intelligence.

LessWrong

LessWrong is an online forum, which is important to the contemporary rationalist community. It was first created by Eliezer Yudkowsky to share his rationalist ideas and to harbour discourse with like-minded individuals. When asked what exactly *LessWrong* is, this response can be found on the website's FAQ[8]: "LessWrong is a community dedicated to improving our reasoning and decision-making. We seek to hold true beliefs and to be effective at accomplishing our goals. More generally, we want to develop and practice the art of human rationality". Additionally, *LessWrong*, while offering a space for rational discourse, also contains many essays and articles published by Yudkowsky and other users which offer a wide range of rationalist topics, from the complete basics to advanced concepts. Regarding its history, *LessWrong* originated form the rationalist movement, which was briefly discussed in the previous chapters, and it was first established by Eliezer Yudkowsky in 2009. Its predecessor was a group blog called *Overcoming Bias*, helmed by Yudkowsky and others, which originated in 2006 and was mostly focused on making conscious attempts to "move one's beliefs closer to reality despite biases such as overconfidence or wishful thinking" (*LessWrong*). Currently *LessWrong* serves as a general hub for the public with interest in rationality. It is also heavily didactic, as it seeks to teach the reader about rationality and explain concepts which anyone can apply to their everyday life in order to start "thinking better". However, this emphasis on rationality and analytical thinking also allows for potential manipulation, as the influence of rational principles and

[8] FAQ = frequently asked question

their presentation as desirable and advantageous can be used to alter an individual's beliefs or actions. This, however, is partly mitigated by Yudkowsky's insistence that one should question everything, including rationality and his teachings. By maintaining a balanced perspective and employing critical thinking, one should be capable of gaining useful insight as well as recognize any manipulative aspects within rational discourse.

When asked about what sets *LessWrong* apart from other discussion forums, the author of the FAQ, an individual that engages within the forum under the name "Ruby" provided an answer comprising of three points, which assert *LessWrong* as a forum that upholds exceptionally high standards for discourse, but also maintains a receptive attitude towards unconventional ideas and achieves intellectual progress by the means of creating a space to share ideas and concepts among its members. Yudkowsky holds a central role within the community, given that his aforementioned work *Rationality: From AI to Zombies* assumes the role of being an essential reading for individuals seeking to be actively involved within the community. While the website's discussion requirements might seem elitist, the fundamental idea of *LessWrong* is to provide a platform for productive discussions about rationality, and to provide a place for progress to occur. On top of this, *LessWrong* contains many important rationality-related readings, and many of the topics mentioned in *HPMOR* can be found on the website to harbour further understanding for those interested learning more about what is mentioned in the fan fiction. However, the *LessWrong* community has also received the reputation of being cultish, mainly due to its strong ideologies, high level of devotion, sense of intellectual superiority, and the general feeling of exclusivity.

Author's Other Works

Other than his fan fiction *Harry Potter and the Methods of Rationality*, Yudkowsky published numerous other books and essays, which concern themselves with the topics of AI alignment and rationality.

Rationality: From AI to Zombies (2015), formerly known as *The Sequences*, is an e-book that explores the topics of human rationality and irrationality within the context of cognitive science. The e-book is organized into six books and encompasses 333 essays, further divided into 26 sequences. As of the writing of this thesis, two books have already been published in 2018 and are available in Kindle and paperback formats. *Rationality:*

From AI to Zombies serves as a comprehensive introduction to the core concepts and key ideas related to rationality, providing readers an in-depth understanding of rational thinking and its application in various circumstances. It is often considered Yudkowsky's most influential rationalist work. It is most appreciated for its approachable depiction of rationalist concepts and use of thought-provoking practical examples to convey complex rationalist ideas such as Occam's Razor, or the difference between scientific evidence, legal evidence, and rational evidence.

Yudkowsky's book, *Inadequate Equilibria: Where and How Civilizations Get Stuck* (2017) is a 120-page analysis of the age-old question: under what circumstances is it rationally sound to behave as though one knows better than the "leading experts" of society? And how does one deal with the criticism which usually follows? Yudkowsky asserts that being in opposition to the status quo is unlikely to succeed, but such a stance can be of merit when authoritative institutions display flawed motivation. The book further delves into the assessment of the effectiveness of certain societal goals and pursuits, while using the market as a tool to demonstrate and discuss the human capability of assessing areas where society may or may not excel. It discusses and examines the thought processes behind ambitious endeavours, such as creating a social media network more profitable than Facebook or thinking that one can easily create a new type of medical treatment. In essence, *Inadequate Equilibria* offers valuable reflections on personal choices and prompts readers to consider the limitations and potential of different institutions.

Yudkowsky's light novel *A Girl Corrupted by the Internet is the Summoned Hero?!* (2016) deserves mention as well. This concise 80-page light novel follows the story of a young girl who has been "corrupted" by the Internet (specifically by the ease of access to adult-themed material) as she is transported to a different fantasy world by a summoning ritual to defeat the Wicked Emperor[9]. The protagonist's meta-awareness of this widely popular story structure lends a humorous and fascinating reading experience, as she often calls attention to narrative devices and other storytelling techniques happening around her.

[9] This is an example of an *isekai* – a type of fantasy literature in which the protagonist gets displaced from an ordinary world into an alternative or unfamiliar world. This unfamiliar world often upholds different rules, magical elements, or cultural contexts than the protagonist's original realm (Pagan). The term originally comes from Japan (*isekai* meaning different world), but the concept is not new, having been around for centuries.

Interestingly, Yudkowsky acknowledges his inability to keep the story light-hearted, highlighting the conclusion of the second chapter, where probability is discussed at length, as proof of his incorporating rationalist elements into the story.

Yudkowsky’s written works exhibit distinct diversity, encompassing a range of topics largely focusing on rationality and artificial intelligence, while also offering seemingly light-hearted and inconsequential reading experiences with rationalist aspects.

3 Practical Part

3.1 Rationality as a Central Theme

As established, the core theme of *HPMOR* is rationality and its application to the actions and mental processes of Harry Potter and subsequently other characters. Rationality is prominently featured throughout the story in an appealing manner, repeatedly showcasing the advantages of rational thinking. This is achieved through the words, thoughts, and actions of various characters, who serve as mediators of rationality in the narrative by providing the reader with a skilful combination of intellectual language and practical examples to convey complex rationalist concepts. Furthermore, this piece of fan fiction introduces an evident conflict between rational reasoning processes and the magical elements of the *Harry Potter* universe. The character of Michael Verres-Evans, Harry's stepfather, demonstrates this by declaring magic as "just about the most unscientific thing there is!" (Yudkowsky 16). Through the stark contrast provided by these two metaphorical worlds colliding, the author is possibly striving to present rationality as something desirable and applicable in various situations. Furthermore, the ultimate goal of *HPMOR* appears to be to assure the reader that rationality is not complicated or impenetrable, but achievable by the common man.

