Consequentializing Deontology

CHAN, Pui Yee June

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Philosophy

in

Philosophy

The Chinese University of Hong Kong

February 2013

Abstract of Thesis Entitled: Consequentializing Deontology

Submitted by   CHAN, Pui Yee June

for the degree of   Master of Philosophy in Philosophy

at The Chiense University of Hong Kong in February 2013

Discussions in basic ethical problems are often framed by the essential differences between consequentialism and deontology – two fundamental theories in normative ethics. Most arguments in those ethical problems are basically reiteration of how the two theories differs from each other: Deontology holds that consequentialism is too lax as it allows all actions that leads to the best outcome, while consequentialism holds that deontology is essentially paradoxical because it forbids agents to act against a constraint even when doing so can avoid more violations.

My interest in this thesis is to examine a possible alternative in characterizing the differences between consequentialism and deontology, namely the doctrine of "consequentializing deontology". This doctrine holds that all deontological theories can be given a representation in consequentialist form.

This thesis consists of four parts. In Chapter 1, I will first examine three essential features of consequentialism, namely it is structurally axiological, teleological and maximizing. Then I will examine various formulations of deontological constraints and argue that they are best formulated as agent-relative reasons for action. In Chapter 2, I will explicate the first attempt to consequentialize deontological constraints by assigning a negative weighing to any violation. I will show that this attempt is not satisfactory because it entails a number of implausible claims. In Chapter 3, I proceed to examine various accounts that explain the normative power of deontological constraints in terms of values. Following Louise (2004), I argue that deontology can be consequentialized by giving a consequentialist representation to deontology, so that when an agent acts upon a deontological constraint, he is maximizing values that are both agent-relative and temporal-relative in nature. In Chapter 4, I will examine possible challenges to consequentializing deontology and respond to them.

論文摘要

　　後果論（Consequentialism）和義務論（Deontology）是規範倫理學裏兩個基本理論。兩個理論之間的差異，經常成為討論基本倫理問題的框架。一方面，義務論認為後果論道德上容許任何能達到最好結果的行為，這是過於寬鬆。另一方面，後果論認為義務論是自相矛盾的；因為它既禁止某些行為，視之為不道德，但又不容許人們以違反義務為手段，整體減少這類行為發生的次數。

　　這篇論文的旨趣是研究用後果論的理論框架，來表達義務論的可能性。這將提供一個新的視角，以了解後果論和義務論的基本理論差異。

　　全文共分四個部分。第一章，我會檢視兩個理論的一些基本特徵。第二章，我第一次嘗試用後果論的理論框架來表達義務論。方法是給違反義務的行為分配一個負面道德價值。然而，這方法不能成功把義務論表達為一種後果論。因為它引申了一些義務論不接受的道德判斷。第三章，我會檢視義務論和道德價值之間的關係；並順著 Louise (2004) 提出的理論，論證後果論的理論框架可以用來表達義務論。方法是把遵從義務的行為視為把時間和行動者相對 (time-relative and agent-relative) 的道德價值最大化。第四章，我將嘗試回應對這理論可能提出的反駁。

Table of Contents

Chapter 1.   Delineating Consequentialism and Deontology

1.1   Consequentializing Deontology: A Possible Alternative to Characterizing the

            Differences between Consequentialism and Deontology

Consequentialism and non-Consequentialism is a fundamental distinction in categorizing normative ethical theories. Consequentialism includes theories like utilitarianism, egoistic hedonism, act-consequentialism and rule-consequentialism. Non-consequentialism includes theories like deontology, contractarianism and virtue ethics. Among various theories, this thesis will limit itself to act-consequentialism and deontology as representing, respectively, the consequentialist and non-consequentialist approach to provide agents with moral reasons for individual actions. [1] In what follows, "consequentialism" means "act-consequentialism" and "constraint" means "deontological constraints" unless specified.

As a typical textbook on the subject will have it, consequentialism and deontology are commonly considered as rivals that attempt to provide alternative answers to basic problems in normative ethics, such as those concerning the fundamental principles of morality. They are thus often defined or characterized in terms of their difference in fundamental respects. Take for example, the relation between the good and the right. Consequentialism is said to hold that the good is prior to the right, while deontology is said to hold that the right is prior to the good. Or take moral constraints for example. Deontology is said to hold that morality includes constraints as a basic feature (or "factor" in Shelly Kagan's word), while consequentialist are characterized as those who see no point in taking a constraint into

---

[1] All theories in normative ethics would address the problem of moral reasons for action. However, both act-consequentialism and deontology approach the problem in a relatively more direct way. Consider rule-consequentialism, contractarianism and virtue ethics. They take rules, contractual agreements and virtues to carry normative status of moral rightness and wrongness. Rightness and wrongness of individual actions are *derived* from their relations to those entities. While utilitarianism and egoistic hedonism both determine normative status of an action directly base on their respective notion of utility, they both commit to claims that are not compatible with other consequentialist theories. Thus, act-consequentialism serves as a better candidate for representing consequentialism in general. The above list does not exhaust all possible theories in normative ethics. But I hope it gives grounds to the choice of act-consequentialism and deontology as the focus in this thesis. This thesis also assumes deontology is "constraint-based". That is, every deontological theory takes the form of a collection of moral constraints, each states a particular action-type is required or forbidden.

account when, all things considered, such constraint interferes with the pursuit of the greatest good.

My interest in this thesis is to examine a possible alternative for seeing the connections and differences between consequentialism and deontology, namely the doctrine of "consequentializing deontology". Proponents of consequentializing deontology hold that all deontological theories can be represented in consequentialist form. This introduces a possible alternate outlook of the dynamics between the two theories. If consequentializing deontology is successful, some classical formulation of their differences may collapse. For example, if a deontological theory can be consequentialized, a committed deontologist and a loyal consequentialist would always *act* perfectly in the same way in identical situations.

The rise of consequentializing deontology is accompanied by the introduction of a distinction between agent-relativity and agent-neutrality. The distinction has continuously gained popularity since it was first introduced by Derek Parfit (1984), who applied it to normative theories. The distinction can also be applied to values, principles and reasons for action. Often, various applications of the distinction are used to formulate differences between consequentialism and deontology.

Parfit applies the distinction to theories and holds that deontology is agent-relative because it gives different agents different aims, while consequentialism is agent-neutral because it gives all agents the same moral aim.

> Since *C* gives to all agents common moral aims, I shall call *C agent-neutral*. Many moral theories do not take this form. These theories are *agent-relative*, giving to different agents, different aims (Parfit 1984, 27).

On the other hand, McNaughton & Rawling (1991) apply the distinction to reasons for action. They hold that deontology gives agents "agent-relative reasons" because each of them is told not to violate moral constraints *themselves*, while consequentialism gives agents "agent-neutral reasons" because they all act for *the good* in consequences, which is the same to all agents.

On one hand, above applications of the agent-relative/agent-neutral distinction in characterizing the differences between normative theories often results in a *dichotomy*

between consequentialism/deontology. On the other hand, there are philosophers who apply the distinction in support of consequentializing deontology and argue against the dichotomy.

For example, Sen (1983) and Portmore (2001, 2007, 2009) apply the distinction to *perspectives*, and argue that consequentialism is compatible with agent-relative perspectives. They hold that it is not necessary for consequentialist to view consequences of actions in an agent-neutral way, i.e. in a way that does not vary from agent to agent. Take for example egoistic hedonism, which holds that an act is right if it leads to the greatest pleasure for oneself. In evaluating different consequences, egoistic hedonist takes an agent-relative perspective and recognizes agent-relative values. But holding an agent-relative theory of good does not by itself make a theory non-consequentialist. The theory still maintains the core consequentialist claim that moral rightness of an action solely depends on its consequence. If it is true that consequentialism is compatible with agent-relative perspectives, they argue, then deontology may be seen as consequentialist theory that requires agents to always minimize their own violations of constraints in consequences. Louise (2004) takes a similar approach to arguing for consequentializing deontology by applying the agent-relative/agent-neutral distinction to *values*. She holds that deontology falls under "the consequentialist umbrella" with a theory of good that recognizes agent-relative values.

If consequentializing deontology is successful, and deontology can be represented in consequentialist theory, there will be a new way of seeing the connection and differences between the two theories. Take, again, the example of moral constraints. As Kagan suggests, constraint is a moral *factor*, which is basic and supposed to be irreducible to other moral factors. If even a steadfastly loyal consequentialist can in principle act morally in just the same way as his loyal deontological counterpart does in any identical situations – in other words, if deontology can be consequentialized, and be seen as "extensionally" the same as consequentialism – then it seems plausible to think that constraints can be cashed into *the good* – provided a correct formula of exchange.

My aim in this thesis is to examine whether consequentializing deontology is plausible, and to explore its conditions, limitations and implications. My thesis is that

consequentializing deontology is plausible, namely, deontology can be represented as a maximizing consequentialist theory in pursuit of agent-relative and temporal-relative values. However, this type of consequentialist theories would be different from traditional act-consequentialism because they recognize agent-relative values, while the later recognizes agent-neutral values only.

Some proponents of consequentializing deontology go so far to claim that *all* moral theories, not only deontology, *are* fundamentally consequentialist. I believe that the consequentializing deontology does not support this claim. To argue that two theories can be represented in the same form is not the same as to argue that they have identical justification models, which is also an important feature to characterize a theory. Furthermore, the approach of consequentializing does not apply to all theories in normative ethics. Take for example those that endorse genuine moral dilemma. They are in contradiction with consequentialism. The former holds that *all* alternatives open to an agent could be morally wrong. But it is a basic consequentialist claim that the action that leads to the best consequence is morally right. However, I will show that these theories that cannot be consequentialized are extreme cases. They do not undermine the significance of consequentializing deontology, if it is successful.

1.2   Thesis Overview

This thesis is composed of four chapters.

Chapter 1 delineates basic characterization of consequentialism and deontology. Following Carlson (1995), I will explicate some of the essential features of consequentialism, namely it is structurally axiological, teleological and maximizing. "Axiological" means that consequentialist theories define moral rightness in terms of *values*. "Teleological" means consequentialist theories have moral rights depend solely on *outcome*. "Maximizing" means consequentialism has theory of right that directs agents to always *maximize values in outcome* they bring about. They serve as necessary condition in theory structure for a theory

to be "consequentialist".

Various consequentialist theories differ in the additional claims they hold. In particular, they vary on what values are morally relevant and should be maximized, or how to define morally relevant outcomes. For example, utilitarianism as a consequentialist theory holds that only pleasures and pains in the consequences are valuable, while other consequentialist theory may hold a different list of values, like knowledge and freedom. Take for example, utilitarianism holds that moral rightness depends on the consequences for all people, while egoistic hedonism holds that what is right depends solely on the consequences for the ego.

Traditionally, consequentialism is understood as "agent-neutral" in three related senses: (1) consequentialism recognizes only agent-neutral *values*; (2) evaluation of consequences does not depend on any agent's *perspective*; and (3) consequentialism only generates agent-neutral *reasons for action*. However, it is arguable that "agent-neutrality" is a necessary condition for a theory in normative ethics to be "consequentialist". Take for example, egoist hedonism. It requires agents to maximize agent-relative values. Whether the values are those objectively "good-for-him" or subjectively "good-to-him", a loyal egoistic hedonist is maximizing values that are relative to him. A theory in normative ethics may be *implausible* due to agent-relative values that it recognizes, but this implausibility does not seem to change the nature of a theory, from implausible consequentialist to implausible non-consequentialist.

Regardless of the many proponents (e.g. Sen, Dreier, Portmore and Louise), the idea of "agent-relative consequentialism" is still controversial. In this thesis, I will try to examine plausibility of consequentializing deontology base on both traditional and new idea of consequentialism. Hopefully, this thesis can contribute to the debate by arguing for a successful consequentializing deontology *based on* agent-relative consequentialism.

I then examine the central feature of deontology, namely "deontological constraints". Deontological constraints restrict agents from acting in a certain way, regardless of their pursuit of personal or common goods. They give agents *deontological reasons for action* to act in certain way. There are various formulations of deontological constraints base on the

agent-relative/agent-neutral distinction. These formulations can be taken as two different approaches to characterize deontological constraints: (i) the general form approach and (ii) the reason-statement approach.

The general form approach is represented by the formulation proposed by Nagel (1970). He suggested a universal formula for reason that "Every reason is a predicate R such that for all persons p and events A, if R is true of A, then p has prima facie reason to promote A." (Nagel, 1970, 47) When an agent acts under a deontological constraint, he is acting on a deontological reason for action. Such reason, as described by Nagel, is a one that when expressed in the form of the basic formula, the predicate R will contain a "free-agent variable" (in Nagel's word), meaning a referencing to the agent. Thus, the reason is a subjective one. Nagel later uses the term "subjective" and "agent-relative" interchangeably.

The other approach is the reason statement approach. It differs from general form approach in the way that it does not determine whether a deontological reason is an agent-relative or agent-neutral base on any "general form" or "principle" of which the deontological reason for action itself represents as a token. The idea to characterize deontological reason in a reason statement approach is to focus on the description of each particular reason for action itself.

The reasons-statement based approach is represented by the formulations of deontological constraints proposed by Pettit (1987) and McNaughton & Rawling (1991). Pettit holds that a deontological constraint is "an agent-relative reason that cannot be fully specified without pronominal back-reference to the person for whom it is a reason." (Pettit, 1987, 75) On the other hand, McNaughton & Rawling (1991) introduced a formal structure to describe reasons for action: (x) (x S […]), translate as "for all x, x should ensure that…" A reason is agent-relative if and only if there is an occurrence of "x" in the square brackets bound by the initial universal quantifier.[2]

---

[2] McNaughton & Rawling acknowledge Stephen Darwall's idea that Nagel (1970) employs a similar technical apparatus. However, they argue, their way of describing reasons allows them to make distinctions more fine-grained than Nagel's, thus enabling them to give a more palatable account of deontology than that entertained by Nagel (1970, Chapter 10). See footnote no. 6 in McNaughton & Rawling (1991). In this chapter, I will illustrate why the account of deontological constraint suggested by them works better than Nagel's and

I will argue that among these accounts, deontological constraints are best accounted as agent-relative reasons for action, in the way proposed by McNaughton & Rawling (1991). Their formal structure of describing reasons for action enables a fine-grained differentiation between three types of agent-relative reasons for action and their agent-neutral counterparts: (i) author-agent-relative; (ii) patient-agent-relative; and (iii) author-patient-agent-relative. This tripartite differentiation allows us to have a deeper understanding of different kinds of deontological constraints.

Having delineated the basic ideas of consequentialism and deontological constraints, we are ready for the study of consequentializing deontology. In Chapter 2, I start examining a simple and direct form of consequentializing deontology base on two fundamental ideas: (1) deontological constraints as agent-relative reasons for actions; and (2) consequentialism as traditional agent-neutral act-consequentialism, which holds that evaluation of consequences does not depend on any agent's perspective.[3] This chapter limits itself to the traditional conception of consequentialism, which excludes agent-relativity in any aspect.

The method of consequentializing deontology in this chapter is to assign a negative weighing to any violation of constraint. Consequences involving a violation would thus rank sub-optimally against its counterpart, which does not involve any violation. This arrangement ensures that actions involving violations of constraints will never bring about the best consequences.

However, as I will show, this method of consequentializing deontology fails because it entails several claims, that are either implausible itself, or not readily accepted by the original deontology being consequentialized. These problematic claims make it difficult for the consequentializing deontology to claim that it successfully generates a representation of the original deontology.

The difficulties faced by this first attempt in consequentializing deontology reveal that

Pettit's.

[3] In Chapter 1, I try to show that for a theory in normative ethics to be counted as "consequentialist", it requires only structurally "axiological", "teleological" and "maximizing". "Agent-neutrality" is not a necessary condition for consequentialist theories. However, since the idea of "agent-relative consequentialism" is controversial, I will first show the plausibility of consequentializing deontology with a more traditional understanding of consequentialism, namely, "agent-neutral consequentialism".

for consequentializing deontology to be successful, consequentialism needs to be compatible with agent-relativity. This leads us to explore the plausibility of the controversial idea of "agent-relative consequentialism". As stated above, traditional consequentialism is understood as "agent-neutral" in three related senses: (1) consequentialism recognizes only agent-neutral values; (2) evaluation of consequences does not depend on any agent's perspective; and (3) consequentialism only generates agent-neutral reasons for action. "Agent-relative consequentialism", if plausible, could also be understood in three related senses: (1) it recognizes agent-relative values; (2) it allows subjective evaluations; and (3) it may generate agent-relative reasons for action.

In Chapter 3, I will examine a different approach of consequentializing deontology in the context of agent-relative consequentialism. As listed above, there are at least three aspects for a consequentialist theory to be agent-relative. Different combinations of the three aspects will generate different forms of agent-relative consequentialism. In order to focus on the aim of this thesis, namely, to examine the plausibility of consequentializing deontology, I would not go over all possible forms of agent-relative consequentialism. Instead, I will only focus on the agent-relativity of values because the other two (agent-relativity of evaluation perspective and agent-relativity of reasons for actions) can be derived from it in theory.

There are two further reasons for this focus. First, as the aim of this thesis is to examine plausibility of consequentializing deontology, it serves the purpose to establish a form of successful consequentializing deontology. To do so, I try to demonstrate that, when consequentialism recognizes agent-relative values, it can generate theoretical representations for all deontology. Second, consequentialism is made up of two parts: a theory of right and a theory of good. It is a common understanding that consequentialism does not limit itself to particular theory of good – as any textbook in the subject may suggest. Introducing agent-relative values, it seems to me, is a plausible move departing traditional consequentialism. It serves our current purpose if such move can be shown as successful in

consequentializing deontology.[4]

In Chapter 3, I will examine consequentializing deontology in the context of consequentialism that recognizes agent-relative values. I will start with different formulations of deontological constraint in relation to values. Nagel (1986), Korsgaard (1993) and Pettit (1989, 1991, 1997) discussed about the relation between deontological constraints and values. Nagel points out a generally accepted idea that deontological constraint is not grounded on agent-neutral values. The reason is that any reasons for action grounded on agent-neutral values would be agent-neutral. However, deontological constraints are agent-relative reasons for actions. Thus, deontological constraint is not rooted in agent-neutral values.

However, it does not entail that deontological constraint is based on agent-relative values. In fact, Nagel has not specified the case. The reason is, reasons for actions rooted in agent-relative values do not seem able to account for the normative power of constrains.

Consider a type of agent-relative reasons, namely "reasons of personal fulfillment". An agent has a reason to climb a mountain or practice playing the piano because the agent *values* the activities. A reason of personal fulfillment is no longer a reason for action if the agent ceases to value those activities. But it is not the case in deontological constraints. By its nature, constraints do not cease to have normative power over an agent's actions even when the agent does not value acting in accordance with the constraints. This is also the difference between "options" and "constraints". Reason of personal fulfillment is a kind of options that agents are free to act on, but constraints are moral requirements.

Korsgaard (1993) offers a different view on deontological constraint and values. She

---

[4] I also choose to focus on agent-reality in "values" rather than agent-relativity in "evaluation perspective" because the effect of allowing "agent-relative evaluation" in consequentialism is by and large taken care of, by recognizing "agent-relative value" in consequentialist theories. Evaluation of consequences is basically a function of values in the consequences. I do not mean that agent-relativity in "value" is equivalent to that in "evaluation perspective". It is open to discussion how agent-relative values in a state of affairs could be evaluated *agent-relatively* in a consequentialist theory. Take, for example, my health is an agent-relative value. To rank consequences that lead to promotion and reduction of this agent-relative value objectively, I should be always promoting what is best to my health. However, to rank consequences subjectively, it would be a second-order question as how much I value my health. This involves a length of discussion that cannot be dealt effectively in this paper. My assumption is, this second order evaluation can be translated into some kind of agent-relative values (say, health of x as valued by x), so that consequentialist evaluation can be consistently agent-neutral. It helps the discussion between different consequentialist theories be limited to their respective theory of goods.

argues that deontological constraint is based neither on agent-relative values, nor on agent-neutral values. She holds that, deontological constraint is inter-subjective reason that grounds on inter-subjective values. I shall argue that the idea of "inter-subjectivity" is flawed in characterizing deontological constraints.

Lastly, Pettit (1989, 1991, 1997) proposes to characterize the differences between consequentialism and deontology as one between promoting and honoring values. According to Pettit, deontology requires agents to personally honor certain values, even in the cost of not promoting it. Following McNaughton & Rawling, I will show that the promoting/honoring distinction collapses. With proper specification, honoring an agent-neutral value is promoting an agent-relative value.

After reviewing different formulations of deontological constraint in terms of values as proposed by Nagel (1986), Korsgaard (1993) and Pettit (1989, 1991, 1997), I will try to show that there is a difficulty in accounting normative power of deontological constraint in terms of values, be it agent-relative or agent-neutral. Neither of them seems to render satisfactory explanation for the normative force of deontological constraints on agents' actions. I will then turn to McNaughton & Rawling's argument that it is not necessary for a reason to be grounded on any kinds of values.

Re-examining the aim of this thesis, I shall argue that the difficulty in grounding constraints in agent-relative or agent-neutral value is not a challenge to consequentializing deontology. The difficulty in formulating deontological constraint as reasons based on values (be it agent-relative or agent-neutral) does not entail that consequentializing deontology is implausible. It is because, as suggested by Louise (2004), the idea of consequentializing deontology is to offer theoretical representation of deontology in consequentialist form. The result, if successful, is that a loyal consequentialist would be acting exactly as a loyal deontologist is required to do. Whether deontological constraint is rooted on agent-relative, or agent-neutral, or neither kind of values, is itself an irrelevant concern for consequentializing deontology.

Following Louise (2004), I will try to show that if consequentialism is compatible with

agent-relative values, it is possible to generate consequentialist representation for any deontology. The idea is to further differentiate two kinds of agent-relative values: temporal-relative and temporal-neutral. Deontological reasons for action can be viewed as requiring agents to maximize values that are both agent-relative and temporal-relative. To support this, I shall illustrate how constraint functions to impose normative requirements on agents: they require agents not to violate a constraint at the moment of moral decision, not to minimize the overall number of violations in their life course. The former is temporal-relative, the later is temporal-neutral.

In Chapter 4, as a concluding chapter, I try to review the limitations and implications of consequentializing deontology presented in Chapter 3. I will discuss cases in which this method is inapplicable. That it falls to represent deontology in consequentialist form in those cases. However, as I shall argue, these cases are rare and do not fundamentally threaten the consequentializing deontology proposed. I will also argue that, with this consequentializing deontology, consequentialist shall be more sympathetic in understanding deontology and not to dismiss them as paradoxical.

The idea of consequentializing deontology is not to claim that all deontology are *fundamentally* consequentialist. Theoretical representation in a certain structural alone is not enough to determine whether a theory is consequentialist or non-consequentialist. However, it does show that when consequentialism recognizes agent-relative values, it can consequentialize individual deontological constraints. Hopefully, this may offer a different way of characterizing the dynamic and differences between consequentialism and deontology.

1.3   What Makes a Theory in Normative Ethics Consequentialist?

In this part, I will explicate some of the necessary conditions for a theory in normative ethics to be considered as "consequentialist". This part serves to delineate a basic understanding of "consequentialism" for further exploration of consequentializing

deontology in the next chapter.

Rawls' classical definition of consequentialism is generally acknowledged as a fundamental definition of the theory. He defines a consequentialist theory as one holding "the good prior to the right" and deontological theory as one holding "the right prior to the good". This way of differentiating consequentialism creates a view that consequentialism and deontology are fundamentally rival theories. I do not intend to argue against here this characterization of consequentialism and deontology. However, I would like to explore in detail some of the necessary conditions for a theory in normative ethics to become a family member in consequentialism.

