

Claremont Colleges

## Scholarship @ Claremont

---

CMC Senior Theses

CMC Student Scholarship

---

2023

### Moralize through Sympathy: An Analysis of Adam Smith's Sentimentalist Moral Theory

Zhutao Li

Follow this and additional works at: [https://scholarship.claremont.edu/cmc\\_theses](https://scholarship.claremont.edu/cmc_theses)

---

#### Recommended Citation

Li, Zhutao, "Moralize through Sympathy: An Analysis of Adam Smith's Sentimentalist Moral Theory" (2023). *CMC Senior Theses*. 3348.

[https://scholarship.claremont.edu/cmc\\_theses/3348](https://scholarship.claremont.edu/cmc_theses/3348)

This Open Access Senior Thesis is brought to you by Scholarship@Claremont. It has been accepted for inclusion in this collection by an authorized administrator. For more information, please contact [scholarship@cuc.claremont.edu](mailto:scholarship@cuc.claremont.edu).

Claremont McKenna College

Moralize through Sympathy: An Analysis of Adam Smith's Sentimentalist Moral Theory

submitted to  
Professor Adrienne Martin

by  
Flora (Zhutao) Li

for  
Senior Thesis  
Spring 2023  
April 23

## **Acknowledgement**

This thesis would be impossible without the help and guidance from Professor Adrienne Martin. She offers tremendous support to help me organize my thoughts and comprehend the philosophical theory that I have never encountered before. I am grateful for not only her inspiring comments on the content of this thesis, but also her patience and confidence in me along the way. It is a great fortune to have her as my thesis reader in my last semester at CMC. Exploring a whole new moral framework is not something I thought I could do at the beginning of the semester, but her support makes it happen.

I am also thankful to my parents who support my decision to transfer to CMC and major in philosophy. You two are the best parents in the world.

I also thank Kylie Chen whose existence makes me want to procrastinate less, and Yaqin Zhang who saves me from mental breakdown for countless times.

## Table of Contents

<b>Introduction</b>	<b>3</b>
<b>Chapter 1</b>	<b>6</b>
Part 1: Expressivism and Its Strengths	7
Part 2: Concerns of Expressivism	17
<b>Chapter 2</b>	<b>20</b>
Part 1: Hume's Theory	20
Part 2: Smith's Theory	26
<b>Chapter 3</b>	<b>34</b>
<b>Conclusion</b>	<b>45</b>
<b>Bibliography</b>	<b>47</b>

## Introduction

Let's think of a situation where you can easily glimpse your friend's answer sheet in an exam. Suppose that passing that exam is so important to you that cheating becomes a truly tempting option. Many people might just cheat without thinking twice — since they would be better off passing the exam. For other people, however, it takes a lot of effort for them to make the conscious decision to discard their honesty, even though doing so helps with grades.

In countless situations like this, acting in accordance with moral rules seems to be natural and unnatural at the same time. On the one hand, it is natural because people want to be trusted and praised for their character and integrity. On the other hand, when an easy but unethical action can best satisfy a desire, persuading people to act ethically requires reasons. To account for this observation, philosophers come up with theories to reveal moral rules and justify their binding power in a variety of ways. Among these theories, human reason, or rationality, usually plays one of the two roles.

In some moral theories, reason plays a twofold role in morality. It presents the external situation to the moral agent and justifies the normative statements people make about the situation. In this case, reason can answer both questions: is an action truly moral, and what makes it moral or immoral? Immanuel Kant bases his framework of morality on the notion of reason. He argues that complying with moral laws is unconditionally applicable to any rational beings (Johnson, "Kant's Moral Philosophy").

Nothing other than the representation of the law in itself, which can of course occur only in a rational being, insofar as it and not the hoped-for effect is the determining ground of the will, can constitute the preeminent good we call moral (Kant 14).

Kant continues to argue that morality is the relation of actions to the universal law comprehended through the autonomy of will, which is a feature of reason (46). A belief or judgment becomes a moral law only if it can be universalized. I will not spell out the whole process of universalization. My purpose is simply to show that this process is executed by reason.

In some other theories, the reason is used as an instrument that only presents the moral agent with facts of external situations and does not justify any normative judgments. Utilitarianism is a classic case. The goal of utilitarian theory is also to provide an action guide to assist the moral agent in deciding how to act. In utilitarian theories, an action is moral if it creates utilities and leads to beneficial consequences. The moral agent employs reason to perceive an accurate picture of the situation and assign utility to different reactions to that situation. Even though a moral rule cannot be justified without reason accurately reporting the utility of possible reactions, reason cannot explain why maximizing utility makes an action moral. In a utilitarian theory, reason is like competent but indifferent software, producing normative judgments mechanically based on external inputs and the moral principles programmed into it.

I focus on the second type of moral theory in this thesis and explore a moral framework that bases normativity on human sentiments. I introduce sentimentalism and examine if it provides a framework that adequately accounts for moral actions and the binding power of moral duties.

In chapter one, I explain expressivism as the metaethical ground of sentimentalism and explain the worry that emotions are too volatile to be relied on in determining moral rules. In chapter two, I introduce David Hume and Adam Smith's

theories of sentimentalism respectively to address this critique. In chapter three, I offer my critique of Smith's theory.

## Chapter 1

### Part 1: Expressivism and Its Strengths

Moral theories explain what people should do. They are guidelines an agent can rely on to make normative statements. Different from descriptive statements, normative statements specify things that we ought to do, providing reasons for the approval or disapproval of certain actions. “Saving a drowning child is dangerous” is a descriptive statement that describes a fact about the action of saving a child from the water.

However, “saving a drowning child is admirable” conveys normativity, from which people can infer that it is desirable to, or when a person can, they *should* save a drowning child.

The difference between these two kinds of statements should be obvious. A person can agree that a descriptive statement is true without granting their agreement to the plausibility of a normative statement about it: I can see the danger of jumping into the water to save a child without being convinced that it is good to perform such an action. From an epistemological perspective, a person can also see the danger of saving a drowning child without even knowing that it is nonetheless an admirable action.

With this distinction, philosophers began to wonder about the nature of normativity. Is normativity a matter of our emotions, or a property that exists independently of what people feel? Some people say saving a drowning child is admirable — is this because they feel good when they see such an action being carried out, or because admirable-ness, just like dangerousness, exists as a worldly property of this action?

The candidates of the nature of normativity are concepts that only exist in the realm of thoughts, or properties that can be instantiated into facts. Understanding



normativity as a concept should be an intuitive move that most people can agree on: people have normative thoughts, such as “x is good,” and “that person is evil.” We seem to grasp the concepts of “goodness” and “evilness” in our minds, and they are readily available when people make normative judgments. Even though it is not necessary that these concepts must also be metaphysically true for us to assign them to objects, it nonetheless makes it tempting to take one step further and commit to the belief that normativity can be instantiated into facts.

Suppose we buy into this theory that there exist normative facts, and we think, for instance, that saving a drowning child is good — we then believe that it is *true* that saving a drowning child is good. Here, for a sentence to be true, it is then factually accurate. The statement that “saving a drowning child is indeed good” means the same thing as “goodness exists in the action of saving a drowning child.” In this case, saving a drowning child is just as good as it is dangerous. It would be a fallacy to think otherwise. I do not intend to take a stance, but I would like to point out that even though the existence of normative concepts seems to be intuitive, whether or not they exist as normative facts or properties requires further examination.

