# 規範邏輯的理由轉向

梁廷瑋

## 中文摘要

本文試圖釐清「理由」（reason）、「規範」（norms）與「行動」（actions）之間的關聯，並建立一套基於理由的規範邏輯（reason-based deontic logic）系統。這裡的「規範」，意旨「規範性命題」（normative proposition），也就是包含了「應該」、「允許」、「禁止」、「免為」等規範詞（deontic word）的命題。規範邏輯沿襲了語句邏輯（sentence logic）的推導形式，將思維的對象聚焦在規範性命題上，由此延伸出了不少至今未解的麻煩與悖論。在後設倫理學近幾十年來的發展中，分析哲學家已經將對於規範的討論，逐漸轉向了對於理由的討論；然而，規範邏輯卻基於歷史的因素，一直囿限在傳統的討論框架當中，無法跟進倫理學的新進展。這一方面是由於：規範邏輯一開始作為模態邏輯（modal logic）的副產品，難以脫離可能世界語意學（possible world semantics）的傳統根源；另一方面是因為：即便後設倫理學發生了「理由轉向」，但對於理由的性質、理由如何「關聯於」行動，仍然是百家爭鳴，更遑論形式化的可能。為此，本文採用了王一奇的「基於差異製造的理由論」，將理由定義為「由行動所導致的差異結果」，並以舊瓶裝新酒，將規範邏輯的架構套用至因果模型語意學（causal modeling semantics）當中，進而摸索出規範邏輯的理由轉向可能性。

本文共六章，各章分述如下：

第一章是**緒論（Introduction）**，本章的目的在於概述後設倫理學與規範邏輯雙方各自的發展。後設倫理學進入到了「理由轉向」之後，對於「規範」與「理由」之間的關係有了更清楚地闡明：若 A 應該做 P，則 A 存在做 P 的理由 R；並且，我們對於「應該」的宣稱，是我們對於種種理由考量的綜合結果。另一方面，規範邏輯劃分出了四種的規範狀態：應該（Obligatory）、允許（Permissible）、禁止（Impermissible）、免為（Omissible）；並通常以「應該」作為初始運算子（primitive operator），定義出其他的規範狀態。

第二章是**規範邏輯及其悖論（Deontic logic and Its Paradoxes）**，本章的目的在於探尋規範邏輯的悖論及其產生原因。第一節（2.1）探討了規範邏輯的緣起，規範邏輯的出現，來源於馮．賴特（von Wright）對於模態邏輯形式上的拓展，使得同一套語意學系統能夠處理多種哲學問題。然而第二節（2.2）所羅列的種種悖論重重打擊了馮．賴特在規範領域的企圖心，本節總結了悖論發生的原因：1、規範邏輯無法脈絡化地處理各種情境下的應然判斷。2、規範邏輯無法區分規範系統中的應然以及真正繫於行動的應然。3、規範邏輯混淆了規範的「存在」（existence）以及「效力」（evaluation）。4、規範邏輯實際上無法窮盡所有的規範

狀態。而這四種原因都導向一個根本問題：語句邏輯的推論形式並不符合實際應然的判斷模式。

第三章是**規範及理由（Norms and Reasons）**，本章的目的在於釐清規範與理由的性質。第一節（3.1）指出，之所以規範邏輯混淆了規範的「存在」以及「效力」，來自於一個最常見的誤解：規範被視為一種理由，從而能對我們的行動產生效力。本節透過語法分析，指出規範性命題實際上是一句陳述句（indicative sentence），目的在於斷言理由的存在；因此，規範的「存在」是指存在規範性命題的斷言，規範的「有效」則是指規範性命題所作的斷言為真。第二節（3.2）則從王一奇的「基於差異製造的理由論」出發，將理由定義為行動到結果的因果鍊。然而，僅僅因果練並不足以解釋行動的意圖面向，因此我在王一奇對於理由的定義當中增補了態度（attitude），並將行為導致態度的各種形式對應了規範邏輯的四種規範狀態，形成了定理 1（Theorem 1）。

第四章是**基於理由的形式化（Reason-Based Formalization）**，本章的目的在於建構一套基於理由的規範邏輯系統。我們採用因果模型（causal models）的進路，其包含結構方程（structural equations）與因果圖（causal graphs），本文將之與期望值（expectations）作結合。第一節（4.1）指出，將規範邏輯化約至條件句（conditionals）的設想，最早溯及至安德森（A. R. Anderson），然而安德森當時尚未出現理由的討論，以至於其想法並未獲得很好的發展。本節重新詮釋了安德森的化約方案，將條件句進一步發展為反事實（counterfactuals），並套入「基於差異製造的理由論」當中。第二節（4.2）則由此給出期望值公式，並以是否戒菸（Quitting Smoking）為例，解析我們對日常正反理由的權衡方式。第三節（4.3）更進一步分析出我們歸咎責任的模式，因果的充分性（sufficiency）是我們歸咎責任的重要考量，並且，因果圖正描摹出了我們對於責任歸屬的理解；因此每一幅因果圖，正代表一種道德直覺。本節以著名的電車難題（Trolley Problem）為例，說明在雙重效果的難題下，因果圖是如何描摹出我們的歸責模式。

第五章是**悖論的解決（Resolution of the paradoxes）**，以理由為基礎的規範邏輯採用了不同的語意學系統，而這樣的語意學表述為<V, E, A>。其中，<V, E>來自於因果模型的表述形式，而 A 則是在規範領域中，基於對規範性命題的態度而給予的賦值（assignment）。定理 1 已經給出了原子語句（atomic sentences）的賦值規則，而對於由布林函數（boolean function）所組成的複合語句（compound sentences），本章則提出了定理 2（Theorem 2）來處理，並透過新的語意學系統回過頭來逐一解決悖論。

而第六章則為**結論（Conclusion）**，本章重新回顧了悖論發生的起因，來自於語句邏輯背後真值函數（truth functions）的預設。真值函數最終只會讓我們關注真假值的結果，而因果模型所採用的結構方程，會讓我們更加關注行動結果的產生。因此，規範邏輯將真值函數替換成結構方程是有必要的。

# Deontic Logic Turning to Reasons

Niû Têng-úi*

## 1. Introduction

The purpose of this paper is to elucidate the relationship between reasons, norms, and actions while striving to develop a formal system of reason-based norms. The term "norm" here expresses a normative proposition that employs deontic words such as "ought to", "permitted", "forbidden", or "exempted". The relationship between normativity and reason has been at the forefront of discussions on normativity in recent years. We often use "reason" to capture the notion of normativity: if $A$ ought to do $P$, then there must be a reason $R$ for $A$ to do $P$. The relationship between a normative proposition and a reason is a sufficient condition, not an equivalence relation; for example, the usage of "ought to" implies the existence of a reason, but the existence of a reason does not directly imply "ought to". This is because reasons can be defeated. The fact that one has a reason to do $P$ does not mean one ought to do $P$ because one may also have a reason not to do it. We will only say that one ought to do $P$ if there is a reason to do $P$ that is not defeated.

On the other hand, normative propositions can largely be captured by deontic logic with the following four normative statuses:

It is obligatory that p, if and only if $Op$
It is permissible that p, if and only if $Pp$
It is impermissible that p, if and only if $Fp$
It is omissible that p, if and only if $Mp$

The normative statuses mentioned above correspond to deontic words such as "ought to", "permitted", "forbidden", as well as "exempted", and deontic operators $Op$, $Pp$, $Fp$, and $Mp$, respectively. The first three are the most frequently cited, and the fourth is often not labeled, but we can derive it from the first normative status by adding the negation $\neg Op$. The rest of the normative statuses can be defined in the same way. Typically, one of the first two is taken as primitive, and the others are defined in terms of it, but any of the first four can play the same defining role. The most prevalent approach is to take $Op$ as primitive and define the rest as follows:

$$Pp \stackrel{\text{def}}{=} \neg O \neg p$$
$$Fp \stackrel{\text{def}}{=} O \neg p$$

* Department of Fundamental Legal Studies, National Taiwan University (r09a21005@ntu.edu.tw).

$$Mp \overset{\text{def}}{\equiv} \neg Op$$

However, since the development of deontic logic predates the discussion of the theory of reason, contemporary discussions of deontic logic tend to ignore the issues with or debates about reason theory due to historical factors. But reasons are the sources of the validity of normative propositions, and all of their claims must arise from specific reasons. Reasons are different from normative propositions in that the former are context-sensitive, while the latter are often "outcomes" derived from the considerations of reasons, making normative propositions abstract and action-oriented in nature. In general, normative propositions are the conclusions of practical reasoning, while reasons are the premises of practical reasoning (Ruey-Yuan Wu 2015: 3). Reasons thus differ from normative propositions while maintaining a significant relationship with them. If, in discussing normative propositions, we ignore any issues with reasons, the deontic logic that results will contain significant gaps, which will lead to paradoxes and will not meet our intuition of moral inference.

Therefore, this paper aims to integrate the discussion of reasoning into the subject of deontic logic and propose a new formalism of moral inference. Progress in this area has been made possible by the recent development of causal models and their application to reason theory by Linton Wang (2015) and Peng-Hsiang Wang (2015), who define reason as a causal path from action to its outcome. Their approach has provided considerable inspiration for this article. Interestingly, today's deontic logic has also moved in the direction of normative propositions to context-sensitive conditionals, which will be traced in this paper to lead to the discussion of reason theory.

## 2. Deontic logic and Its Paradoxes

### 2.1 About Deontic Logic

The emergence of deontic logic can be attributed to the development of modal logic and a continued curiosity about and inquiry into its function. "Modal" denotes an expression with a mode of existence and presentation of things extending beyond the actual world. When we try to express modes of presentations of things or to assert universal or existential, we call such expressions or assertions modal propositions. The logic dealing with these modal propositions is called modal logic. In the narrow sense, a modal proposition includes the phrases "necessarily" and "possibly". Most modern philosophers use the term "possible world" to construe and understand the concept of modality, which defines a modal proposition's "necessarily" or "possibly" by describing the truth status in each possible world. Therefore, in possible world semantics, we define "necessarily $p$" and "possibly $p$" in the following way:

**Necessarily** $\boldsymbol{p}$ Proposition $p$ is necessary; if and only if $p$ is true in all possible worlds.
**Possibly** $\boldsymbol{p}$ Proposition $p$ is possible; if and only if $p$ is true in some possible worlds.