"Consequentialism" is a family of theories, which holds that normative property of certain entity is solely dependent on consequences. Its family members differ from each other mainly in three aspects: (1) the kind of normative property is in concerned; (2) the notion of relevant consequence; and (3) the way that normative property depends on the relevant consequences. Different consequentialist theories would be holding different sets of claims in regard to these three aspects.

For example, "normative property" may refer to rightness and wrongness of an action (as in utilitarianism) or moral rightness of a rule (as in rule-consequentialism). "Relevant consequences" ranges from direct consequences caused by an action, to a totalistic state of affairs after the moment of act. Different theory of goodness determines the ranking of consequences, and thus, how a normative property is linked to them. For example, a consequentialist theory recognizing only pleasure and pain as moral values will give a different evaluation and ranking of consequences, as compared to a consequentialist theory that recognize pluralistic values like knowledge and humanity.

The above explains how different consequentialist theories differ from each other. However, it does not seem to tell what puts these theories together, under the family name "consequentialism". Following Carlson (1995), I will explicate some of the essential features of consequentialism, namely it is structurally axiological, teleological and maximizing. "Axiological" means that consequentialist theories define moral rightness in terms of *values*.

"Teleological" means consequentialist theories have moral rights depend solely on *outcome*. "Maximizing" means consequentialism has theory of right that directs agents to always *maximize values in outcome* they bring about. They serve as necessary condition in theory structure for a theory to be "consequentialist". To close this part, I will discuss whether agent-neutrality is a necessary condition for consequentialism.

These features delineate a general understanding of what is and what is not necessary for a theory to become consequentialist. I do not intend to drill on a complete list of collectively sufficient conditions. Rather, my purpose here is to focus on the features that are both fundamental in consequentialism and are prone to be objected by non-consequentialist. Take for example, the idea of grounding moral rightness of an action solely on its outcome is in conflict with the intuitive non-consequentialist idea that some actions *are* morally bad, regardless of the goodness they bring about. Clarifying the "teleological" feature in consequentialism helps to address some of the objections, and prepare for further discussions.

1.3.1 Conception of Outcome: Future-Oriented is Not a Necessary Principle for Consequentialism

Consequentialism holds that moral status of an action is wholly determined by the value in outcome of the action. Thus, it is "axiological"[5] by its study in values, and "teleological" by its grounding normative status of an action solely on its consequence.

However, this general claim of does not suffice the idea of consequentialism. Take for example, if a doctrine holds a peculiar conception of "outcome", such that the only morally relevant consequence of an action is the states that "such an act has been performed". Even if this doctrine commits to rank "consequences of an action" according to values in them, it is not plausibly a consequentialist theory because its conception of "outcome" is too narrow. There are only two alternatives in every case: either certain action is performed or it is not, without further qualifications. It literally excludes all factors but the action itself in the

---

[5] The term is coined by Vallentyne (1987).

outcome, and thus making an empty notion of grounding moral rightness of an action solely on its "consequence".

This extreme example shows that a plausible conception of outcome is a crucial part in defining consequentialism. Although it is sensible from above that consequentialist should not accept such a narrow conception of outcome, it remains arguable how broadly it should be construed in order to allow a spectrum of consequentialist theories. This question is important because generally, non-consequentialist, especially deontologist, objects consequentialism by pointing out moral factors *other than* outcome of an action. Take for example, justice, physical integrity, or in particular the moral bearing of certain action-types itself. These are common examples raised by non-consequentialist *in contrast to outcome* when they argue against consequentialism. Clarifying a plausible conception of "outcome" would help in a way to characterize distinction between consequentialism and other theories in normative ethics.

There are many aspects to characterize a plausible conception of "outcome" in consequentialism. However, I would focus on the temporal aspect of outcome and seek to clarify at which moment of time does outcome of an action start – does it start *after* the action; *with* the action; or counter-intuitively *before* the action? The first two are "future-oriented", which does not take into account the states of affairs before the moment of action. A commonly accepted conception of "outcome" may accept either of the two future-oriented views and reject the non-future-oriented one. However, I would argue, consequentialist has more reasons to prefer the last view over the first two. I hope to show that a futuristic conception of outcome is deemed too narrow for consequentialism to recognize certain kind of values that must be evaluated together with the past. A conception of outcome broadly construed in the temporal aspect enables consequentialism to recognize a broader range of values, thus representing a more comprehensive spectrum of consequentialist theories.

"Outcome of an action", without much clarification, usually refers to what happens after the action. With articulated, this idea has at least two interpretations:

O1   Oa, the outcome of an action a, is a set of state of affairs (s_1, s_2, s_3…) in which no part of it (i.e. either of s_1, s_2, s_3…) starts before a ends.

O2   Oa, the outcome of an action a, is a set of state of affairs (s_1, s_2, s_3…) in which no part of it (i.e. either of s_1, s_2, s_3…) starts before a begins.

The first interpretation (O1) excludes the action itself, while the second interpretation (O2) includes the action itself as part of the outcome of an action. In common use of the term "outcome", it is usually interpreted in the first way, i.e. excluding the action itself. The first interpretation (O1) was also used by some to characterize the differences between consequentialism and deontology, and to argue that "teleology does not concern about the actions but deontology does".[6] Such dichotomy fails quickly when some consequentialist, such as David Sosa, point out that outcome of an action, as states of affairs that brought about by the action, has to include the state of affair that "such an action is brought about".

> …there is no state of affairs and no agent such that the agent can bring about the state of affairs without bringing it about that he brings about the state of affairs. (Sosa, 1993, 112)

Take for example, a girl has just finished her breakfast. The outcome of her action is not fully explicated with the action-type "someone had breakfast" and the state of affair that "the plate is emptied". It should also include a description of the action-token "she had breakfast". A conception of outcome excluding the action itself is not appealing under scrutiny.

Next we shall examine the second interpretation (O2), which includes the action itself but excludes any state of affairs that happens before the action starts. To complicate the case and for the sake of argument, let us suppose "outcome" of an action *can* legitimately include states of affairs that happen *before* the action starts. Such conception of outcome may take the form of:

O3   Oa, the outcome of an action a, is a set of *future* state of affairs (s_1, s_2, s_3…) in which no part of it (i.e. either of s_1, s_2, s_3…) starts before a begins, and a set of *past* state of affairs (s'_1, s'_2, s'_3…) which includes all happenings until a begins.

---

[6] See e.g. Moore, 1942, p. 559; Bergstrom, 1966, p. 65; Sobel, 1970, p. 398.

In consider that we are exploring a conception of outcome *for* consequentialism, it would be useful to ask, is there more reasons for consequentialist to ground normative status of an action on outcome in the sense of O2 or in the sense of O3. In short, are states of affairs in the past relevant in determining normative status of an action?

The intuitive answer seems to be "no". As in general, it is believed that an action *impacts* the future, not the past. Our psychological makeup is future biased, which leads us to believe that what matters is the *future* state of affairs, instead of the state of affairs that creates the background situation in which the action was bought about. Our future biased psychological trait is best illustrated by a case of physical pain described by Derek Parfit. Imagine that you are going to have a very painful operation in hospital, which is always successful and completely safe. Due to the tremendous pain, patients are made to forget the operation afterwards. This morning your find yourself wake up in the hospital and cannot remember anything after dinner last night. You ask a nurse if the operation is over but she received conflicting information due to a system error. Either (i) your operation took place yesterday night and last for ten hours, or (ii) it will take place this afternoon and will last for an hour.[7] If you can decide between (i) and (ii) by pressing a button, it is most natural that one would prefer (i) to (ii).

Consider two consequentialists, each seeing the above case with a conception of outcome in the sense of O2 and O3 respectively. With O2, a consequentialist would rank (i) over (ii) as the former contain less pain in outcome. But with O3, a consequentialist would rank (ii) over (i) as the former contain less pain in outcome *overall*.[8] It shows that different conception of outcome leads to different consequentialist theories. However, it does not show whether O3 is a plausible conception or if it is a better account for outcome in consequentialism. I would try to show that, if consequentialism limits itself to a conception of outcome narrowly construed as O2, it would be unable to fully appreciate some kind of values in consequences of an action. This particular type of value depends on how the action

---

[7] Parfit, 1984, p.165-166. This is a simplified version of the original example.
[8] For simplicity, I assume equal weighing between future and past state of affairs.

brings about changes to states of affairs in the past. To take a narrowly construed and futuristic conception of outcome is to exclude an important set of values and thus a range of theories that ground moral rightness and wrongness of an action solely on values in its outcome from the consequentialist family.

Judicial punishment is one of the possible areas in which values brought about by an action may be measured *in combined* consideration with the past. Some proponents of punishment argue from the notion of "compensation". They hold that although punishment itself induces pain and thus brings negative effect to value of the world, punishment has a "compensation" effect when the injustice and harm brought about by the crime are also taken into account.

Although it is arguable in theory of punishment whether "compensation" is a ground for judicial punishment, it does show that it is not necessary for consequentialism to exclude certain type of evaluation simply because they take the past into account. A conception of outcome broadly construed to include states of affairs in the past (e.g. O3) is consistent with the core idea of consequentialism, i.e. grounding normative status of an action on values in its outcome. It enables consequentialism to accept a range of theories that commit to values with evaluation dependent on the past. For these reasons, it seems plausible for consequentialist to commit to a *non-future-biased* conception of outcome.

1.3.2   Maximizing is a Necessary Principle for Consequentialism

Generally, consequentialism is regarded as a maximizing theory because it requires agents to always act for the best consequence. This maximizing character of consequentialism has attracted challenges from opponents for its over-demandingness. It is because under the maximizing principle, it would be morally wrong for an agent to act for non-optional consequences base on personal interests, which is morally acceptable in most non-consequentialist theory as long as the act itself does not violate any moral constraints. Some consequentialist try to respond to the challenge by arguing for a new type of

consequentialism, namely "satisfying consequentialism," which does not require agents to always act for the best consequence but sets a bar for "good enough" consequences so to allow agents to choose freely within the range. (Slote & Pettit, 1984)

If the "satisfying" principle is consistent with consequentialism, it follows that "maximizing" is not a necessary principle for a theory in normative ethics to be consequentialist. It also entails that the spectrum of consequentialist theories would be at maximum doubled – because for every maximizing consequentialist theory, there could be a satisfying counterpart that allows agents to act in non-optimal way. In regard to the thesis aim, it means there could be two ways of consequentializing deontology, namely, representing deontological theories in maximizing or satisfying consequentialist form. Thus, it is helpful to examine whether the "satisfying" principle is consistent with consequentialism, before proceeding to the study of consequentializing deontology. In this part, I will first examine Slote's argument on "satisfying consequentialism"; then argue that the "satisfying" principle is inconsistent with consequentialism. Rather, "maximizing" is a necessary principle for a theory in normative ethical to be consequentialist.

According to Slote, it is important to distinguish between two ideas: (1) the rightness of an act depends solely on its consequences; and (2) the rightness of an act depends on its having the best consequences. He holds that traditional consequentialism with the optimal or maximizing principle entails both of the above ideas, but it is not necessary for consequentialism to hold the second claim. He borrows the notion of "rational satisfying" from the study of economics, and applies it in the domain of moral philosophy. As Slote puts it, the doctrine of "rational satisfying" holds that it is not irrational for an agent to be satisfied with non-optimal satisfaction of desires, or to be "modest in desire and needs". An example of "rational satisfying" applied in morality is that commonsense morality allows agents to act in non-optimal way when they exercise beneficence.

Slote illustrates his "rational satisfying" in morality with two examples: a daily scenario in which a man rejects an easily accessible snack even though he *would* enjoy if he has it; and a fairy tale in which when a hero is offered a wish, he asks for a fortune less than he

might have. In both cases, as Slote points out, the person explicitly rejects an offer that brings higher degree of enjoyment and well beings. According to Slote, such decision is *not* because they are motivated by some adverse consequences that may follow the "optimal" option (say for example, having a snack will spoil the dinner). But the decision is made because the agents are modestly inclined on wants and needs. In short, they are satisfied, and their choice is not considered as irrational.

Slote coins "rational satisfying" in morality with below three principles:

1. Agent acknowledges the ranking of consequences according to some principles, e.g. his and his family's well being.

2. Agent rejects the optimal consequence by knowingly choose for a non-optimal outcome.

3. This is a rational choice for the agent.

The principles of "rational satisfying" in morality are conflicting with the maximizing principle because the later would not take rejection of optimal consequence as a moral action. The maximizing character of consequentialism is based on the teleological conception of reasons. That is, if among two possible outcomes, one has more reasons to prefer O_i than O_j, then he has more reason to perform A_i, the action that leads to O_i, instead of A_j, the action that leads to O_j. Slote holds that satisfying consequentialism is an alternative to maximizing consequentialism. For as long as an act is promoting the good, it is unnecessary for a loyal consequentialist to maximizing the good promoted. However, I will try to argue against satisfying consequentialism by showing that it encounters difficulties in setting a reasonable satisfying level.

Satisfying consequentialism has two forms, namely "comparative" and "non-comparative". Comparative satisfying consequentialism holds that rightness of an action depends solely on the valuation of its outcome, relative to that of its best alternative. Non-comparative satisfying consequentialism holds that an action is right if and only if the value of its outcome meets an absolute level. In either case, any satisfying consequentialist theory has to offer a way to determine a "good enough" level. Say for example, in the

comparative case, the satisfying level may be defined as a function of the value brought about in the best outcome.[9] Below I will explore possible ways for each form of satisfying consequentialism to determine the satisfying level.

There are at least two ways for non-comparative satisfying consequentialism to determine the satisfying level: Given an absolute satisfying level x, an act is right if and only if the value in its outcome (1) meets x; or (2) promotes the value in the current state of affairs by at least x.

Non-Comparative Satisfying Consequentialism (1): An act is right if and only if the value of its outcome meets x

As Carlson (1995) argues, this doctrine of determining moral status of an action base on whether the value of its outcome meets an absolute level is implausible. Given any absolute value x, any action would bring about value that either meets/exceeds x; or falls below x. All actions in the former case are morally right. However, it does not seem plausible for consequentialism to hold that all actions in the later case are morally wrong. Consider an action that brings about the most value in its outcome, as compared to all its alternatives. Even if the value it brings about is below x, consequentialism shall hold that it is a morally right action because it *maximizes* the goods. To hold that such action is morally wrong is neither supported by the "satisfying rationality" or the "maximizing rationality".

Thus, under the doctrine of non-comparative satisfying consequentialism, an action could be morally right either if (1) it brings about value that either meets/exceeds x; or (2) it brings about the best consequence among its alternatives when the value in it does not meet x. The first is grounded by satisfying consequentialism, while the later is grounded by maximizing consequentialism.

Given any level of x, the total value in the present state of affairs either meets/exceeds x or falls below x. For simplicity, we can consider two extreme cases. Consider a case in which

---

[9] In determining a satisfying level, either in terms of a function to the values in the best outcome, or in terms of an absolute value, the problem of arbitrariness may occur as it is difficult to determine a satisfying level in numerical term. However, it does not entail that it is impossible to outline the principles on how the satisfying level can be determined.

all available actions bring about *net positive values* to the present states of affairs. It follows that all these actions are morally right. An agent can freely choose to act in any way brings about positive value to the current world, regardless of how minimal the impact is. Consider another case in which all available actions bring about *net negative values* to the present states of affairs. It follows that only one action is right – the one that brings about the best outcome. Below table shows the dynamics between actions and the satisfying level *x*:

	Current state of affairs with value ≥ *x*	Current state of affairs with value < *x*
Action brings about net positive values	All available actions are morally right	i. Any action that brings about a state of affairs ≥ *x* are morally right ii. Otherwise, moral rightness of an action determined by maximizing principle
Action brings about net negative values	Moral rightness of an action determined by maximizing principle	Moral rightness of an action determined by maximizing principle

The above shows table that when all alternative actions bring about net negative values, satisfying consequentialism retreat to maximizing consequentialism in deciding which action is morally right. When value in the current state of affairs already meets/exceeds *x*, any actions that brings about minimal net positive impact would be morally right. This is implausible to consequentialism because it seems to hold that moral status of an action *does not* depend on its outcome – rather, it depends on the satisfying level. Also, given two alternatives that with similar amount of effort but resulting in values that are 100 times different from each other, satisfying consequentialism has to explain why both are morally right. It seems counterintuitive that it is not morally wrong if one choose to bring about minimal overall goods when similar effort can lead to much better result.

It leaves us with the case in which value in the current state of affairs is below *x*. The moral rightness of an action now depends on several factors: (1) how great is the difference between *x* and the value in the current state of affairs; (2) how many positive net value can be brought about by the action itself; and (3) values brought about by its alternative.

Considering the first factor, if satisfying consequentialism sets a satisfying level far below the value in the current state of affairs, it is questionable that any available action could vastly compensate the difference, and produce an outcome with values more than or equal to x. In such case, it is plausible that satisfying consequentialism mostly has to rely on the maximizing principle to determine which action is morally right.

Now if the difference between the satisfying level x and the value in the current states of affairs is relative small, there is a real possibility for an action to overcome the difference and produce a state of affairs with values more than or equal to x. Satisfying principle would hold that all these action would be morally right. Again, consider two alternatives that with similar amount of effort, both lead to outcomes with value above the satisfying level x. If one brings about much more values than the other, satisfying consequentialism has to explain why both are morally right.

In summary, non-comparative satisfying consequentialism with an absolute satisfying level is, in most cases, reducible to maximizing consequentialism. While in the rest cases, it has to offer stronger argument for how "satisfying rationality" can support an agent in choosing a far less optimal case with similar amount of effort.

Non-Comparative Satisfying Consequentialism (2): An act is right if and only if the value of its outcome promotes the current state of affair by at least x

Alternatively, non-comparative consequentialism may hold that given a satisfying level x, an act is right if and only if the value in its outcome promotes the current state of affair by at least x. Slote may be referring to this when he states that "satisfying act-consequentialism can permit a doctor to work in India, even if he could do more good elsewhere, as long as the amount of good he will do in India is judged to be sufficient. (Slote & Pettit, 1984, 158)"

Under this formulation, satisfying consequentialism does not require agents to act for the optimal outcome as long as the *incremental value* is good enough. Slote maintains that this formulation of satisfying consequentialism aligns with "the more commonsensical moral view that makes it permissible to pursue non-optimal personal projects and commitments."

This formulation of satisfying consequentialism partly addresses a major challenge that is faced by the previous formulation, namely holding all actions morally right as long as they bring about net positive values to current state of affairs. This formulation sets a standard for moral rightness by defining a satisfying level of x. However, in case that none of the alternatives produce increment in value that is at least x, it falls back to maximizing principle in determining moral rightness of an action. Arguments against the first version of non-comparative satisfying consequentialism are also structurally applicable to argue against this version. It faces the same difficulty when specifying a fixed level of x to increase.

Comparative Satisfying Consequentialism: An act is right if and only if the value of its outcome meets a level relative to that produced by its best alternative

Both formulations of non-comparative satisfying consequentialism described above have to rely on the maximizing principle to determine moral rightness of an action in a wide range of cases. This undermines the explanatory power of satisfying consequentialism as an ethical theory although it does not show that it is implausible.

Another type of satisfying consequentialism, namely "comparative satisfying consequentialism", may hold that an act is right if and only if the value of its outcome meets a level relative to that produced by its best alternative. Under this doctrine, the satisfying level is a function of the value produced by the best alternative. A simple and direct way of setting this function is by percentage: an action is morally right if it produces at least x% of the value that is brought about by its best alternative.

As Slote argues, satisfying consequentialism is plausible base on the "satisfying rationality", that is, it is not irrational to choose for modest satisfaction of wants and need (in economic cases). When the notion of "satisfying rationality" is applied to the domain of normative ethics, it entails that agents have a moral reason to act for non-optimal cases. The result of satisfying consequentialism, as Slote holds, is that consequentialism aligns commonsense morality by allowing agents (to a certain extend) to pursue personal interests and commitments. However, I will try to argue below that the "satisfying rationality" is

inadequate to support satisfying consequentialism.

According to Slote, "satisfying rationality" holds that it is not irrational to choose for modest satisfaction of wants and need (in economic cases). He tries to illustrate with examples like a man choosing not to have a chocolate before dinner although he loves chocolate; and a hero in fairy tale choosing a less amount of fortune instead of the maximum amount he can get. In both cases, agents choose for non-optimal outcome.

In explaining for their non-optimal choices, Slote holds that the agent could be acting out of modest attitude towards wants or acting for an indirect optimal outcome. Say for example, if a man chooses not to have chocolate before dinner because that would spoil his appetite for dinner, he is making a non-optimal choice (relative to how he values chocolate) for an indirect optimal outcome (relative to how he values chocolate and dinner). Similarly, if a man asks for a less fortune in afraid that more fortune will have negative impact on his well-being, he is making a non-optimal choice (relative to how he values fortune) for an indirect optimal outcome (relative to how he values fortune and well-being). In these cases, the notion of "satisfying rationality" collapses into "maximizing rationality". It is because with careful articulation of agent's wants and needs, his choice is always the optimal one, backed by producing the most of the overall good.

However, agents could also make non-optimal choices simply because of their modest attitude towards satisfaction, *without* any indirect consideration in mind. Say for example, a man chooses not to have chocolate *not* because that would spoil the dinner, but because he is modest towards chocolate. Similarly, a man asks for less fortune *not* because that would affect his well-being, but because he is modest towards fortunes.

In these cases, agents seem to value something that they have reasons not to achieve maximum satisfaction of it. However, it is worth noting to differentiate "satisfying rationality" and "satisfying consequentialism". The former holds that one is not irrational to have modest wants and needs. But it does not entail the later, that an agent always have a reason to choose for non-optional satisfaction of his modest wants and needs. Both "non-optimal" choices of not having chocolate and asking for less fortunes, are *optimal* relative to how they meet the

agents' modest wants. With articulation of values that take the agents' modest attitude into account (say for example, the agent values a state of affair that his modest appetite for chocolate is mildly satisfied), the choice is always an optimal one.

To conclude, satisfying consequentialism reduces to maximizing consequentialism, either by indirectly achieving the most overall values, or by optimally satisfying agents' modest wants. "Satisfying rationality" proposed by Slote is inadequate to justify satisfying consequentialism as a genuine counterpart of maximizing consequentialism.

1.3.3   Consequentialism Essentials Summarized

In Part 1.3, I have discussed the conception of outcome in consequentialism, and showed that a "future-oriented" conception, which excludes states of affairs before an action starts, is not necessary for consequentialism. Instead, it is plausible for consequentialism to hold a conception of outcome broadly construed to take the past into account. The result of that, is allowing consequentialism to include a range of theories in normative ethics that recognize values that should be evaluated with respect to the action's relations to the past.

I have also examined the doctrine of satisfying consequentialism, and showed that Slote's argument for satisfying consequentialism based on the "satisfying rationality" is inadequate because the idea that one is not irrational to have modest wants and needs, does not entail that an agent always have a reason to choose for non-optional satisfaction of his modest wants and needs. Justification of satisfying consequentialism reduces to maximizing consequentialism, either by indirectly achieving the most overall values, or by optimally satisfying agents' modest wants. Thus, I conclude that maximizing principle is essential for a theory in normative ethics to be consequentialist.