From the very beginning, there is a notable difference of narrative style, with *HPMOR* choosing to employ a more analytical approach, using clear, specific, and accurate language to convey logical arguments and principles while not losing the approachable flair to keep the reader engaged. The departure from Rowling's writing style is most apparent in an extensive use of monologues and dialogues, which demonstrate Harry's (and others) thought processes. They usually span over multiple paragraphs and mainly focus on scientific and rational perspectives. An example of such a situation occurs early on in the fan fiction where Professor McGonagall demonstrates the existence of magic to the Evans-Verres household by briefly levitating Harry's stepfather and subsequently transforming into a cat. This causes a profound realization for Harry, putting his understanding of the world into question and forcing him to reexamine the very foundations of physics and the laws he believed to be applicable on the whole universe.

> A blur was coming over Harry's vision, as his brain started to comprehend what had just broken. The whole idea of a unified universe with mathematically regular laws,

> that was what had been flushed down the toilet; the whole notion of *physics*. Three thousand years of resolving big complicated things into smaller pieces, discovering that the music of the planets was the same tune as a falling apple, finding that the true laws were perfectly universal and had no exceptions anywhere and took the form of simple math governing the smallest parts, *not to mention* that the mind was the brain and the brain was made of neurons, a brain was what a person *was*-
>
> And then a woman turned into a cat, so much for all that. (Yudkowsky 23)

Another example of the inclusion of the theme of rationality within the fan fiction is a scene which takes place in chapter 6 titled "The Planning Fallacy" (Yudkowsky 51), in which Harry at length explains the logic behind the planning fallacy, a prevalent cognitive bias which represents the tendency of individuals to underestimate the time and effort needed to complete a task. As rationalists take great interest in cognitive biases, a phenomenon which will be further exemplified in oncoming chapters, Harry's explanation of it proves to be deeply rational and successfully showcases Harry's understanding of these aspects of rationalism. The context of the scene is such that Harry and Professor McGonagall are shopping in Diagon Valley, and Harry asks to purchase a healer's kit. Professor McGonagall, suspicious of any of Harry's actions, questions him as to why he so wishes to purchase one. He, in turn, proceeds to explain the planning fallacy, using an example of a group of students estimating how long it will take them to finish their homework. The experiment Harry mentions proves that when estimating the time it will take to finish any task, humans usually blindly assume everything will go according to plan. They also generally underestimate the complexity of the task, putting them at an even greater disadvantage (Yudkowsky 61). Harry's critical thinking in this example allows him to identify the planning fallacy, as well as successfully explain it to Professor McGonagall, and convince her to let him purchase a healer's kit. This is significant because Professor McGonagall is depicted as a rational character who acknowledges the strength of a compelling logical argument. Therefore, Harry's successful persuasion speaks to his ability to present a convincing argument supported by existing evidence, thus reinforcing the central theme of rationality.

An effective piece of evidence for the central theme of rationality in *HPMOR* is that, generally speaking, characters who act and think rationally often fare well in the story, while those who directly act against rationality receive a sort of punishment. At first, during the first 30 or so chapters, Harry is seen as an anomaly by his peers and professors, not only because of his intellect but also for his maturity, self-awareness, and rational inclinations. However, as the story progresses and Harry showcases the benefits of rationality to other characters, he gradually succeeds in not only teaching them rationality, but also in portraying rationality as something worthwhile.

Chapter 22 contains an important example of the treatment of rationality by Harry's peers at an earlier point of the story. In this chapter, aptly labelled "The Scientific Method", Harry Potter, after successfully convincing Draco to become a Scientist like himself, decides to apply the Scientific Method to one of Draco's core beliefs. This belief is that wizards are gradually getting weaker due to mixing of blood between magical and non-magical individuals. This belief in pureblood[10] supremacy is the basis of ideology for the Malfoy household (and Death Eaters in general), and Draco has been led to believe throughout his whole life that "mudbloods" (a slur for wizards both of whose parents are not wizards) are what's causing the decline of power in the wizarding world.

Due to this, Draco initially struggles to come up with any alternative hypotheses to the blood purist one but is successfully encouraged by Harry to consider other viewpoints and to avoid jumping to conclusions. This highlights the importance of being open-minded and willing to consider different possibilities. During this conversation, Harry also mentions two concepts which tie into rationality, namely Litany of Tarski[11] and Litany of Gendlin[12]. As these concepts encourage believing in what is true and not what one wants to believe, he uses them to persuade Draco to consider other causes for the decline of magic. Harry and Draco then work together to form multiple hypotheses, which they then proceed to test using different methods. This is a direct use of the Scientific Method, which is closely related to

[10] In the world of Harry Potter, there are three major groups of humans. There are Wizards (those fully capable of magic), Squibs (those who cannot do magic but one/both of their parents was/were wizard/s) and Muggles (those who cannot perform magic and are not related to a wizard). Purebloods are those who are born to two wizards and are wizards themselves.

[11] Litany of Tarski is a tool to reinforce the idea that our beliefs should be grounded in reality. The template follows the structure of „If X, I desire to believe that X". An example of applying the template would be: „If the box contains a diamond, I desire to believe that the box contains a diamond" (Tarski qtd. in Yudkowsky).

[12] Litany of Gendlin serves a similar purpose to the Litany of Tarski. The most often used part of the Litany is „What is true is already so. Owning up to it doesn't make it worse" (Gendlin qtd. in Yudkowsky).

rationality. The final result they come to using this method is that wizards are getting weaker not because of interbreeding with Muggles and Squibs, but due to a gradual loss of knowledge. This obviously devastates Draco, as his core belief has just been disproven. At first, a rift forms between the two boys, however, after the initial wave of emotions passes, Draco understands that this is the cost of asking questions, seeking answers based in reality, and being rational.

The narrative effectively explains complex rationalist ideologies to the readers using intellectual language in combination with practical examples, the didactic effect of which will be further exemplified in the following chapters. The storytelling characteristics therefore successfully argue for the central theme of rationality within *HPMOR*.

3.1.1 Rationality Explained and Exemplified

Yudkowsky highlights the role of rationality as an integral part of the narrative, as any action of Harry's is inherently influenced by rational thought processes in overt or covert ways. However, there are also other characters who show interest in rationality, namely Hermione Granger and Draco Malfoy, who learn to use certain aspects of rationality throughout the story. There are ample references to specific academic essays and books focused on rationality. Such is the case of a scene taking place during an introductory discussion between Harry and Hermione on the Hogwarts Express after Harry recognizes Hermione's intellect as comparable to his own. He quizzes her on which books she is familiar with, commenting: "…and I've read *The Feynman Lectures* (or volume 1 anyway) and *Judgment Under Uncertainty: Heuristics* and *Language in Thought and Action* and *Influence: Science and Practice* and *Rational Choice in an Uncertain World*…" (Yudkowsky 128), showcasing Harry's eruditeness, while also suggesting these real academic works to any reader interested in rationality. These specific works cover a wide range of topics related to rationalism, specifically physics, cognitive biases, semantics, and rational decision-making respectively.