Typically, operator $\Box p$ and operator $\Diamond p$ are used for "necessary $p$" and "possible $p$", and they are mutually defined as follows:

$$\Box p \leftrightarrow \neg\Diamond\neg p$$
$$\Diamond p \leftrightarrow \neg\Box\neg p$$

The field of possible world semantics has been developed significantly by Saul Kripke, who defined a model of possible world semantics by a triple <**W, R, V**>. **W** is a non-empty set of possible worlds, and according to the canonical model, a possible world is defined as a maximal consistent set of sentences. **R** signifies an accessibility relation, which is a set of binary relations between possible worlds. Finally, **V** signifies valuation, which assigns truth values to each atomic sentence at each world included in **W**. This type of model is also known as the Kripke model, which allows the axioms of modal logic to be adjusted to particular philosophical positions or paths through the setting of accessibility relations **R**. Moreover, von Wright (1957: 58) found that modal concepts can be extended to other philosophical issues from concepts of "necessarily" and "possibly", allowing many philosophical concepts to be construed with modal logic. Thus, von Wright has classified four main kinds of modal concepts as follows:

| | | |
|---|---|---|
| **Alethic** | necessary | possible |
| **Epistemic** | know/believe | not know not/not believe not |
| **Deontic** | obligatory | permissible |
| **Existential** | universal | existential |

In deontic logic, "obligatory" is construed by analogy with "necessary", and "permissible" is construed by analogy with "possible". Typically, deontic logic uses axiom K $\Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q)$ and axiom D $\Box p \rightarrow \Diamond p$ as its accessibility relation, and replaces modal operators $\Box$ and $\Diamond$ with deontic operators $O$ and $P$ respectively. Meanwhile, other kinds of deontic logic systems select more axioms to construe the institution, which the former cannot fit to form a standard deontic logic (SDL) as follows:

**SDL**

A1. All tautologous formulas of the language

A2. $O(p \to q) \to (Op \to Oq)$

A3 $Op \to Pp$

R1 If $\vdash p$ and $\vdash p \to q$ then $\vdash q$

R2 If $\vdash p$ then $\vdash Op$

A2 and A3 come from axiom K and axiom D. A3 is an axiom specific to deontic logic, which comes from our intuitive understanding of the norm: if $p$ is obligatory, then $p$ is permissible. Even so, the settings of A2, A3, and R2 are contestable, primarily since the setting of R2 is derived from the rule of necessitation in normal modal logic. Still, in this way, it implies that all logical tautologies are obligatory. This rule so seriously violates our intuition that, as Linton Wang (2015) and Niû Têng-úi (2022) point out, a valid reason explanation for a reason must be able to illuminate the counterfactual, and the ability to illuminate the counterfactual is a *necessary quality for an explanation* to have. Logical tautologies, however, merely highlight a constant status quo and cannot be used as reasons for actions. The earliest version of deontic logic proposed by von Wright (1951: 10-11) directly sets the principle of deontic contingency to rule out R2 rule[1]. The main reason is that von Wright believes deontic logic is similar to epistemic rather than alethic and existential status. Just as the alethic status does not affect the epistemic status, the deontic status should not be affected by the alethic status.

## 2.2 The Paradoxes of Deontic Logic

Unlike other logic systems, various counterexamples have challenged deontic logic since its inception. As von Wright says:

> Traditional deontic logic was born as an offshoot of modal logic. Was this a happy birth or a miscarriage? I am afraid we must say that it was not an altogether happy birth. Perhaps a reason why deontic logic has aroused such lively interest has been an implicit feeling that the whole enterprise was problematic. The symptoms of illness were the existence of various anomalies and the fact that doubt could be raised on intuitive grounds about the validity of so many of its formulas (von Wright, 1980: 420-421).

The developments in deontic logic that are taking place presently can be seen as nothing more than administering first aid to this dying premature child. Multiple factors have caused deontic logic's illness. First, it is necessary to examine the paradoxes

[1] In von Wright's (1951: 11) original article, deontic contingency means "A tautologous act is not necessarily obligatory, and a contradictory act is not necessarily forbidden."

surrounding the most controversial R2 rule:

***Case 1*** (Ross's paradox)

(A) It is obligatory that the letter is mailed ($m$).

(B) It is obligatory that the letter is mailed ($m$) or the letter is burned ($b$)

Case 1 is Ross's paradox (Ross, 1944), which can be formalized as $Om \rightarrow O(m \vee b)$. We would think that the derivation of (A) to (B) in Case 1 is counterintuitive, but such a derivation is a theorem in SDL. Because of R2, we can derive the above theorem by the Rule of Addition, as follows:

***Reasoning 1***

(1) $\vdash m \rightarrow (m \vee b)$ (Add)

(2) $\vdash O(m \rightarrow (m \vee b))$ (R2)

(3) $\vdash Om \rightarrow O(m \vee b)$ (A2)

(1) is the so-called rule of addition, a theorem in classical logic. (2) is derived from (1) by R2, and (3) is derived from (2) by A2. In addition to "obligation", there is also a "permission" analog, $Pm \rightarrow P(m \vee b)$, that seems quite bizarre in Ross's paradox, which can be derived as follows:

***Reasoning 2***

(1) $\vdash (p \wedge q) \rightarrow p$ (Simp)

(2) $\vdash O((p \wedge q) \rightarrow p)$ (R2)

(3) $\vdash O(p \wedge q) \rightarrow Op$ (A2)

(4) $\vdash \neg Op \rightarrow \neg O(p \wedge q)$ (Contra)

(5) $\vdash P\neg p \rightarrow P\neg(p \wedge q)$ (DN)

(6) $\vdash P\neg p \rightarrow P(\neg p \vee \neg q)$ (DeM)

(1) is the so-called Rule of Simplification, and (4) is the so-called Rule of Contraposition, which are also theorems in classical logic. In (5), DN means deontic negation, which is derived from the mutual definition of deontic operators as stated above. When we view m as $\neg p$ and b as $\neg q$, then (6) can be the result of the formula $Pm \rightarrow P(m \vee b)$. However, it still goes against our intuition. We can find the same counterintuition by replacing "obligatory" with "permissible" in Case 1.

Moreover, Reasoning 2 will lead to another bizarre paradox similar to the one in Case 1. When we translate (3) into natural language, it is easy to determine which parts of it are counterintuitive. Consider the following case (Prior, 1958):

***Case 2*** (Good Samaritan Paradox)

(C) It is obligatory that Tim rescues Tom ($p$) from drowning ($q$).

(D) It is obligatory that Tom is drowned ($q$).

In Case 2, from (C) formalized as $O(p \wedge q)$, we can derive (D) formalized as $Oq$. However, it hardly seems that if rescuing the drowning man is obligatory, it would follow that he ought to be likewise drowning, as such a conclusion would again go against our intuition. We can consider Case 2 to be another variant of Ross's paradox because the Rule of Addition and the Rule of Simplification can essentially define each other. There is also a "permission" analog, $P(p \wedge q) \rightarrow Pp$, present in Case 2. Thus, we can proceed with the Reasoning from (3) in Reasoning 1 as follows:

***Reasoning 3***

(1) $\vdash Om \rightarrow O(m \vee b)$ ((3) in Reasoning 1)

(2) $\vdash \neg O(m \vee b) \rightarrow \neg Om$ (Contra)

(3) $\vdash P\neg(m \vee b) \rightarrow P\neg m$ (DN)

(4) $\vdash P(\neg m \wedge \neg b) \rightarrow P\neg m$ (DeM)

Also, when we view $p$ as $\neg m$ and $q$ as $\neg b$, then the formula $P(p \wedge q) \rightarrow Pp$ can result in (4). We can also find it to be counterintuitive when replacing "obligatory" with "permissible" in Case 2.

Such paradoxes occur not only because of the theorems of classical logic but also because of the theorems of modal logic, one of the most discussed of which is the Paradox of Epistemic Obligation. Consider the following case (Åqvist, 1967):

***Case 3*** (The Paradox of Epistemic Obligation)

Alexis works as a security guard at a bank. One day, a gang comes to rob the bank, but Alexis is completely unaware of it. We would think that Alexis ought to be reprimanded because:

(E) It ought to be the case that Alexis knows that the bank is being robbed.

(E) is formalized as $OK_A s$, where $K_A$ means "Alexis knows" in epistemic logic, and $s$ means "the bank was robbed". Typically, epistemic logic uses axiom T $\Box p \rightarrow p$ as its accessibility relation, and replaces modal operators $\Box$ with epistemic operators $K$, so $Kp \rightarrow p$ (KT) is the theorem in epistemic logic, which is also a common sense in the field of epistemology: knowing that $p$ means that $p$ is true. We can therefore deduce the following:

***Reasoning 4***

(1) $\vdash K_A s \rightarrow s$ (KT)

(2) $\vdash O(K_A s \rightarrow s)$ (R2)

(3) $\vdash OK_A s \rightarrow Os$ (A2)

According to (3), since it is obligatory that Alexis knows that the bank is being robbed, it follows that it is likewise obligatory that the bank is being robbed; but this is also counterintuitive. All of these paradoxes appear, at least, to depend on R2, and R2 itself also feels so alien to our intuition that some philosophers naturally single it out as an idea that requires further examination. In addition, von Wright, S. O. Hansson (1990; 2001) also developed a non-normal logic system where R2 is not sound. However, other philosophers (Castañeda, 1981; Tomberlin, 1981) consider that R2 is not the real source of the paradox. Castañeda notes that in Ross's paradox, no one would merely utter the statement that one ought to mail or burn the letter, knowing that, in fact, he or she ought to mail the letter.

The occurrence of Ross's paradox can also bypass the derivation of R2, deriving $Op \vee Oq$ from $Op$ simply through the Rule of Addition. In natural language, there is no difference between the meanings of "the letter ought to be mailed or the letter ought to be burned" and "the letter ought to be mailed or be burned". The problem is that, in normative propositions, we tend to understand disjunctive propositions as conjunctive propositions (Ross, 1941). We instinctively understand the following two statements to be equivalent:

(F) You are permitted to participate in the meeting physically or virtually.

(G) You are permitted to participate in the physical meeting, and you are also permitted to participate in the virtual meeting.

(F) focuses on the actor's "choice" while (G) focuses on the "existence" of the norm; the former uses "or" as its conjunction while the latter uses "and", but both express the same "normative proposition". Such equivalence, however, cannot be processed by deontic logic because deontic logic has confused evaluation (right/wrong) with the existence (true/false) of norms.