1.4   Deontological Constraints

Deontology is essentially characterized by its moral constraints. A moral constraint

restricts agents from acting in certain way in pursuit of the good. The concept of moral constraint is common in non-consequentialism, e.g. commonsense morality and deontology. To differentiate, I will use the term "deontological constraint", or in short, "constraint" to refer to moral constraints in deontology.[10]

Deontological constraints restrict agents by holding that certain action-types are morally wrong. Classical examples of deontological constraints are killing or harming innocents. Agents are allowed to do anything except those forbidden by deontological constraints. Different deontological theories may be considered as holding different sets of deontological constraints. Thus, a theory that allows all kinds of actions except killing is a deontological theory with one deontological constraint.[11]

Contrary to consequentialism, deontology are said to have "the right prior to the good". It is because moral rightness and wrongness of any action are determined without considering their individual consequences. If an action-type is morally wrong under a deontological constraint, any action-token of which is morally wrong, regardless of the good brought about by it.

In this part, I will explicate different formulations of deontological constraints. Agreeing on a proper formulation is essential to the study on consequentializing deontology.

1.4.1   Absolute and Threshold Deontological Constraints

As commonly understood, deontological constraints hold certain action-types as morally wrong in nature. Actions that fall under those action-type are morally wrong regardless the situations in which they are performed, or the outcomes they bring about. Generally, there are two types of constraints: absolute and threshold constraints.

---

[10] In general, there is no difference between the way that deontological constraints restrict agent's action, and the way that other moral constraints do.

[11] This is an oversimplified situation because a deontological theory seldom comes with only one moral constraint. Say, for example, Kantian ethics would at least require agents not to treat any human being merely as a mean, on top of other constraints. However, consider that the aim of this thesis is to study possible ways to consequentialize deontology; I would focus on the essential feature of deontological theory, namely, deontological constraints. I assume that a plausible deontological theory would include a set of deontological constraints that forbids agents to act in various ways.

An absolute constraint against killing holds that killing is impermissible in *any possible situations*. Thus, it would be morally wrong to kill someone even when doing so can dramatically "save the world". A threshold constraint against killing holds that killing is impermissible *unless certain conditions applied.* Thus, if the threshold of a constraint against killing is met when "1,000 killings can be avoid", an agent is permitted to kill someone if doing so avoids 1,000 people from being killed.

Threshold deontological constraint is analogical to a dam, with consequence of an action analogical to the water in it. When the water builds up high enough and spells out of the dew, the threshold deontological constraint is no longer "effective". There are few ways to set the thresholds. (1) It can be set at an absolute level of the good brought about by a violation. Say for example, an agent is allowed to kill someone if the total value in its consequence meets an absolute level. (2) It can be set as a function of the deontological constraint itself. Say, an agent is allowed to kill someone if 1,000 killings can be avoided. (3) It can be set as a function of other deontological constraints. For example, an agent is allowed to lie if a killing can be avoided.

It is plausible but not common for threshold deontological constraints to set its threshold in the first way. It is plausible because theoretically, a threshold can be any condition specified, and a deontological theory may take consequences into consideration when the value in it is significant enough. However, it is not common because theoretical structure of deontology does not require a theory of values. To set the threshold with specific theories of values allowed is to create a "hybrid" theory. Thus, in what follows, I will focus on the later two types of thresholds, i.e. either relative to the constraints itself; or relative to other constraints.

1.4.2   *Agent-Relative & Agent-Neutral Reasons for Action: Two Approaches*

Since Derek Parfit introduces the agent-relative/agent-neutral distinction, it has been considered as an important distinction in moral philosophy. The distinction was applied to

values, reasons, principles, rules, and theories. The various applications also introduced new ways of seeing the differences between consequentialism and deontology. For example, Parfit applies the agent-relative/agent-neutral distinction to aims and theories, and hold that consequentialism is agent-neutral theory and deontology is agent-relative theory. Because the former produces the same moral aim for all agents; while the later gives them different moral aims. There are various applications and formulations of the agent-relative/agent-neutral distinction. However, a common and loosely articulated understanding of the distinction is that, it is agent-relative when a notion involves necessary "referencing to the agent"; otherwise it is agent-neutral.

In what follows, I will focus on its application in formulating deontological constraints. Because a constraint forbids agents to act in certain way, it gives agents *a deontological reason for action* to act accordingly. There are mainly two approaches to formulate deontological constraints using the agent-relative/agent-neutral distinction, namely, "general form approach" and "reason statement approach". Both attempt to formulate deontological constraints as *agent-relative reasons for action*, but differ in their approaches. General form approach differentiates agent-relative and agent-neutral reasons for action based on the principle that generate a particular reason for action. If the principle includes a necessary referencing to the agent, the reasons it generates are agent-relative; otherwise they are agent-neutral. Reason statement approach differentiates agent-relative and agent-neutral reasons for action based on the reason itself. A reason is agent-relative if, when fully specified, it includes a necessary referencing to the agent; otherwise it is agent-neutral.

Nagel is representative in proposing the general form approach; while McNaughton & Rawling propose the reason statement approach.[12] I will explicate their formulations and try to argue that the formulation by McNaughton & Rawling is the best way to differentiate agent-relative and agent-neutral reasons. Their formulation is more fine-grained and thus helpful in distinguishing different kinds of deontological constraints.

---

[12] See Ridge (2005) for detailed classification of the two approached, with other philosophers (e.g. Pettit) taken into account. However, I will only focus on Nagel and McNaughton & Rawling as they represent the central idea of the two approaches.

The General Form Approach

Before Parfit introduces the terminology, the concept of distinguishing agent-relative and agent-neutral reasons is already found in Nagel's classical *The Possibility of Altruism* (Nagel, 1970). In his work, Nagel uses the term "subjective" and "objective" to denote two kinds of reasons and adopts Parfit's terminology in his later works.[13]

Nagel (1970) proposes a basic formula of universal reasons and articulates subjective and objective reasons base on this basic formula:

> Every reason is a predicate R such that for all persons p and events A, if R is true of A, then p has prima facie reason to promote A. (Nagel, 1970, 47)

> Formally, a subjective reason is one whose defining predicate R contains a free occurrence of the variable (The free-agent variable will, of course, be free only *within R*; it will be bound by the universal quantification over persons which governs the entire formula.) All universal reasons and principles expressible in terms of the basic formula either contain a free-agent variable or they do not. The former are subjective; the later will be called objective. (Nagel, 1970, 91)

Nagel later adopts Parfit's terminology of "agent-relative" and "agent-neutral", in his work *The View from Nowhere*:

> If a reason can be given a general form which does not include an essential reference to the person who has it, it is an agent-neutral reason…If on the other hand, the general form of a reason does include an essential reference to the person who has it then it is an agent-relative reason. (Nagel, 1986, pp. 152-3)

Both notions of "basic formula" or "general form" holds that whether a reason is agent-relative or agent-neutral depends on a generalized description of the particular reason.

Suppose in every situation and for every person p, there is a reason to be eco-friendly. The general form can be expressed as

(p, A) (If A is eco-friendly, then p has reason to promote A)

A can be anything from an eco-friendly item, technology or action. According to Nagel, A can be an action such that performing A itself is a trivial way to promote A. Since the general form does not contain the free-agent variable p in the antecedent, any reason with

---

[13] It is worth noting that the agent-relative/agent-neutral distinction and the subjective/objective distinction are not interchangeable. For example, an agent-relative reason not to harm innocents (based on a deontological constraint against harming innocents) is not subjective in the sense that its validity solely depends on the agent's endorsement. Rather, it is objective in the sense that it requires all agents to follow.

this general form is agent-neutral. Accordingly, the reason to recycle paper (with A = recycling paper), the reason to promote green technology (with A = green technology), and the reason to purchase a hybrid car (with A = a hybrid car), are all agent-neutral.

On the other hand, suppose in every situation and for every person p, there is a reason to make one's family happy. Its general form will contain the pronominal back reference to p in the predicate. Thus, any reason that falls under this general form is agent-relative:

(p, A) (If A makes p's family happy, then p has reason to promote A)

Nagel's idea is to always check against the general form, which is universally quantified, and determine whether the free-agent variable exists in the reason predicate.

The Reason Statement Approach

This approach is similar to the general form approach because they both rely on the pronominal indexical as an indication of agent-relativity. However, reason statement approach focus on the explication of reason statements, instead of any general form of reasons. In their work "Agent-relativity and the doing-happening distinction", David McNaughton and Piers Rawling offer a formal account of the agent-relative/agent-neutral distinction in reasons for action.

Given any reasons for action,[14] it can be transformed into a standard form (x)(x S […]), which reads as "for all x, x should ensure…".[15] A reason is agent-relative if and only if there is an occurrence of "x" in the square brackets bound by the initial universal quantifier; agent-neutral otherwise. (McNaughton & Rawling, 1991, 175)

For instance, the reason for every parent to take care of their own children can be transformed into: (x)(x S [(x is a parent → x cares for x's children)]). Since there is an

---

[14] McNaughton & Rawling use "reason" and "rule" loosely and interchangeably. Here, I will only use "reason" or "reason for action" for consistency and to avoid ambiguity.

[15] The use of technical apparatus in a standard form looks similar to Nagel's formulation of general form. It may lead readers to believe that they are taking the same approach. On one hand, McNaughton & Rawling acknowledge that their approach is similar to Nagel's, only that it is more fine-grained. On the other hand, Ridge (2005) classifies them as different approaches. I believe that both Ridge and McNaughton & Rawling are right in pointing out the similarities and differences between the two formulations. I follow Ridge's classification mainly to point out that, various formulation of agent-relative/agent-neutral reasons may fall within a spectrum – at one end, it relies on the generalization of reasons to determine agent-relativity, at the other end, it relies on the specification of reasons. Later, I will try to show several problems of Nagel's formulation, which I believe, are common to other generalization approaches while the reason statement approach is not inherently prone to them.

occurrence of "x" in the square bracket, this reason is agent-relative.

The account is structured and elaborated in a way to cover three types of agent-relative/agent-neutral reasons for actions, namely "author agent-relative/neutral", "object agent-relative/neutral", and "author-object agent-relative/neutral". Below table shows the six types of agent-relative/agent-neutral reasons:

	Author-oriented	Object-oriented	Author-object relation oriented
Agent-neutral	(x)(x S [(y) (y does not lie)])	(x)(x S [(y) (y is a child → y is cared for)])	(x)(x S [(y) (y is a parent → y cares for y's children)])
Agent-relative	(x)(x S [x does not lie])	(x)(x S [(y) (y is x's child → y is cared for)])	(x)(x S [(x is a parent → x cares for x's children)])

Author-oriented reasons concern agent's actions regardless of the existence or identity of the "receiver". Say for example, the reason simply against lying falls under this category and applies even when there is no person to hear the lie.

Object-oriented reasons concern actions done to an object regardless of who the doer is. For example, reason concerning whether a child is taken care of falls under this category and applies regardless of who takes care of the child. If there is a reason to ensure that "no children hear a lie", it will take below two forms:

AN   (x)(x S [(y) (y is a child → y hears no lie)])

AR   (x)(x S [(y) (y is x's child → y hears no lie)])

Author-object relation oriented reasons concern actions done by agents in relation to particular objects. Both doer and receiver are specified in the reason. If there is a reason to ensure that "no children hear a lie from their parents", it will take below two forms:

AN   (x)(x S [(y) (y is a parent → y's children does not hear a lie from y)])

AR   (x)(x S [(x is a parent → x's children does not hear a lie from x)])

1.4.3   Deontological Constraints & Reasons for Action

As noted above, deontological constraints restrict agents from performing certain

actions. They generate deontological reasons for actions to agents not to act in certain ways. To prepare for the study in consequentializing deontology, I will first examine which account, the one from Nagel, or the one from McNaughton & Rawling, serves better to formulate deontological reasons. I will argue that McNaughton & Rawling's account is immune to some difficulties encounter by Nagel's account. Their account is also more sophisticated and able to capture some subtle differences between agent-relative and agent-neutral reasons, which Nagel's account fails to grasp.

McNaughton & Rawling suggests that, there is a possible "double reading effect" in determining whether an agent is acting for an agent-relative reason or an agent-neutral reason. Consider Grace, who reads to her children and is a loyal subscriber to the doctrine that parents should look after their own children. According to Nagel's account, Grace is promoting "care for her own children". Translated into general form, it becomes:

Pr: (p, A) (If A gives p's own children care, then p has reason to promote A)

This looks like a perfect agent-relative principle. Grace is acting on an agent-relative reason because she takes care of *her* own children and read to them by *herself*. But McNaughton & Rawling holds that, Grace may on the contrary be acting for an agent-neutral reason promoting the doctrine of parental care:

Pn: (p, A) (If A encourages parents to care for their children, then p has reason to promote A)

In this later case, Grace may consider that reading to her own children is the best way for her to encourage other parents to care for their children, ultimately promoting the overall value of parental care. To complicate the case, even when Grace subscribes to Pr, it does not mean that Grace has to take care of her own children by herself. It actually allows Grace to hire a nanny to take care of her children, so that she is still promoting *her* care for *her* children. In this way, Nagel's general form is not able to distinguish the two kinds of reason.

To describe a reason for Grace to promote "parental care for her children", it may fall under below general form:

Pr': (p, A) (If A allows p to give p's own children care, then p has reason to promote A)

Again, the three general forms above are applicable to describe the reason for Grace's reading to her own children in order to promote parental care. Nagel's account leads to ambiguity in determining whether an agent is acting for an agent-relative or an agent-neutral reason.

It may be responded that McNaughton & Rawling's account faces similar difficulty. Without clarifying Grace's fundamental reason for reading to her children, their account is also unable to differentiate between below possible formulations of Grace's reason for action:

	Author-oriented	Object-oriented	Author-object relation oriented
Agent-neutral	(x)(x S [(y) (y is caring)])	(x)(x S [(y) (y is a child $\rightarrow$ y is cared for)])	(x)(x S [(y) (y is a parent $\rightarrow$ y cares for y's children)])
Agent-relative	(x)(x S [x is caring])	(x)(x S [(y) (y is x's child $\rightarrow$ y is cared for)])	(x)(x S [(x is a parent $\rightarrow$ x cares for x's children)])
Remarks	Grace may read to her children either because it is the best way to promote caring, or because she should be caring.	Grace may read to her children either because it is the best way to promote child care, or because her child should be cared for.	Grace may read to her children either because it is the best way to promote parental care, or because she should care for her children.

However, it is worth noting that, McNaughton & Rawling's account is structurally ready to accommodate the "double reading effect". Under Nagel's account, it is unguided how to clarify and address possible "double reading". There is no guiding principle in how general should a reason be specified, in order to determine whether the reason is an agent-relative one or an agent-neutral one.

Furthermore, it is commonly understood that deontology concerns "agent's actions", rather than "states of affairs". Deontological constraints require each agent to avoid violation *themselves*. They do not require agents to *minimize* violations of a constraint happening in the world. Thus, it is important to distinguish between doing and happening in any study

related to deontology. Consider Nagel's general form again:

> Every reason is a predicate *R* such that for all persons *p* and events *A*, if *R* is true of *A*, then *p* has prima facie reason to promote *A*. (Nagel, 1970, 47)

Here, *A* could be an action or a state of affairs. Say for example, consider eco-friendliness: (*p*, *A*) (If *A* is eco-friendly, then *p* has reason to promote *A*). If *A* is an action "to recycle paper", requiring *p* to promote the action is to require *p* to recycle paper himself. This is an agent-relative reason according to McNaughton & Rawling's account. But it is an agent-neutral reason in Nagel's account because it does not have any referencing to *p*. If *A* is a state of affairs of "having more hybrid cars on road, requiring *p* to promote the state of affairs does not require *p* to buy a hybrid car himself. It concerns not about agent's action, but the state of affairs. By allowing *A* to be filled as either action or state of affairs, Nagel's account does not distinguish doing and happening.

McNaughton & Rawling's account, on the contrary, takes the doing/happening distinction into account. A reason concerns "happening" if it does not specify "who" takes the action. Thus, all three types of agent-neutral reasons concerns "happening". Interestingly, object-agent-relative reasons also concerns happening instead of doing even though it is an agent-relative reason. It is because an object-agent-relative reason requires an agent to ensure something *is done* to an object, which stands in particular relation with the agent himself. This does not specify *who* carries out that action. Say for example, a mother hiring a nanny to take care of her children can be acting to ensure *her* children are cared for. Below table highlights in shade the types of reasons that concern happening instead of doing:

	Author-oriented	Object-oriented	Author-object relation oriented
Agent-neutral	(x)(x S [(y) (y does not lie)])	(x)(x S [(y) (y is a child → y is cared for)])	(x)(x S [(y) (y is a parent → y cares for y's children)])
Agent-relative	(x)(x S [x does not lie])	(x)(x S [(y) (y is x's child → y is cared for)])	(x)(x S [(x is a parent → x cares for x's children)])

To summarize, it is commonly understood that deontological constraints concerns agent's actions. They give different agents different moral reasons for action. The deontological constraint against killing requires *you* not to kill, and *me* not to kill. In this way, we do not have a common aim to minimize killing in the world. If I can only choose between *my* killing and *your* killing, as a loyal deontologist, I would be required to avoid *my* killing. In next chapter, I will start exploring ways to consequentialize deontology, namely to offer all deontological theories a consequentialist representation.

Chapter 2.   Attempt of Consequentializing Deontology without Agent-Relativity

2.1   The Consequentialism/Deontology Distinction & the

           Agent-relative/Agent-neutral Distinction

Before the agent-relative/agent-neutral distinction was introduced, differences between consequentialism are typically characterized in below few ways:

The Right and The Good: Since Consequentialism holds that moral rightness of any action is solely determined by the value in the outcome it brings about, no moral status of any action is determined before evaluating its possible outcome. On the other hand, deontology holds that moral status of any action is determined *prior* to consideration of its outcome. Thus, the consequentialism/deontology distinction is typically characterized as one between "the good prior to the right" and "the right prior to the good".

Doing and Happening: Deontology restricts agents from acting in certain ways by imposing deontological constraints. Each constraint gives different moral requirements to different agents. Say for example, the constraint against killing requires A *himself* not to kill, and B *himself* not to kill. Thus, deontology concerns particular *action* of each agent. On the other hand, consequentialism evaluates an action's outcome by accounting values in the possible state of affairs brought about by an action. A state of affairs *obtains* if and only if the scenario described in it *happens*. Thus, the consequentialism/deontology distinction is often characterized as one between "happening" and "doing".

Objective and Subjective: This distinction is closely related to the doing/happening distinction in three ways. First, two deontologists committed to the same deontological constraint will have *subjective reasons* not to violate the constraint by themselves. However, two loyal consequentialists following the same consequentialist theory will have an *objective reason* to bring about the same state of affairs. Secondly, the deontological *perspective* is subjective in the sense that agents can determine moral rightness of his action solely by counterchecking *his* actions with deontological requirements. He does not need to take an

consequentialist objective view to see the total states of affairs obtained if he acts in certain way, in order to determine moral rightness of the action. Finally, consequentialism is commonly understood as recognizing only *objective values* because the consequentialist perspective sees everyone as the same. By recognizing the objective value of parental care, instead of subjective value of parental care in *your* or *my* families, consequentialism gives all agents a common goal to promote the common good.

After its introduction, the agent-relative/agent-neutral distinction is applied to theories, reasons, perspective, values, etc… in order to characterize the differences between theories in normative ethics. Below examples show how Parfit and McNaughton & Rawling apply the agent-relative/agent-neutral distinction differently:

Parfit applies the distinction to theories. He holds that a theory is agent-relative if it gives different agents different aims; agent-neutral if it gives all agents the same moral aim.

> Since *C* gives to all agents common moral aims, I shall call *C agent-neutral*. Many moral theories do not take this form. These theories are *agent-relative*, giving to different agents, different aims (Parfit 1984, 27).

Base on this formulation, Parfit concludes that deontology is agent-relative and consequentialism is agent-neutral.

McNaughton & Rawling apply the distinction to reasons for actions. Given any reason, it can be transformed into a standard form (x)(x S […]), which reads as "for all x, x should ensure…" A reason is agent-relative if and only if there is an occurrence of "x" in the square brackets bound by the initial universal quantifier; otherwise, it is agent-neutral. Base on this formulation, McNaughton & Rawling also conclude that deontology is agent-relative and consequentialism is agent-neutral.

The above applications of agent-relative/agent-neutral distinction in characterizing the differences between normative theories often results in a *dichotomy* between consequentialism/deontology. However, there are philosophers who apply the agent-relative/agent-neutral distinction to argue against the dichotomy. Some of them apply the distinction in support of "consequentializing deontology", a doctrine holding that all deontological theories can be given a consequentialist representation.

The idea of "incorporating" deontology under consequentialism is not new. Sen (1983) argues for an evaluator-relative consequence-based morality earlier than the introduction of the agent-relative/agent-neutral distinction. Brown points out a typical way of consequentializing deontology as to "build what (deontologist) concerned about into a theory of the good. Then consequentialism can well address (the) concerns." Dreier (1993) goes so far as to suggest that it is simple to consequentialize any given moral theory – simply by taking the features of an action that the theory considers to be relevant, and build them into the consequences.

With the agent-relative/agent-neutral distinction applied, the idea of consequentializing deontology is articulated further in details. Portmore (2001) applies it to *perspectives* and argues that consequentialism is compatible with agent-relative perspective. These philosophers argue that, since it is not necessary for consequentialists to evaluate outcome of an action in the agent-neutral way,[16] it is plausible that deontological moral considerations can be accommodated in consequentialist theories. For example, a way to consequentialize deontology is to view it as a consequentialist theory that asks agents to always minimize their own violations of constraints in consequences. Louise (2004) takes a similar approach to arguing for consequentializing deontology by applying the agent-relative/agent-neutral distinction to *values*. She holds that deontology falls under "the consequentialist umbrella" with a theory of values that recognize agent-relative values.

There are at least two ways to consequentialize deontology: one is to assign a negative weighing to any violations so that actions involving violations will never be the best outcome. This does not require consequentialism to recognize agent-relative values. Thus it is based on traditional consequentialism and is less controversial. Another one is based on agent-relative consequentialism and to argue that it can consequentialize deontology by recognizing agent-relativity of any kind, e.g. values.

---

[16] An example of agent-relative consequentialist theories is egoistic hedonism, which holds that an act is right if it leads to the greatest pleasure for oneself. In evaluating different consequences, egoistic hedonist takes an agent-relative perspective. However, this does not make the theory become a non-consequentialist because it holds that moral rightness of an action solely depends on its consequence.

I will focus on the first approach in this chapter and proceed to the second approach in the next chapter. There are two reasons for this: First, the idea of agent-relative consequentialism, although argued for by a number of philosophers, is still controversial to some. [17] Secondly, distinguishing the two approaches enables us to clarify whether agent-relativity *is* necessary in consequentializing deontology.

In what follows, I will try to explicate how traditional and agent-neutral consequentialism can "build in deontological concerns" (as in the form of deontological constraints). As discussed in Chapter 1, there are two types of constraints: threshold and absolute. I will discuss how to consequentialize them respectively.

2.2 Consequentializing Deontology with Threshold Constraints

Threshold deontological constraint is analogical to a dam, with consequence of an action analogical to the water in it. When the water builds up high enough and spells out of the dew, the threshold deontological constraint is no longer "effective". There are two typical ways to set the thresholds. Type 1: It can be set at a function of the deontological constraint itself. Say, an agent is allowed to kill someone if 1,000 killings can be avoided. Type 2: It can be set at a function of other deontological constraints. For example, an agent is allowed to lie if a killing can be avoided.