An intriguing example of the theme of rationality is the case of Harry undergoing the Sorting Ceremony in chapters 9 and 10. The Sorting Ceremony is an event held before the beginning of every school year at Hogwarts, and, as the name suggests, it serves the purpose of sorting any newcoming pupils into their respective Hogwarts Houses. There are four

Hogwarts Houses available to students, each with certain expectations of characteristics of those who join them. In the original, Harry Potter gets sorted into Gryffindor, a house for those who are courageous and brave. Contrastingly, in *HPMOR*, Harry James Potter-Evans-Verres is sorted into Ravenclaw, which values intellect, wisdom, and wit. However, before Harry's Sorting, it is revealed that in HPMOR, the Sorting Hat is an enchanted object which responds to the psyche of the person wearing it. As such, the first statement spoken by the Sorting Hat after Harry dons it is: "Oh dear. This has never happened before…" *What?* "I seem to have become self-aware" (Yudkowsky 139). What follows is a chapter-long dialogue between Harry and the Sorting Hat, which is mirroring his mind, reflecting his consciousness and perceiving his thoughts as they form. The dialogue contains many concepts related to rationality and philosophy, and even artificial intelligence, another of Yudkowsky's areas of expertise. Sections of note within this dialogue are ones dealing with Harry's newly found awareness of his "dark side", which he at first suspected to be anger management issues but after a specific circumstance realized it is something much deeper. Harry and the Hat discuss this situation, proving the existence of this dark side through evidence and communicating back and forth as to how to mitigate it. Of course, since the Sorting Hat is a mirror to Harry's mind, it explicitly informs him: "I cannot comprehend this matter for you, when you do not understand it yourself" (Yudkowsky 143).

The position of rationality as an integral part of the story is further solidified through the relationship between Draco Malfoy and Harry Potter. After having made their introductions, Harry and Draco become reluctant friends and decide to form a secret group titled *Bayesian Conspiracy.* The title of the group comes from Harry's and Draco's shared ambition of conducting experiments using the scientific method and utilizing Bayes' theorem to update their beliefs as they progress. *Bayesian Conspiracy* formed between the unlikely pair specifically because of Draco's interest in learning science after Harry had offered Draco something he believed science could provide – power. When questioned by Draco whether they would be equals or not, Harry responds that for the time being they cannot. To demonstrate his point, Harry explains that it is not only Draco's knowledge that is preventing them from being equal for the time being.

> "The problem isn't that you're ignorant of specific science things like deoxyribose nucleic acid. *That* wouldn't stop you from being my equal. The problem is that you aren't trained in the methods of rationality, the *deeper* secret knowledge behind how those discoveries got made in the first place. I'll *try* to teach you those, but they're a lot harder to learn. Think of what we did yesterday, Draco. Yes, you did some of the work. But I was the only one in control. You answered some of the questions. I asked all of them. You helped push. I did all the steering by myself. And without the methods of rationality, Draco, you can't possibly steer the Conspiracy where it needs to go." (Yudkowsky 429)

Throughout a significant part of the story, Harry's and Draco's interactions are mostly of didactic nature, as Harry teaches Draco the basics of rationalism. This allows the reader to learn along with Draco, successfully adapting the mindset of a rationalist over time.

Additionally, Harry has a similarly collaborative relationship with Hermione, conducting experiments with her and honing her methods of rationality at a faster rate due to their comparable minds. Although this might seem manipulative at first, this is mostly due to Hermione's capability of effectively memorising large quantities of information and her and Harry's shared passion for learning. In contrast, Harry and Draco have significant differences in values and goals, and Harry, while attempting to prevent himself from turning Dark, is seeking to do the same for Draco, who has showcased knowledge of Dark magic and holds morally questionable beliefs taught to him by his father, Lucius Malfoy. The beginning of the relationship between Harry and Hermione can be observed as early as in the 8th chapter, in which Harry does something seemingly impossible and encourages Hermione to attempt to find an explanation. Although Hermione ultimately fails, Harry realizes that Hermione has an impressive amount of knowledge, but a painful lack of praxis. In the following chapters, such as in chapter 22 or chapter 28, Harry and Hermione work together to test hypotheses and explore the strengths and weaknesses of magic.

As has been alluded to, Harry is not the only person exhibiting a rationalist mindset. Yudkowsky's Professor Quirinus Quirrell has undergone a significant change in characterization when compared to his counterpart in the canon. In *HPMOR*, Quirrell serves as a mentor to Harry, challenging his beliefs and asking relevant questions in accordance with rationalism. He also exhibits an extremely rational and calculating mindset, constantly weighing the costs and benefits of various actions as well as skilfully manipulating those around him, including Harry. Although he does turn out to be the main antagonist of the fan fiction, as he is revealed to be possessed by Lord Voldemort, his actions, even if evil in nature are no less rational than those of Harry. A significant demonstration of his rational thinking is at play in chapters 51-55. These chapters consist of Harry and Quirrell working together to break Bellatrix Black out of Azkaban. Due to Azkaban's reputation as the place no one can escape from, Quirrell's plan is extremely detailed, consisting of steps which were unlikely to fail. Unfortunately, what Quirrell did not anticipate is Harry's actions, and during an encounter with an Auror inside the prison, Harry intervenes despite being told not to, as he fears that the Professor would kill the Auror. As a result, the plan falls apart, as Quirrell's and Harry's spells meet and cause an interference, rendering Quirrell unconscious. Harry is then forced to analyse the current situation and to proceed without Quirrell, which he does successfully until the Professor finally awakes and executes the remainder of the plan, breaking Bellatrix Black out of Azkaban. While this storyline certainly shows the power of a rationalistic mindset, some plot points such as this one can be regarded as unnaturally elaborate and hard to identify with.

Depiction of Bayes' Theorem and Bayesian Reasoning within the Fan Fiction

Bayes' theorem and Bayesian reasoning play a significant role in *HPMOR*, as the author, Eliezer Yudkowsky, incorporates these concepts as crucial elements of rationality and rational decision-making. As established in the theoretical part, Bayes' theorem is the basis for Bayesian reasoning, which serves to update one's beliefs based on evidence and new information. It could be argued that most of Harry's beliefs adhere to this schema, having an underlying effect on his thoughts and actions all throughout the story without having to be mentioned explicitly. There are, however, instances of direct reference to the Bayes' theorem and Bayesian reasoning, which impact the story in significant ways and directly affect Harry's decision-making processes.