In the Good Samaritan Paradox, drowning is not considered a choice of behavior but rather a background condition that triggers the validity of the norm. Further, in the Paradox of Epistemic Obligation, "ought to know" means "ought to do something to make oneself know as much as possible", because "to know" is a state rather than an action and is also a background condition for the validity of the norm. These paradoxes confuse the context in which the norm applies and translate it into formal and abstract logical symbols. As outlined in my next inference, such confusion is a crucial factor from which all these paradoxes arise, so we need to find a new way to formalize "contexts". Before that, it is necessary to review one more puzzle caused by the failure to clarify the specific context in which the norm applies. Let us consider the following four premises (Chisholm, 1963):

***Case 4*** (Chisholm's Puzzle)
(H) It is obligatory that Jones goes to assist his neighbors ($p$).
(I) It is obligatory that if Jones goes ($p$), then he tells them he is coming ($q$).
(J) If Jones does not go ($\neg p$), then he ought not to tell them he is coming($\neg q$).
(K) Jones does not go ($\neg p$).

For Case 4, we can have three logical interpretations, the first of which is the most straightforward one, while the second and third come from P. McNamara and Frederik Van De Putte (2021):

| | **Interpretation 1** | **Interpretation 2** | **Interpretation 3** |
|---|---|---|---|
| Premise (1) | $Op$ | $Op$ | $Op$ |
| Premise (2) | $O(p \to q)$ | $p \to Oq$ | $O(p \to q)$ |
| Premise (3) | $\neg p \to O\neg q$ | $\neg p \to O\neg q$ | $O(\neg p \to \neg q)$ |
| Premise (4) | $\neg p$ | $\neg p$ | $\neg p$ |

According to Interpretation 1, Premise (1) and Premise (2) can lead to $Oq$, and Premise (3) and Premise (4) can lead to $O\neg q$; however, $Oq$ and $O\neg q$ are contradictory, which means it is obligatory and likewise impermissible that Jones tells his neighbors that he is coming. Such a contradiction often occurs when there is a hierarchy present in the priority of norms. This is because, in deontic logic, norms of different orders are placed on the same plane of judgment, leading to such contradictions. Such a contradiction is also triggered by the following: "Although you are permitted to take time off, you ought to attend the meeting", also roughly interpreted as $P\neg p \land Op$ in SDL.

Nevertheless, we can try to avoid the contradiction by altering the scope of the deontic operator, which is the case for Interpretation 2 and Interpretation 3. For Interpretation 2, whether Jones goes to assist his neighbors or not is just a condition, which is a fact that will trigger a specific norm. For Interpretation 3, the entire conditional relationship is considered a norm. However, in the premises of Interpretation 2, Premise (2) follows from Premise (4), and in the premises of Interpretation 3, Premise (3) follows from Premise (1), which shows that the premises do not have joint independence (P. McNamara and Frederik Van De Putte, 2021). The derivations of Interpretation 2 and Interpretation 3 show that all paradoxes based on conditionals likewise occur in deontic logic, which makes the paradoxes of deontic logic more perplexing. The paradox of conditionals, as it occurs in deontic logic, is known as The Paradox of Derived Obligation, which is formalized as follows:

(L) $Op \rightarrow O(q \rightarrow p)$
(M) $O\neg p \rightarrow O(p \rightarrow q)$

Both of these are theorems in SDL. (L) means that if $p$ is obligatory, then it is also obligatory that $p$ if anything, and (M) means if $p$ is impermissible, then it is likewise obligatory that anything if $p$. An example of the former can be given as "if it is obligatory that John helps others, then it is also obligatory that John helps others if killing his parents". Likewise, an example of the latter can be given as "if it is impermissible that the people rob a bank, then it is obligatory that the people kill the security guard if robbing a bank[2]". Both inferences are contrary to common sense.

In sum, the factors that cause illness of deontic logic can be roughly narrowed down to the following. First, deontic logic is unable to contextually symbolize situations that deal with conditions, hierarchies, and alternatives among the various norms and jumbles them all within the same logical structure; thus, when triggering conditions of norms or conflicts of obligation and choice, deontic logic will have its limitations.

Second, deontic logic is limited in its ability to differentiate among norms based on their underlying reasons, as well as distinguish between "obligations within a normative system" and "real obligations involving actions". It also fails to depict the "process of consideration" among these obligations, making it impossible to deal with conflicting obligations.

Third, in terms of value, deontic logic also fails to distinguish between the evaluation

---

[2] In the face of Ross's paradox challenge, von Wright defended deontic logic by saying:

"$Op$ entails $O(p \vee q)$. This is so because if one demands that it be the case that $p$ then one cannot consistently with this also allow that $\neg p$ be the case in conjunction with something else ($P(\neg p \& \neg q)$). If I order a letter to be mailed I cannot consistently with this allow it not be mailed but burnt. Therefore by demanding it to be mailed I have demanded it to be mailed or burnt!...Since that one norm entails another one cannot mean that if the first is true, then the second is necessarily also true, - what can it mean? We have already provided the answer: It means that the first norm is inconsistent with the negation-norm of the second. But, retorts our conservative logician, this notion of inconsistency does not mean what is normal in logic, viz. the necessary falsehood of a conjunction. The notion was explained in terms of a notion of rationality which relied on considerations about the purpose of norm-giving activity." (von Wright, 1991: 276)

However, in this example, what von Wright means is that it is irrational that government has enacted a law to prohibit people from killing a security guard while robbing a bank because it has already enacted a law to prohibit people from robbing banks, which leads to another difficult problem.

and existence of norms, which confuse right/wrong with true/false. Deontic logic thus cannot distinguish the difference between (F) and (G), leading to Ross's paradox.

Finally, deontic logic does not actually exhaust all normative statuses. In SDL, the four normative statuses have exhausted all logical possibilities by defining each other, so that logically $Op \vee Pp \vee Fp \vee Mp$ is a theorem; yet in reality, we find that there are many facts in an undefined status. Undefined statuses are neither obligatory nor omissible, and neither impermissible nor permissible. However, there is no way to symbolize such a state in SDL, because SDL is a deontically closed system, i.e., either impermissible or permissible, and either obligatory or omissible. A deontically closed system can only describe normative statuses within an ideal normative system (e.g., a legal system). However, it cannot capture normative statuses involving actions in reality, leaving a gap between the norms and actions.

## 3. Norms and Reasons

### 3.1 Norms

The most critical reason for the occurrence of such a paradox lies in what was mentioned in 2.2: deontic logic fails to distinguish between the evaluation and existence of norms—that is, when a norm exists, it does not mean that the norm is valid. Such a claim seems odd considering we have not made a good distinction between the nature of reasons and norms, or clarified how they relate to one another. When a norm exists, its very existence asserts that there is a reason for it, and when that assertion is true, the norm is valid. In other words, a normative proposition is actually an indicative sentence that aims to assert the existence of reasons. When I say: "You ought to...", I am actually asserting: "You have reasons to..." This assertion has a truth value because my assertion that you have reasons may be true or false. Therefore, we call the propositions expressed by norms "normative propositions", which are easily confused with a similar syntax, namely, imperatives. Let us consider the following example:

***Case 5*** (Grammatical Person of The Imperative)
***English***
(N) You open the door.
(O) Open the door.
(P) Let us open the door.
(Q) I open the door.*
(R) He opens the door.*

***Ancient Greek***

(S) λέγ-ε. (say-2SG.PRES)
(T) λέγ-ετε. (say-2PL.PRES)

***Taiwanese***
(U) Gún khui-mn̂g.* (We-1PL.EXCL open door)
(V) Lán khui-mn̂g. (We-1PL.INCL open door)

***Case 5*** (Grammatical Tense of The Imperative)
(W) Let's have gone.*
(X) Let's went.*

The subject of an imperative sentence must be in the second person, regardless of singular or plural. For example, in (N), "you" can be singular or plural. If we compare (Q) and (R) with (N), we will find that the sentences of (Q) and (R) will not be regarded as imperatives but as indicatives. Therefore, it is grammatically incorrect to use (Q) and (R) as imperative sentences, which are labeled with * to indicate grammatically incorrect sentences here. In addition, the subject of imperative sentences is mostly omitted, such as (O) is an example of omitting the subject. Where the subject is omitted in a sentence is assumed to be the second person. We would thus consider (O) to be an omission from (N). Moreover, the grammatical person will be manifested on verbs again when the language has a person and number marking, such as (S) and (T), where 2SG means second-person singular and 2PL means second-person plural.

Inclusive first-person plural pronouns that include the second person can also be used to express the imperative mood, such as in (P), where "us" actually includes the second person, but because there is no distinction between exclusive and inclusive first-person plurals in English, we can't see the grammatical difference, but if expressed in Taiwanese, such a grammatical difference can be clearly expressed in (U) and (V), where EXCL means exclusive and INCL means exclusive. The former is not regarded as an imperative but as indicative, which is also labeled with * to indicate the grammatically incorrect sentences here. On the contrary, the latter can be used as an imperative since the subject includes the second person.

In addition to grammatical person, there are important differences in grammatical tense. In imperative sentences, since the event has not been completed, the verb must be in the present tense, so in English, it would be ungrammatical to translate (P) into the sentences in Case 5.

None of the above restrictions exist for normative propositions, which are freer in terms of person and tense, such as:

***Case 6*** (Normative Proposition)
(Y) He ought to obey the law.
(Z) Smoking marijuana was once forbidden here.

These examples all show that a normative proposition is more like an indicative sentence, which aims to assert the existence of reasons. “You ought to...” means “you have reasons...”. For example, (Y) asserts that he ought to obey the law because the law itself provides a reason. In contrast, (Z) asserts that smoking marijuana was once forbidden here because the law prohibiting smoking marijuana has been abolished. Normative propositions thus have truth values because the assertion may be true or false. On the contrary, an imperative sentence is more like an action, which cannot be evaluated as true or false but can be evaluated pragmatically as happy or unhappy[3]. For example, in (O), if the door is open already, it will become unhappy to say this sentence. In addition, the action object of the imperative sentence is very clear, that is, the dialogue object of the speech, which is why the imperative sentence is limited to the second-person present tense, while the normative proposition can be used more freely.

Normative propositions are so different from imperative sentences that they cannot serve as reasons for action. We can say, “I quit smoking” because, “my girlfriend asked me to quit smoking”, but if we say, “I quit smoking” because “I should quit smoking”, this does not seem to explain anything. Sometimes we indeed use “…should…” to explain our actions. However, it does not mean that such propositions are really used as reasons but imply the existence of a reason because speakers and listeners share background knowledge. In the case of unclear context, such an explanation will also not hold true. If someone asks:

(a) “Why are you here?”

Then the answer is:
(b) “Because I should be here.”