Consequentializing Type 1 Threshold Constraints

To consequentialize a deontological constraint with threshold relative to the constraint itself, consider a constraint against killing with a threshold set at five killings avoided. The goal for consequentializing deontology here is to ensure that agents are not required to kill before the threshold is met. In other words, is to ensure that the best consequence does not involve killing *unless* five lives can be saved from being killed. We can name the ratio of

---

[17] For example, Mark Schroeder is skeptical of the doctrine of "agent-relative teleology", which holds that "For all agents *x*, *x* ought always to do that action that will bring about the most of what is good-relative-to *x*." See Schroeder (2007).

five as the "threshold ratio".

To consequentialize this type of threshold constraints, or in other words, to build this deontological consideration into consequence, a typical way is to give a negative value (or "weighing") to violation of the constraint,[18] *in the threshold ratio* to the positive value achieved by avoiding one violation of the constraint.

For example, to consequentialize the deontological constraint against killing with a threshold set at five killings to be avoided, below values will be applied:

$$| \text{Value of killing once} | \quad = \quad \text{Threshold ratio} \quad x \quad | \text{Value of avoiding killing once} |$$

$$\underbrace{\qquad\qquad} \qquad\qquad\qquad\qquad\qquad \underbrace{\qquad\qquad\qquad}$$

$$\text{Negative} \qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{Positive}$$

This arrangement creates a theory of good that takes deontological constraint into account in the following cases:

(I)   Given any action available to an agent, if it involves violation of the constraint, its outcome will not be ranked as the best. With the impact of a negative weighing, an alternative action that does not involve violation will lead to a better consequence with more values in it.[19] In this case, both deontology and consequentialized deontology would hold that it is morally wrong for the agent to violate the constraint.

(II)   Given any action available to an agent, if it involves violation of the constraint and leads to a number of avoidances, the negative value in violation will be outweighed accordingly. As long as the number of avoidances is below the threshold ratio, its alternative action that does not involve violation will still lead to a better consequence. In this case, both deontology and consequentialized deontology would hold that it is morally wrong for the agent to violate the

---

[18] Because this chapter limits itself to consequentializing deontology base on traditional and agent-neutral consequentialism, the negative value here applies to "violation of constraint" *in general*, not the agent's violation of constraint. This limitation makes it difficult to offer a successful account in consequentializing deontology. Discussions on possible challenges can be found in later part.

[19] I assume that a case without alternative action available is not a genuine case for moral decisions.

constraint.

(III)   Given any action available to an agent, if it involves violation of the constraint and leads to a number of avoidances that either meets or exceeds the threshold ratio, the negative value in violation will be balanced off. In this case, deontology would hold that it is permissible for the agent to violate the constraint. On the other hand, consequentialized deontology would have two consequences ranking the same: one without violation, and one with violation leading to five avoidances of violation. With all other things being the same, the agent would be allowed to act in either way.

Above three cases exhausts possibility actions in relation to violation and avoidance of constraint. Case (I) applies when an available action involves violation of constraint but does *not* lead to any avoidance. Case (II) applies when an available action involves violation of constraint that brings about a number of avoidances *below* the threshold ratio. Case (III) applies when an available action involves violation of constraints that brings about a number of avoidances *at or above* the threshold ratio. In all three cases, consequentializing deontology does not require or allow agents to violate the constraint *until* the threshold is met. When the threshold is met, deontology holds that it is permissible for agents to violate the constraint and consequentializing deontology also allows agents to act in either way.

Consequentializing Type 2 Threshold Constraints

To consequentialize a deontological constraint with threshold set relative to other constraints, consider a constraint against lying with a threshold set at one killing avoided. The goal for consequentializing deontology here is also to ensure that agents are not required to lie before the threshold is met. In other words, it is to ensure that the best consequence does not involve lying *unless* a life can be saved from being killed. Here, we use a lexical ranking, instead of a ratio, to determine the relative values of violation and avoidance of the constraint against lying, and the constraint against killing. In this case, the lexical order applies as below:

1.		Avoidance	of	killing
2.	*is better than*	Avoidance	of	lying
3.	*is better than*	Violation	of	lying
4.	*is better than*	Violation	of	killing

Similarly, this arrangement creates a theory of good that takes deontological constraint into account in the following cases:

(I) Given any action available to an agent, if it involves lying, its outcome will not be ranked as the best because its alternative that does not involve lying will lead to a better consequence. In this case, both deontology and consequentialized deontology would hold that it is morally wrong for the agent to lie. (Similarly when applying to constraint against killing.)

(II) Given any action available to an agent, if it involves lying and leads to avoidance of killing, it would be a better consequence than its alternative that does not involves lying at all. In this case, both deontology and consequentialized deontology would hold that it is morally required for the agent to lie. (Similarly applying to constraint against killing, in which if killing leads to avoidance of lying, it is morally wrong for agents to lie, according to both theories.)

We can further complicate the case by adding, for example, that an agent is permitted to lie if it can avoid 10 lies from being told; and that an agent is permitted to kill if 1,000 lives can be saved from being killed, etc. However, I shall stop illustrating further as the principle of consequentializing is the same.

2.3   Consequentializing Deontology with Absolute Constraints

Absolute constraints restrict agents from acting in certain way *in all cases*. That is, there is no threshold over which an agent is permitted to violate the constraint. It is questionable whether any plausible deontological theory would take such an extreme position and hold that all its constraints are absolute. However, we may still apply the method of lexical

ranking to consequentialize deontology if the theory has *only one* absolute constraint. The lexically negative priority given to the violation of this absolute constraint would ensure that its consequence will not be the best among its alternatives. However, difficulties arise when there are two or more absolute constraints.

Deontological theories holding two or more absolute constraints make it possible for agents to encounter moral dilemma. In a moral dilemma, all available actions are morally wrong. It is the case when agents can only choose between two alternative actions, but each of which involves violation of an absolute constraint.

There are two problems in consequentializing deontology with more than one absolute constraint. First, the method of lexical ranking cannot determine the lexical priority between violations of absolute constraints. Consequentializing deontology thus cannot determine the ranking between outcomes of different violations. Moreover, it is not plausible to maintain that violations of both absolute constraints are lexically equal. It is because according to the deontology that accepts the existence of moral dilemma, both violations are morally wrong. If consequentializing deontology maintains that both violations are lexically equal, it entails that their consequences, with other things the same, are equally good. It would be morally right to choose for either violation.

Some proponents of consequentializing deontology recognize that they cannot accommodate moral dilemmas due to a consequentialist principle. Since the action that leads to the best consequence is morally right, it follows that there *must be* one morally right action. Some, like Brown (2010), even holds failing to accommodate moral dilemmas is of advantages to consequentialism because it is implausible to accept moral dilemma.

I would not drill on arguments concerning moral dilemma. However, I would like to point out that deontologist would not hold an agent *blameworthy* of violating a constraint in a moral dilemma. The idea of moral dilemma in above cases is that: all available actions are morally wrong; and each of them gives the agent a deontological reason to avoid. However, it is implausible for deontologists to hold that violations in moral dilemma are *just as morally wrong* as violations not in moral dilemma. Being indifference to the two cases is

inconsistent with recognizing moral dilemma as a morally significant case. If a deontological theory has more than one absolute constraint and accepts existence of moral dilemma, it is plausible for it also to accept the agent choosing from the morally forbidden actions in face of dilemma. Although this point does not address the asymmetry between deontological and consequential moral status, it shows that both deontology and consequentialized deontology give agents *a moral reason* to choose freely in face of moral dilemma.

2.4   Limits and Problems with Consequentializing Deontology

Having explicated different methods to consequentialize deontology with threshold and absolute constraints, I would argue below that above methods fail because they entail implications that are implausible and are not held by the original deontology. The goal of consequentialized deontology is to give every deontological theory a representation in consequentialist form, so that a committed consequentialist would have moral reasons for actions that support him to act in the same way as his deontologist counterpart does – in identical situations. If there are implausible implications entailed by consequentializing deontology, leading an agent to act in different way as his deontological counterpart does, consequentializing deontology would fail in its own term.

2.4.1   Perfect calculus leads to implausible implications

Recall the method of consequentializing deontology by giving a negative weighing to violation of a constraint relative to its threshold:

| Value of killing once[20] | =   Threshold ratio   x | Value of avoiding killing once |

$\qquad\qquad$ Negative $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ Positive

---

[20] The negative value is given to "violation of constraint against killing" instead of "agent's violation of constraint against killing" because this chapter limits itself to agent-neutral consequentialism. See footnote 19.

With above arrangement, it is said that consequentializing deontology is successfully in creating a representation for its deontological counterpart in all possible cases:

Case (I)   applies when an available action involves violation of constraint but does *not* lead to any avoidance.

Case (II)   applies when an available action involves violation of constraint that brings about a number of avoidances *below* the threshold ratio.

Case (III)   applies when an available action involves violation of constraints that brings about a number of avoidances *at or above* the threshold ratio.

In all three cases, consequentializing deontology does not require or allow agents to violate the constraint *until* the threshold is met. When the threshold is met, deontology holds that it is permissible for agents to violate the constraint. In this case, consequentializing deontology also allows agents to act either way.

However, this arrangement leads to at least two implications that deontologist may not readily accept. The first one is illustrated below.

Suppose t_0 is the moment *just before* an agent decides for action; and the value in total states of affairs at t_0 is α. Consider a deontological constraint against killing with threshold set at five killings being avoided. Below table shows how moral status of a violation changes along with the number of avoidances it bring about.

No. of avoidance achieved by a violation	Moral status of a violation according to deontology	Evaluation of states of affairs according to consequentializing deontology
0	Not justified	$\alpha - 5x$
1	Not justified	$\alpha - 5x + x$
2	Not justified	$\alpha - 5x + 2x$
3	Not justified	$\alpha - 5x + 3x$
4	Not justified	$\alpha - 5x + 4x$
5	Justified	$\alpha - 5x + 5x$
6	Justified	$\alpha - 5x + 6x$

When a violation of constraint brings about *just* five avoidances, the threshold is met. According to deontology, agents are permitted to violate the constraint in this case. On the other hand, base on the negative weighing arrangement, consequentializing deontology

would hold that the states of affairs brought about by a violation has ($\alpha - 5x + 5x$) amount of values. Since the negative weighing is balanced off, the net value in outcome is α. It follows that if an agent is presented with two alternative actions: one does not involve killing (without consequence); the other involves killing one to safe five being killed. The agent can choose either action because they lead to outcomes with the same amount of values. But it is counterintuitive to be indifferent between these two cases.

The fact deontology allows agents to violate a constraint when a threshold is met, does not imply a loyal deontologist would take a state of affairs with one killing to save five, as good as a state of affairs with no killing (without consequence). If these two alternatives are open to a deontologist, it is more plausible that he acts on later option.

The second implication is a similar, with one killing saving six lives, its outcome will have ($\alpha - 5x + 6x$) amount of values. With a net value of ($\alpha + x$), this action brings about a state of affairs that is *better* than a state of no killing (without consequence). When an agent is presented with these two alternatives, consequentializing deontology will require that he act for former option. However, it does not seem plausible that a deontologist will choose to kill in this case simply because it brings about one *net* saving of life.

2.4.2   *Agent-relativity in Deontological Constraints*

As stated earlier in this chapter, there are two major ways of consequentializing deontology. One is based on *broad* consequentialism which is compatible with agent-relativity; another is based on *traditional* consequentialism that is agent-neutral in every aspect (e.g. reasons for action, perspective, and values). In this chapter, I focus on agent-neutral consequentialism in explicating methods to consequentialize deontology.

The basic method to consequentialize deontology is by giving negative weighing to violation of a constraint as below example:

$$| \text{Value of killing once} | \quad = \quad \text{Threshold ratio} \quad \text{x} \quad | \text{Value of avoiding killing once} |$$

$$\underbrace{\phantom{| \text{Value of killing once} |}}_{\text{Negative}} \qquad\qquad\qquad\qquad \underbrace{\phantom{| \text{Value of avoiding killing once} |}}_{\text{Positive}}$$

The negative weighing is given to "violation of a constraint" instead of "agent's violation of a constraint". This arrangement is to make it consistent that consequentializing deontology in this chapter is undertaken in an agent-neutral context.

Given a deontological constraint against killing with threshold set at five lives saved. Consider below two cases:

Case 1: Adam is an adventurer who travels in The Amazon Rainforest. He is caught by a tribe which is having a trial of five criminals. Since he is a special guest, the leader offers him with two choices: either he kills a criminal so everyone is saved, or all of them are killed.

Case 2: Adam is caught together with his partner Billy. Since they are special guests, the leader of the tribe offers each of them with three choices: either he or Billy kills a criminal so everyone is saved, or they *make their partner* kills a criminal so everyone is saved. If no one kills a criminal, all of them will be killed.

Suppose Adam is a deontologist who commits himself to the constraint stated above. In Case 1, he is permitted to violate the constraint against killing because more than five lives will be saved. Since he will die if he does not kill, also killing is permitted in this case, Adam will kill a criminal and save all five lives (including himself).

Suppose Adam is a consequentialist who commits himself to the consequentialized deontology. In Case 1, he would also choose to kill a criminal, but for a different reason. Consequentializing deontology gives a negative value to killing, which is in the ratio of five (i.e. the threshold ratio) to the positive values of killing avoid. If Adam does not kill any criminal, it will cause a total of six killings. If Adam kills a criminal, it will cause one death with five killings being avoided. Since the later state of affairs is better, Adam, as a consequentialist in this case, will kill a criminal.

Now consider Case 2. It is permissible for Adam to kill a criminal because it could save more than five lives from being killed. Between killing a criminal by himself and making others to kill, a deontologist would be inclined to prefer the later one, with other things being equal. It is because a deontological constraint gives different agents different moral aims. The constraint against killing would require Adam and Billy themselves not to kill. With an option of *making others kill*, Adam as a deontology will prefer it over violating the constraint himself.[21]

If Adam is a consequentialist following the consequentialized deontology, he would evaluate the consequences of all three alternatives. The consequence that everyone dies is the worst. However, the state of affairs in which *his* killing that saves all, would be as good as the state of affairs in which *Billy's* killing saves all – because they involve the same number of killing and lives being saved.[22]

Thus, Case 2 shows that agents have certain agent-relative reason for choosing *other's* violation over their own violation, while consequentializing deontology cannot capture the deontological reason for action in consequentialist term.

2.5   Responding to Challenges

2.5.1   Perfect calculus leads to implausible implications

A basic method to consequentialize deontology is by assigning a negative weighting to violation of constraint. This method, however, leads to two implications that deontologist may not readily accept. They are: (1) a state of affairs in which the threshold of a constraint is *just met*, is as good as a state of affairs in which there is no violation and no avoidance of

---

[21] This option may be taken as "using others as a means". We can assume here that both Adam and Bill are not actively forcing another to kill, but by delaying the killing so that the other may give up and kill an innocent to save everyone, as is in the game of chicken. The idea here is to point out the agent-relativity in normative force of deontological constraint on an agent.

[22] This is a simplified view. Since states of affairs *may or may not* obtain, they are bearers of probabilities. Consequentialism requires agents to act in a way that *brings about* the best states of affairs. Thus, the degree of probability for a state of affairs to obtain has to be taken into consideration. In this case, if Adam chooses not to kill a criminal himself, risk that Billy also does not kill any criminal and that everyone will be killed. This minimal probability will affect consequentialist calculations. For simplicity, it is ignored in this case.

violation at all; and (2) a state of affairs in which the number of avoidances brought about is *more* than the threshold, is *better* than a state of affairs in which there is no violation and no avoidance of violation at all.

These two implications affect how a committed consequentialist acts under consequentialized deontology. Consider again, a constraint against killing with threshold set at five killings avoided. Suppose an agent is presented with two alternatives: one that he kills one and save five; and one that he does not kill and no one dies. A consequentialist will see these two consequences as good as each other. He would be morally right if he choose either one. However, a deontologist may not be inclined to kill and save five in this case.

Suppose an agent is presented with a third alternative: that he kills one and save six. A consequentialist will see this as a *better* consequence then the other two. He would be morally *required* to bring about this outcome. However, a deontologist would be morally *allowed* to choose between any of the three alternatives.

The key to explaining for above differences between deontology and consequentialized deontology, I believe, lies in the asymmetry in moral status of an action between deontology and consequentialism. And that consequentializing deontology can only respond by weakening its position.

Deontology holds any action restricted in deontological constraints are "forbidden", and any action *not* specified in a constraint is "allowed" as an option to the agent. As long as an agent does not violate any constraint, he literally is morally permitted to do anything in pursuit of personal goals. On the other hand, consequentialism holds that moral status of the action leading to the best outcome is "right". Any action that leads to non-optimal outcome is "wrong". Thus, there is an asymmetry between moral status of actions in consequentialism and deontology.

Consequentializing deontology, by assigning negative weighing to violation of constraints, ensure that agents are not required or allowed to violate a constraint before its threshold (if applicable) is met. This ensures actions forbidden by deontological constraints will never be "right" in consequentialist term. However, the decision of which action to take,

among morally allowed alternatives, is subject to agent's personal interests and commitments. Thus, the claim that consequentializing deontology can make a committed consequentialist act in the *same* way as his deontological counterpart does, is too strong and shown implausible with above cases.

Deontologist considers an agent's personal interests and commitments as a kind of *moral factors*, and allows for *options*. This is incompatible with agent-neutral consequentialist because the later does not accept agent-relativity in any kind (e.g. perspectives and values).

Initially, consequentializing deontology made a strong claim that all deontological theories can be represented in consequentialist form so that a committed consequentialist will act "extensionally" the same as his deontological counterpart in identical situations. Without taking agent-relativity of personal interests and commitments into account, the consequentializing deontology falls to "imitate" agent's choice among *options*. Consequentialist deontology can defense with a weak claim: that all deontological theories can be represented in consequentialist form so that a committed consequentialist will not violate a constraint before its threshold is met. Thus, in every case that a loyal deontologist avoids violating a constraint, its consequentialist counterpart will do the same.

This significantly weakens the position of consequentializing deontology. It becomes a doctrine about consequentializing deontological constraints, instead of deontological theories. However, as I shall ague below, not even this weak position is sound enough.

2.5.2   *Agent-relativity in Deontological Constraints*

Recall the cases of Adam and Billy in the Amazon Forest.

Given only two alternative actions, that is, either he kills a criminal so everyone is saved, or all of them are killed, Adam as a deontologist will kill a person because he is morally allowed so, in order to save a number of lives over the threshold. However, given a third alternative that he can indirectly cause his partner Billy kill a criminal and save all persons, it

is plausible that, as a deontologist, he would choose for the third alternative.

Consider Billy, who is committed to the consequentialized deontology. Given three alternative actions just like Adam is, Billy will see that a state of affairs in which he kills a criminal and saves all, is as good as a state of affairs in which Billy kills a criminal and saves all. It is because the negative weighing was given to "violation of constraint against killing", but not "agent's violation of constraint against kill". With this arrangement, consequentialized deontology has a theory of good that makes Billy indifferent to who kills.

Consequentializing deontology may reply that its weaker version can also respond to this challenge. In the above case, the threshold is already met. Agents are *morally allowed* to violate the constraint against killing. Thus, even when Adam, as a deontologist would choose not to kill, it is not necessary for consequentializing deontology to ensure that Billy follows through and does not kill as well. The case that Adam is acting out of *moral option* is irrelevant to make an objection.

However, I believe, the case and the objection is relevant. It is because when deontologist chooses among alternatives that do not violate any constraint, his decision is normally subject to personal interests and preferences that is not *given* or guided by deontological principles. But in the above case, Adam's inclination to *not* killing anyone by himself is grounded on his commitment to deontology. Although agents are justified to violate a constraint when the threshold is met, it is worth differentiating between morally avoiding violation (even when the threshold is met), and choosing out of personal preferences. Adam's case falls in the former category.

Since Adam has a deontological reason to prefer Billy's violation of constraint over his own violation of constraint, and consequentializing deontology cannot capture this different and represent the deontological reason in consequentialist term, it is plausible to conclude that even a weaker version of consequentialized deontology fails.

2.6   Summary: Consequentializing Deontology Fails

As explicated in Chapter 1, it is commonly acknowledged that deontological constraint is *agent-relative* as it gives different agents different moral aims. To "build" this deontological concern into consequence is not as simple as some philosophers suggested. Above explication has shown that when consequentialism is not broadly construed to accept agent-relativity, it is implausible for it to consequentialize deontology.

We also learn from above that, representing a deontological theory in consequentialist term is *more* than "building deontological concern into consequences". It has to allow rooms for moral options or it is deemed to be weak and trivial.

The differences between consequentialism and deontology are in one way characterized as recognizing different moral factors like individual well-being, justices and personal aspiration. These moral factors are commonly considered as irreducible to each other, and cannot be fully recognized if "consequence" is the only moral consideration – as it is in consequentialism.

The doctrine of consequentializing starts as a "logical interest of a new structural possibility".[23] It tries to explore possibility of viewing deontology in a consequentialist structure. Since it has been concluded that, without allowing agent-relativity, it is not plausible to consequentialize deontology. In the next chapter, I will explore other methods to consequentialize deontology in a different context.

By explicitly allowing agent-relative consequentialism (as Sen, Portmore, Louise do), proponents of consequentializing deontology tries to validate the possibility that all deontological theories can be given a consequentialist representation – so that a committed consequentialist is morally required to act in the same way as his deontological counterpart does in identical situation. The significance of trying to validate possibility of consequentializing is that, if such "extensional identity" exists, it provides a ground to question whether "moral factors" that are typically considered as non-consequentialist can be

---

[23] As Donald Regan commented on Sen's "Rights and Agency".

reduced to "consequence".

Chapter 3.   Deontology and Agent-Relative Values

3.1   Consequentializing Deontology and Agent-Relativity in Values

In previous chapter, I have shown that consequentializing deontology fails to give all deontological theory a consequentialist representation, with straight agent-neutrality. In this chapter, I will explicate a method to consequentialize deontology with agent-relativity allowed. The idea of agent-relative consequentialism is not new. Sen, for example, argued for an "evaluator-relative consequence-based morality"[24] before Parfit coined the terminology of "agent-relative and agent-neutral".

Agent-relative consequentialism holds that consequentialist reasoning is a theory of right that grounds moral rightness of action solely on values in its consequences. This logical structure is compatible with any theory of goods that defines what is valuable in consequences. An example of agent-relative consequentialism is egoistic hedonism, which holds that an act is right if and only if it brings about the most pleasure and least pain for the agent.

There are many possible ways to incorporate agent-relativity in consequentialism because the agent-relative/agent-neutral distinction can be applied to a range of entities: principle, perspective, rules, reasons for actions and values. In this chapter, I will limit agent-relativity applied to values for two main reasons. First, the theoretical structure of consequentialism is composed with a consequentialist theory of right and a theory of value. Typically, different consequentialist theories vary from each other by what they count as the good. Secondly, I believe agent-relativity in perspective and reasons for actions can be accounted with agent-relativity in values, in the context of consequentialism. Take Sen for example, in arguing for his "evaluator-relativity", Sen points out that "states include actions done", thus "the moral evaluation of a state by an agent may respond to the fact of his agency of his actions". If a state of affairs is valued from *the doer*, it follows that the

---

[24] See Sen A. (1982) and (1983).

evaluator would take the fact that it is *his action* into account. If a theory *requires* agents to value states of affairs *indifferently* as any other observers do, this theory should provide a special defense, not the doctrine of agent-relativity. To Sen, evaluator-relativity" is closely related to relativity in values.

3.1.1   The Deontology/Consequentialism Distinction and The Agent-Relative/Agent-Neutral Distinction in Values

The agent-relative/agent-neutral distinction has been applied in different ways to characterize the differences between deontology and consequentialism. As Louise (2004) summarizes, "the distinction can be drawn in terms of values (Pettit), aims (Parfit), reasons (Korsgaard) or rules (McNaughton & Rawling)."

