Deontic Theory by Hierachical Unification

Ryan Hebert

1 The Need for Stratification

Consider the following case.¹

**Capture**: $S$ happens by a burning building. An innocent child is trapped inside. Naturally, $S$ is obligated to do see to it that the child is rescued from the fire. $S$ considers the two morally permissible, exhaustive, mutually-exclusive alternatives: (i) rush into the building and personally save the child or (ii) call the firefighters. The deontic ranking—best to worst—of the state of affairs brought about by each alternative is (i) and (ii), respectively. Being a horrendous profiteer, $S$ regularly captures children and sells them for profit on the international slave market (knowing full well that it is evil to do so). $S$ personally rescues the child solely on the basis of selling the child into slavery.

This case is interesting because it illustrates a serious defect in the present literature. Let’s begin by asking: if $S$ is in the wrong, why? I want to say that $S$ is blameworthy for being moved by wicked reasons. But this presents a conundrum. $S$ has morally excellent reasons to do what $S$ does. I would explain this fact in the following way: (i) is permissible because (i) occurs in a deontically ideal world. You could alternatively explain it by saying that (i) is what a deontically ideal agent would do or that (i) is what a deontically ideal observer would laud. Whatever the details, it is in virtue of those reasons that (i) is permissible. But what $S$ actually does—in virtue of what $S$ actually intends—is very far indeed from a deontically ideal world or agent or anything of the kind. When we focus on the evaluation of the agent, it is tempting to say that $S$ acts impermissibly; the wicked reasons could not be anything but an impermissible basis of action. But how can $S$ act

¹I want to thank Erin Beilstein-Wedel, Justin Caouette, Joseph Keim Campbell, Jeremy Fantl, and Ish Haji for many thoughtful discussions on this topic.

¹Ishtiyaque Haji introduced me to Capture.
impermissibly if (i) is permissible? If (i) is permissible, mustn’t $S$ act be permissible? If $S$ acts impermissibly, mustn’t (i) be impermissible?

The correct answer, and the way to solve the puzzle, is to deny that equivalence of deontic attributions of agents, acts, and propositions. This resolution, however, requires deontic stratification; that is, we need a hierarchical theory of obligation (permission, justification) that preserves the threefold distinction between propositions, actions, and agents. The conditions that make e.g. a proposition obligatory (permissible, justified) for an agent must be different than those that apply to an act or agent, and so on.

The cumulative efforts of what I call the “Massachusetts school”² strike me as the best attempt to adumbrate an adequate framework. Unfortunately, it fails to offer anything like a hierarchy. I refine this framework by introducing stratification. The result is a general formal hierarchical theory.

The proposed theory is general in the sense that it functions just as well as a theory of moral or prudential or etiquettal justification. Moreover, it is formal in the sense that it offers schematic analyses. The analysis is schematic because it does not presuppose or depend upon any particular kind of substantive theory. A formal theory provides a framework for substantive theory (e.g. an axiology) such that any theory that satisfies some basic, general constraints can avail itself to the formal theoretical structures. Metaphorically, substantive theory can be “plugged into” or “imported into” formal theory. A formal theory is largely blind to substantive axiological issues. While I do not pretend to present a rigorous analysis of the distinction, it should be clear that formal theory is simultaneously more general and of a ‘higher order’ than substantive theory.

I present deontic stratification in the language of justification. My project, of which this essay is a kind of prospectus, is to offer a unified account of epistemic justification.

²The main framework is developed by Feldman [1]. Elaborations and refinements of the basic theoretical framework can be seen in e.g. Feldman [3], [2]; Haji [8], [7], [6], [4], [5]; McNamara [16], [15], [14], [12], [13]; Zimmerman [30], [28], [27], [26], [25].
2 Massachusetts Formal Framework

“There is a magnificent old idea,” says Feldman, “according to which the concept of obligation can be understood by appeal to the concepts of possibility and goodness” [1: 3].

Roughly, the idea is that something is obligatory if and only if it is the best of the possibilities. This idea appears in very simple guise in the popular maxim “you ought to do the best you can” [1: 3].

Feldman’s notion of possibility can be understood in terms of an accessibility relation whereas goodness can be defined from the output of a valuation function.