We must think that such an answer does not explain anything, and we will certainly ask for further: “Why?” Perhaps a clearer example as follow can identify the differences in language intuition:

***Case 7***
(c) “Marry me!”
(d) “You should marry me.”

3 For the meaning of happy and unhappy in linguistic pragmatics, see Austin (1962 :14).

In sentence (c), the answer “Yes!” or “No!” would be slightly more unhappy than in sentence (d), which is more to say “Fulfill my will!” but (d) is really to say “There exists a reason for you to marry me.” Therefore, a more appropriate answer to sentence (d) should be “Why?”, even if there are some occasions where you answer “Yes!” or “No!”; but it is actually because they are both used in the same context and because the answerer knows that this is purely a request, that is, the normative proposition is used as a request here.

It is easy to mistakenly think that normative propositions can be directly used as reasons. Such misunderstandings all come from confusion about the existence and effectiveness of norms. Hans Kelsen (1941: 50), for example, clearly stated that “if we say that a norm ‘exists’ we mean that a norm is valid. Norms are valid for those whose conduct they regulate”. In the discussion of reason theory, however, such a statement has been sublated, as John McDowell (1978: 14) pointed out that “the reason is not expressed by the ‘should’ statement itself. The reason must involve some appropriate specific consideration which could in principle be cited in support of the ‘should’ statement”. A valid norm means there is a true normative proposition, or a reason for action asserted by the normative proposition exists.

Confusion about the validity and existence of norms can be extremely problematic in deontic logic. Such confusion seems to imply that all normative propositions derived from logic will be sound, i.e., valid. However, the question is, does the derivation of reasons follow the path of the sentence logic? Suppose the derivation of the reason is different from the logic of the sentence. In that case, deontic logic will not be sound in possible world semantics, and other formalizations must be found. Nevertheless, we must first ponder whether the reason can be formalized.

### 3.2 Reasons

To show how the formalization of reasons is possible, it is necessary first to answer:

(1) What the reason is.
(2) How a reason supports a norm.

Regarding (2), T. M. Scanlon (1998: 3) asserts that “[p]ractical reasons for actions are facts that count in favor of these actions.” However, it must be asked, how do these facts count in favor of actions? Scanlon (1998: 17) argues that there is no further explanation. In other words, “‘[b]y providing a reason for it’ seems to be the only answer.” As he said:

…I am quite willing to accept that “being a reason for” is an unanalyzable, normative, hence non-natural relation. …(Scanlon, 1998: 11)

Such a statement is called Reason Primitivism. According to this statement, we cannot find a basis to formalize our reason inference. On the contrary, John Broom takes a different view, arguing that reason is the explanation for action, but what is the relationship between explanation and action? Broom provides an ambiguous explanation:

> A perfect reason, therefore, need not be a unique canonical reason. Suppose you ought not to drink home-made grappa because it damages your health. The fact that home-made grappa damages your health explains why you ought not to drink it, so it is a perfect reason for you not to drink it. Another explanation of why you ought not to drink home-made grappa is that it contains methyl alcohol. This is not a rival explanation; it is consistent with the first. So a perfect reason for you not to drink homemade grappa is that it contains methyl alcohol. Now we have two distinct perfect reasons for you not to drink home-made grappa. This is confusing, but only because the individuation of explanations is confusing. We need not fuss about it. (Broom, 2004: 35)

Broom does not actually offer a clear definition of the explanation, and he argues that there can be more than one explanation for the action. In the above quotation, Broom points out that the normative proposition "you ought not to drink home-made grappa" has two explanations, "home-made grappa damages your health" and "home-made grappa contains methyl alcohol", and these two explanations are mutually consistent. Linton Wang, however, criticizes that these two facts are related to the normative proposition by completely different paths, and only one of the facts has reached the real explanation. If the meaning of "explanation" is not clarified, the properties of these two facts will be confused.

According to Linton Wang, the explanation can be achieved because it must point out the relevant difference-making. Generally speaking, when we explain a fact, explanans $P$ explains explicandum $Q$ if and only if $P$ "causes" $Q$. The so-called cause, to define the concept in the broadest and most commonly used way, is the definition of the *counterfactual*. Linton Wang uses "canonical difference-making" to capture the simplest and most central counterfactual concept:

***Definition 1*** (Canonical Difference Making)
A fact $P$ canonically makes a difference $Q$; if and only if,
(1) If $P$, then $Q$
(2) If not $P$, then not $Q$

Compared with the explanation of facts, which is to find the cause of events, a normative explanation requires an *explanatory inversion*, that is, to find out what outcomes the regulated facts will lead to (Linton Wang, 2015: 112). In Broom's case, for instance, when we say, "you ought not to drink home-made grappa", one of the reasons is that "home-made grappa damages your health". This reason points to the outcome of regulated actions, and moreover, it also gives us counterfactual assumptions: If you did not drink home-made grappa, you would not damage your health. This explanation points to the relevant difference-making and thus serves as a reason for "you ought not to drink home-made grappa". In addition, Broom also pointed out another perfect reason is that "home-made grappa contains methyl alcohol", but this fact does not point to any counterfactual outcome, so it does not really explain the regulated facts. Linton Wang claims that this is not a "reason fact", but a "reason-giving fact". Reason-giving fact does not explain actions but explains how differences in outcomes arise. Linton Wang (2015: 111) then defines "reasons" and "reason-giving fact" as follows, which he calls *the difference-making-based theory of reasons*:

***Definition 2*** (The Difference-Making-Based Theory of Reasons)
**Reason-Fact** $R$ is a reason for $Q$ ($R$ is a reason-fact); if and only if,
(1) $R$ is a difference-making fact, which shows that $Q$ can make a difference to some outcomes $X$.
(2) The difference that $Q$ makes has a feature $F$.

**Reason-Giving Fact** $P$ gives one a reason to $Q$ ($P$ is a reason-giving fact); if and only if, $P$ can make a difference to $R$, which is a reason to $Q$.

Regarding (1), a reason for action is just a causal path from action to its outcome. So, in Broom's case, the real reason is "home-made grappa damages your health", while "home-made grappa contains methyl alcohol" only explains why "home-made grappa damages your health". That is to say, it explains the difference, not the action. However, we always intuit that the reason-giving fact can explain the normative proposition. It is because the explanation of the difference in outcomes implies the causal path that the regulated fact will lead to such difference, and therefore, the reason-giving fact implies the existence of reasons. In some places where the context is unclear, the reason-giving fact cannot explain the normative proposition and action, as is the case in the following examples:

(e) Alice: "Why are you here?"
(f) Alison: "Because you are here."

(f) is a reason-giving fact, as same as (b), which seems not to explain anything.

Although it suggests that there is a difference in outcomes, and the difference is related to "Alice is here."; yet we do not understand the background knowledge of Alison's answer so much that we do not know what difference would have happened had Alison not gone to the place where Alice is. In cases where the context is unclear, the difference in explanation between "the reason" and "the reason-giving fact" can be clearly identified. In addition, whether a fact of a causal path can be used as a reason depends on the setting of the background conditions of the causal path, that is, on the truth value of reason-giving facts. Linton Wang (2015: 122) calls this dependence *circumstance sensitivity of reason-giving fact*.

Back to Case 7, (c) as a requirement is not actually a reason but a reason-giving fact. I would agree to or refuse another person's marriage proposal because I love him/her or other considerations, but (c) as a requirement triggers the establishment of the real reason. On the other hand, (d) as an assertion informs that the reason exists. Therefore, (c) is a kind of *triggering reason-giving*, and (d) is a kind of *epistemic reason-giving*[4].

Although Linton Wang's argument distinguishes the difference between cause explanation and reason explanation, he ignores the attitude towards the difference in action outcomes. If the reason is only regarded as the objective causal path from actions to outcomes, according to Niû Têng-úi (2022: 8-9), the more important aspect of the reason will be missed, i.e., intention. We can imagine the following example:

(g) If one has sex without contraceptives, one will get pregnant.

(g) pointed out the causal path from sex without contraception to pregnancy. However, this causal path does not support whether "we should use contraceptives" or "we should not use contraceptives". This causal path alone cannot support any normative conclusion. If one partner does not want to conceive, this causal path will become a reason for "they should not use contraceptives"; on the other hand, if both parties want to conceive, this causal path will become a reason for "they should use contraceptives". That is to say, the attitudes we display toward differences in outcomes are also very important factors in the explanation for action, and, precisely because of attitudes, an action out of reason is intentional. Therefore, it is necessary to redefine the reason as follows:

***Definition 3***

**Reason-Facts** $R$ is a reason for $Q$ ($R$ is a reason-fact); if and only if,

(1) $R$ is a difference-making fact, which shows that $Q$ can make a difference to

[4] For the meaning of triggering reason-giving and epistemic reason-giving, see Enoch (2011).

some outcomes $X$.

(2) The difference that $Q$ makes has a feature $F$.

(3) Agent $A$ has a desired/undesired attitude towards feature $F$.

However, will the addition of attitude shake the neutrality of the difference-making-based theory of reasons? Linton Wang incorporates his "difference-making-based theory of reasons" into three mainstream viewpoints of reason theory: a desire-based theory of reason, a value-based theory of reason, and an agency-based theory of reason[5], which seems that he does not intend to propose a new theory of reason that can be paralleled with the other three, but makes a general analysis of reason. Any substantive definition of attitudes, however, will shake the neutrality of "the difference-making-based theory of reasons". So, this paper refrains from the substantive controversy in reason theory and regards attitude as "a way of classification" of difference: attitude is a set of differences rather than a specific entity. I categorized the differences as follows:

***Definition 4***

**Desired** The difference caused by $Q$ having a feature $F$ is *desired* $F = e$; if and only if, when $Q$ causes the difference $F = e$, we consider $Q$ to be obligatory.

**Undesired** The difference having a feature F is *undesired* $F = s$; if and only if, when $Q$ causes the difference $F = s$, we consider $Q$ to be impermissible.

Accordingly, all of the normative statuses can be derived from Definition 4:

***Theorem 1***

**Obligatory** We consider $Q$ to be *obligatory*; if and only if,

(h) The difference caused by $Q$ having a feature F is desired $F = e$ or,

(i) The difference caused by $\neg Q$ having a feature F is undesired $F = s$.

**Permissible** We consider $Q$ to be *permissible*; if and only if,

(j) The difference caused by $\neg Q$ having no feature $F$ is desired, i.e., $F = e'$ or,

(k) The difference caused by $Q$ having no feature $F$ is undesired, i.e., $F = s'$.

**Impermissible** We consider $Q$ to be *impermissible*; if and only if,

(l) The difference caused by $\neg Q$ having a feature $F$ is desired $F = e$ or,

(m) The difference caused by $Q$ having a feature $F$ is undesired $F = s$.

**Omissible** We consider $Q$ to be *omissible*; if and only if,

(n) The difference caused by $Q$ having no feature $F$ is desired, i.e., $F = e'$ or,

[5] For an introduction to these three viewpoints of reason theory, see Ruey-Yuan Wu (2015: 27-34).