When the distinction is applied to values, there are two major rival accounts concerning how the distinction applied in values can characterize the differences between deontology and consequentialism. The first holds that deontology recognizes values that consequentialism cannot recognize. Nagel (1986) and Emet (2010) are proponents of this stance. The basic reason behind is that: deontological reasons are agent-relative and consequentialist reasons are agent-neutral. Since agent-neutral value can only give rise to agent-neutral reasons for action, it is implausible to hold that consequentialism can recognize agent-relative values. This entails a dichotomy between three distinctions: (i) the deontology/consequentialism distinction; (ii) the agent-relative/agent-neutral distinction in reasons for actions; and (iii) the agent-relative/agent-neutral distinction in values.

On the other hand, Sen (1983), Dreier (1993), Portmore (2009) and Louise (2004) argue that consequentialism is not restricted to agent-relative values. As proponents of agent-relative teleology, which holds that consequentialism is not restricted to agent-neutrality, they argue that consequentialism can take agent-relative values into account.

Korsggard and Pettit each take a different approach to characterize the differences between consequentialism and deontology with the notion of "agent-relativity". Korsggard

argues that deontological constraints are neither agent-relative nor agent-neutral. She introduces the notion of "inter-subjectivity" and holds that deontological reasons, as well as values that give rise to them, are inter-subjective.

Pettit tried to draw the distinction between consequentialism and deontology in terms of *how* each theory responds to values. He introduces a distinction between honoring and promoting, and argues that deontology asks agents to *honor* values, while consequentialism asks agents to *promote* values. Under this theory, it is theoretically plausible for both deontology and consequentialism to recognize the same set of values, but requires agents to respond in different ways.

To explore possible ways to consequentialize deontology with agent-relative consequentialism, I will first, in this chapter, explore different accounts of deontological constraints in terms of values proposed by Nagel, Korsggard and Pettit.

Nagel explicitly argues that deontological constraints are not rooted in agent-neutral values. For if they are, reasons for action arise from these agent-neutral values should be agent-neutral as well, not agent-relative as constraints are. On the other hand, Nagel points out the difficulties in accounting deontological constraints in terms of agent-relative values. It is because constraints, not like options, are moral rules that require all persons to follow. When moral options are accounted based on agent-relative values, it is understood that they are subject to agent's personal interests and commitments. According to Nagel, it does not seem clear how agent-relative values of this kind can explain for the normative force of deontological constraints.

Korsgaard takes a different approach. She argues that deontological constraints are neither agent-relative nor agent-neutral reasons for actions. Instead, she holds, they are inter-subjective reasons, and the values that give rise to deontological constraints are inter-subjective values. Pettit's approach is of yet a different kind. He proposes a distinction between honoring and promoting, in order to characterize the differences between deontology and consequentialism.

I will argue that both Korsgaard's notion of inter-subjectivity and Pettit's

honoring/promoting distinction collapses. I will try to show that the notion of "inter-subjectivity" suggests a tripartite distinction in values and reasons for actions: agent-relative, agent-inter-subjective and agent-neutral. However, given that the "sharing" of values and perspectives can be viewed as a spectrum going from a single person (extreme agent-relative) to all persons that ever and will exist (extreme agent-neutral), it is unclear how Korsgaard may define inter-subjectivity in between.

I will also show that Pettit's honoring/promoting distinction fails because of three reasons. First, McNaughton & Rawling (1992) argues while all values can be, not all values can be honored. Secondly, the idea of "honoring" a value is ambiguous. An agent may be recognizing or endorsing the value, or he may be instantiating the value. In both case, it is plausible to hold that an agent honors a value. However, if "honoring" a value can be done without "instantiating" a value, an agent may not be following deontological constraint himself. That is, the honoring/promoting distinction cannot capture the differences between deontology and consequentialism.

Thirdly, if Pettit limits "honoring" to meaning "instantiation", it follows that the agent maybe "promoting" an agent-relative value. In fact, Pettit himself acknowledge that honoring of agent-neutral value is equivalent to promoting of agent-relative value. It makes the honoring/promoting distinction collapses into one between promoting agent-relative values and promoting agent-neutral values. The later distinction is also less ambiguous.

Different accounts of deontological constraints in terms of values seems to reveal a dilemma in accounting deontological constraints in terms of values, either as agent-relative or agent-neutral. On one hand, agent-neutral values seem to raise agent-neutral reasons for action only. They require agents to maximize agent-neutral values in consequences. On the other hand, agent-relative values are subject to particular perspectives. It seems implausible that values varies from agents to agents can be a source of normative power for deontological constraints.

This dilemma supports a claim by McNaughton & Rawling. They hold that "there are reasons which are not grounded in the promotion of any sorts of value, whether

value-for-the-agent (which grounds options) or impersonal value (which, on the consequentialist picture, grounds agent-neutral reasons). Appeal to value of either kind cannot ground agent-relative duties…" (1995, 37) The idea of consequentializing deontology with agent-relative values is challenged.

3.1.2   Consequentializing Deontology with Agent-Relative & Temporal-Relative Values

In the last chapter, I proceed to review the aim of consequentializing deontology. Following Louise, I will argue that whether deontological constraints are *indeed* grounded on agent-relative or agent-neutral values, or neither of them, is *irrelevant* to consequentializing deontology. Reviewing Nagel's account of deontology, I argue that the notion of "intent" is important in characterizing the *practice* of deontological constraints. I will explicate Louise's argument that deontological constraints can be represented in consequentialist term as maximizing values that are both agent-relative and temporal-relative. I will try to show that this way of consequentializing deontology is advantageous in two areas.

First, it is compatible with moral options in the sense that any consequentialized deontological theory can integrate with agent's personal interests and commitment, like any deontological theory does. Consequentializing deontology in this way does not entail implausible views concerning moral options, and it does not have to hold a weak version as it did in previous chapter.

Secondly, it helps clearing the paradoxical nature of deontological constraint in a consequentialist way. Represented as maximizing values that are both agent-relative and temporal-relative, deontological requirement can be viewed as consistent with consequentialist reasoning – that with a theory of good, an agent should be allowed to pursue the most of it. Consequentialist may still disagree with deontology and argue that constraints are implausible. But it may now point to the implausibility in theory of value, instead of holding that constraints are paradoxical.

3.2   Nagel: Agent-Relative Reasons and Agent-Relative Values

A central theme of Nagel's classical work *The View from Nowhere* is "objectification", a process of gradual detachment by putting a previous point of view into sight. Nagel argues that there is a form of "objective blindness" in science, philosophy of mind, and ethics. That "objective blindness" is to take the point of view from extreme detachment to account for the world. To Nagel, the completely objective point of view "inevitably leads to false reductions or to outright denial that certain patently real phenomena exist at all." (Nagel, 1986, 7)

Nagel's project is to seek a proper account for "objectivity" from which genuine yet subjective reality could be fairly taken into account to avoid "objective blindness", which takes different forms in science, metaphysics, and philosophy. When it comes to ethics, he believes it is a "persistent tendency... to seek a single complete objective account of reality," and "in the area of value that means a search for the most objective possible account of reasons for action: the account which engage us from maximally detached standpoint". (Nagel, 1986, 162)

Nagel accounts for such tendency towards maximum objectivity in ethics and values with agent-neutrality *and* consequentialism as he immediately puts afterwards:

> This idea underlies the fairly common moral assumption that the only real values are
> impersonal values, and that someone can really have a reason to do something only if
> there is an agent-neutral reason for it to happen. That is the essence of traditional forms
> of consequentialism: the only reason for anyone to do anything is that it would be better
> in itself, considering the world as a whole, if he did it. (Nagel, 1986, 162-3)

Nagel   characterizes   the   differences   between   consequentialism   and non-consequentialism as follows: The former concerns about what happens; it yields agent-neutral reasons and holds that the only real values to be concerned in ethics are agent-neutral values. The later concerns about what one does; it yields agent-relative reasons and holds that there are real values that are agent-relative. Such view also implies dichotomizing agent-relativity in values and reasons: agent-relative reasons are grounded on agent-relative values; while agent-neutral reasons are grounded on agent-relative values.

According to Nagel, there are mainly three types of agent-relative reasons, if they exist: reason of autonomy, reason of deontology and reason of obligation. Reason of autonomy, as Nagel states, "stems from the desires, projects, commitments, and personal ties of the individual agent, all of which give him reasons to act in the pursuit of ends that are his own." When an agent pursues something of his interest *without* violating any deontological constraints, he is acting out of reason of autonomy. I will use the term "reason of personal interests" interchangeably.

Reason of deontology "steams from the claims of other persons" that they are "not to be maltreated in certain ways". It takes the form of deontological constraints that restrict agents from performing certain kinds of actions. Constraints limit what agents are permitted to do in pursue of personal interests and commitment.

Lastly, reason of obligations "stems from the special obligations we have to whom we are closely related: parents, children, spouses, siblings, fellow members of a community or even a nation". It gives us reasons to act *partially* in favor of those we are closely related. However, it is seldom considered as *perfect duty* as reason of deontology is. That means, agents may act *partially* out of reasons of obligations, but they are not *morally forbidden* if they act otherwise.

In this part, I will only explore Nagel's account on the first two in details. It is because reason of obligation does not generate moral requirements *in the way as* reason of deontology does. As a kind of agent-relative reasons, reason of obligation shows agents are justified to act partially, if doing so does not violate moral constraints. This in a way resembles reason of autonomy as they both are *moral options*.

According to Nagel, reason of personal interests is given rise by agent-relative values. That is, whether there is a reason to climb a mountain for *A*, depends on whether *A* is interested in climbing a mountain. If *A* changes his interest to playing the piano, there is no longer any reason for him to climb a mountain. However, Nagel argues, it is not clear how deontological constraint can obtain its normative force from agent-relative value. It is because constraints are rules, something that requires every agent to follow. Holding that

constraints are given rise from agent-relative values makes it difficult to explain for their normative power.

For example, if the constraint against harming innocents steams from the claims of other persons that they are not to be maltreated in certain ways, it seems plausible to hold that the constraint against harming innocents is given rise by the disvalue in physical pain suffered by victims. However, there are two possible ways for interpretation: either it is the agent-neutral value of physical pain-in-general; or it is the agent-relative value of pain in a particular victim. If it is the former, the reasons given rise shall be an agent-neutral reason to *minimize* the total amount of pain-in-general being suffered. If it is the later, it seems theoretically difficult to explain why agent-relative value of pain in particular person can give rise to agent-relative constraint that governs another agent's action.

3.2.1   *Reasons of Autonomy*

Nagel argues that there are legitimate reasons for action that ground on personal interests. In his account, there are two types of personal interests: one concerning physical pleasure and pain; another concerning personal projects and preferences. The former type gives rise to agent-neutral values for promoting pleasure and minimizing pain; the later type gives rise to agent-relative values and reason of autonomy. Examining Nagel's account of how and why different types of personal interests stem different types of values gives a better understand of his view on the agent-relative/agent-neutral distinction.

The difference between personal interests (like pain and pleasure) that have claims on everyone's actions, and personal interests (like climbing the mountain) that do not, is an old one. The classical mountain climber case and Scanlon's god worshiper case, both stress on how different types of personal interests give rise to agent-relative and agent-neutral values.

Personal interests like satisfaction of personal projects, preferences, desires, generally bears only agent-relative values and ground agent-relative reasons only. *My* personal projects, if they have any values, are generally agent-relative values only, unless I take up "promoting

overall happiness" as my personal project. I have agent-relative reasons to work on my projects, but I do not have agent-neutral reasons for any others to help me in achieving these projects.

Physical pain, however, bears negative agent-neutral value that gives everyone a reason to stop that pain occurs to anyone. My pain, as mild as a headache, would give anyone a *pro tanto* reason to stop it if that is an alternative open to them.

Nagel tries to illustrate the difference with the notion of "objectification", which is a process of progressively putting previous point of view under the scene. A more objective perspective is formed, from which the previous point of view can be observed. Thus, a more objective account could be given to what is viewed and believed under the more subjective viewpoint.

Fsor example, when we see the sunrise and compare the size of the sun we observe to the size of a coin we hold in hand, we may conclude that "a coin is larger than the sun". However, when we proceed to a more objective point of view, and put this "sunrise-seeing and coin-in-hand" event into scene, we see that the belief is formed subjectively: the judgment is made base on the unequal observation distances. Thus, a more objective point of view falsifies the conclusion that "a coin is larger than the sun".

Nagel explains how objectification is applied to values. Many things we pursue in life, e.g. our personal goals, projects and concerns, are *optional*. As Nagel puts it, "they acquire value only because of the interest we develop in them and the place this gives them in our lives, rather than evoking interest because of their value." When we are the very persons who possess these desires, we see reasons to satisfy these desires – from our perspective.

However, when we start with the process of "objectification" and put these specific perspectives under a more objective point of view, we come to see that these desires are optional. We do not necessarily prefer football to baseball, or studying philosophy to playing music. Our personal goals, projects and concerns are in a great deal influenced by past experiences and can change over time. More importantly, when objectification includes perspectives of *other* individuals, it shows that our personal interests are limited to ourselves;

other individuals do not necessarily take up and share the same interests.

On the other hand, according to Nagel, it is a different case when objectification is applied to immediate personal interests like physical pain and pleasure. We do not disvalue physical pain by accident. The negative valuation of it is also shared between individuals. Consider a case in which Adam suffers from headache and Ann suffers from stomach pain. From a subjective point of view, both of them have agent-relative reasons to stop their own pain. From an objective point of view, their experience of pain can be generalized so that it is not subjected to particular person or experience. It gives all persons an agent-neutral reason to stop the pain.

To consequentialize deontology, consequentialist has to take reason of autonomy into account. As shown in previous chapter, consequentializing deontology by assigning an agent-neutral negative value makes it difficult to explain for agent's *choice* upon moral options. That is, it fails to give a consequentialist representation when a loyal deontologist acts for reasons of personal interests. To respond, consequentializing deontology retreats to a weak version, claiming that it only attempt to offer a consequentialist representation in which agents are not *required to violate a constraint* before any threshold is met. However, as argued in previous chapter, this response also fails. If consequentializing deontology can take reason of autonomy into account, it is plausible that it can maintain a strong version, which requires any agent to act *in the same way* as his deontological counterpart does.

3.2.2   *Reasons of Deontology*

In Chapter 1, I explicate different accounts of deontological constraints from Nagel (1986), Pettit (1987) and McNaughton & Rawlings (1991). I argued that the tripartite account for constraints as agent-relative reasons for action proposed by McNaughton & Rawlings is the most detailed and well-structured account to understand different types of constraints. In Chapter 2, I show how consequentialism attempted and failed to consequentialize deontology by assigning an agent-neutral value to violation of constraints. In this part, I will show a more established account of constraints by Nagel, which accounts

for constraints in terms of agent-relative values.

Deontological constraints restrict agent's actions in pursuit of the good, either it is personal or impersonal. Deontology recognizes a wide range of constraints: e.g. restrictions against lying and betrayal; the prohibitions against violating various individual rights, e.g. rights not to be killed, injured… (Nagel, 1986, 176) Nagel holds that these reasons of deontology, "stems from the claims of other persons not to be maltreated in certain ways". (Nagel, 1986, 165)

Constraints restrict agents from certain kinds of actions by giving each agent different moral aims. Say for example, the deontological constraint against murder gives *you* the aim that *you* do not murder; and it gives *me* the aim that *I* do not murder. It is in this sense that constraints are agent-relative reasons for action. It concerns what agents *do*, rather than what *happen*. On the other hand, agent-neutral reasons give all agents a single aim. In the case of murder, the single and common aim is to minimize total number of murders. Agent-neutral reasons would require agents to murder if that could reduce the total number of murders.

Nagel holds that if deontological reasons exist, they "cannot be explained simply in terms of neutral values, because the particular relation of the agent to the outcome is essential." (Nagel, 1986, 176) Similar to the case in reasons for autonomy, Nagel confines agent-relative values to agent-relative reasons; and agent-neutral values to agent-neutral reasons. This implies a dichotomized view on agent-relativity, between application in reasons and values.

Nagel tries to argue for a deontological reason against harming innocents with below case:

> You have an auto accident one winter night on a lonely road. The other passengers are badly injured, the car is out of commission, and the road is deserted, so you run along it till you find an isolated house. The house turns out to be occupied by an old woman who is looking after her small grandchild. There is no phone, but there is a car in the garage, and you ask desperately to borrow it, and explain the situation. She doesn't believe you. Terrified by your desperation she runs upstairs and locks herself in the bathroom, leaving you alone with the child. You pound ineffectively on the door and search without success for the car keys. Then it occurs to you that she might be persuaded to tell you where they are if you were to twist the child's arm outside the bathroom door. Should you do it? (Nagel, 1986, 176)

By stressing on the dilemma, he argues that there must be "a special reason against *doing* such a thing." For if saving a life outweighs some mild pain in a child, which makes twisting the child's arm *morally permissible*, there will be no dilemma at all. There is a reason *for* the agent *not* to twist the child's arm.

Below I will explore two possible accounts for values that evoke reasons of deontology: is it agent-neutral value, or agent-relative values? Following Nagel, I will first argue that deontological constraints do not stem from agent-neutral values. All agent-neutral values *per se* only give rise to agent-neutral reasons for action. Then, I will explicate Nagel's argument that deontological constraints ground on agent-relative values. As we shall see, reason of deontology and reason of autonomy differ in a great deal: the former is *required* while the later is *optional*. Thus, any account trying to rest deontological reasons on agent-relative values must further explain why some agent-relative values give rise to constraints, while some give rise to options. While Nagel do acknowledge and tries to address this issue, I will argue that his explanation is flawed.

Deontological Constraints and Agent-Neutral Values

As stated above, Nagel holds that reasons of deontology "stems from the claims of other persons not to be maltreated in certain ways". When these interests of *others*, e.g. their right not to be killed or injured, are generalized, they give rise to a set of agent-neutral values. Deontological constraints do not forbid harming of a particular person, or particular group, but it is a general requirement to everyone not to harm anyone. Thus, it may seem plausible to account normative power of constraints in terms of agent-neutral values.

However, such an account cannot explain for a basic character of constraints: agents are required not to violate a constraint even by doing so can avoid more violations by the agent himself in the future. If an agent is motivated purely by agent-neutral values, he would be taking up an agent-neutral perspective. From which, he can see the total number of murders *happening*. When guided by the agent-neutral disvalue of killing, an agent would try to minimize total number of murder, and not to avoid killing when he knows avoidance would

cause more killings to occur.

Thus it is difficult to rest deontological constraints on agent-neutral values. As Nagel puts it, "The relative character of the reason cannot come simply from the character of the interest that is being respected, for that alone would justify only a neutral reason to protect the interest." (Nagel, 1986, 178)

Deontological Constraints and Agent-Relative Values

Agent-relative values are acquired "only because of the interest we develop in them and the place this gives them in our lives." From his account for reasons of autonomy, Nagel holds that agents value their personal projects and thus have agent-relative reasons to work on it. Something is valued agent-relatively means it could have been otherwise and it is not necessarily shared by the others.

If a theory argues that deontological constraints stem from agent-relative values, it must also explains for the difference between reason of autonomy and reason of deontology – both are given rise by agent-relative values, but the former is optional, while the later is required. As Nagel puts it, "the peculiarity of deontological reasons is that although they are agent-relative, they do not express the subjective autonomy of the agent at all. They are demands, not options." (Nagel, 1986, 181) The difference between rationales grounding constraints and options is also pointed out by Scheffler, in his *independence thesis*, where his use the terms "agent-centred restrictions" and "agent-centred prerogatives":

> I believe that there is an underlying principled motivation for an agent-centred prerogative,… And I believe that this motivation is independent of any rationale there may be for agent-centred restrictions, in the sense that someone who is motivated in this way to accept a prerogative can at the same time consistently refuse to accept such restrictions. I will call this *the independence thesis*. (Scheffler, 1982, 81-2)

An account is required for the peculiarity of constraints as agent-relative reasons. Instead of identifying the kind of agent-relative values that grounds deontological reasons, and going further to differentiate them from options, Nagel offers a more detailed look on deontological requirements by pointing to the role of intent in deontological constraints.

Constraints concern what agents do, rather than what happen. Thus, it is commonly taken that constraints forbid agents from acting in certain way, *period*. As long as the *particular action* itself is not a token of the action-type restricted by a constraint, the agent is not violating a constraint. On the contrary, if the particular action is an action token of the forbidden action-tape, the agent is violating the constraints. However, Nagel points to the notion of "intent", with the principle of double effect to characterize deontological constraints:

> I believe that the traditional principle of double effect… provides a rough guide to the extension and character of deontological constraints… The principle says that to violate deontological constraints one must maltreat someone else intentionally. The maltreatment must be something that one does or chooses, either as an end or as a means, rather than something one's actions merely cause or fail to prevent but that one doesn't aim it.

Examples illustrating the principle of double effect show that, an agent may not be held as violating a constraint even if his action falls under an action-type forbidden by a constraint. Consider below Airplane Case and its implications:

A pilot is trying to land a plane fully loaded with passengers. The runway in the middle should be cleared, but there appears a truck with an oil tank. On the left, there is a few acre of unused land – with some children playing in the bushes. On the right, there is the seaside. To avoid hitting the oil tank or crashing into the water, the pilot decides to turn left, attempting also to avoid the bushes as much as possible.

Apparently there is deontological reason for the pilot *not* to turn left because of the foreseeable killing. However, this Airplane Case, and any other cases that demonstrate the principle of double effect, is meant to show that the act is permissible even from deontological perspective. Applying Nagel's view regarding the principle of double effect on this case, as the pilot does not aim to kill anyone during the landing *in order to* save lives in the plane, he does not violate the constraint against murder. It is the same conclusion when McIntyre (2011) summarizes implications of the traditional trolley case, that "it would be permissible to divert a runaway trolley onto a track holding one and away from a track

holding five: in that case one foresees the death of the one as a side effect of saving the five but one does not intend it."[25]

Constraints do not forbid a particular type of actions *per se*. Instead, they forbid intentional maltreatment of others, either as a mean or an end. Having clarified the character of constraints, I will proceed to examine how it is related to agent-relative values. Besides pointing to the intentional aspect of constraints, it is unclear that Nagel has specified any agent-relative values that support constraints in his *The View from Nowhere*. However, based on his view on agent-relative and agent-neutral reasons and values, as well as his characterization of constraints, we could try to figure what kind of agent-relative value is there behind constraints.

Nagel acknowledges the paradoxical character of constraints. Agent-relative constraints require agents not to violate even doing so could avoid more violations, either by others or the agent himself in the future. This is paradoxical because if something is worth an agent to avoid, it should always be better that more instances of which are *avoided.* However, Nagel points out that, if we are not to deny existence of deontological reasons, such paradoxical nature needs to be explained. We have already noted that pointing to negative agent-neutral value in violation of a constraint cannot account for this paradoxical nature – rather, it *shows* that it is paradoxical not to allow agents to violate a constraint in order to avoid more violations.