One overarching theme of Harry applying Bayes' theorem to his decision-making is his opinions and beliefs regarding his mysterious mentor, Professor Quirinius Quirrell. Throughout the story, Harry's opinion of Quirrell changes drastically. At first, he is unaware of Quirrell's true nature (his being possessed by Tom Riddle, a.k.a. Lord Voldemort) and regards the man with interest and respect, admiring his intelligence and knowledge, seeing similarity between Quirrell and himself. As the story progresses, Harry updates his opinions of Quirrell based on increasingly suspicious events happening during his school year, such as when Hermione is framed for an attempt at Draco's life, or when she is killed later by a conveniently placed troll. Both of these significant events, and many more, had a plausible way of being done by Professor Quirrell. Although Harry used Bayesian reasoning very often and it served him well in most cases, his use of the framework regarding Professor Quirrell was not successful until it was far too late. The reveal regarding Quirrell's possession happens in chapter 105, where Harry finally questions the sheer impossibility of the synchronicity of various events, and everything falls into place.

As mentioned before, Quirrell is too a character exhibiting rationalism, in thoughts, words, and actions. An interesting passage dealing with the concept of Bayesian reasoning, (and by extension, Bayes' theorem), this time mentioned by Professor Quirrell, occurs when he and Harry discuss the happenings within a class of Defence Against the Dark Arts (or Battle magic, as it comes to be referred to). In this class Quirrell sought to prepare his students for combat, starting this endeavour off by teaching them to lose before teaching them to win. Quirrell, being aware of a severe fierce fight Harry had earlier that day with Professor Snape, decided Harry was the best candidate for this lesson.

Professor Quirrell thought it best to teach Harry how to lose a dominance contest by practical example, specifically by having Harry stand in the middle of a classroom and calling on other willing students to take their turns shoving him and berating him. This technique, although highly questionable, seemed to have imparted an important lesson on Harry, as after the ordeal has finished, Harry considers the experience beneficial to managing his unwillingness to lose, as demonstrated by his verbal fight with Professor Snape. Harry then, in a calculated move, speaks to the rest of the students, forgiving them for their deeds, and asking his peers not to take revenge on the offending students. Afterward, Quirrell questions him: "Is the Sun still in the sky? Is it still shining? Are you alive?" (Yudkowsky 329) and then sends him to a separate room to rest. However, Harry's forgiveness intrigued Quirrell, and led to a conversation with the undertones of Bayesian reasoning. After conversing for a while, Harry inquired whether he truly was off the path of becoming a Dark Lord. In the following excerpt, Quirrell considers the probability of Harry's forgiveness being sincere against his previous experience, ultimately rendering the act improbable.

> "There is nothing you can do to convince me because I would know that was exactly what you were trying to do. And if we are to be even more precise, then while I suppose it is barely possible that perfectly good people exist even though I have never met one, it is nonetheless *improbable* that someone would be beaten for fifteen minutes and then stand up and feel a great surge of kindly forgiveness for his attackers. On the other hand it is *less* improbable that a young child would imagine this as the *role to play* in order to convince his teacher and classmates that he is not the next Dark Lord. The import of an act lies not in what that act *resembles on the surface,* Mr. Potter, but in the states of mind which make that act more or less probable." (Yudkowsky 335)

Within this excerpt, Professor Quirrell directly mentions probability, which is closely associated with Bayesian reasoning, as Bayesian reasoning deals with the evaluation of possibilities. Rational decision-making also makes an appearance, the last sentence of the monologue portraying the mental processes taking place before Harry's display – Harry has presumably determined the desirability of different outcomes based on his actions and settled on the one with the highest level of practical appeal.

More Bayesian reasoning can be observed in an unfortunate encounter between Harry and Headmaster Dumbledore in chapter 39. This chapter features Harry being called up to the Headmaster's office in order to provide council for a specific situation. Professor Quirinus Quirrell wishes to bring a Dementor into Hogwarts as a target for his students to practice the Patronus charm on. Dumbledore, being suspicious of Quirrell due to past events, asks Harry to tap into his dark side, and to explain to him what this event truly is, to reveal what Quirrell is planning. After Harry classifies the event as a distraction, Dumbledore and Harry speak further and Dumbledore poses another question. "Tell me, Harry," said the Headmaster (and now his voice sounded simply puzzled, though there was still a hint of pain in his eyes), "why do Dark Wizards fear death so greatly?" What ensues is an intense conversation about death and the afterlife. Dumbledore, being a staunch believer in the afterlife, is puzzled by Harry's requirement of concrete evidence for such a claim. Harry, holding a firm stance of rationalism, defies Dumbledore's attempts at convincing him of the existence of the afterlife and points out how easily disprovable Dumbledore's evidence is. As proof of the afterlife, Dumbledore mentions The Veil, a stone archway kept in the Department of Mysteries. This archway is believed to be a gateway into the land of the dead, to which Harry responds with exasperated interest, proclaiming that he shall hear out Dumbledore's evidence, but only for the simple reason that "that is what a scientist does" (Yudkowsky 681). Harry also stresses the importance of empirical evidence to the Headmaster, telling him to share what he has *seen*, not what he believes. After hearing the description of The Veil, Harry coldly retorts: "That doesn't even sound like an *interesting* fraud", immediately thinking of ways such a thing could be fabricated. In this instance, although severely emotionally charged, Harry's use of the Bayesian framework is still present, as this discussion between Harry and Dumbledore incorporates elements of comparing prior probabilities, empirical evidence and considering alternative explanations for various phenomena. However, it can be argued that Harry's constant questioning of authority and absolute refusal to suspend disbelief in critical situations (even if this behaviour is later rewarded) can sometimes occur at the expense of readability.

It is clear that Yudkowsky's work makes a unique contribution to the understanding of rationalist concepts by portraying them through a universe and characters many know well. Although the similarities between *HPMOR* and the original dwindle as the story progresses, Yudkowsky does a fine job making the original story his own and adding depth to underdeveloped elements while still including aspects many *Harry Potter* fans hold near and dear to their hearts. Even then, there are times during the story where the amount of knowledge presented to the reader can be considered disruptive, aggravating, and contrived, in turn making the reader feel preached to and alienated if they do not share Yudkowsky's seemingly radical perspective.

3.1.2 The Rational Character and Actions of Harry Potter

Harry Potter has been indisputably chosen by Yudkowsky as the central source of rational concepts in *HPMOR*. Throughout the story, Harry Potter consistently demonstrates his commitment to rationality through his actions and thought processes. However, the changed character of Harry Potter is one of the main sources of *HPMOR*'s polarizing nature, as many find Yudkowsky's rendition of Harry unbearable and self-important and cannot enjoy the story with a protagonist they perceive to have a superiority complex. For example, in contrast to his canonical counterpart, Harry James Potter-Evans-Verres is portrayed as highly intelligent and rational[13] to a fault. He engages in critical examination of established norms and practices, employing aspects of science in his investigations. He also often seeks out evidence instead of blindly accepting what he is taught and told. In the following example, this process of questioning, experimenting, and knowledge-seeking allows him to invent a new type of Transfiguration. Transfiguration is an extremely complex and dangerous process of changing the form and appearance of objects, so to discover a new type of Transfiguration is a notable feat which has never been achieved before.