Accessibility is a relation that holds between an agent and two worlds at a given time. A world might be thought of as a complete ‘life-history’ of an agent comprising of the total set of true propositions. A world, \( w' \), is accessible to an agent, \( S \), at a time, \( t \), from a world, \( w \), if and only if it is possible, as of \( t \), for \( S \) to see to it that \( w' \) occurs (is actual, is accessed) at \( t \) from \( w \). Thus accessible two worlds, \( w \) and \( w' \), are alike in all respects except that \( p \) is true in \( w' \) and \( \neg p \) is true in \( w \). Thus \( S \) sees to it that \( w \) occurs if \( S \) brings about \( p \). Conversely, if a proposition, \( p \), is metaphysically or physically impossible for an agent to bring about at a time, there is no world where the agent brings about \( p \)—i.e. there is no \( p \)-world—accessible to the agent as of that time.

Let \( V_\times \) be the valuation function that returns an integer representing the \( \times \)-valuation of its argument such that \( V_\times(w) \) returns the \( \times \)-valuation of a world, \( V_\times(\varphi) \) returns the \( \times \)-valuation of a \( \varphi \)ing\(^3\), \( V_\times(p) \) returns the \( \times \)-valuation of a proposition, and so on. Let \( \times \) distinguish classes of axiological valuations such that \( V_D \) returns deontic values, \( V_E \) returns epistemic values, \( V_U \) returns prudential values, and so on.

A common element that supervenes on the valuation of something—whether a proposition, act, agent, world, etc.—is the set of axiologically-relevant properties. There is the sticky question of what constitutes the set of axiologically-relevant properties. I do not presently defend an axiology.\(^4\)

\(^3\)I understand action—\( \varphi \)ings—inclusively throughout. I treat assertings, believings, ‘bringings about’, decidings, doings, and ‘seeings to it’ as modes of action.

\(^4\)The set of deontically-relevant properties, for example, is a matter for normative ethicists to decide. I take it that Kantians, broadly, parameterize the quality of an agent’s will for each act. Consequentialists, broadly, parameterize the net hedonic or intrinsic value of the states of affairs brought about by the agent for each act. Virtue ethicists, broadly, parameterize the agent’s degree of exemplification of the operative virtues for each act.
It suffices for my purposes that the valuation function returns integers that represent the measure of the $\times$-goodness of its argument.

The concept of goodness relevant for the Feldmanian analysis of obligation is a feature of worlds. A world is a best for an agent as of a time if and only if there is no better world accessible to the agent as of that time.

as of $t$, $w'$ is a $\times$-best world for $S$ at $t' =_{df}$ as of $t$, $w'$ is accessible to $S$ from $w$ at $t'$ and there is no world $w''$ such that $w''$ is accessible to $S$ from $w$ at $t'$ and $V_x(w'') > V_x(w')$.

Bestness analytically implies accessibility. There are no best worlds that are inaccessible to the agent. However, bestness does not imply uniqueness. There may be any number of bests for an agent at a time. If $w'$ and $w''$ are bests for $S$ at $t$, neither world is evaluated of higher value to $S$ at $t$ than the other. Consequently, $V_x(w') = V_x(w'')$.

By appeal to the accessibility and bestness of worlds, obligation and permission is defined as follows.

as of $t$, $p$ is $\times$-obligatory for $S$ to $\varphi$ at $t' =_{df}$ as of $t$, $S$ $\varphi$s that $p$ in every $\times$-best world at $t$.

as of $t$, $p$ is $\times$-permissible for $S$ to $\varphi$ at $t' =_{df}$ as of $t$, $S$ $\varphi$s that $p$ in a $\times$-best world at $t$.

3 Outlining the Deontic Theory of Justification

I propose the deontic theory of justification that holds between an agent $S$, propositional object $p$, and action $\varphi$ broadly understood—including, but not limited to, assertings, believings, decidings, and doings. The fundamental thesis, $J = P$, holds that justifications and permissions are equivalent. $J = P$ is nicely summarized by Littlejohn [11].

Justifications show that you were in the right. You have a justification when you meet your obligations, but not if you do not. [...] [T]he point of justification is to distinguish what would be right or permissible from what would not be [11: 4].

3.1 Propositional Justification

Coupled with the Massachusetts analysis of permissions, $J = P$ yields necessary and sufficient conditions for propositional justification. A proposition is justified for an agent to assert, believe,
bring about, decide—in a word, ϕ—just in case the agent so acts in a best world. The official definition is as follows.

**Propositional Justification:** for any agent S, proposition p, act ϕ, and times \( t \leq t' \): as of \( t \), \( p \) is (propositionally) \( \times \)-justified for \( S \) to \( ϕ \) at \( t' \).  