When we hold that constraint is paradoxical *because* it prohibits murder even if it is necessary in order to prevent more murders, we are committed to a view that *generalizes* different cases of murder. It takes generalization of murders by different agents, as well as

---

[25] Thanks to Professor Li Hon Lam for pointing out that Scanlon, in his work *Moral Dimensions*, argues against the doctrine of double effect. Scanlon differentiates "meaning" and "permissibility" of an action: the former refers to agent's reasons for action; the later refers to justifiability of an act. Scanlon holds that while the former depends on an agent's intent, the later, in general does not. For most cases in which permissibility of an action depends on the agent's intent, he argues, are cases in which "permissibility" depends on "meaning". I believe the debate on the doctrine for action does not fundamentally undermine my about consequentializing deontology. It is because the doctrine of double effect raise here is to illustrate how deontological constraints *operate* to determine whether an agent violates a constraint, rather than arguing for justification of deontological constraints. The doctrine of double effect, as argued by Scanlon, *may not* support the claim that permissibility of actions depends on agent's intent. However, it shows how deontology determines whether an agent violates a constraint. That serves our current purpose in looking for possible ways to consequentialize deontology, i.e. to give a consequentialist representation of deontology so that a consequentialist will act in the same way as its deontological counterpart does.

murders by the same agent at different times, in order to reach the conclusion that prohibiting *a particular* murder even if it is necessary to prevent *other* murders, is paradoxical.

Nagel tries to explain for the "paradoxical flavor" by pointing to the intentional aspect of an action and how it determines whether an agent has violated a constraint. In every case of intentional maltreatment the agent aims to violate other's right. In cases like twisting a child's arm in order to get the car keys and save your friend, such maltreatment is taken as a mean to achieve the good. Nagel comments on these cases,

> You are pushing directly and essentially against the intrinsic normative force of your goal, for it is the production of his pain that guides you. It seems to me that this is the phenomenological nerve of deontological constrains. What feels peculiarly wrong about doing evil intentionally even that good may come of it is the headlong striving against value that is internal to one's aim. (Nagel, 1986, 182)

Recall the case where you are twisting a child's arm in order to get the car keys and save your friend. If the child does not cry or his grandmother does not open the door and give you the car keys, you will twist his arm harder because creating harm is a mean to an end. This is what Nagel means by saying that the action is guided by the pain in child. I believe what Nagel tries to show in this case is that deontology and constraints ultimately concern the intentional aspect of an action, not an action *per se*.

My interests here is not whether the strong resentment in agents towards doing wrong, even in order to achieve greater good, can itself justify constraints. What I am trying to do is to clarify what constraint is about and how it operates in moral reasoning. If Nagel is right, the agent-relative values that can capture this intentional feature must be different from agent-relative values that give rise to options. I will leave this part for now, and revisit Nagel's account again after exploring other accounts of constraints in terms of values. I will close this chapter by proposing that if it is reasonable to account for constraints in terms of value, that value must be both agent-relative and temporal-relative.

3.3   Korsgaard: Deontology and Inter-subjectivity

Korsgaard argues that applying the agent-relative/agent-neutral distinction is not the best way to capture and account for normative force of values. She holds that deontology recognizes values that are "neither subjective nor objective, but rather are *inter-subjective*," as they "supervene on the structure of personal relations". In order to illustrate her point, Korsgaard introduces a further differentiation between two views of agent-neutral values: Objective Realism and Inter-subjectivism. (Korsgaard, 1993, 28) The former holds that all agent-neutral values are "good-absolutely", regardless of any existence of or appreciation by agents; while the later holds that agent-neutral values exist for all rational agents, but would not exist in a world without them. (Korsgaard, 1993, 27-8) Such distinction creates a differentiation between three types of values: objective agent-neutral, inter-subjective agent-neutral and agent-relative. I will show that such distinction between objective realism and inter-subjectivism is irrelevant in the discussion on agent-relative/agent-neutral values and deontological constraints.

Korsgaard's argument against the agent-relative/agent-neutral distinction in values is mainly a response to Nagel's example in *The View from Nowhere*: the dilemma of twisting a child's arm in order to save a friend who is seriously injured. As pointed out by McNaugton & Rawling, Korsgaard has two major concerns. First, the child as a victim has the right to object being harmed, and the right to object being used. However, Korsgaard believes "the theory that deontological reasons are agent-relative or only subjectively normative cannot accommodate the victim's right to complaint". (Korsgaard, 1993, 45) She holds that constraints framed as agent-relative reasons are normatively powerful only in a subjective manner. They govern an agent's conduct, but do not grant the victim any right to complain if other agents act against these reasons. Second, since a relative reason is something *for* that particular person who has it only, *others* can only acknowledge that there is such a reason for the possessor, but they can never have the reason to ensure that he acts on it, nor to complaint if he does not. (McNaughton & Rawling, 1995, 41)

These two ideas come from the same line of thoughts: constraints as agent-relative reasons are reasons *for* individuals alone. Neither victim nor any third party could appeal to these constraints as an objection against a person when he violates the constraint. Korsgaard argues, if these deontological reasons are given rise from agent-relative values of pain suffered by victims, victims would have agent-relative reasons to *want* the wrong-doers to stop, but they have no reason to *require* the wrong-doers to stop. It is simply because the victims and the wrong-doers do not share the same point of view, same agent-relative values or reasons.

> According to Objective Realism, subjective values are derived from objective ones: an individual comes to value something by perceiving that it has (objective) value. Our relation to values... is epistemological. ...According to Inter-subjectivism, objective values are... constructed from subjective ones. ...our relation to value is one of creation or construction. (Korsgaard, 1993, 28)

Korsgaard holds that on an inter-subjectivist interpretation, "neutral reasons are shared, but they are always initially subjective or agent-relative reasons," and this "makes it natural for an inter-subjectivist to deny that values can be added across the boundaries between people." For example, my happiness is something valuable for *me*, and your happiness is valuable for *you*. Korsgaard holds that, because we share this *valuing* of *our* happiness, the "happiness" itself is collectively constituted as an objective value, in the sense of *inter-subjectivism*, not *objective realism*. It is important to Korsgaard that the two senses of agent-neutral values are differentiated because the distinction captures the differences between consequentialist and deontological in their approaches to values:

My happiness is good for me and yours is good for you, but the sum of these two values is not good for anyone, so inter-subjectivist will deny that the sum as such is a value.[26] But an objective realist, who thinks that values are suited in objects rather than in their relation to the subjects, is indifferent to individual happiness. Base on this distinction, Korsgaard holds that consequentialism is based on objective values in the sense of objective realist, not

---

[26] Note that Korsgaard is not to deny the value that is "good for us". She only claims that such value, if exists, would not be the result of addition. It is collective, not aggregative goods. (c.f. footnote 13, Korsgaard, 1993)

inter-subjectivist because consequentialism is indifferent to values across agents. [27]

Korsgaard also rejects the possibility[28] that, deontological constraints could be stemming from agent-neutral values in objective realism sense. However, Korsgggard's argument involves two major ideas that are questionable. I shall explain each of them below.

In her work titled "The Reasons We Can Share: An Attack on the Distinction between Agent-Relative and Agent-Neutral Values", Korsgaard introduces the idea of inter-subjectivity to account for reasons and values because to her, the agent-relative/agent-neutral distinction[29] is not good enough to account for different types of reasons and values. She holds that there are reasons and values "supervene on the structure of personal relations". (Korsgaard, 1993, 25) In other words, they are neither agent-relative nor agent-neutral, but are "shared".

As explicated above, Korsgaard argues that if deontological constraints are agent-relative reasons for action that arises from agent-relative values, others, especially the victim, would not have any reason to stop the wrong-doings. McNaughton & Rawling argue that Korsgaard misses the distinction between options and constraints, the two senses in which a reason is personal. Reasons are personal in the sense as *options*, when they concerns personal *aims* that normally others do not share; but they are personal in the sense as *constraints*, when they are relative to personal point of view. Only reasons generated by options, and not those generated by constraints, are personal in the sense that they are solely the business of the possessor. (McNaughton & Rawling, 1995, 41-42)

McNaughton & Rawling further distinguish two levels of moral rules as to explain why victims and others can appeal to a shared reason to try to ensure that the wrong-doer act on *his* reason to avoid violation of constraints. The first-order level is deontological moral rules that restrict a particular person's actions. However, they hold that these agent-relative

---

[27] Here, Korsgaard made a disclaimer in footnote No. 14 that "the array of logically possible position goes far beyond the two that are schematically described in the text.... I am not concerned to discuss all possible theories of neutral values, but only the two I find most natural." So in fact, she is awarded of the possibility that an Intersubjectivist   holding that values can be added across the boundaries of persons, while an Objective Realist denying that values can be added. But she rests on the two theories that she finds *most natural*.
[28] See footnote No. 4 above.
[29] In Korsgaard, 1993, this is interchangeable as the subjective/objective distinction

reasons, although constituting the crucial part of deontology, are not the *only* kind of reasons that deontological theories can appeal to. There are second-order level reasons that are *shared* and agent-neutral. Consider "a deontology which said that each of us has reason not to harm the innocent, but that none of us has any reason whatever to ensure that others do not harm the innocent," as McNaughton & Rawling put, it "would indeed be very unattractive." Theses shared reasons are there to "promote the observance and even enforcement of the rules." Under these second-level rules, members of a society would have shared reasons to "educate, encourage and even discipline people into following the first-order rules." (McNaugton & Rawling, 1995, 41-3)

"Inter-subjectivity" is introduced to account for normative power of values that give rise to constraints. Their objectivity does not stem from being good-absolutely because they are valued subjectively in the first place. McNaughton & Rawlings tried to argue against Korsgaard that a deontological theory would not be a good one if it lacks second-order rules to promote or even enforce compliance of first-order rules by individual agents. But this only addresses part of Korsgaard's concern because she can still hold that normative power of those second-order rules are better accounted, indirectly, by "inter-subjective values". The nexus of the issue, I believe, is that the notion requires more detailed explications before turning a plausible account for constraints.

Recall that in Korsgaard's account, inter-subjectivism bears a kind of objectivity that differs from objective realism in the way that values are given rise from entity: it may get its value *from* the way it affects us, or that we *confer* value on it by liking, either collectively or individually. (Korsgaard, 1993, 34-35) Korsgaard argues for inter-subjective by explicating the structure of values in personal ambitions and deontological constraints. She holds that in both cases, neither agent-relative values nor agent-neutral values are good account. In the former case of personal ambitions, if we are to accept that the pursuit of it is solely justified by agent-relative values, we cannot explain why we would have reasons to support others in pursuit of *their* projects, unless we also appeal some objective or agent-neutral values like satisfaction of personal aspiration. While in discussion of deontological constraints, if we are

to accept that a wrong-doer only has some purely agent-relative reasons to avoid violating a constraint *himself*, we cannot explain why we would have reasons to stop him, unless we also appeal to the agent-neutral value of pain. Worse, if the victim only as an agent-relative reason to want a wrong-doer to stop torturing him, he would lack the ground to stop the wrong-doer, unless he also appeals to the agent-neutral value of wrong-doing.

Throughout the analysis, Korsgaard presents a dilemma between accounting constraints as rooted in agent-relative and agent-neutral values. In an example, she describes her "ambition to write a book about Kant's ethics that will be required reading in all ethics classes",

> Following Nagel's analysis, we will say that this ambition is agent-relate, since it gives me a reason to try to bring it about that my book is required reading, but it does not give anyone else a reason to require my book... But this way of describing the situation... suggests that my desire to have my book required is a product of raw vanity... This does not correctly reflect the structure of my ambition... It is an ambition to do something good... (Korsgaard, 1993, 37)

According to Korsgaard's analysis, an ambition without at least some traces of agent-neutral values in it, for example, to do good objectively, would lead us to "deny that it provides any kind of a reason, *even an agent-relative reason for its agent*." She claims that this is a plausible way to deny any ambition to have one's own statue on campus, because it would be simply be an *unmotivated* desire out of vanity.

Korsgaard recognizes possible objection that not every ambition is an ambition to do or produce something good, e.g. to climb a mountain, as if the action itself has some intrinsic value. And she feedbacks as follows,

> But neither does it seem right to say that those who pursue such projects are in the grip of unmotivated desire... There are reasons for caring about these things, reasons which are communicable and therefore at least potentially shareable. Ask a mountain climber why she climbs and she need not be mute: she may tell you things about the enlarged vistas,… She does not take the value of the climb to be conferred on it simply by her desire to do it. (Korsgaard, 1993, 39)

Korsgaard struggles between two ends of agent-relativity and agent-neutrality as applied

on values. On one hand, she struggles to offer some grounds for the all *personal* agent-relative value so that the value can account for ambitions that, intuitively worth others' support. On the other hand, she struggles to offer some grounds to account for values that agents confer collectively, in order to avoid theoretical requirement to claim that agent-neutral values are all good-in-itself. The struggle leads Korsgaard to reject the problematic agent-relative/agent-neutral distinction and find a new way to account for values.

However, the above explication shows that the struggle is rooted in a false dilemma between accounting values with agent-relativity and agent-neutrality. I believe the "inter-subjectivity" posit fails because (i) categorizing all values recognized by deontology as inter-subjective blurs the distinction between options and constraints; and (ii) even in respective to Korsgaard's aim to account for normative power of values in the cases she analyzed, "inter-subjectivity" does not seem a better way to fulfill the aim, as compared to agent-relativity and agent-neutrality.

Parfit has pointed out that in a sense, agent-neutral reasons can be interpreted as agent-relative because all reasons can be relative. It is because when two persons are acting for one common aim to promote something, they may still act in different ways. Since one of them may not be able to act to promote, or in a situation that it brings about a better result if he is not acting to promote the common aim.

> I should explain further the sense in which reasons can be relative. In one sense, all reasons can be relative to an agent, and a time and place. Even if you and I are trying to achieve some common aim, we may be in different casual situation. I may have a reason to act in a way that promotes our common aim, but you may have no such reason, since you may be unable to act in this way. Since even agent-neutral reason can be, in this sense, agent-relative, this sense is relevant to this discussion. (Parfit, 1984, 143)

Similarly, under Korsgaard's explication of inter-subjectivity, all values that ground deontological reasons, either as options or as constraints, are inter-subjective. Korsggard may not be confusing options and constraints as McNaughton & Rawling suggest, but when she argues that both personal ambitions and physical pain are inter-subjective, it is reasonable to

expect a detailed account of how inter-subjective values may ground both options in some cases, and constraints in another.

To argue against the idea of agent-relative and agent-neutral values, Korsgaard takes the extreme interpretation for both agent-relativity and agent-neutrality. For example, agent-relative values need not be all personal, private and unmotivated as if to simply saying "I *just* want to", as Korsgaard puts it, while agent-neutral values need not be all good-absolutely, as having no string attached to humanity. Even if we grant that, for the sake of argument, Korsgaard is well-grounded in holding that deontological reasons are rooted in inter-subjective values, instead of agent-relative values, it is still questionable how this may account for the normative power of deontological constraints.

The idea that inter-subjectivity differs from objective realist is that the former is subjective in the beginning. An inter-subjective value was appreciated subjectively by all, thus becoming an objective value (in the sense of being inter-subjective). Thus, the objectivity is *derived*. Below, I try to explicate how inter-subjective values operate in deontological constraints.

When an agent violates a constraint and hurts an innocent, the victim experience subjective value of pain.[30] Different subjective values of pain, which could be comprehended by all, collectively make up the inter-subjective value of pain. This inter-subjective value of pain, according to Korsgaard, grounds the constraint against harming others. Because it is inter-subjective, the victim, the agent, and others, all can appeal to the same deontological constraint and stop the violation. In this way, Korsgaard's concern about the victim not having any right to complain, or others not having any reason to stop the instance, are resolved.

However, this does not address how inter-subjective values would give rise to different "inter-subjective" constraints that restrict *you* and *me* from harming, respectively. Korsgaard may reply that it is because the objectivity in inter-subjective values is fundamentally

---

[30] I use "subjective" here instead of "agent-relative" in align with Korsgaard's terminology. She uses the terms interchangeably.

subjective. It entails that the different subjective values acknowledged by different agents are not commensurable. Thus, Korsgaard may argue, while different agents may appeal to the same deontological constraint, the fundamental subjectivity in values makes the constraint relative to different agents.

This response is not satisfactory. Constraints are requirements that restrict agent's actions. It is difficult to explain how a constraint restricts an agent from harming others when their subjective values of pain are *in conflict*, if its normative power grounds on the inter-subjective value of pain, which ultimately are individual subjective values of pain. The idea that the value is shared and objective would not address the problem because Korsgaard relies on the fundamental subjectivity in it to account for the relativity in constraints. She appeals to the "shared" feature of the values when arguing for constraint that everyone can appeal to, and appeals to the fundamental subjective feature of the values when accounting for the relativity of constraints. A more detailed account is needed to explain the relation between these two features and how they interplay to ground normative power of constraints.

In Korsgaard's analysis of ambition, she attempts to use the *same* inter-subjective value to explain for agent's reason for action, as well as others' reason to acknowledge his action. According to Korsgaard, if an ambition isn't *unmotivated* desire out of vanity like having one's own statue, it must be motivated by some values that is shareable by others. While others, motivated by the values we share, have the *pro tanto* reason to acknowledge my ambition.

However, it seems a flawed association between values, motives, and reasons for action. Consider a case that I am *motivated* to climb high mountains because of fame. I decide to climb the Everest and to raise money for a worldwide famous charity because that is an effective way to achieve fame. Thanks to the power of the internet and all kinds of social networks, my story is reported all over the world. Moved by my determination to risk my life and to endure all the pain in an unselfish manner, a lot of people, in fact over 5 million people around the globe, made donation to the charity. The project is a huge success.

In the above case, what motivates me is vanity fame, but it does not undermines the

value to raise money for charity by climbing the Everest, neither does it has any direct relation to others' acknowledgement on my action. Korsgaard attempts to *unify* the value that motivates an agent with the value that gives reasons to others to acknowledge and support the agent. But different agents could have different motives, values and reasons even they are acting for the same goal.

3.4   Pettit: The Honoring/Promoting Distinction as the Deontology/Consequentialism
        Distinction

Volumes of literatures have been written on how to characterize the differences between consequentialism and deontology. Traditionally it has been characterized in terms of the priority between the right and the good. With the introduction of the agent-relative/agent-neutral distinction, the distinction between the rival theories can be framed in a new way: deontology is agent-relative and consequentialism is agent-neutral. As shown in last chapter, the two theories were differentiated in terms of agent-relative and agent-neutral reasons. Pettit suggested yet a different way to characterize the distinction: each theory tells us a different way to respond to values. (See Pettit, 1989, 1991 and 1997) In this part, I will explicate the account and argue that it collapses into the distinction between agent-relative and agent-neutral values.

The "honoring" and "promoting" distinction is proposed by Pettit as to distinguish deontology and consequentialism. According to Pettit, the two theories *may* agree on the same set of values, but they differ in their way in respond to those values. Consequentialism asks agents to promote the values as much as possible, so agents are required to act in ways that maximize actualization of values. Deontology asks agents to honor the values wherever possible, so agents are told to act in ways that they personally perform (or in Pettit's word "instantiate", which means "being an instance of") the values.

Pettit uses several appealing examples to illustrate the distinction in different aspects. Roughly speaking, to promote a value is to maximize its overall realization. So to promote

honesty is to do what one can to ensure that there is as much honesty in the world as possible. To honor a value, by contrast, is to *be* honest even it cannot bring about as much honesty as possible. (Pettit, 1997, 125-127)

According to Pettit, honoring a value may also be in a sense, promoting a value. However, it is to promote the value in a perfect world, as if everyone else is also honoring the value. Pettit also uses Bertrand Russell as an example to show what promoting a value is like. In the First World War, the British philosopher opposes British involvement in the war even it got him into prison. The same promoting attitude towards "peace" led him to support British involvement in the Second World War because of his different estimation what brings about "peace". (Pettit, 1997, 125-127)

The account is intuitively appealing with a list of commonly known values: honesty, family value, friendship, physical integrity of others, etc. For each of these values, the honoring/promoting distinction can fully account the differences between deontology and consequentialism. One who honors friendship would not betray his friend even it promotes overall friendship in the world. One who honors physical integrity of others would not harm others even it avoids more people from being harmed. This coheres what constraints require from agents: no violation even it minimizes overall violations.

Both McNaughton & Rawling (1992) and Louise (2004) reject the distinction and hold that it collapses, but they differ in their stances and arguments. McNaughton & Rawling give three reasons why the honoring/promoting distinction collapses. First, although it works for some values, it does not work for all. While all values can be promoted, not all values can be honored. An example of which is the value of happiness. As the sole value in utilitarianism, happiness can be promoted. But McNaughton & Rawling argues that, contrary to Pettit's claim, there is not natural way of generating a single non-consequentialist counterpart to utilitarianism. Second, the distinction fails to capture all aspects that consequentialism and deontology differ. For example, there are some agent-relative rules which are plausible deontological principles, but cannot be described in terms of honoring. Third, the main strand of deontological thinking is that rightness is prior to value. Accounting deontology as

a theory of honoring values is distorting this central idea of deontology and putting it inevitably vulnerable to consequentialist attack. In short, McNaughton & Rawling (1992) believe that the honoring/promoting distinction cannot capture the differences between consequentialism and deontology; and deontology should not be accounted in terms of values. However, Louise (2004) only recognizes the first of these three arguments, and argues against the later two. Same as McNaughton & Rawling, she holds that the honoring/promoting distinction collapses. But she argues differently that deontology can be captured in terms of agent-relative values.

In what follows, I will argue that the honoring/promoting distinction collapses and the differences between consequentialism and deontology cannot be captured by the distinction. First, following McNaughton & Rawling, while all values can be promoted, not all values can be honored. Secondly, the idea of "honoring" is unclear and in some cases, it can hardly be distinguished from promotion of a value. Thirdly, following Louise, I will argue that the honoring of an agent-neutral value is equivalent to promoting an agent-relative value.

3.4.1 Values that cannot be honored

Base on Pettit's account, the honoring/promoting distinction represents two different attitudes towards values. Under this account, it is possible that consequentialism and deontology agree on the same theory of value and endorse the same set of values in morality, but differ in their attitude towards those values. It implies that there would always be a deontological counterpart for any consequentialist approach to promote a value. However, as McNaughton & Rawling point out, although all values can be promoted, not all values can be honored. An example is the value of "pleasure", which is the only value that utilitarianism recognizes. Utilitarianism, as the paradigm case of consequentialism, requires agents to bring about as much pleasures as possible. If Pettit is right in suggesting that the honoring/promoting distinction represents the respective approach towards values by deontology and consequentialism, there would be a deontological counterpart of

utilitarianism which requires agents to honor pleasures. Put in other words, it requires agents to be as happy as possible themselves. The resulting theory is egoistic hedonism, and can hardly serve as a deontological theory.

When we apply McNaughton & Rawling's formalization of agent-relative and agent-neutral rules, the honoring/promoting distinction works well to differentiate consequentialism and deontology in the value of honesty and loyalty to friends. (McNaughton & Rawling, 1992, 837)

Where S stands for "should [ceteris paribus] ensure that",

AR	$(x)(xS[(x \text{ is honest})]$;	(1)
AN	$(x)(xS[(y)(y \text{ is honest})])$.	(2)
AR	$(x)(xS[(y)(y \text{ is } x\text{'s friend} \rightarrow x \text{ is loyal to } y)])$;	(3)
AN	$(x)(xS[(y)(z)(z \text{ is } y\text{'s friend} \rightarrow y \text{ is loyal to } z)])$.	(4)

The first pair represents agent-relative and agent-neutral reasons for action related to the value of honesty. The second pair represents agent-relative and agent-neutral reasons for action related to the value of loyalty to friends. Both agent-relative reasons (1) and (3) require agents to act and exemplify the values. According to Pettit, agents are *honoring* the values. Both agent-neutral reasons (2) and (4) require agents to ensure as many actualization of the value as possible, regardless of whether himself act to exemplify the value or not. According to Pettit, agents are *promoting* the values.