Many philosophers, such as Blackburn reject that normativity exists as a fact. In short, Blackburn argues that normativity exists only as a concept, and all normative statements that sound like “assertions,” such as “x is good”, are not really assertions or beliefs, but merely expressions of sentiments or attitudes, such as approval, disapproval, aversion, or subscription to a norm (Copp, “Normativity in Metaethics”). When I say “x is good,” I do not express the belief that I see goodness exists in x — I am expressing my approbation of x. This view is also known as expressivism.

In the rest of this chapter, I elaborate on Blackburn's theory and its two strengths. I then propose three lines of attack. To ground his expressivist approach, Blackburn first lays out three characteristics of the process of moralizing.

First, Blackburn argues that ethics, the theory of making normative judgments, is fundamentally practical. It is impossible to argue against the existence of ethics in practical life, or in his own words: "There is no getting behind ethics. (2)." Moral theories provide a guideline of how people should react to situations, characters, and behaviors of other people. "To develop an ethical personality is to become sensitive to different aspects of things, and to be disposed to use them to influence or determine attitudes, emotions, and choices (1)." Under this framework, it is possible to object to a certain type of ethics, but impossible to reject the concept altogether. Normative concepts must exist, even though there is no consensus regarding the form of their existence. Even though Marx and Engels argue that ethics are fundamentally tools of the powerful to keep their citizens obedient to their governance, and Nietzsche rails against ethics based on Christian tradition and argue for a proper aristocratic pride, they just moralize differently — substituting another set of values for then-current ones that they found problematic. Their seeming withdrawal from ethics is still a move within the field of ethics because fundamentally they simply just propose other ways of living.

Second, normative judgments are produced through an input-output system. The common inputs are actions, situations, or characters of other people; the outputs are the agent's attitudes, choices, and actions. Adopting this should not lead to the conclusions that 1) it does not presuppose how much rationality will be involved and 2) the input/output only functions in one direction and not the other. For 1), the degree of the

involvement of reason varies from case to case. In some cases, the agent just responds “without thinking,” whereas in other cases it requires more observation and imagination to represent a situation that suggests certain judgments. For 2), the way that a situation is represented can also be determined or influenced by the agent’s characters and values. Splitting the moral judgment into two parts creates a way to better assess people’s reactions and moral judgments in an intellectual way. The input is usually facts that cannot be commented on other than being true or false, but the way people react to those facts invites moral judgments. Without a distinction between these two factors, there is little room to discuss the normativity of moral judgments themselves.

Third, to give a scale between pure preferences and moral judgments, Blackburn proposes a staircase of practical and emotional ascent (9). At the bottom, the agent simply expresses likes and dislikes. Neither their response nor the situation they react to proposes an issue of ethics. Oppositely, when the agent not only emotionally reacts to a situation themselves, but also feels disposed to encourage others to share the same sentiment, it is more likely that they treat the situation as a moral issue that should require public concern. This scale is not only measured through the intensity of emotion, but also the degree of engagement, and readiness to deploy pressures on other people to conform or to change (10). Under this framework, a follow-up question is how far up, and how quickly it is appropriate to climb up the staircase and make an issue about morality. It is surprising how much aesthetic evaluations are intermingled with moral judgments if we insist on using emotional responses as a parameter to distinguish them. Blackburn discussed an example of a hypothetical proposal of printing the surface of the moon, so people will see a Coca-Cola logo when they look up at the sky at night. Some people may

like this idea, but others will strongly oppose it. On the one hand, it seems like the appearance of the moon is about personal taste and aesthetic preference, but on the other hand, it involves people's different values regarding what constitutes a legitimate reason to change the order of nature.

To sum up, the three characteristics construct the background of the discussion on the nature of normativity. First, there must exist normative concepts that can be manifested practically. Second, the agent's reactions and the situations that they react to are separate. Even though the agent is affected by a situation, their disposition to feel certain emotions in that situation is shaped over time and subject to conditioning. This fact makes the agent's emotional reactions accountable for normative judgments of other people because the agent's emotions are not completely involuntary. Third, there is no qualitative difference between the judgment of morality and aesthetic taste. It is fundamentally the judgment of the same kind, and the specific scope varies from culture to culture.

With the context being set up, Blackburn starts to lay out the theory of expressivism. As mentioned above, the normativity of an action is different from its descriptive features. An action can exist and be understandable even with no normative judgments assigned to it. For example, saving a child from a running river is itself a natural, complete action — the ethics of this action is not an intrinsic part of it. Because the ethics of an action does not just pertain to the situation itself, but also to how an agent should respond to it. Therefore, a successful theory of ethical thought must reconcile, or account for, both the natural and the normative elements of an object to which we assign moral values.

The expressivist's strategy to solve this problem is to treat ethical propositions as counters in people's transactions with each other's values (Blackburn 66). For example, if I say, or think to myself that "saving a drowning child is good," I do not assert the value of saving children; this moral judgment only expresses the attitude that I value the action of saving drowning children. It is fundamentally a description of my state of mind, rather than the concerned action. In other words, for a subject S to think that X is good, it is for S to value it, and S's value is explained in natural terms. Following this strategy, expressivists solve the problem of reconciliation by treating the normative element of X as a natural action of S.

Expressivism requires a naturalistic story of the state of mind regarding valuing something. A twofold conclusion can be supported based on empirical observations and an a priori principle of interpretation (API), which says it is analytic that creatures with beliefs, desires, and other states of mind, behave in ways that (best) make sense (and not in ways that make no sense), given those states of mind (Blackburn 55). For example, given that I am in a state of mind craving ice cream, I must behave in a way that makes sense to me with that desire affecting me: I may go to buy some ice cream to satisfy that desire or avoid looking at pictures of ice cream to suppress it. In other words, to have the desire for F is to be disposed to behave in any of the variety of ways that make sense when having the desire for F (Blackburn 57). Based on this theory, the concept of belief or desire, or other states of mind, is identified using normative terms. For example, if I buy a cup of iced latte, this action implies that I have a desire for coffee because this is what makes sense to do if I do desire coffee. My desire does not manifest itself by the mere act of me buying a latte, but by the fact that it is what makes sense to do. Suppose in

an alternative universe that the only way to address the desire for coffee is to think about it, and holding a cup of coffee in hand can offer no help, purchasing coffee can no longer offer any clues about whether or not I desire coffee.

On the other hand, the presence of these states of mind is identified empirically in terms of the causal structures visible in the actions that the subject performs, and those they would perform in other circumstances (Blackburn 58). If these two conclusions hold, a person's desires can always be traced through that person's actions. However, there also exists mismatches and situations where desires and actions do not harmonize. Below are expressivists' replies to these kinds of cases.