Transfiguration as a whole is portrayed much more in depth in *HPMOR* than in the canon. Due to the nature of rationalism, this magical process of changing one object into another has been depicted with much more complexity and is, as mentioned, also highly dangerous. In chapter 15, during Harry's first Transfiguration lesson with Professor McGonagall, Harry is vigorously taught that learning Transfiguration is a very delicate process, which requires a high level of discipline and adherence to very specific rules, one of the rules being that one could only Transfigure whole objects into other whole objects. Professor McGonagall does not fail to mention that should any student break any rule at any point, they will simply not learn anymore Transfiguration for the rest of their studies at Hogwarts. Despite this, Harry dares to experiment with Transfiguration in chapter 28, and based on a complex mental process creates something referred to as Partial Transfiguration. He successfully mentally breaks the object down beyond atoms or quarks to a "gigantic *factor* in a wavefunction that *happened to factorize*", which then allowed him to only Transfigure a part of an object, thus successfully breaking one of the rules of Transfiguration

[13] This can be attributed to the fact that Lord Voldemort has imprinted a part of his conscious on Harry's mind during the night of the Potter's murder, however, Harry‘s step-father's expertise in the realm of science, which Harry indisputably adapted throughout his youth, also played a significant role in developing his brilliance. It is worth noting that the entirety of the story, along with its complex and intellectually challenging elements illustrated by Harry and others, takes place during Harry's first year at Hogwarts. This too speaks to his high intellect, as Harry is fully capable of understanding and teaching complicated concepts at 11 years old.

(Yudkowsky 513). This was possible due to his mental process of reductionism[14]. Only thanks to his in-depth knowledge of physics, experimental spirit, and the willingness to challenge preconceived notions did Harry manage to invent a new type of Transfiguration at 11 years old.

Harry's knowledge of rationalist concepts is also evident in his awareness of cognitive biases during the story, one such example of this awareness being in chapter 5, where Harry and Professor McGonagall set out to gather all the necessary items and material for Harry's stay at Hogwarts. During this, an unknown man approaches Harry and regards him with awe, inquiring whether he truly was Harry Potter. Harry, in turn, began questioning his own identity, discussing whether he truly is the person which saved everyone from Lord Voldemort. At first, possibly in jest, he responds to the man with scepticism, pointing out that technically speaking, people could simply find a different orphan and raise him to believe that he was Harry Potter, but after the conversation ceased and Harry was briefly chastised by Professor McGonagall, Harry introduces the fundamental attribution error. Due to the nature of the fundamental attribution error, it becomes clear that people generally attribute his victory over Lord Voldemort to his ability and heroicness, while in truth, Harry was only 15 months old at the time, which strongly suggests that his triumph was most likely nothing more than, in his own words, "contingent environmental circumstances" (Yudkowsky 43).

As established, experimentation is an important aspect of rational decision-making. A direct example of Harry testing the boundaries of magic using the scientific method is in chapter six, where Harry obtains a Mokeskin pouch with an enchantment allowing the holder to call forth any item from the pouch. Harry almost immediately begins testing the capabilities of the enchantment, asking for a bag of gold within the pouch using different methods, such as referring to the item as *bag of element 79*, *bag of okane*[15]*,* and further experimenting with the enchantment. Ultimately, Professor McGonagall explains the phenomenon away as simply magic, but still Harry demonstrates a commitment to

[14] Reductionism in philosophy conveys the idea that things can be understood or explained by breaking them down into simple or smaller things of different kind. The idea that physical things are specific collections of atoms is an example of reductionism (*Encyclopaedia Britannica*).

[15] okane is the Japanese term for money (*WordSense Dictionary*)

understanding the magical world through rational inquiry, evidence-based reasoning, and logical deduction.

Furthermore, most of the examples given within this thesis could serve as proof of Harry's inherent rational nature. Although Harry is not infallible, he still seeks to get as close to the truth as his cognitive abilities allow him. It could also be argued, that Harry's rationality is the main driving force behind the plot of the whole story, as *HPMOR* is made intriguing by the simple fact that "rationalist!Harry" is willing to ask questions the canon Harry did not, and therefore the author has the opportunity of developing aspects of the story which were left unexplored in the original. Of course, "rationalist!Harry" carries with him a significant advantage in his intellect and reasoning abilities beyond his years, however, due to the high level of adversity he faces, Harry does not come across as a flawless character simply coasting through life. The story places Harry in a complex world full of dilemmas and challenges and facing these challenges humanizes him. His facing adversity occasionally shows Harry's shortcomings, but also makes it possible to display growth, as well as present rationality as a valuable tool, instead of a fix-all solution.

In summary, the portrayal of Harry James Potter-Evans-Verres and his rationality add depth and intrigue to the story. Although it can sometimes be perceived as excessive, his constant pursuit of the truth and his refusal to accept the easiest answer provide tension and suspense, engaging the readers as they accompany him on his quest of world optimization. His rational decision-making also supplies the plot with complexity and depth, further underlining the importance of analytical thinking and decision-making, showcasing the power that lies in rationality. Alongside this, the contrast between Harry's mostly logical demeanour and the more emotional disposition of other characters adds a layer of intricacy and provides the story with interesting conflict.

3.1.3 Cognitive Biases in *Harry Potter and the Methods of Rationality*

As established in the theoretical part of this essay, the prevalence of cognitive biases in a wide of scope situations is undisputable. In their nature, cognitive biases prevent individuals from accurately capturing the truth and cloud their judgement in the context of decision-making. For this reason, cognitive biases are undesirable when one aims for rational decision-making. Step one of preventing cognitive bias is to simply recognize its existence.

Once we realize our judgment is being influenced by biases, be it due to a heuristic or not, we can now take direct action to mitigate it best to our ability.

Cognitive biases are also of high interest to rationalists, as they aim to perceive the world around them as objectively as possible. As such, there are numerous mentions of different biases throughout the fan fiction. The ones explicitly mentioned in *HPMOR* are for example the planning fallacy, fundamental attribution error (both of which have been addressed in the previous chapters), confirmation bias, and many more. There are many points throughout the story where Harry Potter falls victim to a bias but does not fail to eventually realize it and attempts to mitigate it to the best of his ability. The consequent chapters of this thesis will focus on examples of confirmation bias and anchoring bias within the fan fiction, as these two biases occur multiple times at key points of the story, be it explicitly or implicitly.

Depiction of Confirmation Bias within the Fan Fiction

Confirmation bias occurs at numerous points within the fan fiction. As explained, confirmation bias is the tendency to seek out and understand evidence in accordance with our current beliefs. In *HPMOR*, the characters depicted, including the main character Harry James Potter-Evans-Verres, are not immune to the effects of biases, including the tendency to interpret things so that they best serve the character's beliefs rather than disprove them. Although this bias can be found in the story multiple times, it does not dimmish the intellect of the characters, rather it serves as a reminder of possible issues one can encounter when aspiring to be a rationalist.