(o) The difference caused by $\neg Q$ having no feature $F$ is undesired, i.e., $F = s'$.

In Theorem 1, each normative status has a pair of definitions. One derives from the desired value, and the other derives from the undesired value. Linton Wang, on the other hand, slightly modified the definition of reason in his recent paper:

***Definition 5***
**Reason-Facts** $R$ is a reason for one to φ ($R$ is a *reason-fact*) just in case that $R$ is a value-revealing difference-making fact, which shows that φ-ing can make a difference to some valuable outcome. (Linton Wang, 2022: 9)

However, in Definition 5, the term "valuable" may fall into the substantive controversy of reason theory, which may be confused with "the value" of "the value-based theory of reason", then lose its neutrality. Furthermore, is the term "valuable" sufficient to capture all aspects of the attitude? That is, could the reason revealing the difference only to values can exhaust all the normative statuses?

Definition 5 reveals the two deontic operators, $O$ and $P$, are interdefinable, which takes $Op$ as primitive and defines the rest of the normative statuses, then make normative statuses defined in Theorem 1 only leave (h), (j), (l) and (n) defined by the desired attitude. However, in debates within deontic logic, there are two main stands on the disputed questions. One perspective suggests that permission is simply the lack of prohibition. Thus, it can be precisely called impermissible. According to the opposed view, permission is something "over and above", absence of prohibition, that is, when we permit one to do something, we have some *reasons*. This point of view takes $O$ and $P$ are all primitives and uninterdefinable. In the former view, such permission is considered to be *weak*, and in the latter view, such permission is considered to be *strong*. von Wright (1980: 414) gives a new axiom to capture the concept of strong permission

(p) $Pq \rightarrow \neg O \neg q$

while the opposite $\neg O \neg q \rightarrow Pq$ is not valid in this new system.

This paper, though, does not want to fall into the debates within deontic logic, yet obviously, Definition 3 can do more than Definition 5. Theorem 1 divides each normative state into two categories: one defined by desired and the other by undesired. Unlike mainstream debates, Theorem 1 takes (h) and (m) as primitives, denoted as $Oq$ and $Fq$. When $Fq$ is considered to be stronger than $O\neg q$, then

(q) $Fq \rightarrow O\neg q$
(r) $F\neg q \rightarrow Oq$

is valid, and

(s) $O\neg q \rightarrow Fq$
(t) $Oq \rightarrow F\neg q$

is not valid. Therefore, $\neg O\neg q \rightarrow \neg Fq$, which is equivalent to (p), can be derived from (q). It shows that Theorem 1 contains both *strong* and *weak* permission, but in Definition 5, only the permission defined by the desired attitude, that is, the strong version of the permission, can be derived.

Definition 3, which defines the concept of reason, and Theorem 1, which defines normative statuses with reasons, provide an opportunity to reformulate reason-based deontic logic. This reformulation has been made possible through the use of causal models, a new discipline developed in the 21st century from the fields of statistics and computer science. Causal models incorporate both structural equations and causal graphs, which results in a more context-sensitive and nuanced approach to reason-based formalization. However, prior to the introduction of reason-based formalization, various context-sensitive logical systems were developed within deontic logic to address paradoxical problems. These historical developments are worthy of attention, and I refer to them as "conditional-based formalization."

## 4. Reason-Based Formalization

### 4.1 From Conditionals to Counterfactuals

SDL constructing the general obligation leads to whatever ought to be the case and causes paradoxes like contrary-to-duty imperatives and Chisholm's Puzzle. Moreover, the paradoxes based on conditionals make the problems more perplexing. To deal with them, von Wright (1956) proposed that the concept of an obligation to do something given certain conditions is not definable in terms of a concept of obligation simpliciter. He set a conditional obligation $O(A/B)$ as primitive, expressing "A is obligatory under condition B", then used it to define "conditional permission" $P(A/B)$ and "conditional prohibition" $F(A/B)$, to compose a conditional-based deontic logic. Von Wright's approach is a considerable advance since it shows that deontic logic can also be reasoned in context. However, von Wright only illustrates the "correlation" between regulated actions and conditional facts in terms of "conditional relations", There is no way to analyze further how conditional facts are related to regulated actions, we thus cannot make a more precise analysis of the reasons for the normative proposition.

A. R. Anderson (1957), in addition, reduced deontic logic to alethic modal logic. He argued that it is important to relate both norms and sanctions to possibilities for action.We thus need a method of relating norms to the system of social sanctions or penalties that support them. (Anderson, 1957: 14) Anderson introduced a propositional constant s as “sanction” or “something undesired” into his new deontic logic system, and defined each normative status as follows:

$$Op \overset{\text{def}}{=} \Box(\neg p \rightarrow s)$$
$$Pp \overset{\text{def}}{=} \Diamond(p \land \neg s)$$
$$Fp \overset{\text{def}}{=} \Box(p \rightarrow s)$$
$$Mp \overset{\text{def}}{=} \Diamond(\neg p \land \neg s)$$

Anderson's approach that reduces deontic logic to conditionals still cannot avoid the paradoxes caused by conditionals themselves, neither can contextually symbolize situations that deal with conditions, hierarchies, and alternatives among the various norms. Therefore, in his later writings (Anderson, 1967), he applied his work on relevance logic to his alethic reduction of deontic logic. He used a relevant form of implication, $\Rightarrow$ to express the reduction as: $Op \overset{\text{def}}{=} \neg p \Rightarrow s$, which means “if not p then a sanction is imposed”. Anderson argues that the consideration of norms cannot be separated from the support of “conditional sentences”, that is, “if…then…”, as he stated:

> As I have mentioned elsewhere, when someone is told that he ought to do something, and asks why, the reply very often has the form: “Well, if you don't, then...,” where the lacuna is filled in with a description of some state-of-affairs which the questioner is expected to recognize as bad. (Anderson, 1967: 346)

For Anderson, conditionals are common justifications for normative propositions, and he thus turned the study of deontic logic to relational logic. This idea is almost equivalent to the difference-making-based theory of reasons. In Anderson’s quotation, Anderson pointed out that the explanation of “why we should do $p$” is “if we don’t do $p$, then…”, which is clearly a counterfactual assumption. Therefore, this explanation can be changed into the more precise “if we should do $p$, we will…”. In Anderson’s time, however, the discussion of reason theory has not yet started, so this kind of thinking lacks proper development. In the difference-making-based theory of reasons, conditionals are replaced with counterfactuals. If we compare the deontic operator defined by Anderson with Theorem 1, we will find that there are similarities between them, in which $Pp$ and $Mp$ may be different because of the definition of conditional. And Anderson’s permission is a weak version; that is, permission means no prohibition, while Linton Wang’s permission derived by Definition 5 is strong permission. Both versions of the permission are included in Theorem 1.

Counterfactuals are a very important concept for norms. As I have mentioned elsewhere (Niû Têng-úi, 2022: 8), the ability to illuminate the counterfactual is a *necessary quality for an explanation* to have. Consider the following case:

***Case 8***
(u) We should run the red light.
(v) Because everyone runs the red light.

In Case 8, obviously, the fact that (v) "Because everyone runs the red light" cannot be the reason for (u) "We should run the red light"; otherwise, it would violate Hume's guillotine. However, when we use reasons to justify normative propositions, we always use factual premises to support normative propositions, but why can't (v) support (u)? This is because (v) itself is the status quo stated by (u), and it does not point out what difference will result if we do not run a red light. The valid reason for each one trying to maintain the status quo should be: "If the status quo were not maintained, what would happen?" And such a claim is also a counterfactual claim.

Therefore, we must move justification for norms from conditionals to counterfactuals. Causal models, in this sense, were developed by Judea Pearl (2010). He developed the Bayesian network and divided our thinking structure on causality into three levels from low to high: prediction, intervention, and counterfactuals. Prediction is to observe the data correlation between the preceding event and the succeeding event and calculate the conditional probability of "if preceding event $A = a$, then succeeding event $B = b$ will also follow", formalized as $P(B = b|A = a)$. Intervention is to actually control the value of preceding event $A = a$, then estimate the conditional probability of succeeding event $B$'s value $B = b$, denoted as $P(B = b|do(A = a))$. Pearl (2010: 70) introduces a new operator $do(A = a)$ to express "to set variable $A$'s value to $A = a$. Counterfactuals are to trace the time back to the point when the preceding event $A = a$, and modify the preceding event's value to $A = a'$, then follow the process under normal circumstances to estimate the conditional probability of changing the value of succeeding event $B = b'$, denoted as $P(B_{A=a'} = b'|A = a, B = b)$. High-level calculations must be based on low-level calculations and contain low-level concepts, so counterfactuals are the most important core in causal models.

One of the biggest highlights of the causal model is the combination of the causal graph, which represents our causal interpretations of event associations. Such interpretations come from our pre-understanding, background knowledge, and theories on special expertise. They are depicted as a graph consisting of a set $V$ of vertices and a set $E$ of edges that connect some pairs of vertices. Each edge in a graph can be either directed, undirected, or, in some applications, bidirected to denote confounders. But for the convenience of discussion, the causal graphs used here will be all directed. In the field of norms, causal graphs can be considered as our intuitions about norms and decision-making, where $V$ represents our selects of discrete events from the continuous world, and $E$ represents counterfactual relations between discrete events. In the case of relation to descendants of action, according to Definition 3, such counterfactual relations can be considered as reasons. In this way, we can construct a way of context-sensitive formalization.

### 4.2 Expectations for Normative Statuses

A causal model **M** is an ordered pair <**V, E**>. **V** is a set of *variables* and **E** is a set of *structural equations*, which represent dependence relations among variables. Such dependence relations are expressed by functions $Y = f_Y(PA_Y)$[6], where $PA_Y$ stands for the parents of $Y$. For short, we will use capital letters $X, Y, Z \ldots$ for variables and lowercase letters $x, y, z \ldots$ for specific values taken by corresponding variables.

An action with reasons, in the sense of Pearl (2010: 108), is contemplated decision making after deliberation involving comparison of outcomes. Therefore, it looks forward and cannot be predicted by data. According to Niû Têng-úi (2022: 8), an action can be interpreted in two aspects: one is *causal interpretation*, which is to find causes of action from the physiological function, such as organic, neural, or psychological diseases, etc. The other is *rationalized interpretation*, which is to find reasons for action from the agent's purpose: what difference does action make? And what are the attitudes of the agent towards this difference? Pearl (2010: 108) refers to what is interpreted with causes (causal interpretation) as an act, and what is interpreted with reasons (rationalized interpretation) as an action. In this sense, we can apply the do-operator $do(x)$ to the action, which amounts to removing equation $x = f(pa_x)$ from the model and set $X = x$ in the remaining equation, then calculate the expectation of the action with a formula:

(w) $U(x) = \sum_y P(y|do(x))u(y)$ [7]

Where $u(x)$ is the utility of outcome $x$.