The first type of case is when a person says they value A, but do X, which contradicts, or is not something that makes sense to anyone who desires A. These cases can often emerge through self-deception or hypocrisy. However, this should not cause any problem to the theory above, because what the person says can only reflect their desire of "wanting to say they value A" rather than A itself. The second type of case is where a person values something without being disposed to any actions to address that value. For example, a person without a sense of taste can sincerely value a glass of good wine through the brand of the wine, rather than the delicious taste. In this case, valuing this wine still motivates some actions, such as serving the public function of grading the wine, or encouraging others to enjoy its smooth taste. In some other cases of this kind, people may want to resist their desires or feel frustrated when they act or must act in ways that hinder them to address their desires. Nonetheless, in both kinds of cases, the causal link between values and actions still exists, and values, if they really exist in a person's mind, will always manifest themselves directly or indirectly.

One strength of this approach is that it accounts for the motivational force of normative statements. A person may desire, or will be better off by performing immoral actions, but they may act otherwise and comply with their moral duties. A successful moral theory should be able to explain the source of this force that changes people's actions.

Blackburn first lays out the structure of human motivation. To determine the scope of what counts as values and desires. The term "concerns" is used to refer to almost everything that the agent cares about in a situation. Some of these concerns are serious and more important, and intuitively desires and values are these concerns rather than trivial ones. These strongest concerns engage emotions, triggering changes in human behaviors sometimes even to an extent to which they are immune to higher-order control, such as panic attacks. For other concerns, they lack motivational power because they engage in fewer emotions.

Blackburn objects to this approach because he believes concerns that seem to involve fewer emotions are not so. Long-term goals often equally engage the agent's emotions, such as fear, vanity, and pride. Even a person with "cold, calculating anger" who bends to a long-term revenge plan must have this disposition to manifest this anger if the circumstance is right. Therefore, the degree to which a concern can excite certain emotions cannot be observed or measured accurately to be useful in determining the real motivational force.

He also rejects that the motivations must come from the most reasonable concerns. Reason indeed plays a role in presenting aspects of situations that are of concern to us, and can sometimes change the agent's desires. Realizing the fact that I will

be the designated driver tonight may suppress my desire to have a drink. My reason presents a situation where I can weigh the different concerns that I have, but this cost-benefit analysis does not touch on the normativity of the action. Blackburn argues that empirically the agent does not consciously carry out this kind of analyzation in their mind, nor does this analyzation supplant the emotional basis of decision-making. The concerns appear to us as costs and benefits simply because they are likely to yield pleasant or undesirable outcomes.

Lastly, Blackburn rejects psychological egoism: human beings always act out of self-love. He argues that first, not all desires are selfish. The second-order desires of wanting to have my desire fulfilled stem from self-interest, but for these desires to exist, there have to be some first-order desires that are directed at something external.

Secondly, he makes the distinction between having a desire X and desiring the pleasure of having X fulfilled. In some cases, the latter kind of desire will not be fulfilled until the agent forgets about their own self-interest. In this situation, because the agent will gain pleasure in satisfying someone else's desires, they will be motivated to do it.

The second strength of expressivism is that the motivational force explained above lies in the agent's natural desire to perform actions and possess the quality that they love to see in other people. If moral actions are fundamentally actions the agent approves of, they will naturally be motivated to perform such actions. However, this does not mean that there is no distinction between a normal, pleasant action and a moral action, even though both actions lead to pleasure. Blackburn argues that the motivational force behind moral actions is the desire for praiseworthiness. People would like to be and to be seen as someone they admire. Let me go back to the cheating example mentioned at



the beginning of this chapter. Cheating may seem to be a natural action to perform since it provides instant utility at low risk. However, a refusal to cheat is also natural if no student who is liked or trusted by their classmates and teachers would cheat in an exam. In this case, it is unnatural for the agent to cheat, because people tend to avoid actions that they know are disliked by other members of their community. Everyone has the tendency to perform admirable actions and win other people's approbation, so if performing morally is a way to achieve this goal, it is natural for people to feel happy in acting morally. When an agent refuses to cheat, their pleasure will arise from the fact that if there were an invisible person quietly observing them, this person must admire the agent's action very much.

Because of this reason, the expressivist moral frameworks all argue that moralizing is an action of the agent reflecting on their action from another person's perspective, as an action that illustrates some qualities (Blackburn, 1998). In Smith, the "impartial spectator" is the symbol of this self-reflection process. This is also allied to our ability to occupy "the common point of view" proposed by Hume.

Under Hume's "common point of view" theory, there are two elements worth noting. First, Hume believes that "in assessing a character or the actions of a character, we should consider the interests of the person himself whose character is examined, or that of persons who have a connection with him (Blackburn, 1998). Society should aim to ensure that people seek praiseworthiness rather than just praise, and avoid blameworthiness rather than just blame, through education. Thus, it is a contingent result of socialization that most people adopt the impartial view. The content of morality is fundamentally contingent, and is a social achievement, rather than a set of rules produced

through certain processes. Second, this theory involves an ideal of civility: the requirement that in a conversation with others, we find common ground with them (Blackburn, 1998). Hume assumes that we all have an interest in pursuing conversations with others and seeing ourselves as justified.

To sum up, expressivists defend that normativity exists only as concepts, and normative statements are not assertions or beliefs, but merely expressions of sentiments or attitudes, such as approval or disapproval of a norm. Normative statements are fundamentally expressions of emotions, and reason only presents us with the situation and cannot lead us to a moral judgment alone.

## Part 2: Concerns of Expressivism

To this theory, two concerns need to be addressed. First, it is important to note that people understand external situations differently, and this difference in interpretation influences the agent's moral judgment. According to the input-output system, normative judgments are the agent's emotional reactions to external situations. The agent has at least some, if not full, control over their emotions, and this fact holds them morally accountable for their actions. The agent will be responsible for failing to express appropriate emotions in a situation where they should, but they are not responsible for what happened in that situation. Reason only concerns what happens in a given situation, so it is not what's being judged when we evaluate an agent's morality. However, if the agent fails to react to a situation with the appropriate emotion, it can be caused by either a failure to control their emotions or a misunderstanding of the situation. This distinction is important because being able to locate the source of conflicting moral opinions enables moral agents to communicate and evaluate each other's judgments.

For example, most people can agree that murders are wrong, and wrongful killings usually trigger anger. However, people also hold vastly different views regarding what kind of killings count as murder. Suppose that there is a woman who is a victim of domestic violence, and one day she kills her husband when she feels like her husband wants to abuse her again. There are two people who know the exact same details about this killing. Person A might consider such killing as an act of self-defense, while person B would think it is a straight murder. They understand the situation differently, which leads to their different emotional responses to this incident. If they would like to have a conversation about their moral judgments, excluding the possibility of a different

understanding of the situation can make them simply misjudge the other person's moral character, overlooking their consensus that murder is irritating, and harm done by self-defense can be excused.

To reply to this concern, Expressivists may argue that the situation above is just a case of agents having conflicting concerns. To person A, the concern that arises from the killer being abused may override the other concern that the killer killed, so to them, they believe that the killing is justified, and by saying that they express their approval of killings as such. Another person might weigh these concerns differently and arrive at a different conclusion.