A notable case of confirmation bias also occurs in chapter 31. This chapter has a complex lead-up, as Professor Quirinius Quirrell offered all students the opportunity to join extracurricular activities to further expand on their knowledge of Battle magic. One such activity involves teaching the students how to conduct themselves during large-scale battles in case there is ever a war. To create the suitable conditions for this to take place, Professor Quirrell split up all interested students from the same year into three groups – three armies, each helmed by a different General and taught them all a sleep spell to ensure they don't harm each other with other spells. Harry Potter and Draco Malfoy become the Generals of their armies, Chaos Legion and Dragon Army respectively. The unexpected General of the third army to join the fray turned out to be Hermione Granger, deciding to call her battalion Sunshine Regiment. Both Harry and Draco were incredulous upon finding this out, bringing up their doubts about Quirrell's decision to allow this directly to the Professor, who, in turn, warned the boys not to underestimate Hermione.

Ultimately, upon the beginning of the first fight, after all three armies have been instructed to strive for victory, Harry and Draco swiftly find out Hermione had seemingly split her army in half, each half heading for the other two armies. As soldiers of the Sunshine Regiment turned out to be simple to deal with at first, Draco and Harry's armies turned to fighting each other. Once Harry's army was down to six soldiers and Draco's was only made up of two, Hermione's Sunshine Regiment pounced and secured a victory. All of this could have been avoided if Draco and Harry had not underestimated Hermione's capabilities and strategic thinking, assuming that her soldiers are defeated based on surface-level observations. This biased perspective led to them overlooking the possibility that Hermione had devised a clever plan to deceive both the Dragon Army and the Chaos Legion. Instead of critically evaluating the situation, seeking additional evidence, and not disregarding Hermione's high intellect and resourcefulness, they relied on their preconceived notions, and in turn had their assessment of the battlefield severely hindered.

Another instance of confirmation bias, one where it is explicitly mentioned, is in chapter 46. This chapter speaks of an aforementioned event, where Professor Quirrell had wished to bring a Dementor to Hogwarts as it would allow his students to field-test their Patronus charm and give a chance at successfully casting it to those who were unable up until that point. Unfortunately, not all would go according to plan, as Harry would undergo too long of an exposure to the effect of the Dementor and become temporarily demented as result. After being rid of this effect, Harry recuperates, considers the nature of the Dementor and afterward successfully destroys it, a feat that has never been achieved before. Once this admittedly disastrous affair had concluded and Harry refused to share the secret of his triumph over the Dementor, he and Quirrell walk and emboldened by the victory, Harry asks Quirrell, that if this incident were to be a distraction, what the was the true plot behind it. After humouring his inquiry, Quirrell asks a question in return, wondering whether Harry has considered the possibility of Dumbledore being the culprit.

Harry then comes to a revelation, realizing his confirmation bias regarding the Headmaster. "[Harry]'d really been supposed to know better than that already. Confirmation bias was a technical term; it meant, among other things, that when you chose your information sources, there was a notable tendency to choose information sources that agreed with your current opinions" (Yudkowsky 721). From this point onward, Harry makes a conscious effort to seek knowledge regarding the Headmaster, even if it doesn't align with his beliefs or opinions.

In summary, it is safe to say that whenever Confirmation bias makes an appearance in *HPMOR*, it always carries negative repercussions with it. Be it Harry and Draco's ultimate loss during the fight between the Armies, or Harry's inclination to regard Dumbledore not as a threat, but more so as a genius feigning madness. By recognizing this cognitive bias (with the help of Quirrell), Harry can now overcome it and approach Dumbledore with a more rational and unbiased mindset. In regard to the Armies situation, both Harry and Draco realize the error of their ways and regard Hermione as equally cunning as them. Confirmation bias is therefore portrayed as a negative element and all the characters who realize it and combat it are better for it.

Depiction of Anchoring Bias within the Fan Fiction

Although anchoring bias, unlike confirmation bias, is not explicitly mentioned within *HPMOR*, it still plays a role within the story. There are multiple instances where anchoring bias influences the plot, a significant one, which can also be found in the canon, is the case of the Hogwarts Houses. As partially explained in a previous chapter, there are four Hogwarts Houses in total, each representative of a specific set of characteristics. Gryffindor, a house Harry Potter and Hermione Granger are sorted into in the canon, is a house for those who exhibit bravery, courage, and determination. Slytherin, a house unfortunately associated with evil and wrongdoing, values ambition, cleverness, and resourcefulness. Ravenclaw, the house *HPMOR*'s Harry Potter-Evans-Verres and Hermione Granger are sorted into, fits those who exude wit, wisdom and intellect, and at last Hufflepuff, a house often underestimated by all others, welcomes hard workers with a sense of justice and patience. These characteristics are important to the whole of Wizarding Britain, exemplified by the fact that the Sorting Hat, as discussed, simply mirrors the mind of the child which wears it.

This proves, that the Sorting Hat places the child in the house it most identifies with based on the aforementioned characteristics. There are also instances of entire families expecting their children to be sorted into a specific house due to an imagined prestige, such as the Malfoy family. Therefore, although maybe in the past the characteristics were used to describe the students, now it stands that the students consciously or subconsciously shape themselves based on the house they are sorted into even before the Sorting Ceremony. Thus, in this instance, the anchor of the anchoring bias lies within the expectations of each House, specifically the fact that those who are to be sorted often initially perceive a specific house as more favourable than the others and then struggle to have their opinions changed. A good example of this is the discussion between Harry and the Sorting Hat, where the Sorting Hat offered Harry to join any of the four houses, but warned him that if he were to join Ravenclaw or Slytherin, his coldness (i.e. his dark side) would be strengthened, while if he were to join Gryffindor or Hufflepuff, he would gain warmth in turn. Harry, after having witnessed Hermione being sorted into Ravenclaw and wishing to be in the same house, as well as identifying with the characteristics of Ravenclaw the most, is still determined to join Ravenclaw, comparing the other houses with the one he wishes to go to instead of evaluating each house on their own merits.

Another example of anchoring bias is Harry's perception of Professor Severus Snape. Harry's first indirect introduction of the Professor occurs in chapter 18 before Harry's first Potions class. He is warned by Ernie Macmillan on the behest of Neville Longbottom.