However, what (s) calculates is the expected value of the action, or in other words, the desired value. In Theorem 1, normative statuses are defined both by desired and undesired attitudes, and we must modify (s) to formulate Obligatory and Impermissible in Theorem 1 as:

---

[6] Due to the convenience of discussion, Pearl's function is simplified here. In Pearl's works, He uses $Y = f_Y(PA_Y, U)$, where $U$ is background variables.

[7] Pearl (2010: 108)

$$\text{Obligatory} = \begin{cases} O(q) = \sum_{f_i} P\big(f_i \big| do(q)\big) e(f_i) \\ F(\neg q) = \sum_{f_i} P\big(f_i \big| do(\neg q)\big) s(f_i) \end{cases}$$

$$\text{Impermissible} = \begin{cases} O(\neg q) = \sum_{f_i} P\big(f_i \big| do(\neg q)\big) e(f_i) \\ F(q) = \sum_{f_i} P\big(f \big| do(q)\big)\, s(f_i) \end{cases}$$

Where $e(f_i)$ is the desired value and $s(f_i)$ is the undesired value of feature $f_i$ which the difference in outcomes has. Obligatory and Impermissible can derive Permissible and Omissible as follow:

$$\text{Permissible} = \begin{cases} \neg O(\neg q) = \sum_{f_i'} P\big(f_i' \big| do(\neg q)\big) s(f_i') \\ \neg F(q) = \sum_{f_i'} P\big(f_i' \big| do(q)\big) e(f_i') \end{cases}$$

$$\text{Omissible} = \begin{cases} \neg O(q) = \sum_{f_i'} P\big(f_i' \big| do(q)\big) s(f_i') \\ \neg F(\neg q) = \sum_{f_i'} P\big(f_i' \big| do(\neg q)\big) e(f_i') \end{cases}$$

Where $s(f_i')$ is the undesired value and $e(f_i')$ is the desired value of feature $f_i'$ which is the state apart from the difference in outcomes. Formulae $O(q)$ to $\neg F(\neg q)$ correspond in an orderly fashion (h)-(o) in Theorem 1. $\neg O(\neg q)$ to $\neg F(\neg q)$ shows that both permission and omission are something "over and above" absence of prohibition and obligation. $\neg O(\neg q)$ points out, for instance, that something undesired will happen if we do not permit $q$; that is to say, $\neg q$ forms a counterfactual relationship with $f_i'$ then constitutes a reason for permission. The same explanation can be applied to other formulae, in other words, both versions of permission or omission are stronger than the absence of prohibition or obligation. We have some reasons to support them. Moreover, $\neg O(\neg q)$ to $\neg F(\neg q)$ is equivalent to the following formula:

$$\text{Permissible} = \begin{cases} \neg O(\neg q) = \sum_{f_i} \Big(1 - P\big(f_i \big| do(\neg q)\big)\Big)\, s(f_i') \\ \neg F(q) = \sum_{f_i} \Big(1 - P\big(f_i \big| do(q)\big)\Big)\, e(f_i') \end{cases}$$

$$\text{Omissible} = \begin{cases} \neg O(q) = \sum_{f_i} (1 - P(f_i | do(q))) s(f_i') \\ \neg F(\neg q) = \sum_{f_i} (1 - P(f_i | do(\neg q))) e(f_i') \end{cases}$$

This makes it unnecessary for us to search from data for counterfactual relationships from actions to $f_i'$.

Among the four normative statuses, if $O(q) = F(\neg q)$, $O(\neg q) = F(q)$, $\neg O(\neg q) = \neg F(q)$ and $\neg O(q) = \neg F(\neg q)$ is considered to be valid, that is, the above values only have quantitative differences, then normative logic will have only one primitive. In contrast, if the above equation is considered to be invalid, then there will be two kinds of primitives in deontic logic, and these two kinds of primitives and their derivation will be qualitatively different, but how (q) and (r) can be made mathematically possible remains to be further studied.

To explain the above formula, consider the following case:

***Case 9*** (Quitting Smoking)

John is deciding whether to quit smoking ($\neg s$). In his consideration, he listed the following reasons (Assume that he has exhausted all reasons):

| **Disadvantages of Smoking** | **Advantages of Smoking** |
|---|---|
| (Φ) Smoking causes lung cancer.<br>(Ψ) Smoking causes poverty. | (φ) Smoking satisfies his addiction to tobacco.<br>(ψ) Smoking allows him to fit into the friend circle. |

He gives undesired values to differences in outcomes caused by these reasons:

s(Φ)=10　s(Ψ)=1

which means that he will get something undesired if he continues smoking, while

s(¬φ)=5　s(¬ψ)=7

mean that he will get something undesired if he quits smoking. Also, if he is sure that smoking does not cause lung cancer *or* poverty, quitting smoking will make him:

s(Φ’) = 5+7-1 = 11
s(Ψ’) = 5+7-10 = 2

If he is sure that not smoking does not make him satisfy his addiction to tobacco *or* fit into the friend circle, continuing smoking will make him:

$$s(\neg\varphi') = 10+1-5 = 6$$
$$s(\neg\psi') = 10+1-7 = 4$$

Suppose that we now have the data:

$$P(f|do(\Phi)) = 0.85 \qquad P(f|do(\neg\varphi)) = 1$$

$$P(f|do(\Psi)) = 0.2 \qquad P(f|do(\neg\psi)) = 0.8$$

Then, we can calculate:

$$F(s) = 0.85 \times 10 + 0.2 \times 1 = 8.7 \qquad F(\neg s) = 1 \times 5 + 0.8 \times 7 = 10.6$$

$$\neg F(s) = 0.15 \times 11 + 0.8 \times 2 = 3.25 \qquad \neg F(\neg s) = 0 \times 6 + 0.2 \times 4 = 0.8$$

To make a comparison of expectation, values with the same object but contradictory statuses must be subtracted from each other:

$$F(s) = 8.7 - 3.25 = 5.45$$

$$F(\neg s) = 10.6 - 0.8 = 9.8$$

Therefore, the conclusion after John weighs the reasons is: *he should continue smoking.*

### 4.3 Causal Graphs of Reasons

A causal graph consists of a set V of vertices and a set E of edges that connect some pairs of vertices. We can think of the vertices in causal graphs as variables in causal models, and the edges in causal graphs as the structural equations in causal models. Such compatibility has been proven feasible by Pearl (2010:16).

Case 9 can be depicted as follows:

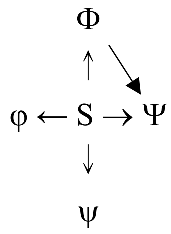


**Figure 1**

Figure 1 means that smoking S leads to four outcomes, which are the four reasons listed in Case 9. Causal path $\Phi \rightarrow \Psi$ means that the medical expenses incurred by lung cancer will lead to poverty, which has been included in the formula $P\big(f \big| do(\Psi)\big)$.

The causal graph cannot be obtained from the probability function because the same probability function may depict multiple equivalent causal graphs. Causal graphs that are equivalent cannot be distinguished without resorting to manipulative experimentation or temporal information; that is, we must have pre-understandings of things to be able to draw causal graphs. Thus, in deontic logic, we can think of these pre-understandings as our moral intuitions, biases, and background knowledge about how things relate. The task of the causal graph is to show the moral intuitions of agents through formalization clearly and analyze what kind of reasons work under such moral intuition, then figure out how can we revise our moral intuitions if new evidence emerges.

Any moral evaluation has to start with a causal analysis of the situation. We usually think of *sufficient probability* as the basis for the attribution of responsibility. For example, we would not hold a murderer's mother responsible for the murder of her child, even though "if the mother had not given birth to the murderer, the victim would not have died." Due to the lack of sufficient conditions, the mother's childbirth is not considered the "cause" of the victim's death. Therefore, when we want to clarify the responsibility, the expectation formulae of the normative statuses can be modified as follows:

$$\text{Obligatory} = \begin{cases} O_S(q) = \sum_{f_i} P\big(f_q^i \big| q', f_i{}'\big) e(f_i) \\ F_S(\neg q) = \sum_{f_i} P\big(f_{\neg q}^i \big| \neg q', f_i{}'\big) s(f_i) \end{cases}$$

$$\text{Impermissible} = \begin{cases} O_S(\neg q) = \sum_{f_i} P\left(f^i_{\neg q} \middle| \neg q', f_i'\right) e(f_i) \\ F_S(q) = \sum_{f_i} P\left(f^i_q \middle| q', f_i'\right) s(f_i) \end{cases}$$

$$\text{Permissible} = \begin{cases} \neg O_S(\neg q) = \sum_{f_i'} P\left(f^i_{\neg q}{}' \middle| \neg q', f_i'\right) s(f_i') \\ \neg F_S(q) = \sum_{f_i'} P\left(f^{i}_q{}' \middle| q', f_i'\right) e(f_i') \end{cases}$$

$$\text{Omissible} = \begin{cases} \neg O_S(q) = \sum_{f_i'} P\left(f^{i}_q{}' \middle| q', f_i'\right) s(f_i') \\ \neg F_S(\neg q) = \sum_{f_i'} P\left(f^i_{\neg q}{}' \middle| \neg q', f_i'\right) e(f_i') \end{cases}$$

where $f^i_q$ expresses $P\left(f_i \middle| do(q)\right)$, and $f^{i}_q{}'$ expresses $P\left(f_i' \middle| do(q)\right)$.

To explain how causal graphs reveal our moral intuitions, let's take the trolley problem (Foot, 1967), the most-studied moral dilemma, as an example:

***Case 10*** (Trolley Problem)

(x) You observe a trolley out of control is headed toward five people standing on a railroad track. If you don't do anything, the trolley will kill the five people on the track. However, if you throw a switch, the trolley will change to the course onto a side track. As it turns out, one person is standing on the sidetrack. If you throw the switch, the five people on the main track will survive, but the one person on the side track will die. If you don't throw the switch, the five people on the main track will die, but the person on the side track will survive. Is it morally obligatory or permissible to throw the switch?

(y) This time, you are on a bridge crossing the railroad track, and the only option you have for stopping the trolley is to push a large man off the bridge onto the track. This will stop the train but kill the large man. Is pushing the large man off the bridge morally obligatory or permissible?