Second, if this theory stands, it must be true that the content of moral rules can only be produced by the agent's emotional reactions to external situations. Expressivists claim that moral rules are contingent, and they are justified when people feel positively about them when they think from a common point of view. However, emotion comes and goes fast, and people have dramatically different emotional responses even to the same instance. Moral rules, on the other hand, seem to be quite stable over time and do not change with people's feelings. It is not the case that on day one I feel good to save a drowning child, so it becomes a moral action, but on the next day I do not have the same feeling about this type of incident, so it is no longer moral. Stability makes moral rules objective and convincing. Emotions, however, are short-lasting, subjective, and vary dramatically from person to person. If moral rules are fundamentally a special kind of expression of emotions, they will lack the stability that moral rules seem to have. The variety of ways people experience and express emotions also makes it difficult for them to discover the same set of moral rules.

To reply to this concern, earlier in this chapter, I explained the shared desire people have for praiseworthiness. Even though emotional reactions may vary from person to person, the two emotions: resentment and approbation, are in fact quite stable when the agent witnesses other people's behaviors. For example, suppose that I am in a good mood today, and I feel good when I see a person saving a drowning child from the river. If tomorrow I no longer have such a mood, it will be absurd that suddenly I no longer consider saving a drowning child an admirable action. It is possible that I will not feel the intense gratitude to the saver and relief seeing the child safe as I do today, but my approbation so such an action does not change with my mood. Disposition to resentment and approbation should be stable enough to serve as the basis for moral judgments. The stability worry remains at an interpersonal level. We know that a person has a stable emotional response to things they approve or disapprove of, but it is still unclear whether or not the approbation and resentment of different members of society will be excited by the same set of actions. If one action appears approvable to me, will it also be approvable to other people? Most moral rules and values are shared by a majority of society, so people should share similar sets of values to certain actions. How can expressivists account for this similarity in values from the interpersonal perspective?

In the next chapter, I introduce two sentimentalist theories through the frameworks proposed by Hume and Smith respectively. I then examine if their theories sufficiently account for the stability of emotion-based moral rules among different members of society.

## Chapter 2

This chapter is divided into two parts. In the first and second part, I present Hume's and Smith's theories respectively to as examples of expressivist moral theories. In the third part, I examine if Smith's theory adequately addresses the concern in chapter one.

### Part 1 Hume's Theory

In *Treatise*, Hume assumes that nothing is ever present in the mind but its perception, including actions such as seeing, hearing, thinking, hating, and so on (456). In our minds, perceptions resolve themselves into either impressions or ideas. The question following this assumption is whether, by means of ideas or impressions, we distinguish vice and virtue.

Hume rejects that moral judgments can be derived from reason (457). In *A Treatise of Human Nature*, Hume argued that reason is perfectly inert, and has no motivational power to cause any changes in human emotions or actions. To Hume, the only role of reason is the discovery of truth or falsehood, and nothing beyond:

Truth or falsehood consists in an agreement or disagreement either to the *real* relations of ideas, or to *real* existence and matter of fact. Whatever, therefore, is not susceptible of this agreement or disagreement, is incapable of being true or false, and can never be an object of our reason (Hume 458, emphasis added).

It follows that if moral rules are determined by reason, these rules must be judged in terms of their accuracy to facts. Moral rules will be treated as facts, and they will possess eternal fitness or unfitness of things or appears the same to every agent considering it

(457). However, this cannot be true because moral rules should influence our passions and actions, and to go beyond the calm and indolent judgments of the understanding (Hume 457). On the other hand, based on empirical observations, morality indeed affects human passions and actions naturally, motivating them to admire certain dispositions and dislike others. This change in human attitudes can never be explained through reason alone. For example, if I see a drowning child, the fact that there is indeed a child struggling in water cannot motivate me to do anything without me also having values such as “an adult should always help children get rid of dangerous situation.” Reason provides me only with the fact that there is a drowning child, but morality is the value I hold about this fact, which is not supplied by reason. Hume continues to argue that, as a practical, rather than speculative philosophy, morality not only does but *should* influence people’s emotions and behaviors (457). The source of moral judgments must provide a motivational force for humans to fulfill their duties and deter them from injustice. Therefore, Hume concludes that reason cannot be what human beings rely on to articulate moral judgments.

Two conclusions can be derived from Hume’s argument. First, actions do not derive their merit from their conformity to reason, nor blame from a contrariety to it. Briefly speaking, Hume perceives reason as something inactive and only reacts to things that are exposed to disagreements of these two kinds (458). Emotions and actions, however, are original facts, and thus do not involve any ambiguity that invites reason to get involved. For example, suppose that I feel worried seeing a child struggling in the water, so I jump into the water to save them. The facts that I feel worried and I jump into the water are not ambiguous. Reason has no problem telling us that my emotions and

actions are real. However, if a person wants to judge my reaction from a moral perspective, they need to go beyond the fact-checking and see if I should feel and act in the way I do in this situation. If my action is genuinely moral, it is so not because I actually perform it, but it is what should be performed in this situation. Therefore, this spectator cannot tell whether or not my reaction is moral through their reason alone. To generalize this conclusion, people's judgments regarding emotions and actions, such as praiseworthy or blameworthy, are irrelevant to reason.

Secondly, because moral judgments never involve reason, reason cannot be what people rely on when they make such judgments. It is true that people can derive passion in situations where reason plays a dominant role: scientists feel excited when they discover some true relations or facts. However, this cannot support the point that there are some moral passions that can be excited by reason. As mentioned above, reason only reacts to two kinds of disagreements concerning truths, and neither is what moral judgments concern. It would be absurd to say an action is praiseworthy just because it truly happened. One action is laudable or blamable, therefore, is different from it being true or false. The latter cannot prove or justify the former. In cases where the injustice is caused by ignorance or wrong beliefs, people would cease to consider the person having a defective moral character. This person's lack of certain knowledge can never be the original source of immortality. For example, suppose that in a foreign country, the word "hi" is profanity. If I am a foreigner who does not know about it and I say hi to greet a local resident, I should not be held morally accountable in this situation. If a local resident said hi to insult, their action should be considered differently from mine.



It is important to keep in mind that morality does involve reason during the discovery of morality. Morality is a special relationship between an incident and a certain moral standard (Hume 463). One may argue that reason plays a role in discovering that such a relationship exists in a moral context. However, Hume rejects this view because the discovery of such a relation requires there exists a moral standard in the first place (464). It is impossible to reject the relation to a standard that is not explained. “In such a manner of fighting in the dark, a man loses his blows in the air, and often places them where the enemy is not present (Hume 464).”

Based on the discussion above, Hume concludes that since vice and virtue cannot be discovered by reason, it must be by means of some impression or sentiment they occasion, that we are able to make the distinction between them. The question following is what the nature of these impressions and sentiments is. In his theory, Hume believes that moral good or evils are nothing but particular pleasure or uneasiness (468). To understand it from the other perspective, having a sense of virtue is then to feel the satisfaction of a particular kind from the contemplation of a character.

The pain or pleasure, which arises from the general survey or view of any action or quality of the mind, constitutes its vice or virtue, and gives rise to our approbation or blame, which is nothing but a fainter and more imperceptible love or hatred (Hume 614).