> "Neville thought I should warn you," Ernie said in a low voice. "I think he's right. Be careful of the Potions Master in our session today. The older Hufflepuffs told us that Professor Snape can be really nasty to people he doesn't like, and he doesn't like most people who aren't Slytherins. If you say anything smart to him it… it could be really bad for you, from what I've heard. Just keep your head down and don't give him any reason to notice you." (Yudkowsky 297)

This piece of information given by Ernie Macmillan is the anchor of Harry's opinion of Professor Snape. Although this warning does turn out to be appropriate, as Harry and Professor Snape engage in a heated argument after Harry does not stop himself from calling attention to Snape's unacceptable attitude towards students, Harry's opinion regarding

Snape remains negative even after being given various reasons to trust him, such as Dumbledore having full confidence in Professor Snape, or Professor Snape coming to Harry with a case of bullying, which he strongly encouraged Harry to deal with. Harry's continuous apprehension is proof of anchoring bias.

In conclusion, although anchoring bias is never explicitly mentioned in *HPMOR*, it can still be observed in certain aspects of the story. Its presence within the story, along with confirmation bias and other mentioned biases serve as a reminder that even highly intelligent and rational individuals like Harry aren't flawless, therefore successfully bringing awareness to the issue of cognitive biases to readers as well.

4 Conclusion

The primary aim of this thesis is to examine and analyse elements of rationality and rationalism within Eliezer Yudkowsky's fan fiction *Harry Potter and the Methods of Rationality* (*HPMOR*), specifically in the context of cognitive psychology. The thesis also aims to investigate how these aspects shape the plot the fan fiction and their role in advocating for rationality to the reader. By examining rationality from a cognitive psychology perspective, this research seeks to explore its narrative impact and its potential to influence reader's perspective. In order to conduct the analysis, the concept of fan fiction was established to showcase its importance for contemporary literature, which was then followed by an introduction to the specific type of fan fiction *HPMOR* is. To provide the opportunity to identify and exemplify aspects of rationality in *HPMOR*, multiple rationalist concepts were subsequently explored within the theoretical part using academic monographs and articles, along with Yudkowsky's essays, books, as well as posts from his *LessWrong* forum. After comparing results from academic sources with Yudkowsky's writing, Yudkowsky can be considered a reliable source of information on the topic, as he mostly uses preexisting concepts and recontextualises them for the rationalist framework. The concepts mentioned are the general meaning and goals of rationality, Bayesian reasoning and Bayes' theorem, cognitive biases and heuristics, confirmation bias, and anchoring bias. To provide further context, the premise of *HPMOR*, and Yudkowsky's background along with other of his works, including the *LessWrong* forum, are also explored.

The practical part of the thesis applies the knowledge established in the theoretical part to analyse the role of rationality in the fan fiction, including how the aforementioned rationalist concepts are portrayed and explained by Yudkowsky, as well as how they influence the plot and affect the reader's stance toward rationality. Through this exploration, it becomes evident that *HPMOR* serves as a thought-provoking medium for examining rationality and rationalism by unconventional means. The rationality exemplified throughout the fan fiction challenges conventional beliefs and promotes critical thinking, further inspiring the readers to question their own beliefs and gain awareness of their biases, while simultaneously adding depth and complexity to the storytelling. However, it is hard to definitively determine the extent to which Yudkowsky popularised the topic. Although there were mentions of *HPMOR* in multiple popular magazines such as *The Atlantic*[16] or *The Guardian*[17], Yudkowsky's impact varies across communities. While some communities formed around rationality, cognitive science, and fan fiction may acknowledge *HPMOR*'s influence on the popularization of rationality, the reach and recognition of Yudkowsky's impact may not be as significant in mainstream academic spheres.

The positive depiction of Bayes' theorem and Bayesian reasoning further encourages probabilistic thinking and rational decision-making, proving that along with a commitment to overcoming cognitive biases, the character of Harry Potter successfully provides the reader with an introduction to methods of rationality. Furthermore, the presence of confirmation bias and the anchoring bias highlights the need for self-awareness and rational thinking in order to avoid these biases. The story serves as a reminder that even highly intelligent and rational individuals aren't immune to these biases and that constant evaluation of one's own beliefs is essential.

Yudkowsky's endeavour generally received positive feedback and was met with high praise around the time of its gradual publishing. As written by David Whelan in a *Vice* article

[16] Daniel D. Snyder writes in a *The Atlantic* article that „*Methods of Rationality* remains one of the most popular stories on FanFiction.net with more than 13,000 reviews. But more importantly, it demonstrated the extent to which *Potter* fans have expanded the universe beyond Rowling's original designs and helped amplify the series' popularity".

[17] In a *The Guardian* interview with Ben Wikler, an American politician, when asked what podcasts he listens to, Wikler mentions the podcast version of *Harry Potter and the Methods of Rationality*, dubbing it „the #1 fan fiction series of all time" (Raptopoulos).

from 2015, *HPMOR* was regarded as „the most popular Harry Potter book you've never heard of". However, there is potential for the reader to interpret Yudkowsky's use of a beloved character to strongly advocate for rationality as moralizing or even demeaning, as the constant push for rationality and seemingly never-ending barrage of knowledge tend to make the reader feel overwhelmed and inadequate. In addition, rationality, even if it is seemingly a positive trait, can be quite demanding and time-consuming, as well as inconsiderate of emotions. Yudkowsky does explicitly frame rationality as considerate of emotions, but due to the logical nature of rationality, many often erase empathy from the equation of their rationalist thought processes. Additionally, although Yudkowsky uses compelling language and practical examples to explain complex concepts, occasionally the amount of jargon gets in the way of comprehension and the reader gains nothing but confusion. In spite of this, the overall impact of *HPMOR* in promoting rational thinking and challenging biases remains significant within specific circles. The concepts explored in the story are placed within a well-developed context so that even if the reader experiences fatigue from the constant learning and repetition, they are often still invested enough to continue reading in order to find out what happens next.

In conclusion, this thesis demonstrates an in-depth exploration of rationality and rationalism within *HPMOR*, showcasing the potential of fan fiction to engage readers with intellectual topics, therefore fostering a deeper understanding and even appreciation for rational reasoning. The plot is provably made interesting by the deep integration of rationalist ideas into the narrative, and although some readers might not enjoy the highly didactic nature of the fan fiction or the significant changes their favourite universe has undergone by Yudkowsky's hand, the author still manages to write a fascinating story, which emphasizes the importance of critical thinking in navigating complex issues and encourages readers to adopt a rational mindset for decision-making and personal growth.

5 Sources Used

5.1 Primary Sources

Yudkowsky, Eliezer S. *"*Harry Potter and the Methods of Rationality*." FanFiction.Net*, 28 Feb. 2010, www.fanfiction.net/s/5782108/1/Harry-Potter-and-the-Methods-of-Rationality.

5.2 Secondary Sources

Bronwen Thomas. "What Is Fanfiction and Why Are People Saying Such Nice Things about It??" *Storyworlds: A Journal of Narrative Studies*, vol. 3, 2011, pp. 1–24. JSTOR, https://doi.org/10.5250/storyworlds.3.2011.0001. Accessed 10 June 2023.

Lipton, Jacqueline D. "Copyright and the Commercialization of Fanfiction." Hous. L. Rev. 52 (2014): 425.