To depict Trolley Problem, we use a set of variables **V**1= $\{\mathrm{A}, \mathrm{B}, \mathrm{C} \dots\}$ to represent Case 10, which has the following interpretation:

$\mathrm{A} = a$ if the trolley is out of control, $\neg a$ if otherwise.
$\mathrm{B} = b$ if you throw a switch, $\neg b$ if otherwise.
$\mathrm{C} = c$ if the five people on the main track are killed, $\neg c$ if the five people on the

main track survive.

$\mathrm{D} = d$ if the one person on the side track is killed, $\neg d$ if the one person on the side track survives.

$\mathrm{E} = e$ if you push the large man off the bridge onto the track, $\neg e$ if otherwise.

$\mathrm{F} = f$ if the large man is killed, $\neg f$ if otherwise.

Then (x) and (y) can be depicted as the following causal graphs:

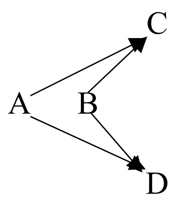


**Figure 2**

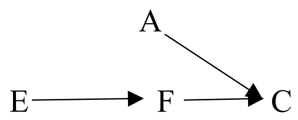


**Figure 3**

Figure 2 and Figure 3 respectively represent (x) and (y). The usual understanding of Figure 2 and Figure 3 yields the following functional relationships:

In Figure 2:

$$c = a \wedge \neg b$$

$$d = a \wedge b$$

Figure 3:

$$c = \neg f \wedge a$$

$$f = e$$

Testing Figure 2:

$$D_b(a) = d \qquad C_{\neg b}(a) = c$$

$$D_b(\neg a) = \neg d \qquad C_{\neg b}(\neg a) = \neg c$$

$$C_a(\neg b) = c \qquad C_{\neg a}(\neg b) = \neg c$$

$$D_a(b) = d \qquad D_{\neg a}(b) = \neg d$$

This shows that B's options lack sufficiency for people's death in the case of $\neg a$, while A is always sufficient for people's death no matter what the value of B.

Testing Figure 3:

$$F_e(a) = f \qquad C_{\neg e}(a) = c$$

$$F_e(\neg a) = f \qquad C_{\neg e}(\neg a) = \neg c$$

$$C_a(\neg e) = c \qquad C_{\neg a}(\neg e) = \neg c$$

$$F_a(e) = f \qquad F_{\neg a}(e) = f$$

This shows that both A and E are sufficient for people's death. Judging from the above two sets of tests, it is enough to prove that pushing the large man off the bridge is worse than throwing the switch, because the former is more sufficient for death than the latter. Assume that the probability of the trolley going out of control is $P(\text{A}) = 0.01$, then we have obtained reasonable probability data as follows:

| Throwing a switch | No Throwing a switch |
| --- | --- |
| $P(c_b \mid \neg \text{c}, \neg b) = 0$ | $P(d_{\neg b} \mid \neg d, b) = 0$ |

**Form 1**

| Pushing the large man off the bridge | No pushing the large man off the bridge |
| --- | --- |
| $P(f_e \mid \neg f, \neg e) = 1$ | $P(c_{\neg e} \mid \neg c, e) = 0.01$ |

**Form 2**

Form 1 and Form 2 respectively represent (x) and (y) in Case 10. Next, we give undesired values to these differences in outcomes:

$$s(c) = 50 \qquad s(d) = 10 \qquad s(f) = 10$$

This means that we will get something undesired if we choose which action to take, and if we are sure that the actions we avoid do not cause something undesired, then avoiding the actions will cause:

$$s(c') = 10 \qquad s(d') = 50 \qquad s(f') = 50$$

Therefore, we can calculate the expectation of (x):

| Throwing the switch | No throwing the switch |
|---|---|
| $F_S(b) = 0 \times 10 = 0$ | $F_S(\neg b) = 0 \times 50 = 0$ |
| $\neg F_S(b) = 1 \times 50 = 50$ | $\neg F_S(\neg b) = 1 \times 10 = 10$ |

$$F_S(b) = 0 - 50 = -50$$

$$F_S(\neg b) = 0 - 10 = -10$$

In (x), the conclusion is: *To throw the switch is more permissible than not.*

Expectation of (y):

| Pushing the large man off the bridge | No pushing the large man off the bridge |
|---|---|
| $F_S(e) = 1 \times 10 = 10$ | $F_S(\neg e) = 0.01 \times 50 = 0.5$ |
| $\neg F_S(e) = 0 \times 50 = 0$ | $\neg F_S(\neg e) = 0.99 \times 10 = 9.9$ |

$$F_S(e) = 10 - 0 = 10$$

$$F_S(\neg e) = 0.5 - 9.9 = -9.4$$

In (x), the conclusion is: *it is impermissible to push the large man off the bridge and is permissible not to do.*

Some scholars advocate the use of counterfactuals to capture the difference between (x) and (y)[8]. Inverting the causal process from decision to action makes us able to infer the agents' mental states from their actions, and then make judgments about responsibility. In (y), if the large man hadn't been pushed off the bridge, then the five people would not have been saved. The survival of the five depends counterfactually on the death of the one. Thus, there is a causal path from F to C in Figure 3, which

[8] For a discussion of counterfactuals in the trolley problem, see D. A. Lagnado and T. Gerstenberg (2017).

means that pushing the large man off the bridge was a means for saving the five. In contrast, in (x), the five on the main track would have been saved even if there had been no person on the sidetrack. The survival of the five was not counterfactually dependent on the death of the one. Thus, there is no causal path from $D$ to $C$, which means that the death of the person on the sidetrack was a side effect, rather than a means for saving the five.

Counterfactuals can help us to capture the difference between intended effects and those that were merely foreseen, yet they fail to grasp the key essence. Pushing the large man off the bridge as a counterfactual condition for saving the five only points out the way one could save five people but does not specify the attribution of responsibility. If we connect the track with one person to the track with five people and take a loop, making each side form a counterfactual relationship with the other side, counterfactuals will be considered not to make a useful explanation here[9]. It is usually thought that pushing the large man off the bridge is worse than throwing the switch because the former has a high sufficient probability of the large man's death, but the latter is not sufficient to cause death. It is the out-of-control trolley that kills the men, not the person who throws the switch.

## 5. Resolution of the paradoxes

According to the approach in 4.2 and 4.3, it can be found that the causal model of reasons is different from that of causes. The former only calculates the probability, while the latter also involves the expectation. Therefore, a causal model with only an ordered pair <**V, E**> is not sufficient to deal with normative propositions that include deontic words such as "ought to", "permitted", "forbidden", or "exempted". We also need assignment rules to assign the validity from reasons to normative propositions, while <**V, E**> as an inference framework between facts is just a *structure* of reasons. Theorem 1 has assigned validity to atomic sentences; we still need an assignment rule for compound sentences. In the causal model, the *Boolean function* is most commonly used to show the relationship between variables, so we assign the Boolean function as follows:

***Theorem 2***

(α) In the equation of the form $p \vee q = r$; $r$ is obligatory if and only if $p$ is obligatory and $q$ is obligatory.

(β) In the equation of the form $p \vee q = r$; $r$ is impermissible if and only if $p$ is

[9] "The loop variant" was first proposed by J. J. Thomson (1985) to refute the Kantian interpretation of the trolley problem that the key is to treat people as an end or means.

impermissible and $q$ is impermissible.

(γ) In the equation of the form $p \vee q = r$; $r$ is permissible if and only if $p$ is permissible and $q$ is permissible.

(δ) In the equation of the form $p \vee q = r$; $r$ is omissible if and only if $p$ is omissible and $q$ is omissible.

(ε) In the equation of the form $p \wedge q = r$; $r$ is obligatory if and only if $pq$ is obligatory.

(ζ) In the equation of the form $p \wedge q = r$; $r$ is impermissible if and only if $pq$ is impermissible.

(η) In the equation of the form $p \wedge q = r$; $r$ is permissible if and only if $pq$ is permissible.

(θ) In the equation of the form $p \wedge q = r$; $r$ is omissible if and only if $pq$ is omissible.

Given these findings, we decided to add the assignment **A** to the causal model <**V, E**> to form a causal model of reason <**V, E, A**>, where **V** signifies a set of *variables*, **E** signifies a set of *structural equations*, and **A** signifies an *assignment* following Theorem 1 and Theorem 2. The ordered triple is the semantics of what this paper calls "Reason-Based Deontic Logic".

By examining this semantic set, we can explore the paradoxes of deontic logic that arise from confusion about the context of events and abuse of sentence logic. With the help of causal graphs, we can clarify the connections among events to resolve these paradoxes and problems.

**Ross's Paradox**

In reason-based deontic logic, we must set a purpose for the action; that is, we need to find the difference in outcomes caused by the regulated action. In Case 1, however, only information (A) and (B) cannot point out the purposes of these actions. For the convenience of proof, we assume p and q to be the purpose of normative propositions Om and $O(m \vee b)$, and depict their causal graph as follows:


**Figure 4**

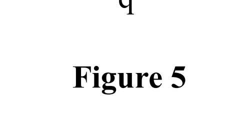


**Figure 5**

Figure 4 and Figure 5 are represented respectively by the Boolean function as:

$$\text{In Figure 4, } p = m$$
$$\text{In Figure 5, } q = m \vee b$$

For $q$, both $m$ and $b$ are highly sufficient. So, as long as either $m$ or $b$ is true, their counterfactually correlated outcomes will follow so that $q = m \vee b$ can be formed. Then we assign the validity to these equations:

$$Op = Om$$
$$Oq = Om \vee Ob$$

In the case of $p \neq q$, no rule can deduce from $Om$ to $Om \vee Ob$. Neither can the case of $p = q$. In the case of $p = q$, however, $Om \vee Ob$ can be deduced to $Om$ according to (α) in Theorem 2, and $Op = Om$ becomes a special case of $Oq = Om \vee Ob$. Similarly, $Oq = Om \vee Ob$ also can be deduced to $Oq = Ob$, becoming a special case of $Oq = Om \vee Ob$. That is why normative propositions containing "or" can be viewed as a "choice". When we say, "it is obligatory that m and b", it means that there are two reasons $m \rightarrow p$ and $b \rightarrow q$. If $p = q$, then these two reasons can be combined into the form of Figure 2. Hence (F) is equivalent to (G).