Notice that Hume does not say an action is immoral because it causes negative feelings to its observers. People react emotionally to cases that are not subject to moral judgments, so we must separate emotional reactions that are moral judgments from those that are not. For example, an ugly, dangerous monster tends to make its observers feel unpleasant, but

the monster is not morally evil simply because it causes negative feelings to the observers. In this case, the spectator's disgust fails to support that the monster is evil. Even though virtuous actions excite pleasure and approbation in people, an action being pleasant does not guarantee that it is also virtuous. Another way to reply to this worry is that Hume separates the pleasure derived from, species and appearances of objects from the pleasure derived from approbation (617).

Based on what's discussed so far, Hume argues that moral good and evil are distinguished by our sentiments, and not by reason (589). Hume's theory of moral judgments is thus a sentimentalist theory. It accounts for the morality of an action based on the sentiments of its spectators. Whenever there is a moral action, its spectator (factually or imaginatively) must feel pleasant and joy in observing it. To avoid the spectator having a certain feeling because of personal reasons, the spectator's sentiment can count as moral approval only if they contemplate action from the common point of view (Hume 590).

... therefore, to prevent those continual contradictions, and arrive at a more stable judgment of things, we fix on some steady and general points of view; and always, in our thoughts, place ourselves in them, whatever may be our present situation (Hume 581).

In a nutshell, justifying a moral statement is a three-step process. When the agent is about to make any moral judgment, first, they need adequately understand the situation by sympathizing with the actor. Second, the agent needs to consider if they can still sympathize with the actor's sentiments if they take a common point of view. Last, if the spectator approves of the sentiment through successfully sympathizing with the actor, the

agent's emotional reactions can be taken as a legitimate moral judgment. Let me put it in an example. Suppose one day Elly is punched in her face, and she believes she is the victim of violent behavior. To understand her situation, I, as the agent, need to sympathize with her resentment, and think how I would feel if I were the person who is punched. This is step one. From here, I may feel resentful as well, but my sentiment at this moment is not justified as a legitimate moral judgment yet. I need to take one step back, and think whether a third person, who is not a friend of Elly or me and has nothing at stake with this incident, can sympathize with my resentment. This is step two. When taking the common point of view, if the spectator sympathizes with Elly and me, her resentment is affirmed, and I express my disapproval of the incident she suffers through sympathizing with her resentment, which is itself a moral judgment that Elly is indeed wronged in this situation. In summary, an action or character is moral only if a spectator, taking a common point of view, can approve of it.

## Part 2: Smith's Theory

Before Smith, Hume's theory was dominant in the field of sentimentalism. Smith agrees with Hume that moral judgment is discovered and supported through the sympathy of the spectators. However, he also rejects some aspects of Hume's theory.

First, Smith argues that the spectator's sympathetic sentiment is not "passed" from the actor, but emerges from the agents' own imagination. According to Hume, the spectator's sympathy for others comes from direct observations and inferences based on the cause of the actor's emotions.

As in strings equally wound up, the motion of one communicates itself to the rest; so all the affections readily pass from one person to another, and beget correspondent movements in every human creature (Hume 575).

As for Smith, however, he argues in *The Theory of Moral Sentiments* that the agent can sympathize with the actor only through their own imagination of what they would feel if they were put in the same situation (12). This distinction is important because first, it is possible that the situation that the spectator reconstructs mentally can be completely different from the real situation encountered by the actor, which pushes us to reconsider the appropriate way to moralize through sympathy. Smith points out that people can sympathize with those whose situation is impossible to know. He raises the example that people sympathize with the dead (12). The spectator may feel miserable about that person's death, because the situation they reconstruct in their mind is a person lying in the cold grave, losing their ability to move, and deprived of sunlight — which is so unfortunate for a living human (13). However, it is also obvious that no one knows what an afterlife is actually like. Smith's example may be an extreme case. In real life, because

we all more or less share some similar experiences with the person who we sympathize with, our imagination can put together a situation very similar to the actual situation encountered by the actor. However, Smith's point indicates that for a spectator to sympathize with another person, the spectator does not need to know what the situation is as long as they can imagine one that they believe is the actual situation. Therefore, when a spectator cannot sympathize with the actor, the actor's sentiment does not to be illegitimate — since it is also possible that the spectator lacks adequate imaginative resources. This difference in the source of sympathy requires Smith to articulate a different theory from Hume's that focuses on the propriety of sympathy.

Smith starts *The Theory of Moral Sentiments* with a discussion about selfishness. He claims that even the most selfish person has interests in happiness belonging to other people, even though this person gains for himself nothing “except for the pleasure of seeing it (9).” Based on this statement, he rejects that 1) human beings are indifferent to each other's well-being by nature, and 2) human beings do not sympathize unless they could benefit themselves in doing so. Here, “selfishness” is both a moral and an epistemic issue (Griswold 77). From within a moral horizon, we ask if people care about others regardless of if doing so will earn us benefits. From an epistemic point of view, we ask if people really successfully enter into each other's world (Griswold 78). Therefore, a selfish person is either someone who cares for other people only for the sake of benefiting themselves, or a person who fails to understand another person's needs.

Smith also acknowledges that people are fundamentally separate, and we do not have direct access to another person's feelings and experiences (138). Nonetheless, we demand others' sympathy with our pleasure and pain, and we recognize in ourselves

experiences that others have undergone or that we imagine they have undergone. People also find pleasure in sympathizing and being sympathized.

The agreeable passion of love and joy can satisfy and support the heart without any auxiliary pleasure. The bitter and painful emotions of grief and resentment more strongly require the healing consolation of sympathy (Smith 15).

This "feeling-with" is what Smith calls sympathy. It is true that one can sympathize with any passion, it must be possible to "sympathize" with someone and not approve of them (Griswold 85) Therefore, not all sympathizing entails moral approval.

Even though people sympathize with others through their own imagination. Smith rejects that sympathy must be a selfish action (25). He points out that in order for the spectator to enter into the actor's world, it is not enough to just imagine a change in circumstances. Instead, they must also change in persons and characters (317). For example, in order to feel a father's pain of losing his son, the spectator does not just imagine what they, with this and that identities and personalities, would feel if they lose their son. Rather, they imagine how they would feel in this situation if they were the actor, who lost his son. Suppose that the spectator is a detached person who does not really care about their family members, while the actor does, the spectator needs to enter the actor's situation as someone who treasures their family. Only if the spectator does so, their grief is truly upon the actor's account. Because of this distinction, it seems that to sympathize with a person's own account and to sympathize with the actor's account are different. Yet, in the second half of TMS, Smith insists that one cannot sympathize with another person without doing so on their account (316). A father sympathizes with his

wife's pain when giving birth to their child, yet he could never feel this pain as himself; thus, as long as he still sympathizes, he must do so on his wife's account (Smith 316).

After laying out the nature of sympathy, Smith introduces the idea of the impartial spectator as the "standard" of how people should sympathize with external situations, besides doing so on the actor's account. Notice that here he does not assume that this concept exists as a property or a matter. It exists as a special perspective people take when they moralize — expressing values and sentiments (Griswold 129). In other words, an "impartial spectator" can be anyone who has the attitude to objectively enter into the actor's situation. I will unpack what it means to be objective below. For the sake of concision, I will just use "impartial spectator" as a shorthand for "a person who possesses such an attitude" in the following discussion.