McNaughton & Rawling point out that this kind of pairing between agent-relative and agent-neutral reasons works well in representing the honoring/promoting approaches to values in character traits. However, when it comes to values like health, wealth and happiness that are not character traits, it is difficult to see how an agent could honor the values themselves.

For example, when it comes to "pleasure", the only value recognized by utilitarianism, the formalization of agent-neutral and agent-relative reasons gives below:

AR	$(x)(xS[(x \text{ is as happy as possible})]$;	(5)
AN	$(x)(xS[(y)(y \text{ is as happy as possible})])$.	(6)

While (6) is a representation of utilitarianism, (5) formulates egoistic hedonism, which can hardly serve as a plausible deontological theory, not to mention as a deontological counterpart for utilitarianism.

3.4.2 *It is not always clear what counts as "honoring" a value*

"Honoring" differs from "promoting" by requiring the agents always to exemplify a value themselves even though it does not promote the overall value. However, as Louise argues, there are more than one way to exemplify a value: either I do whatever would exemplify the value *now*, or I do whatever would exemplify the value overall. It is unclear that Pettit has provided an account on which is the better way to honor a value and why.

Consider this case: in a campus that honor free speech, there comes a politician who spreads extreme Nazi beliefs. Which is better way to honor free speech: to allow the politician to give talks in the campus or not? In both ways, it is reasonable to say that the campus is exemplifying the value of free speech. Even when it rejects the politician to make public speech, it could be that the campus is honoring free speech by avoiding the value of free speech from being undermined in the campus. It is unclear why rejecting the politician to give public speech is not honoring the value. In short, it is unclear in some cases why promoting a value (not in the perfect world as suggested by Pettit) is not also honoring it.

One may respond that, since deontology requires agent not to violate a constraint even when it can avoid more of his *future* violations, it should be concluded that honoring a value requires agents to exemplify the value *now*. However, this is begging the question. The honoring/promoting distinction was proposed to differentiate the deontological and consequentialist attitudes towards values. It would be begging the question if we attempt to clarify the meaning of honoring a value by referring to deontology.

Further, by introducing the idea of temporal relativity, we can further distinguish the two possible ways to exemplify values in one's life: one is to always exemplify a value now or the other is to exemplify a value with one's life as a whole. They differ in whether

temporal relativity is concerned. In the former case, one is to maximize a value relative to him and relative to now, thus it is promotion of a value that is both agent-relative and temporal-relative. In the later case, one is to maximize a value relative to him but neutral to time. He acts as to promote the total value exemplified by himself in the long run. Thus, it is promotion of a value that is agent-relative but not temporal-relative.

This shows that when the value that is being "honored" is described in details, it can be characterized as a kind of promotion. This idea of *representing* honoring as promoting of agent-relative values leads us to explore if deontological constraints can be *represented* as promotion of certain kind of agent-relative values, with proper qualifications.

3.4.3   *Honoring is promoting agent-relative value*

Louise (2004) suggests that honoring of agent-neutral value is really promotion of agent-relative value. Reconsider these pairs of agent-relative and agent-neutral reasons of actions. Under the honoring/promoting account, agents acting on the agent-relative reasons are said to honor the values.

AR (x)(xS[(x is honest)];   (1)
AN (x)(xS[(y)(y is honest)]).   (2)
AR (x)(xS[(y)(y is x's friend -> x is loyal to y)]);   (3)
AN (x)(xS[(y)(z)(z is y's friend -> y is loyal to z)]). (4)
AR (x)(xS[(x is as happy as possible)];   (5)
AN (x)(xS[(y)(y is as happy as possible)]).   (6)

However, it is also true that agents following agent-relative reasons for action are acting in a way to ensure exemplifications of the value by themselves as much as possible. In other words, they seek to maximize some agent-relative values. In each of the agent-relative reasons for action above, agents are promoting *his* honesty, *his* loyalty to friends, and *his* happiness. The honoring/promoting distinction thus collapses into promotion of different types of values. As Louise holds, "when the value is specified precisely enough to be meaningful, the 'honoring' response is merely a form of promoting." (Louise, 2004, 533)

Louise (2004) introduced the idea of temporal relativity to distinguish two kinds of agent-relative values. An agent acting to promote a value that is both agent-relative and temporal-relative would always maximize the value *by himself* and *now*. An agent acting to promote a value that is agent-relative but temporal-neutral would always maximize the value *by himself* in the long run. He would choose to bring about less value *now* if he knows that would bring about more values by himself in the long run.

Below table shows different values under the agent-relativity and temporal relativity matrix:

	Temporal-neutral	Temporal-relative
Agent-Neutral	1. Agent-neutral, temporal-neutral	3. Agent-neutral, temporal-relative
Agent-Relative	2. Agent-relative, temporal-neutral	4. Agent-relative, temporal-relative

Below are different ways that agents act to promote the values correspondingly:

1. Agent always acts to maximize the values brought about in the consequences of his action, even when the action itself has negative impact on the value.

2. Agent always acts to exemplify a value himself as in the long run. He would sacrifice any current chance of exemplification if he knows that it will lead him to more explications in the future.

3. Agents always acts to maximize the values brought about *to all* in considered of the immediate impact of action taken, and not the consequences in long run.

4. Agent always acts to exemplify a value himself, now, regardless of the consequences or his own explications in the long run.

According to Parfit, any theory which claims incomplete relativity, i.e. agent-relative but temporal-neutral or temporal-relative but agent-neutral, should be rejected. (Parfit, 1984, 137-48) Base on Parfit's argument, Louise holds that both values in boxes 2 and 3 above do not truly exist.

However, I believe that Parfit's argument is irrelevant in our current aim to see whether

constraints can be consequentialized. It is because the above four ways of promoting values can be clearly differentiated from each other. And each of them is distinctly conceivable as a plausible action. In case deontological constraints can be represented in one of the forms with incomplete relativity, it becomes a challenge to the consequentialized theory to address the incomplete relativity. The idea that all theories with incomplete relativity should be rejected could be right based on Parfit's idea of personal identity. However, this idea does not by itself undermine the plausibility of consequentializing deontology.

McNaughton & Rawling hold a strong position against ascribing deontological reason to promotion of values of any kind.

> It is important to be clear, however, that there are reasons which are not grounded in the promotion of any sort of value, whether value-for-the-agent (which grounds options) or impersonal value (which, on the consequentialist picture, grounds agent-neutral reasons). Appeal to value of either kind cannot ground agent-relative duties (or type 1 agent-relative reasons). (McNaughton & Rawling, 1995, 37)

There are two reasons for this. First, there are plausible deontological constraints that cannot be described as honoring a value, which is promoted in agent-neutral reasons for action. The implication is that, while a person is acting under a constraint, his deontological reason is not grounded on any values, even with detailed explications. Second, the definitive feature of deontology as against consequentialism is that the former has the right prior to the good, while the later has the good prior to the right. Thus, deontological reasons are neither aimed to promote nor rooted in values.

To illustrate the first reason, McNaughton & Rawling consider "the suggestion that each parent has a special responsibility for the education of his or her children, in addition to any general responsibility we may all share to ensure that all children are educated" (McNaughton & Rawling , 1995, 842) They formulate the rules as:

AR   (x) (x S [(y) (y is a child of x -> y is educated)])   (A)

AN   (x) (x S [(y) (y is a child -> y is educated)])   (B)

They argue that although someone who follows the agent-neutral reason for action is

promoting the education of children, "it makes little or no sense to describe an adherent of agent-relative reason as honoring the education of children in her own life. Rather, she is promoting the education of *her* children." In this case, the agent is acting with the deontological reason that "parents should take care of their children's education". But she is not exemplifying the agent neutral value of education of children. The asymmetry in this agent-relative/agent-neutral pair of reason shows that the deontological reason is not grounded on any value.[31]

However, as Louise suggests, the agent-relative reason above prescribes that all should ensure their own children are educated. The agent-neutral counterpart of this rule is not that everyone should ensure the children in general are educated, but that everyone should ensure that all parents ensure that their own children are educated. So it should be formulated as:

AN   (x) (x S [(y)(z) (y S [z is a child of y -> z is educated])])

Louise argues that the agent-neutral value is that all parents ensure that their children are educated. Someone who follows the agent-relative reasons (A), like the agent in our example above, thus can plausibly be seen as honoring this agent-neutral value, or promoting the agent-relative value relative to her.

The second point McNaughton & Rawling suggest is that, since the definitive feature of deontology lies in the priority of the right and the good, deontological reasons cannot be consequentialized because they are not given rise by any kinds of values. Louise responds by arguing that the issue of priority is not really relevant to the discussion about consequentializing strategy. It is important to distinguish between the idea that deontology is "rooted in" or "given rise from" values, and the idea that the theory can be "represented" as a consequentialist theory, which will give the same moral requirements to agents as its non-consequentialist counterpart. In short, the problem of justification shall not be confused with the problem of theoretical representation.

---

[31] One may object that, since the agent is promoting the education of her own children, the deontological reason could be grounded on the agent-relative value of parental care. That is exactly Louise's response to the argument. However, we may appreciate McNaughton & Rawling's idea that, if a deontological constraint as agent-relative reason cannot be fairly described has having an agent-neutral counterpart that promotes a particular value, it shows that the deontological constraint is not grounded on agent-relative version of that value.

Louise's clarification is an important move in consequentializing strategy. When we go through arguments from Nagel and Korsgaard, we see difficulties in resting constraints on any kinds of values. While certain kinds of agent-relative reasons like options can be explained by agent-relative values, there is theoretical difficulty to account constraints in terms of values of any kind. The difficulty lies in the dilemma that on one hand, constraints as reasons for action *are* agent-relative in nature. It gives different agents different moral aims. On the other hand, when we seek to explain it with values, it seems to be something common, something that can be "shared" as in Korsgaard's term. But to explain deontological constraints in terms of agent-neutral values is incompatible with its agent-relative nature as reasons for action. McNaughton & Rawling argues that constraints are simply not based on any kinds of values. And this leads to a major block in the consequentializing strategy when the attempt is made to argue that all theories are teleological in nature but only differ in their theory of values.

However, Louise sets the argument of consequentializing strategy in the right direction by clarifying that whether a theory is *grounded on* promotion of values has nothing to do with whether the theory can be *represented as* promotion of values. The discussion is saved from being side-tracked to validate the paradoxical nature of constraints, or to offer justification for constraints in terms of values. These are irrelevant considerations to the consequentializing strategy. The only questions should be, whether deontology could be represented as teleological in promotion of some kinds of value.

In the next chapter, I will explore consequentializing strategy by Louise, Portmore and Sen. With reference to Nagel's insightful idea that the essential nature of constraints lies in agent's intention, I will argue that Louise's differentiation between temporal-neutral and temporal-relative values is the best way to represent deontological constraints in consequentialist form.

Chapter 4.   Consequentializing Deontology – Its Limits and Implications

4.1   Thesis Project Recap

Consequentializing deontology argues that all deontological theory can be given a consequentialist representation so that a committed consequentialist will act exactly the same as his deontological counterpart does in any identical situations. This thesis aims to clarify and evaluate different methods of consequentializing. If consequentializing deontology is successful, it gives a new way to seeing the difference between consequentialism and deontology.

Traditionally, consequentialism and deontology are taken as rival theories in normative ethics. Their differences are used to illustrate opposing stance in some basic ethical problems. For example, concerning how far should ethical theory direct one's action, deontologist argues that consequentialism is *too lax* because it allows agents to do whatever that leads to the best consequence. But deontologist believes that some actions are morally impermissible by nature.

Another example is related to moral reasons. Consequentialist argues that deontology is paradoxical because it forbids violation of constraints even when doing so can avoid more violations. From the consequentialist perspective, if violating a constraint once is intrinsically bad, violating a constraint more than once would be worse. Thus, it is implausible that an agent is forbidden to act against certain deontological constraint when it can avoid more violations.

Consequentializing deontology attempts to show that, when deontological theories can be represented in consequentialist term, the difference between the two theories can be framed differently.

The debate in consequentializing deontology is intertwined with the agent-relative/agent-neutral distinction. This formal distinction can be applied to various entities, e.g. reasons, principles and values. Philosophers from both sides of the argument

apply the distinction to illustrate and support their ideas. Nagel and Parfit use the distinction to formulate differences between consequentialism and deontology, and hold that the two ethical theories remain as rivals. Other philosophers like Portmore and Louise, apply the distinction to perspectives and values, and hold that agent-relativity is compatible with consequentialism, and all ethical theories can be consequentialized.

To clarify and review various consequentializing strategy, I first delineate core features of consequentialism and deontological constraints in Chapter 1. Classical consequentialism combines various distinct claims which may not be essential to consequentialism *per se*. For example, classical consequentialism takes consequences as opposed to actions, and hold that the normative status of the later depends on the former. However, as Carlson (1995) argues in his *Consequentialism Reconsidered*, the idea of "consequence" in consequentialism should be taken broadly as to include the action itself. A broadly construed conception of "consequence" is compatible with the idea of consequentialism, and it gives the theory a more defensible position. Following Carlson, I argue that consequentialism essentially consists of a broadly construed conception of "consequence", axiological principle and maximizing principle. I also examine various formulations of deontological constraints suggested by Nagel, Parfit and McNaughton & Rawling. I argue that the tripartite formulation proposed by McNaughton & Rawling best describes the essence of constraints: that they are agent-relative reasons for action.

In Chapter 2, I proceed to examine the first consequentializing strategy. The strategy is based on two ideas: (1) constraints are agent-relative reasons for action; (2) intrinsic moral wrongness of actions can be taken into account in consequence broadly construed. To start with a less controversial version of consequentialism, discussion in this chapter is carried out in the context of traditional, agent-neutral consequentialism. There are two types of constraints: absolute constraint and threshold constraint. The former can never be violated regardless of any morally relevant factors like consequence of the action. The later can justifiably violated if the threshold is met. The consequentializing strategy here is basically to assign a negative value to violation of constraint so that any consequence that involves

violation would not top the list of possible alternatives before threshold of the constraint is met. For example, if an agent is allowed to violate a constraint only when it can avoid at least five violations, a negative value of 5x can be assigned to an actual violation of constraint, where x is the value of each violation avoided. When five or more violations can be avoided through one violation, the negative value of an actual violation is balanced off. Thus the consequence that involves one violation and five or more avoidances of violation could top the list of alternatives in ranking of values. In the original deontology, the agent is *allowed* to violate the constraint because its threshold is met. In its consequentialized counterpart, the agent is not required to violate the constraint before the threshold is met.

I argue that this kind of consequentializing strategy is flawed. First, it faces certain theoretical difficulties by inferring ethical claims that are not held by or incompatible with the original deontology. For example, the theoretical perfectness of assigning an integer to represent where the threshold lies leads to an implausible claim that the consequence of one violation with five avoidances is as good as (and as bad as) the consequence of no violation with no avoidance at all. Second, it cannot capture the agent-relative nature of constraints as the negative value of violating a constraint is a constant, regardless of the identity of agents. Putting aside the flaws, this consequentializing strategy sets out with limited ambition: it aims only to provide a consequentialist representation of deontology, so that agents would not be required to violate a constraint before its threshold is met. It is questionable whether negatively ensuring an agent is *not* required to do something, is a fair criterion for consequentializing deontology, which claims to create a consequentialist counterpart for all possible deontological theories.

In Chapter 3, I deepen the discussion by examining how different applications of the agent-relative/agent-neutral distinction relate to each other. Since its introduction, the distinction has gained attentions and become central in ethical discussions within a short period of time. However, little literatures cover how its various applications relate to each other. In particular, it is unclear whether agent-relative and agent-neutral values must give rise to agent-relative and agent-neutral reasons for action respectively in dichotomy. Nagel,

Korsgaard and Louise all try to formulate the differences between consequentialism and deontology by applying the agent-relative/agent-neutral distinction in values, but they arrive in different conclusions. Nagel holds that constraints are not given rise by agent-neutral values. Although he does not further describe the kind of agent-relative values that give rise to constraints, he points to the intentional aspect of an action as the peculiar and central feature of constraints. Illustrating with traditional principle of double effect, Nagel holds that it is the intention that determines whether an agent is violating a constraint or not. This idea is related to our later argument on consequentializing constraint in terms of values, which are both agent-relative and temporal-relative, as proposed by Louise (2004).

Before introducing temporal relativity, I continue to examine Korsgaard's argument that constraints are grounded neither on agent-neutral nor on agent-relative values. She introduces the idea of inter-subjectivity and holds that constraints are grounded on inter-subjective values. I argue against Korsgaard and maintain that the idea of inter-subjectivity needs to be further specified in order to give a comprehensive account for constraints. I explain that there is a dilemma between accounting constraint on agent-relative or agent-neutral values. If agent-relative constraints are given rise by agent-relative values, it reminds questionable how it acquires its normative power as a requirement to *all* agents. On the other hand, there is theoretical difficulty to explain why agent-neutral values can give rise to agent-relative constraints. To explain for both agent-relative and agent-neutral nature of constraints by positing "inter-subjectivity" does not seem satisfactory.

Finally we come to Pettit's honoring/promoting distinction in order to differentiate deontology and consequentialism. This pair of distinction is introduced as to differentiate the two rival theories in terms of their response to values. Under Pettit's theory, consequentialism and deontology could agree to the same set of values but differ in their way of approaching those values. Deontology requires agents to honor the values, while consequentialism requires agents to promote the values. Following Louise and McNaughton & Rawling, I argue that the distinction ultimately collapses into promotion of agent-relative and agent-neutral values. For example, an agent is honoring honesty when he does not lie,

even if it can avoid more people from lying. The act of honoring can also be described as promoting agent-relative value of honesty. When what is indeed honored by the agents is clarified, it results in a form of promotion.

To close the previous chapter, I introduced Louise's idea of temporal relativity to differentiate 4 kinds of values. Below table shows different values under the agent-relativity and temporal relativity matrix:

	Temporal-neutral	Temporal-relative
Agent-Neutral	1. Agent-neutral, temporal-neutral	3. Agent-neutral, temporal-relative
Agent-Relative	2. Agent-relative, temporal-neutral	4. Agent-relative, temporal-relative

When we take a closer look of how agents responds to different values above, it gives us a better picture of what deontological constraints are amount to:

1. Agent always acts to maximize the values brought about in the consequences of his action, even when the action itself has negative impact on the value.

2. Agent always acts to exemplify a value himself as in the long run. He would sacrifice any current chance of exemplification if he knows that it will lead him to more explications in the future.

3. Agents always acts to maximize the values brought about *to all* in considered of the immediate impact of action taken, and not the consequences in long run.

4. Agent always acts to exemplify a value himself, now, regardless of the consequences or his own explications in the long run.

It seems that deontological constraints require agents to act under the forth category, in which agents always avoid violating a constraint by himself at the very moment, instead of pursing maximum agent-relative values in his whole life as corresponding to the second category.

In what follows, I will try to answer the question I set out in the first chapter, whether deontology can be consequentialized. With the above insight by Louise, I will argue that

when agents act in accordance to deontological constraints, it can be given a consequentialist representation that they are trying to maximize certain kinds of value. Next I will delineate the limitation of such representation. Some consequentialist, like Portmore, Dreier and Louise argue far reaching that *all* moral theories can be given a maximizing teleological representation, which only differ from each other by their theory of values. However, I will argue that at least two kinds of ethical theories that cannot be legitimately consequentialized with the above strategy. They are theories which hold that, at least in some situation: (1) none of possible alternative could be morally right; (2) more than one alternative could be morally right although they bring about different values. The first type includes theories that accept genuine moral dilemmas. The later type recognizes agent's option not to maximize values in consequences, while denying that the value of option can be captured in consequences of an action.[32] In response, I try to dismiss these two cases as extreme and do not fatally undermine the plausibility and applicable scope of consequentializing.

Then, I will address two possible challenges to consequentializing strategy. First, the consequentialized representations are made for particular constraints only, and not for deontological theories. Many deontological theories do not have theory of values itself, and they literally allow the agent to do anything but the actions governed by constraints. When given a consequentialized representation, an agent is always required to act for the best outcome. This is not held true in the original deontological theory. Second, giving a consequentialized representation is merely gimmicky. It trivializes consequentialism and distorts deontology.

I will argue against the first challenge by showing that a consequentialized representation mirrors the original deontology both in terms of moral relevance and deontic status of an action. I will address the second challenge by showing how successful consequentializing contributes to the discussions between consequentialism and deontology.

---

[32] It is important to state this qualification because if the value of options can be captured in consequences (like the idea of opportunity cost being commensurable with other economic values), we can argue that the ranking of consequences would be able to take options into account, thus not creating any challenge to consequentializing deontology.

Hopefully by doing so, it helps ease the worry that consequentializing is merely a play of word.

I will conclude this thesis by stating what successful consequentializing implies. First, it undermines the significance of the classical distinction between consequentialism and deontology. The distinction remains valid in meta-ethical discussion about relationship between values, motives and reasons for actions. However, in normative ethics, if consequentializing deontology is successful, the two theories would both seem to guide agents through value maximization. Second, although consequentializing strategy is usually and understandably rejected by deontologists because it seems to erode their uniqueness, representing deontological requirements in terms of agent-relative and temporal-relative values actually offers them reasons to defend against the long accusation that deontology is fundamentally paradoxical. The theory does not require agents to "act against himself". It consistently requires agents to act for the best, now. By embracing the Parfitian concept of agent with fully relativity, deontology requires agents to promotion of values that are both agent-relative and temporal-relative. Deontologists will be better positioned to argue for their theories if they open up to the consequentialized representation. Under the consequentialized representation, the two rival theories come to a platform that they "speak the same language". Traditional consequentialism can no longer dismiss deontology as mere paradoxical.

4.2   Consequentializing deontology and agent-relativity

Recall the different consequentializing approaches in Chapter 2 and 3. The first is to give constraint an agent-neutral value. It takes constraints into account only with their "agent-neutral" impact on values in outcomes, leaving the agents behind. We discover that this approach fails not only because it cannot accommodate agent-relativity, but also the approach faces some theoretical difficulties itself. For example, by assigning a negative weighing to violation of a constraint as to represent the point at which a threshold is met, the consequentializing strategy made an implausible implication: when an agent meet the

threshold to justifiably act against a constraint, the value in the state of affairs will just the same as that in which no violations or avoidance of constraints are involved at all. For example, below table shows how the consequentializing strategy assigns a weighing of 5 to represent the threshold at which an agent is allowed to violate a constraint if five avoidances can be achieved:

No. of avoidance achieved by a violation	Moral status of a violation according to deontology	Evaluation of states of affairs according to consequentializing deontology
0	Not justified	$\alpha - 5x$
1	Not justified	$\alpha - 5x + x$
2	Not justified	$\alpha - 5x + 2x$
3	Not justified	$\alpha - 5x + 3x$
4	Not justified	$\alpha - 5x + 4x$
5	Justified	$\alpha - 5x + 5x$
6	Justified	$\alpha - 5x + 6x$

Having shown that the first consequentializing strategy fails, I proceed from focusing on constraints as agent-relative reasons, to examining what kind of values give rise to constraints. It is because if we can clarify the values that ground constraints, it seems promising that a consequentializing strategy would be successful. To attempt consequentializing constraints, a natural way is to put them under consequentialist framework and look for a plausible account of them. A basic consequentialist structure consists of a theory of right that requires agents to always maximize the good, and a theory of good that defines what values to maximize. Since the first attempt fails, I try to explore what values ground constraints, and hopefully will arrive at a theory of good by that.