**Good Samaritan Paradox**

Good Samaritan Paradox is a typical triggering reason-giving problem. If reason-giving fact $q$ happens, then $q$ can make a difference to $r$, which is a reason to $p$. In Case 2, $q$ denotes Tom is drowning, $p$ denotes Tim rescues Tom, and $r$ denotes Tom's death. It can be depicted as a causal graph as follows:

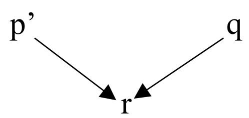


**Figure 6**

This can be represented as $r = p' \wedge q$ by the Boolean function. Then, we assign the validity to these equations:

$$Fr = Fp'q$$

In this case, only $p'$ or $q$ is insufficient for their outcomes $r$, but $p'q$ is. After the assignment, $p'q$ is impermissible, but only $p'$ or $q$ is not impermissible. From this equation, it is easy to see that $p'$ is impermissible as long as $q$ happens, while $p'$ is not impermissible if $q$ does not happen. Conversely, if we are unsure whether $p$ will happen, the best attitude is to desire $q$ not to happen, which matches our intuition. In addition, no rule in Theorem 2 can deduce from $Fr = Fp'q$ to $Oq$.

**The Paradox of Epistemic Obligation**

The Paradox of Epistemic Obligation is another type of triggering reason-giving problem. The difference, however, is that, here, "ought to know" is not a state, but an action, which means "Alexis ought to prepare something to make him know of the event when it happens". The reason might be "if Alexis knows what is happening, then he can prevent the bank from being robbed". We denote "Alexis prepares something" as $p$, "Alexis knows the bank is being robbed" as $k$, and "Alexis prevents the bank from being robbed" as $q$, then we can depict a causal graph as follows:

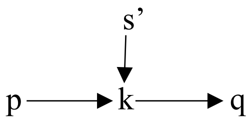


**Figure 7**

Then it can be represented by the Boolean function as:

$$q = k$$
$$k = p \vee s'$$

When we substitute $q = k$ into the expression for $k = p \vee s'$, we obtain a simple conjunction:

$$q = p \vee s'$$

This conjunction is the same as the case in Figure 6. Then, we assign validity to these equations:

$$Oq = Op \vee Os'$$

Thus, it follows that Alexis ought to prepare something to make him know, or the bank ought not to be robbed, which matches our intuition again. Also, no rule in Theorem 2 can deduce from $Oq = Op \vee Os'$ to $Os$.

**Chisholm's Puzzle**

The case of Chisholm's Puzzle is more complicated, but this complexity highlights the contextuality of reasons. To depict Chisholm's Puzzle, we use a set of variables $\mathbf{V_2}$ = {P, Q, R…} to represent Case 4, which has the following interpretation:

P = p if Jones goes to assist his neighbors, p' if otherwise.

Q = q if Jones tells them he is coming, q' if otherwise.
R = 0 if Neighbors haven't changed anything, r if neighbors get help, r' if neighbors get bothered.

Then we depict a causal graph as follows:

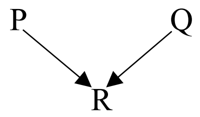


**Figure 8**

According to Case 4, we can list three situations:

$$0 = p' \wedge q'$$
$$r = p \wedge q$$
$$r' = p' \wedge q$$

In Case 4, outcome $r$ is desired, while $r'$ is undesired. We can thus assign validity to these equations:

$$Or = Opq$$
$$Fr' = Fp'q$$

These two equations show that $pq$ is obligatory and $p'q$ is impermissible. Jones first ought to assist his neighbors because it is obligatory that his neighbors get help, and $q$ is also obligatory. However, if Jones is sure that he will not assist his neighbors, then next, he ought not to tell them he is coming so as not to cause them to feel bothered. This order is clearly reflected in the Boolean function and the causal graph.

**Kant's Law**

Kant once made the postulate of practical reason, "one could do something if it were required that one should do it" (Kant, 2015 :28), which is renowned as Kant's Law. However, does Kant's law fit our intuition? There seem to be many counterexamples that challenge Kant's law. For example, I should pay back the money but I'm broke now, or we should predict earthquakes to avoid major disasters but technically cannot. It is generally believed that it is vital to clarify the different levels of possibility, including logical possibility, metaphysical possibility, physical possibility, technological possibility, and personal possibility. However, it is also useful to consider how such a distinction could be possible in possible world semantics.

In the semantics of a reason-based causal model, this problem can be solved with a concise approach. I still have the obligation to pay back the money even if I go bankrupt because I can still work and make money, so I ought to work hard to make my repayment possible. Therefore, it forms a causal chain: *work* → *money* → *repayment*. In the same way, the government should invest lots of money in research to make it technically possible to predict earthquakes. So, a causal chain for such a case would be: *invest money*→ *research* → *earthquake prediction*. One of these causal chains could go on and on until the metaphysical possibility is at its end, as the difference-making-based theory of reasons is a metaphysical claim.

Except for the logical possibility, all kinds of possibilities are on the same causal chain, just as:

*metaphysical possibility*→ *physical possibility*→ *technological possibility*→ *personal possibility*

For normative propositions, how far this causal chain can be traced depends on the importance of difference in outcomes and the consideration of cost. For example, should we devote all our efforts to earthquake prediction? Or should our state invest all funding in research into HIV cures? These eventually come back to the problem of calculating the expectation.

## 6. Conclusion

Looking at the paradoxes of deontic logic, we can find that these paradoxes originate from the mistakes of the research objects. Traditional deontic logic adopts the derivation mechanism of sentence logic, which often follows F. L. G. Frege's viewpoint that treats a sentence as a function that derives true or false. However, we soon discover that normative inferences do not follow sentence logic, and deontic logic seems to be misplaced under the semantics of the alethic model.

As modern reason theory continues to develop, we see that our normative inferences are now often accompanied by concrete reasons. Normative propositions are in fact just the result of our comprehensive consideration of reasons. However, methods for formalizing the reason and its consideration is a major focus of in this paper. Following Linton Wang's claim that reason is a causal path from action to its outcome, this research has incorporated it into a causal model, revealing an array of more contextualized solutions. Causal models use *structural equations* instead of *truth functions* to infer reasons, so they will pay more attention to the establishment of the outcomes than the truth value of sentences. Hence, the paradoxes can be solved.

However, the causal model <**V, E**> only focuses on the inference between facts, but has no way to handle how we evaluate facts and how this evaluation is defined in the causal model. Therefore, I have taken the traditional causal model <**V, E**> as the structure of reaon-based deontic logic, and taken Theorem 1 and Theorem 2 as the assignment **A**, to compose the semantics of reason-based deontic logic <**V, E, A**>.

**Reference**


Anderson, A. R. and Moore, O. K. (1957). The Formal Analysis of Normative Concepts. *American Sociological Review* 22: 9-17.

Anderson, A. R. (1967). Some Nasty Problems in the Formal Logic of Ethics. *Noûs*, 1(4): 345-360.

Åqvist, L. (1967). Good Samaritans, Contrary-to-Duty Imperatives, and Epistemic Obligations. *Noûs*, 1(4): 361-379.

Austin, J. L. 1962. *How to do things with words*. Oxford: Clarendon Press.

Broom, J. 2004. Reasons. Pp. 28-55 in *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, edited by R. J. Wallace, P. Pettit, S. Scheffler, and M. Smith. Oxford: Clarendon Press.

Castañeda, H.-N. (1981). The Paradoxes of Deontic Logic: The Simplest Solution to All of Them in One Fell Swoop. Pp. 37-85 in *New Studies in Deontic Logic*, edited by R. Hilpinen. Synthese Library, vol 152. Springer, Dordrecht.

Chisholm, R. M. (1963). Contrary-To-Duty Imperatives and Deontic Logic. *Analysis*, 24(2): 33-36.

Enoch, D. 2011. Reason-Giving and the Law. *Oxford Studies in the Philosophy of Law* 1:1-38.

Foot, P. (1967). The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review*, 5:5-15.

Hansson, S. O. (1990). Preference-Based Deontic Logic (PDL). *Journal of Philosophical Logic*, 19(1): 75-93.

Hansson, S. O. (2001). *The Structure of Values and Norms.* Cambridge: Cambridge University Press.

Kant, I. (2015). *Critique of Practical Reason*, translated by M. Gregor. Cambridge: Cambridge University Press.

Kelsen, H. (1941). The Pure Theory of Law and Analytical Jurisprudence. *Harvard Law Review*, 55: 44-70.

Lagnado, D. A. and Gerstenberg, T. (2017). Causation in Legal and Moral Reasoning. Pp. 565-601 in *The Oxford Handbook of Causal Reasoning*. edited by M. R. Waldmann. Oxford: Oxford University Press.

McDowell, J. (1978). Are Mo ral Requirements Hypothetical Imperatives? *Proceedings of the Aristotelian Society, Supplementary Volumes*, 52: 13-29.

P. McNamara and Frederik Van De Putte (2021). *Stanford Encyclopedia of Philosophy*. Retrieved from Deontic Logic: https://plato.stanford.edu/entries/logic-deontic/#ChisPuzzSDL

Pearl, J. 2010. *Causality Vol. 2*. United Kingdom: Cambridge University Press.

Prior, A. N. (1958). Escapism: The Logical Basis of Ethics. Pp. 135-146 in *Essays in Moral Philosophy*, edited by A. I. Melden. Washington: University of Washington Press.

Ross, A. (1944). Imperatives and Logic. *Philosophy of Science*, 11(1): 30-46.

Scanlon, M. T. 1998. *What We Owe to Each Other*. Cambridge, Mass: Harvard University Press

Tomberlin, J. E. (1981). Contrary-to-Duty Imperatives and Conditional Obligation. *Noûs 15(3)*, pp. 357-375.

Thomson, J. J. (1985). The Trolley Problem. *The Yale Law Journal*, 94(6): 1395-1415.

von Wright, G. H. (1951). Deontic Logic. *Mind*, 60: 1-15.

——. (1956). A Note on Deontic Logic and Derived Obligation. *Mind*, 65: 507-509.

——. (1957). *Logical Studies.* London: Routledge and Kegan Paul.

——. (1980). Problems and Prospects of Deontic Logic -A Survey. In E. Agazzi (ed.), *Modem Logic -A Survey* (pp. 399-343). London: D. Reidel Publishing Company.

——. (1991). Is There a Logic of Norms? *Ratio Juris*, 4(3): 265-283.

Wang, L. (2022). *Normative Magic Demystified: A Causal Account*. 歐美所 50 周年慶 系列研討會(4)：2022「法律哲學的新開展」研討會（2022 年 12 月 2 日）。

王一奇（2015），理由與提供理由的事實，收於：理由轉向：規範性的哲學研究，頁 106-139，臺北：臺大出版中心。

王鵬翔（2015），規則的規範性，理由轉向：規範姓之哲學研究，頁 325-356，臺北：臺大出版中心。

吳瑞媛（2015），導論，收於：理由轉向：規範性之哲學研究，頁 3，臺北：臺大出版中心。

梁廷瑋（2022），法律構成理由的方式，第二十四屆「基法復活節」學術研討會，臺灣大學法律學研究所主辦（2022 年 5 月 13 日）。