First, the impartial spectator does not have the same emotional "investment" in the situation that the actor does, so they are relatively more detached and critical compared to the person affected (Griswold 102 and Smith 154). However, it does not mean that an impartial spectator must be indifferent or unemotional. Smith brought the example of an art critic who would fail to give a fair assessment of a drama without emotionally engaging it (Smith 35). The only kind of emotions that an impartial spectator lacks are those that will interfere with good judgments. Here, we can observe that Smith does not consider emotions as something that must be suppressed to achieve impartiality — rather, they should be educated to be able to respond to external situations to assist us in forming moral judgments.

Second, the term impartiality is often brought out by contrasting with the partiality of self-love; however, taking the impartial spectator's perspective does not

require us to act as though we had no knowledge of ourselves, or on a veil of ignorance. Smith rejects the notion that we are just a part of the moral universe or we should always sacrifice our own interests to pursue the greater interest of the state, even though many people choose to do so out of their love of virtue (116).

We are pleased, not only with praise, but with having done what is praiseworthy.

We are pleased to think that we have rendered ourselves the natural objects of approbation, though no approbation should ever actually be bestowed upon us (Smith 117).

I argue that this is one advantage of Smith's theory: it accounts for the naturalistic perspective of moral rules and their inherent motivational force. Smith believes that the greater good is not the business of ordinary individuals, and what we should focus on is the ordinary, natural moral rules that focus on specific "selves" (instead of the abstract "whole"). People's love of virtue explains the binding power of moral rules, and Smith will need no explanations and assumptions to round the circle of why people should perform selfless actions. An impartial judge will approve of our partiality toward those who properly fall within our narrow circles of sympathy (Griswold 142). Moderate partiality in some cases does not lead to moral harm and it avoids the whole moral framework being too demanding. This flexibility also does not place this theory onto a slippery slope and will one day extend itself to any cases, because the impartial spectator will always stand as a paradigm exemplifying the best case of sympathetic understanding (Griswold 149).

Hume proposes in his work that the agent should always hold a "steady and general point of view" to avoid biases due to personal experiences. Smith takes one step



further and argues that the emotions that are suitable to their objects are those an impartial spectator would sympathize with. In Smith's theory, impartiality functions as an attitude that affects the way in which we enter into the actor's situation; it does not provide its own standards or rules (Griswold 323). Even though the perspective that an impartial spectator should take belongs to a specific individual — the agent, there is still a notion of universality in Smith's theory: every person, who enters into the agent's perspective impartially, will feel what the agent feels (Griswold 197). The impartial spectator is not a morally perfect individual, but an attitude or an ideal condition when the agent wants to make any moral judgments.

Smith connects sympathy with individual dignity. Suppose any impartial spectator would sympathize with the agent's resentment for an evil action, X, to another person. The wrongdoer, when taking the impartial perspective, must also be able to realize that performing X will be disapproved by the agent. If the wrongdoer nonetheless performs it, it follows that they must unduly disregard the agent's feelings and interests by treating themselves as the moral exception. This disregard constitutes an insult to the dignity of the victim who suffers from the actor's wrongdoing, which gives them the ground to claim justice and equal respect from the actor (Smith 137).

Smith also accounts for the motivational force of moral actions. As mentioned above, Smith assumes that people are natural lovers of praiseworthiness but not merely praise (117). He claims that "the most sincere praise can give little pleasure when it cannot be considered as some sort of proof of praise-worthiness (Smith 116)." Therefore, when people are judged inadequately, by partial spectators, the merit or harm caused by the spectator's praise or blame is not that meaningful to the actor. Oppositely, an

impartial spectator sympathizes only with appropriate objects, so people have the motivation to act in a way that an impartial spectator would approve of. This love of virtue and tendency to seek affirmation from impartial spectators motivates people to follow a set of collectively accepted moral rules and justifies the impartial spectators as a standard in making moral evaluations. Moreover, moral judgments are fundamentally expressions of approbation — which leads to pleasure and pride — the agent's judgment is itself a motivational statement, since it is natural for human beings to pursue what's pleasant and honorable. The process of reflecting the actor's action through an impartial spectator can be applied to an agent's own actions, too (Griswold 108). The actors form their own moral consciousness when they act as spectators and approve of other people's actions through sympathy. Therefore, the ultimate goal of moral education is not to seek constant affirmation from other "impartial spectators," but to be able to internalize this perspective and sympathize with oneself.

Smith successfully addresses the worry that emotions are too volatile to be relied on in the context of morality. First, as mentioned above, not all emotions are relevant to moral judgments. The mere pleasure or disgust cannot be used to support any moral judgment. Under Smith's framework, only sympathy that the agent feels when taking the impartial spectator's perspective can be used to support moral judgments. This ensures that a person will not change their moral values as their moods change. Second, Smith also points out that everyone in the same society, when taking an impartial perspective, should be able to reach a consensus regarding if an action is morally acceptable. This makes it less likely that people of the same community will have completely different moral codes from one another. Third, people do have the natural motivation to take the

impartial spectator's perspective. If as Smith argues, it is true that everyone has an interest in being loved and receiving approval from other members of society, people will naturally consider how their actions will be perceived by other people and what kind of actions are praised or condemned. It explains why people are motivated to perform moral actions, and why moral rules can be imposed on people as duties.

One following question is that, is it possible for people to actually take the impartial spectator's perspective, and if doing so will always enable people to claim justice and never bury or create injustice. I will unfold this concern in Chapter 3.

### Chapter 3

In this chapter, I introduce two critiques on Smith's sentimentalist theory. First, accepting the impartial spectator as a standard to moralize reinforces the injustice buried in social norms. Second, the empathy of an impartial spectator is not sufficient to account for the natural, non-instrumental concern people have for other people. In part 2, I present Smith's replies to these criticisms.

Let me start by providing a recap of Smith's theory in a nutshell. Smith believes people have a natural tendency to care for other people's attitudes and actions. Even though we have no direct and immediate experience of what other people feel, we can address this concern through imagination — projecting ourselves in the situations of other people and “feeling with” them. Through the spectator's empathy in these imaginative states, they approve of others' sentiments if they share the same feelings as the actor. To determine if the spectator's subjective judgment can be counted as a moral judgment, the spectator must take the perspective of the impartial spectator and consider whether they would approve or disapprove the actor's conduct. The impartial spectator thus personifies a perspective free of bias and self-interest (Golemboski, 2015), serving as a justification of moral judgments. An impartial spectator should be able to empathize with any sentiments that are appropriate to the situation. Their sentiments are sincere but not overly intense to a degree that prevents them from properly understanding the situation.

The first critique is that adopting impartial spectator as a standard buries and reinforces the injustice consisting in social norms, which often include unjustified

stereotypes and internalized epistemic injustices. Two characteristics of this process make it vulnerable to this critique.

First, Smith argues that an important motivation of an agent to behave morally is to receive approbation from oneself and other people of society. This will make the values of the agent subjective to the socially accepted values. Smith is fully committed to the viewpoint that moral values are products of socialization: a person learns to shape her behavior on the basis of not only her own desires, but also the evaluation of others around her (Golemboski, 2015). When a person judges their own actions, they do so by imagining how their actions appear in other people's eyes based on the agent's interactions with other people in the society. Through this reflection process, people worry that the agent's moral values will be shaped by prevailing social norms. For example, in a society where most people deny that there exists sexual assault between a married couple, the uneasiness and pain I feel in the sexual behavior with my partner can hardly be empathized. In this society, if I share my experience with people around me, and all of them, who process my case not as my friend but as an impartial spectator, fail to empathize with my distress, it will motivate me to deny my sentiments and never reconsider the morality of unwanted sexual behavior in marriage. Under the name of impartiality, the problematic social norms are buried and supported by the beliefs belonging to the majority of the society.