Hellekson, Karen, and Kristina Busse, editors. *Fan Fiction and Fan Communities in the Age of the Internet: New Essays*. McFarland & Company, Inc., 2009.

holliequ. "[Prompt Challenge] Round 38: June 2023." *Reddit*, 1 June 2023, www.reddit.com/r/FanFiction/comments/13xg3qh/prompt_challenge_round_38_june_2023/.

Jewels5. "The Life and Times." *FanFiction.Net*, 8 Jul. 2009, www.fanfiction.net/s/5200789/1/The-Life-and-Times. Accessed 10 Jun. 2023.

Derecho, Abigail. „Archontic Literature: A Definition, a History and Several Theories of Fan Fiction.“ *Fan Fiction and Fan Communities in the Age of the Internet: New Essays*, edited by Karen Hellekson, Kristina Busse, McFarland & Company, Inc., 2009.

Rowling, J.K. *Harry Potter Series*. Bloomsbury Publishing, 1997-2007.

"Rational Fic." *TV Tropes*, tvtropes.org/. Accessed 5 July 2023.

"Rational Fiction." *Goodreads*, www.goodreads.com/list/show/100705.Rational_Fiction.

"Characteristics of Rationalist Fiction." *Reddit*, old.reddit.com/r/rational/wiki/index#wiki_characteristics_of_rationalist_fiction.

Alwayslily22, Des98. “Through the Quiet Emeral Eyes (The Philosopher’s Stone).” *Archive of Our Own*, 5 Jun. 2018, https://archiveofourown.org/works/14852573/chapters/34383290. Accessed 10 Jun. 2023.

Yudkowsky, Eliezer S. “More Info.” *Harry Potter and the Methods of Rationality*, hpmor.com/info/. Accessed 14 June 2023.

Vanzo, Alberto. "Empiricism and Rationalism in Nineteenth-Century Histories of Philosophy." *Journal of the History of Ideas*, vol. 77 no. 2, 2016, p. 253-282. Project MUSE, doi:10.1353/jhi.2016.0017.

Hardin, Russel. “Rationalism.” *Routledge Encyclopedia of Philosophy: Questions to Socibiology*, edited by Edward Craig, Taylor & Francis, 1998.

Bristow, William. "Enlightenment." *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta, Fall 2017 ed., Stanford University, 2017, plato.stanford.edu/archives/fall2017/entries/enlightenment/.

Yudkowsky, Eliezer S. “What Do We Mean By “Rationality”?” *LessWrong*, 16 Mar. 2009, https://www.lesswrong.com/s/5g5TkQTe9rmPS5vvM/p/RcZCwxFiZzE6X7nsv.

Yudkowsky, Eliezer S. “The Martial Art of Rationality” *LessWrong*, 22 Nov. 2006, https://www.lesswrong.com/posts/teaxCFgtmCQ3E9fy8/the-martial-art-of-rationality.

Efron, Bradley. "Bayes' theorem in the 21st century." *Science* 340.6137 (2013): 1177-1178.

Thagard, Paul. "cognitive science". *Encyclopaedia Britannica*, 17 Mar. 2023, https://www.britannica.com/science/cognitive-science. Accessed 6 July 2023.

Barnes Jr, James H. "Cognitive biases and their impact on strategic planning." *Strategic Management Journal* 5.2 (1984): 129-137.

Ehrlinger, Joyce, Wilson O. Readinger, and Bora Kim. "Decision-making and cognitive biases." *Encyclopedia of mental health* 12.3 (2016): 83-7.

“Confirmation Bias.” *The Decision Lab*, thedecisionlab.com/biases/confirmation-bias. Accessed 14 June 2023.

Koslowski, Barbara, and Mariano Maqueda. “What Is Confirmation Bias and When Do People Actually Have It?” *Merrill-Palmer Quarterly*, vol. 39, no. 1, 1993, pp. 104–30. JSTOR, http://www.jstor.org/stable/23087302. Accessed 31 Mar. 2023.

Oswald, Margit E, and Stefan Grosjean. “Confirmation bias.” Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking and Memory, edited by Rüdiger F. Pohl, *Psychology Press*, 2005.

de Wilde, Tim RW, Femke S. Ten Velden, and Carsten KW De Dreu. "The anchoring-bias in groups." *Journal of Experimental Social Psychology* 76 (2018): 116-126.

Tversky, Amos, and Daniel Kahneman. “Judgment under Uncertainty: Heuristics and Biases.” *Science*, vol. 185, no. 4157, 1974, pp. 1124–31. JSTOR, http://www.jstor.org/stable/1738360. Accessed 14 June 2023.

Ruby. “LessWrong FAQ.” *LessWrong*, 14 June 2019, www.lesswrong.com/posts/2rWKkWuPrgTMpLRbp/lesswrong-faq.

Yudkowsky, Eliezer S. “Rationality: From AI to Zombies.” *LessWrong*, 1 Apr. 2015, www.lesswrong.com/tag/rationality:-from-ai-to-zombies.

Yudkowsky, Eliezer S. “Inadequate Equilibria: Where and How Civilizations Get Stuck.” *Inadequate Equilibria*, equilibriabook.com/about/. Accessed 14 June 2023.

Yudkowsky, Eliezer S. A Girl Corrupted by the Internet is the Summoned Hero?!. E-book ed, 2016.

Pagan, Amanda. "A Beginner's Guide to Isekai." *The New York Public Library*, 15 Jul. 2019, www.nypl.org/blog/2019/07/15/beginners-guide-isekai-manga. Accessed 10 Jun. 2023.

Yudkowsky, Eliezer S. “The Meditation on Curiosity.” *LessWrong*, 6 Oct. 2007, https://www.lesswrong.com/posts/3nZMgRTfFEfHp34Gb/the-meditation-on-curiosity.

The Editors of Encyclopaedia Britannica. “Reductionism.” Edited by Brian Duignan, *Encyclopædia Britannica*, www.britannica.com/topic/reductionism. Accessed 22 June 2023.

“Okane.” *WordSense Dictionary*, www.wordsense.eu/okane/. Accessed 20 June 2023.

Whelan, David. "The Harry Potter Fan Fiction Author Who Wants To Make Everyone a Little More Rational." *VICE*, 2 Mar. 2015, www.vice.com/en/article/gq84xy/theres-something-weird-happening-in-the-world-of-harry-potter-168.

Snyder, Daniel D. "'Harry Potter' and the Key to Immortality." *The Atlantic*, 18 July 2011, www.theatlantic.com/entertainment/archive/2011/07/harry-potter-and-the-key-to-immortality/241972/.

Raptopoulos, Lilah. "Listen to This: Ben Wikler and Aaron Swartz's The Good Fight." *The Guardian*, 11 July 2014, www.theguardian.com/culture/2014/jul/11/ben-wikler-aaron-swartz-good-fight-politics-activism.