Both Nagel and Korsgaard explore the question of what values ground constraints. They arrive at a dilemma between accounting the normative power of constraints on agent-neutral values and agent-relative values but address it in different ways. If constraints are given rise by agent-neutral values, it can hardly explain why agents would be required to act for different moral aims. If constraints are grounded on agent-relative values, it is difficult to explain for its normative power that governs all agents. Instead of specifying any

agent-relative value that grounds constraints, Nagel stresses on the intentional aspect of an action as to determine whether an agent violates a constraint. While Korsgaard introduces the notion of inter-subjectivity and argues that it, but not agent-relative or agent-neutral values, accounts for deontology constraints.

After I show that the idea of inter-subjectivity entails ambiguity in distinguishing constraints and options, I proceed to examine Pettit's argument to different deontology and consequentialism in terms of their different approaches to values, namely, honoring and promoting. Following Louise and McNaughton & Rawling, I argue that the distinction between honoring and promoting collapses into promoting different kinds of values. As Louise suggests, when the horning requirement is delineated in details, it can always be described as promoting some agent-relative values.

I then revisited Nagel's idea that, apart from the action itself, the intentional aspect of an action matters in determining whether an agent has violated a constraint. Nagel illustrates this idea with the principle of double effect. The importance of this idea is that, the essential agent-relative nature of a constraint could possibility reside in the aim of an agent at the moment of action. Deontological constraints are agent-relative in the sense that they govern agent's actions "individually". But from the deontological perspective and with the principle of double effect, performance of an action that leads to alleged violation of a constraint is not considered as a violation if the agent lacks the aim to act against the violation.

If honoring can be represented as promoting agent-relative values, and if Nagel is right in pointing to the intentional aspect of an action as essential in conforming or violating constraints, it seems that Louise's approach of representing constraints as promoting values that are both agent-relative and temporal-relative is promising.

Louise makes it clear that the justification of deontological constraints in terms of values is irrelevant to the problem of formal representation. Whether deontological reasons are grounded on agent-relative values or not, as long as agents are rightly characterized as "promoting values that are agent-relative and temporal-relative" when they follow constraints, deontology is successfully consequentialized. The shift from justification

problem to representation problem avoids the difficulties in trying to *account* the normative power of constraints in terms of values.

The attempt to account normative power of constraints in terms of values grounds on the assumption that once it is successful, a theory of value can be established for deontology, and consequentializing deontology will be successful. However, this rationale presumes some claims that require a lot of validations. For example, it assumes all reasons are grounded in the promotion of some sort of values. As McNaughton & Rawling hold, it is not the case.

> It is important to be clear, …, that there are reasons which are not grounded in the promotion of any sort of value, whether value-for-the-agent (which grounds options) or impersonal value (which, on the consequentialist picture, grounds agent-neutral reasons). Appeal to value of either kind cannot ground agent-relative duties. (McNaughton & Rawling, 1995)

The focus shifts from finding the values that justify deontology to capturing how constraints operate to govern agent's action and motive. The justification problem becomes irrelevant to whether a consequentialist representation can be successfully outlined.

Nagel points out that, constraints do not only govern agent's actions but also their intentions. It seems to me that the idea is in support of Louise's argument that constraints can be given a consequentialist representation that promotes values that are both agent-relative and temporal-relative. Recapture the four kinds of values in below table and their corresponding actions by agents:

	Temporal-neutral	Temporal-relative
Agent-Neutral	1. Agent-neutral, temporal-neutral	3. Agent-neutral, temporal-relative
Agent-Relative	2. Agent-relative, temporal-neutral	4. Agent-relative, temporal-relative

1. Agent always acts to maximize the values brought about in the consequences of his action, even when the action itself has negative impact on the value.

2. Agent always acts to exemplify a value himself as in the long run. He would

sacrifice any current chance of exemplification if he knows that it will lead him to more explications in the future.

3. Agents always acts to maximize the values brought about *to all* in considered of the immediate impact of action taken, and not the consequences in long run.

4. Agent always acts to exemplify a value himself, now, regardless of the consequences or his own explications in the long run.

When constraints operate to restrict agents from certain actions, it is not requiring agents to minimize violations in his life as a whole. It is requiring agent to avoid violation *every time* he when violation is an alternative, as Category 4 specified. If an agent acts under Category 2, and decide to violate a constraint because it can avoid more violations by himself in the future, he is aiming to violate the constraint. For example, if an agent can only avoid further killings by driving on a child, he would make a second attempt if he missed the killing in the first time.

I will end this section by showing the values that are maximized by the three types of constraints captured in McNaughton & Rawling's tripartite account of agent-relative reasons: author-oriented, patient-oriented and author-patient-oriented rules. Below are examples of respective values that are being maximized in consequentialist representation:

Values Maximized	Author-oriented	Object-oriented	Author-object relation oriented
Agent-neutral	Agent-neutral value of honesty	Agent-neutral value of child caring	Agent-neutral value of parenting
Agent-relative	Agent-relative value of honesty	Agent-relative value of child caring	Agent-relative values of parenting
Agent-relative & Temporal-relative	Agent-relative, temporal-relative value of honesty	Agent-relative, temporal-relative value of child caring	Agent-relative, temporal-relative value of parenting

The temporal-relativity rests in the essential idea that deontology requires agents not to violate a constraint even it minimizes their violations throughout life time. Instead, they are asked to always avoid violation *now*, right at the moment of decision. The acute compelling moral aim for agents to always avoid violating the constraint at the moment of decision, is

well captured by consequentialist representation with maximization of values that are both agent-relative and temporal-relative.

4.3   Extended application of consequentializing to other ethical theories

Some philosophers argue that *all* moral theories can be consequentialized. For example, Portmore (2007, 2009) argues for a simple formula for consequentializing all non-consequentialist theories by creating a *deontic equivalent* consequentialist counterpart. I have argued against this approach in Chapter 2. Below is his idea of consequentializing all moral theories:

> Take whatever considerations that the nonconsequentialist theory says determines the deontic statuses of actions and insist that those considerations determine how their outcomes rank. In this way, the consequentialist can produce a ranking of outcomes that when combined with her criterion of rightness yields, in every possible world, the same set of deontic verdicts that the nonconsequentialist theory yields, such that, for any deontic predicate ('permissible', 'impermissible', or 'obligatory'), the resulting consequentialist counterpart theory and the original nonconsequentialist theory will be in perfect agreement as to the set of actions that are in the extension of that predicate. Such theories are deontically equivalent. (Portmore, 2009)

 Other philosophers, like Louise holds that all moral theories fall under "the consequentialist umbrella" that they can all given a consequentialist representation to promote value of some kind:

> …the categorization of theories ought to take into account not just agent-relativity/neutrality, but also temporal relativity/neutrality. Finally, although four distinct types of value could result from the combination of agent-relativity/neutrality and temporal relativity/neutrality, all plausible moral theories are likely to have only fully neutral or fully relative values. Since we are now all under the consequentialist umbrella, the question now becomes not whether we should be consequentialists or not, but whether we should be value-neutralists or value-relativists. (Louise, 2004)

However, if consequentialism is rightly characterized to be teleological, axiological and maximizing as I explicated in Chapter 1, it seems questionable that all moral theories can be

given a consequentialist representation. There are at least two kinds of ethical theories that cannot be legitimately consequentialized with the above strategy. The first type includes ethical theories that recognizes genuine moral dilemma. They hold that at least in some situations, none of the possible alternatives could be morally right. The second type includes ethical theories that allow options. They hold that at least in some situations, more than one alternative could be morally right.

I start with the first type of theories. A consequentialist representation structure entails that there is always one and only one action that is morally right. The idea that agents are doomed to act morally wrong in cases of moral dilemma, is apparently inconsistent with the maximizing feature of consequentialism. However, I believe this does not critically undermine the idea of consequentializing. In a moral dilemma, all actions available to an agent is morally wrong, however, the agent will normal not be held *blameworthy in the same way as* one who acts wrongly under no moral dilemma. This differentiates moral rightness of an action and blameworthiness of an agent. A deontological theory may hold that both lying and causing death of an innocent are *absolutely* forbidden by constraints. However, in face of dilemma between the two, it is still questionable whether deontologist would be indifferent to agent's choice of action. It is plausible that agent would be held less blameworthy if he violates the constraint of lying, in order to avoid the death of an innocent. Thus, a consequentialist representation seems still application to capture what essentially is required from the agent in terms of value. Only in extreme case when a theory recognizes genuine moral dilemmas *and* is indifferent to whatever the agent does in the situation, that consequentialist representation may not be inapplicable.

Now we turn to the second type of theories that recognize agent's option not to maximize values in consequences. It is generally believed that consequentialism is incompatible with options because it requires agent to always maximize values in consequence. However, I believe further clarification is required. A maximizing consequentialist representation can incorporate the value of agent-relative options as agent-relative and temporal-relative values. The idea is that, given any alternatives, an

agent's preference on an action has some moral bearing on determining what is morally right. For example, in a consequentialist representation that an agent's preference bears no moral relevance in evaluation of state of affairs, agents would always be required to act for the best outcome. However, if an agent's preference has some moral relevance in evaluation of state of affairs, he is still be required to act for the best outcome, but the ranking of alternatives may differ from previous case depending on the extend that personal preferences may impact evaluation of state of affairs. It is out of scope here to delineate the plausible ways to incorporate values of personal preferences. The idea is that it is plausible to mirror how options operate to take agent's personal preference into account. Consequentializing may still apply even when a theory recognizes options, unless the theory also denies that the value of option can be captured in outcome of an action. I would keep this open and not dismiss it as implausible before examining the nature of options, which deserves the length of another paper. For now, as I've shown that constraints can be consequentialized, I believe the same strategy can be applied in options, and the objection does not undermine plausibility and applicable scope of consequentializing strategy.

4.4 Fragmentation of values and consequentializing

Nagel (1979) examines a problem which he coin with the term "fragmentation of values", namely there are fundamental values that conflict with each other and cannot be reduced into a single value, i.e. they cannot *outweigh* each other in the straightest sense. These five types of values are: obligations, rights, utility, perfectionist ends, and private commitment. (Nagel, 1979, 131)

Fragmentation of vales poses challenge to consequentialism because in order to generate a unified ranking of states of affaires, it is believed that consequentialism has to assume that all morally relevant values are commensurable. If there are fundamental values that conflict with each other, this assumption certainly fails. Below, I will explicate four typical responses from consequentialism. I will not argue for single one but would like to

show that consequentialism can try to address the challenge in several ways.

The first one is to recognize that there are fundamentally different values. It is not necessary for consequentialism to assume that all values are commensurable. For example, Sinnott-Armstrong (1988, 81) and Railton (2003, 249-91) argue that this position enable them to recognize possible moral dilemmas.

The second is by appealing to practical wisdom. This strategy is applied by various philosophers like Nagel. The idea is that the conflicts between fundamental values are resolved in a way through agent's personal judgment. Practical wisdom is good judgment without total justification, yet is not irrational.

> Provided one has taken the process of practical justification as far as it will go in the course of arriving at the conflict, one may be able to proceed without further justification, but without irrationality either. What makes this possible is judgment – essentially the faculty Aristotle described as practical wisdom, which reveals itself over time in individual decisions rather than in the enunciation of general principles. (Nagel, 1979, 135)

Practical wisdom is not only advocated by consequentialist to address the problem of conflicting values, it is also used by deontologist when fundamental considerations cannot outweigh each other.

It shall not be confused that appealing to practical wisdom in making moral decision entails that the theory also recognizes irresolvable moral dilemmas. A theory may hold that our understanding in moral justifications may not always bring about justified and undisputable moral decisions in daily life, and this is where practical wisdom is called for. It is not the same as holding that there are genuine moral dilemmas that no general principle could possibly resolve.

The third is proposing a lexical ranking that governs various values. The idea is that, even though these values are not *reducible,* they are *commensurable* through a scale that specifies their relative order of importance. For example, "never infringe general rights, and undertake only those special obligations that cannot lead to infringement of anyone's right..." (Nagel, 1979, 131) Or more fine-grained scale like only certain level of personal

commitment may outweigh special obligations.

This method has gained a wide range of support from philosophers like Williams (1985), Scheffler (1982) and Portmore.

Lastly, it is by personal perspective as suggested by Jackson (1991). He borrows the notion of subjective probability from decision making theory and argues that consequentialism should not be requiring agents to act for what is *objectively* the best. In his example, a physician has three drugs to treat her patient. Drug A is very likely to relieve the condition but will not completely cure it. One of drugs B and C will completely cure the condition, while the other would kill the patient. Jackson argues that objective consequentialism as proposed by Railton (1984) would require the physician to *either* use drug B or C because that is what would lead to *in fact* the best result. Taking subjective probability into account, neither prescribing drug B or C is the intuitively right answer.

Jackson argues that in making moral decisions, consequentialist has reasons to prefer the "sector plan" to the "scatter plan", that is, to act partially towards those few that we are close and those projects that have great influences on our life for most of the time. Consequentialist can reply the challenge of incommensurability by holding that, in most cases, agents have more reasons to act on the things that they *can* bring about a positive impact.

I shall close this part with Nagel's mirroring between the situation in moral pursuit and that in scientific explorations:

> To look for a single general theory of how to decide the right thing to do is like looking for a single theory of how to decide what to believe. Such progress as we have made in the systematic justification and criticism of beliefs has not come mostly from general principles of reasoning but from the understanding of particular areas, marked out by different sciences, by history, by mathematics. Theses vary in exactness, and large areas of belief are left out the scope of any theory. …We are familiar with this fragmentation of understanding and method what it comes to belief, but we tend to resist it in the case of decision. (Nagel, 1979, 135-6)

The nature of value is still to be clarified and resolved. However, it shall not stop other

investigations in moral decisions and justification. While fragmentation of values may constitute a challenge to consequentialism because it is an axiological theory, it does not fundamentally undermine the idea of consequentializing with above possible ways to address the challenge.

4.5   Conclusion: How successful consequentializing contributes to moral discussions

It may be objected that consequentializing trivializes consequentialism on one hand, and distorts deontology on the other. By taking away essential features in *plausible* consequentialist theory, such as universality and agent-neutrality, proponents of consequentializing claim that even hedonistic egoism is a consequentialist theory as it holds promotion of egoistic pleasure as morally right. On the other hand, they present deontology as requiring agents to evaluate alternatives, which is not what deontology actually *do*.

Regarding the distortion of deontology, McNaughton & Rawling argue that not all reasons should be grounded on values. In particular, normative power of constraints cannot be accounted in terms of values. They constantly try to help deontology resist the consequentialist "vacuum sucker" by stressing on the essence of deontology: the right prior to the good. As indicated before, I believe that it is helpful to differ between two approaches of consequentializing. The first is trying to show that agent-relative constraints are given rise by certain kinds of values, and to consequentialize is to maximize these values. The second is trying to offer a maximizing consequentialist representation of what agents do when they are governed by agent-relative constraints. The first approach touches on *justification* of constraints while the second approach does not. By taking the first approach, consequentialists are essentially trying to justify constraints in consequentialist terms. As I have shown in Chapter 3, that approach leads to a dimmed end. However, the question of justification is irrelevant if consequentialists take the second approach.

For the trivializing of consequentialism, it seems that by introducing much implausible "kin" under the umbrella term, the notion of consequentialism is shattered. It is worth to

clarify that, existence of some implausible consequentialist theories does not fundamentally harms the plausibility of consequentialism. Take hedonistic egoism as an example, it must be clearly delineated where its implausibility resides. It seems to root in its theory of value by taking egoistic pleasure as solely valuable and morally relevant. Its implausibility has nothing to do with its consequentialist structure. As argued in Chapter 1, a consequentialist structure does not bound to agent-neutrality. Since it is not obvious that for a moral theory to be plausible, it can only recognize agent-neutral values, I believe that as long as consequentialist representation has the essential consequentialist skeleton features such as teleological, axiological and maximizing, it does not trivialize consequentialism in any sense.

I will conclude this thesis with two important implications of successful consequentializing. First, successful consequentializing undermines the significance of the classical distinction between consequentialism and deontology. The distinction between consequentialism and deontology has long been taken as the central and only relevant schema in understanding moral theories. The consequentialism/deontology distinction has contributed an insightful framework in normative ethical discussion. However, I believe that seeking a new framework for discussion is in itself philosophically interesting. Throughout the paper, it has been shown that the agent-relative/agent-neutral distinction is a useful distinction and taxonomy in setting a new framework for the discussion. As Driers (1993) puts it,

> Moral theory can progress by illuminating the variety of structures and contents that normative schemes can have, improving our map of the overall landscape. Or, it can progress by attempting to reduce the variety of plausible or justifiable options, weeding out incoherent ones or trying to drive us to one spot or another by appealing to our settled convictions…. I believe that the contrast between agent neutral and agent centered theories, and not the contrast between consequentialist and nonconsequentialist theories, is a fundamental part of the most useful taxonomy. (Drier, 1993)

When deontology and consequentialism are both fairly represented as promotion of certain values, discussion between the two traditionally rival theories could be broaden. The

discussion becomes more dynamic when it does not limit itself to priority of rightness verses goodness; nor different kinds of reasons for action. A deeper understanding of how agent-relative and agent-neutral values are related to our moral life now directly benefits the discussion between the two traditionally rival theories.

Second, although consequentializing strategy is usually and understandably rejected by many deontologists as it seems to erode their uniqueness, representing deontological requirements in terms of agent-relative and temporal-relative values actually offers them reasons to argue against the old accusation that deontology is fundamentally paradoxical. I believe the traditional consequentialist approach of dismissing constraint totally as paradoxical has not done its best to interpret constraints in its strongest position. It hastily judges that the only way to incorporate constraints in an axiological schema is to take the impact of its conformity and violation agent-neutrally. By approaching the agent-relative constraints with the principle of charity, a consequentialist representation of constraints is the best effort to "make sense of" deontology in a consequentialist framework.

As I have argued, constraints are best described as promoting values that are both agent-relative and temporal-relative. It consistently requires agents to act for the best, now. With such representation, the Parfitian concept of agent with fully relativity could be taken as offering an auxiliary support of why agent-relative and temporal-relative values matter. Under the consequentialized representation, the two rival theories come to a platform that they "speak the same language". Traditional consequentialism can no longer dismiss deontology as mere paradoxical.

My aim in this thesis is to make use of the agent-relative/agent-neutral distinction and explore possible ways to close the gap between consequentialism and deontology. There are numerous important texts on justifications for each theory, and I feel inadequate to contribute in that area. However, by arguing for a consequentializing strategy that gives constraints a fair consequentialist representation, I believe I have contributed to inspire a new perspective to tackle the long discussion between the two moral theories.

Bibliography

Brown, C. (2010). Consequentialise this. (Unpublished). University of Edinburgh, UK.

Carlson, E. (1995). *Consequentialism Reconsidered.* Dordrecht: Kluwer Academic
    Publishers.

Dreier, J. (1993). Structures of normative theories. *Monist, 76*(1), 22-41.

Emet, S. F. (2010). Agent-relative restrictions and agent-relative value. *Journal of Ethics &
    Social Philosophy, 4*(3).

Jackson, F. (1991). Decision-Theoretic Consequentialism and the Nearest and Dearest
    Objection. *Ethics, 101*(3), 461-482.

Kagan, S. (1989). *The Limits of Morality.* Oxford: Clarendon Press.

Korsgaard, C. M. (1993). The reasons we can share: An attack on the distinction between
    agent-relative and agent-neutral values. *Social Philosophy and Policy, 10,* 24-51.

Lenman, J. (2009). Reasons for action: justification vs. explanation. *The Stanford
    Encyclopedia of Philosophy.* Retrieved from
    http://plato.stanford.edu/entries/reasons-just-vs-expl/

Louise, J. (2004). Relativity of value and the consequentialist umbrella. *The Philosophical
    Quarterly, 54*(217), 518-536.

McIntyre, A. (2011). Doctrine of double effect. *The Stanford Encyclopedia of Philosophy,
    Fall 2011 Edition*, Edward N. Zalta (ed.). Retrieved from
    http://plato.stanford.edu/archives/fall2011/entries/double-effect/

McNaughton, D., & Rawling, P., (1991). Agent-relativity and the doing-happening
    distinction. *Philosophical Studies, 63,* 167-185.

    --- (1992). Honoring and promoting values. *Ethics, 102*(4), 835-843.

    --- (1993). Deontology and agency. *Monist, 76*(1), 81-101.

    --- (1995). Value and agent-relative reasons. *Utilitas, 7,* 31-47.

Moore, G. E. (1903). *Principia Ethica.* New York: Cambridge University Press.

Nagel, T. (1970). *The Possibility of Altruism.* Oxford: Clarendon Press.

--- (1979). *Mortal Questions*. New York: Cambridge University Press.

--- (1986). *The View from Nowhere*. New York: Oxford University Press.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Clarendon.

Pettit, P. (1987). Universalizability without utilitarianism. *Mind, 96*(381), 74-82.

--- (1989). Consequentialism and respect for persons. *Ethics, 100*(1), 116-126.

--- (1991). Consequentialism. In P. Singer (Ed.), *A Companion to Ethics*. Cambridge,

Mass., USA: Blackwell Reference.

--- (1997). The consequentialist perspective. In Baron, M. W., Pettit, P., & Slote, M.,

Three Methods of Ethics: A Debate. Malden, Mass.: Blackwell.

Portmore, D. W. (2001). Can an act-consequentialist theory be agent relative? *American

Philosophical Quarterly, 38,* 363-377.

--- (2003). Position-relative consequentialism, agent-centered options, and

supererogation. *Ethics, 113,* 303–332.

--- (2005). Are moral reasons morally overriding? (Unpublished). Arizona State

University, USA.

--- (2007). Dual-ranking act-consequentialism. *Philosophical Studies, 138,* 409–427.

--- (2009). Consequentializing. *Philosophy Compass 4*(2), 329–347.

Ridge, M. (2005). Reasons for action: agent-neutral vs. agent-relative. *The Stanford

Encyclopedia of Philosophy, (Winter 2011 Edition),* Edward N. Zalta (ed.). Retrieved

from http://plato.stanford.edu/archives/win2011/entries/reasons-agent/

Scheffler, S. (1982). *The Rejection of Consequentialism: A Philosophical Investigation of the

Considerations Underlying Rival Moral Conceptions*. Oxford: Clarendon Press.

Sen, A. (1982). Rights and agency. *Philosophy and Public Affairs, 11*(1), 3-39.

--- (1983). Evaluator relativity and consequential evaluation. *Philosophy and Public

Affairs, 12*(2), 113-132.

--- (1993). Positional objectivity. *Philosophy and Public Affairs, 22*(2), 126-145.

--- (2000). Consequential evaluation and practical reason. *The Journal of Philosophy,

97*(9), 477-502.

Sinnott-Armstrong, W. (1988). *Moral Dilemmas*. Oxford: Blackwell.

--- (2011). Consequentialism. *The Stanford Encyclopedia of Philosophy, Winter 2011 Edition*, Edward N. Zalta (ed.). Retrieved from

http://plato.stanford.edu/archives/win2011/entries/consequentialism/

Slote, M. & Pettit, P. (1984). Satisficing Consequentialism. *Proceedings of the Aristotelian Society, Supplementary Volumes, 58,* 139-163+165-176.

Smith, M. (2003). Neutral and relative value after Moore. *Ethics, 113,* 576–598.

Sosa, D. (1993). Consequences of consequentialism. *Mind, 102*(405), 101-122.

Vallentyne, P. (1987). The teleological/deontological distinction. *The Journal of Value Inquiry, 21*(1), 21-32.

Williams, B. (1985). *Ethics and the Limits of Philosophy Cambridge*. MA: Harvard University Press.