In other cases, the unpopular viewpoints are not only suppressed but also ignored. As explained above and in chapter two, sympathizing is fundamentally a process of imagination — the spectator mentally reconstructs the situation before they reflect on how they would feel if they were the person principally affected. In this process, it is

inevitable that they must treat the constituents of the situation differently. For example, if I see Edie crying at her sister's funeral, and I want to empathize with her sorrowness. I then imagine what I would feel if I were at my sister's funeral. When I mentally reconstruct the situation, the facts that I am at a funeral and my sister is dead are far more important than the weather on that day or the color of the attendants' shoes, even though the latter two elements are also present in the situation. An impartial spectator must be able to prioritize the important factors from those trivial ones, or feel strongly only to the factors that deserve such reactions. Going back to the sexual assault example, an impartial spectator will take the fact that I am married into account when they reconstruct the situation, when in another society this factor is irrelevant.

So far, we can see that Smith's theory leaves behind two important characteristics of an acceptable moral standard. First, approval of other members of the society is not a necessary condition for an action to be morally acceptable. One reason is because social norms and values are not established democratically. The powerful and authorities have an unjustifiably large influence on the content of them. In China, there has been no wins of women who charged their partners with sexual violence or rape in court. On the one hand, most of the representatives in Congress and judges are men who cannot relate to the sentiment of women who feel oppressed and insulted at home. They empathize more with men worrying that their wives will cry rape as a threat. On the other hand, because of the traditional culture and mindset that sexual assault can only be done to single women, many married women have no language to phrase their sentiments, which hinders their communication with other members of the society. Therefore, even people

try very hard to empathize with these women, they cannot relate to them, and these women fail to claim their rights and are criticized as being not objective.

Second, some moral rules are inherently perspectival. The injustice of these immoral behaviors can be adequately comprehended only through a non-objective, particular perspective. For example, the fear that women have when they walk alone at night is not the same as the common fear of robbery that is shared by every member of the society. The injustice of this fear not only comes from the inadequate city security, but also from the fact that women are more vulnerable to this kind of violence. If a woman were robbed on the street, it would be insufficient to sympathize with her objectively — the spectator must be able to enter into her situation as a woman. A spectator must be able to not only feel her fear as a victim of robbery, but also her vulnerable position of living as a single woman. In this case, the spectator needs to be “biased” in favor of women to base their normative judgment on a complete picture. Another example is affirmative action. If the spectator has no knowledge about the history of racial or gender oppression, affirmative action may be seen as treating people differently on the basis of race or gender. If a white person argues that they are treated unfairly under the affirmative action policy, an impartial spectator should understand the not empathize with their resentment. In cases like this, the spectator not only needs to distance themselves from the actor, but also take the correct perspective: the one that can best address the normativity of an action.

From the critique above, using the impartial spectator as the sole justification for normative judgment risks falling into the trap of systematic injustice. A Smithean reply may look like this: when people from the dominant class are not aware of this injustice

or have no interest in changing it, it is hard for the ordinary people to picture the full context of the situation. Problematic social norms block the proper functioning of sympathy, preventing the moral agents from successfully taking the impartial spectator's perspective. Smith is not unaware of this critique, he notes that there is a chance that social standards will diverge from true moral standards (Golemboski, 2015).

In estimating our own merit, in judging of our own character and conduct, there are two different standards to which we naturally compare them. The one is the idea of exact propriety and perfection, so far as we are each of us capable of comprehending that idea. The other is that degree of approximation to this idea which is commonly attained in the world, and which the greater part of our friends and companions, of our rivals and competitors, may have actually arrived at (Smith 247).

To address this possibility, Smith argues that “the wise and virtuous man” concerns himself with the true moral standards, if “so far as we are each of us capable of comprehending that idea” (Smith 247). In other words, Smith can reply to this critique by revealing possibility that people who seems to take the impartial perspective are, in fact, still partial. In the example above, the husbands occupy a more advantaged position in family through stripping women of their legal rights and bodily autonomy. Therefore, males have an interest in maintaining the status quo and stay indifferent to the oppression faced by females. This fact does not change no matter whether the male spectator benefits from or is aware of their privilege. Because of this reason, the husband's perspective cannot be taken as an impartial perspective, and the impartial perspective is not indifferent to gender in this case. In order to be impartial, the moral agents must take the



perspective of someone who is not invested in the system. A spectator's "biased" sympathy does not render their perspective as partial.

However, in a society governed by the powerful that praises wealth and ranking, it is very difficult to enter into the impartial spectator's perspective. First, the system of domination prevents relevant us from impartially sympathizing with the marginalized group. Their voice is oppressed, which blocks ordinary people from accessing important facts of injustice. I argue that people can realize that the perspective of status quo is inherently partial after they discover facts that were overlooked in the past. These ignored facts will change their mental representation of the situation, thus motivating them to feel differently and arrive at a different normative judgment. Let us go back to the example of the robbery suffered by single women. When the first spectator reconstructs the situation and judges their reactions from the perspective that seems to be impartial, they may feel resentful only to the fact that the government and the police fail to protect its citizens. However, as such cases happen more frequently, the spectators will observe the pattern that females are more likely to be the target of injuries and bullies. With this piece of information, when they witness another robbery, they will not only resent the robbery, but also the situation that females are more vulnerable to crimes simply because of their gender.

Suppose some spectators still do not believe that this pattern would require them to take the gender of the victim into account: because they think the robbers do not target women, they target anyone who is physically weaker than themselves. Gender injustice is still buried even when the situation is perceived differently. However, the situation will certainly be understood differently when they discover that women who are physically

strong are still attacked on the street. If such gender-ed injustice does exist, it will be exposed in cases where there is no interference of other factors. When most people have no excuses to refuse to face this sexual injustice, they will agree that the gender of the victims matters when they make moral judgments. Only if there are more people who can discover the complete context of such injustice, these situations can be adequately reconstructed by the spectators.

On the contrary, in a society where women occupy a low social status and are not treated equally as men, the society will be less likely to pay attention to the lives of female. The fact that females are more often victims of illegal conducts may never be discovered or taken seriously in a moral context. Without such awareness, it is hard for people to update their background knowledge regarding the context of this injustice — it is then impossible to take the impartial spectator's perspective and properly sympathize with the person principally affected. Furthermore, if the perspective taken by the powerful and majority of the society is never scrutinized, and treated as the impartial perspective. Those people who rightly doubt this value will be stigmatized as partial, which makes it even harder to claim justice through sympathy.

The second critique on Smith's theory is that he assigns a double duty to the series of mental operations that issue in empathy, treating it as the source of both our understanding of other people and our non-instrumental concern for them (Bailey, 256). In the rest of this chapter, I present Bailey's critique and Smith's reply to it.

Birley argues that the empathetic mechanism Smith proposes does generate an accurate understanding of the other, but this understanding will not be able to generate a

selfless concern for them. The sympathy the agent have to the actor will also not be appropriate to use in a moral context (256).

To begin with, Smith explicitly states that empathy is indeed the faculty we rely on to understand others' situations and motivates us to have selfless concern for them. In TMS, he states that “as we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation (Smith 9).” This is one crucial difference between Smith’s and Hume’s sentimentalism theories. In the section right after, Smith also states that spectators can empathize with the passion of other people, so they are naturally motivated to know and care about the emotions of others. Their pleasure makes us happy, and their pity equally affects us and makes us want to address this unpleasant feeling even when we are not technically trapped in the same situation:

However selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it. Of this kind is pity and compassion, the emotion which we feel for the misery of others, when we either see it, or are made to conceive it in a very lively manner (Smith 9).

According to Smith’s theory, “pity and compassion” arise when the spectator witnesses other people’s happiness and misery, and the spectator conceives these emotions vividly. Smith describes that the pity and compassion are “words appropriated to signify our fellow-feeling with the sorrow of others... to denote our fellow-feeling with any passion whatsoever (Smith 11).” From there, Smith proposes the impartial spectator to make our

empathy appropriate and can be relied on to make moral judgments. Therefore, empathy indeed has two duties at the same time: facilitate our understanding of other people's situations and motivate us to care for their well-being.

The critique is that the empathetic feeling we have when caring for others cannot be used to adequately understand other people's situations or relied on to justify any moral judgments. To put it differently, the empathy in us when we take the impartial spectator's perspective is not the same as the empathy we feel when we care for other people selflessly (Bailey 262).

The kind of empathy appropriate for making moral judgment must be made through the impartial spectator's perspective, and it is not an actual feeling, but a conception of it (Bailey 261). However, to adequately empathize with another person, the spectator has to not only imagine what if it is themselves who are in the same situation, but also what the actor, with their personalities and history, would feel in that situation. For example, I saw Edie crying at her sister's funeral, mourning her sister's death, and I want to empathize with her misery. To do this, I should not only imagine what I would feel if I lost my sister, but also if I were Edie, how I would feel if I lost my sister. The empathy in my chest when I substitute myself with Edie is different from the empathy I know I should feel when taking the impartial spectator's position. The former is my actual feeling, which is required for me to have basic concern for others and is not an appropriate sentiment to support any moral judgment (Bailey 262). However, the latter is a conception of a feeling, which justifies moral judgments, but is inadequate to express my care and concern to Edie.

The actual feeling and the conception of it are two different emotions. Going back to the funeral example, I suppose that Edie is sad about her loss of sister. In misery, Edie may think: I lost my sister. If I were an impartial spectator, to empathize with her, I may think: Edie lost her sister. However, these two feelings are not completely interchangeable, because for Edie, what is important is that Edie lost *her* sister, not the fact that someone has lost their sister. Suppose that in an alternative universe, Edie forgets that her name is Edie. “I lost my sister” will still be her losing her sister. However, “Edie lost her sister” will have a different meaning to her. These two thoughts and sentiments do not differ only in terms of their intensity, but one is actual experience, but the other is the product of imagination.

Smith may push back this critique by making the distinction between common sympathy and sympathy as a moral sentiment. In the example above, the spectator’s sympathy Edie’s sorrow is not a moral sentiment. Only the sympathy that is from the perspective of an impartial spectator can be used to indicate moral approvals. When Edie have suffered a great loss, what they expect is not my moral approval, but me sharing with her feeling as an emotional support.

Furthermore, this critique is not applicable to moral sentiments. In Chapter 2, I offer an example of Elly who is irritated because she believes that she is wrongly harmed. According to Bailey’s framework, to sufficiently show my care, I need to sympathize with her resentment and think what if “I am the victim of a violent behavior.” To enter into the impartial spectator’s perspective, I need to sympathize by imagining “someone is the victim of a violent behavior.” In both cases, my sentiment toward unjustified violence is the same, and it does not change no matter who is the victim of such injustice. In this

case, by expressing my resentment to the violent behavior, I handle Elly's resentment with care, and I also successfully enter into the impartial spectator's perspective.

In conclusion, the standard of impartial spectator does adequately address the stability and objectivity concerns people have to sentimentalist theories, but it comes with a price of burying systematic injustice. The impartial perspective is also particularly hard to enter into in a society where people's moral sentiments are corrupted by vanity.

Smith's theory does offer an explanation to the second critique. The moral sentiment people naturally have toward other people's circumstances is no different from the moral sentiment of the impartial spectator.

## **Conclusion**

In this thesis, I present Hume's and Smith's sentimental theories as examples of a special type of moral framework that seeks normativity from human sentiments. They believe that a moral rule is fundamentally a statement of moral approval, which is a human emotion. Every normative judgment is an expression of approbation or resentment, and this statement becomes a moral rule if an impartial spectator can sympathize with the moral agent's sentiments.

Based on volatile human emotions, sentimental theories account for the stability of moral rules through separating common emotions, such as joy and hate, from moral sentiments. A person's mood changes, but their disposition to feel certain moral sentiments stays stable. Members of society express and experience emotions differently, but they come closer to the situation through sympathy and a natural care of others.

Moral rules are convincing and possess a motivational force if they are approved of by an impartial spectator, who is free of bias and understands the situation thoroughly. Their judgments have the intrinsic motivational force because of people's natural love of virtue. The desire to be trusted and admired motivates people to act in a way that they would like to see in others. In ideal situation, an agent will be able to scrutinize their own actions and sentiments impartially, so they always know what actions are moral, and they feel joy in acting in accordance with their moral code.

The concept of impartial spectator also faces the challenge of reinforcing systematic injustice. Due to the power imbalance, the legitimate interests of the marginalized group are often ignored by the general public. The influence of the dominant class, and their vested interest in continuing this oppression prevents people

from accessing the full context of many situations. This hinders people from entering the impartial spectator's perspective, and thus lead to problematic moral judgments.

Furthermore, because of people's pursuit of wealth and power, this vanity corrupts the moral sentiment of many people. They are no longer disturbed by injustice suffered by other people, which pushes them further away from understanding the beauty and honor in morality.



## Bibliography

Bailey, Olivia (2016). Empathy, Care, and Understanding in Adam Smith's Theory of Moral Sentiments. *The Adam Smith Review* 9.

Blackburn, Simon. *Ruling Passions: A Theory of Practical Reasoning*. Clarendon Press, 2009.

Copp, David and Justin Morton, "Normativity in Metaethics", *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2022/entries/normativity-metaethics/>.

Griswold, Jr, C. (1998). *Adam Smith and the Virtues of Enlightenment* (Modern European Philosophy). Cambridge: Cambridge University Press.  
doi:10.1017/CBO9780511608964

Hume, David. *A Treatise of Human Nature*. Clarendon Press, 1739.

Johnson, Robert and Adam Cureton, "Kant's Moral Philosophy", *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2022/entries/kant-moral/>.

Smith, Adam, et al. *The Glasgow Edition of the Works and Correspondence of Adam Smith*. Liberty Fund, 1982.