Propositional justification does not imply that the agent ϕs in the actual world. A proposition may be justified for an agent to, e.g., believe even if the agent does not actually so believe. It does imply, however, that it is still possible for the agent to, e.g., believe; there is an accessible world where \( S \) believes \( p \), namely in a best.

Propositional justification cuts across the distinction between reasons-for and reasons-had; that is, the reasons there are for \( S \) to \( ϕ \) and the reasons \( S \) has to \( ϕ \), respectively. The typical analysis of reasons-had is what Schroeder [18] calls the **Factoring Account**. Let’s stipulate that \( p \) is a reason for \( S \) to \( ϕ \) just in case \( p \) counts in favor of \( ϕ \)ing for \( S \). The Factoring Account states that \( p \) is a reason \( S \) has to \( ϕ \) just in case \( p \) is a reason for \( S \) to \( ϕ \) and \( S \) has \( p \). Accordingly, the reasons-had relation is a subset of the reasons-for relation.

Schroeder’s view, referred to here as the **Duality Account**, understands reasons-had to be conceptually distinct from reasons-for. The Duality Account holds that \( p \) is a reason \( S \) has to \( ϕ \) just in case \( S \) believes that \( p \) is a reason to \( ϕ \) and if \( S \)’s beliefs hold true, holding fixed their content, \( p \) would be a reason for \( S \) to \( ϕ \) [18: 67]. The distinction between reasons-for and reasons-had corresponds to objective and subjective senses of reasons, respectively.

The deontic theory of justification endorses neither the Factoring Account nor the Duality Account. It is nevertheless consistent with both analyses of the reasons-for/reasons-had distinction.

### 3.2 Praxistic Justification

A theme of the proposed theory is that justification is analyzed in terms of praxis in a best world. I propose the term “praxistic justification” for the justification attributed to actions broadly understood. Since belief is a type of action, doxastic justification is understood as a species of praxistic justification.
Praxistic justification is distinguished from propositional justification. The mere fact that \( p \) is \( \varphi \)-ed—i.e. asserted, believed, brought about—in a best world is necessary and sufficient for \( p \) to be justified for \( S \) to \( \varphi \). Not so for praxistic justification. \( \varphi \)-ing may fail to be justified for \( S \) despite the fact that \( p \) is justified for \( S \) to \( \varphi \). If, for example, \( S \) has excellent reasons—perceptual evidence, say—to believe that ‘here is a hand’, but believes ‘here is a hand’ on the basis of a tea leaf reading, \( S \)’s belief is not justified despite the fact that the proposition believed is justified. If \( S \) has excellent reasons—moral obligation, say—to bring about that ‘the drowning child is rescued’, but \( S \)’s act of rescuing the drowning child is motived solely by the desire to torture the child to death, \( S \)’s act is unjustified despite the fact that the proposition brought about by the act is justified.

My conception of praxistic justification assumes that, for any \( \varphi \), \( \varphi \)-ing is justifiable if and only if it is basis-apt: there must be a corresponding belief, desire, intention, reason—something—in virtue of which the agent acts or can serve as the basis of the agent’s action. Bases needn’t be propositional. You may regard givenist foundationist doctrines as specifying a basis (in my sense) for foundationally-justified beliefs. This is not to say that anything will do. If you were to believe a Wittgensteinian hinge proposition, for example, your belief would not be basis-apt and so could not be justified—you would have a literally unjustifiable belief.

Justified bases are those which form the basis of an agent’s \( \varphi \)-ing in a best world. The set of all justified (bestly, permissible) bases relevant for \( \varphi \)-ing (that \( p \)) consist of the elements of all the bases \( S \) has where \( S \varphi s \) (that \( p \)) in each of the best worlds. The justification attributed to a basis is, therefore, a species of propositional justification.

**Basis Justification:** for any agent \( S \), proposition \( p \), act \( \varphi \), basis \( b \), and times \( t \leq t' \): as of \( t \), \( b \) is a (propositionally) \( \times \)-justified basis for \( S \) to \( \varphi \) that \( p \) at \( t' \) =df. as of \( t \), \( S \varphi s \) that \( p \) on the basis of \( b \) in at least one \( \times \)-best world at \( t' \).

Rarely does the basis of \( S \)’s \( \varphi \)-ing consist solely in the set of reasons or properties in virtue of which the \( \varphi \)-ing is justified for \( S \). If \( S \) saves a drowning child, for example, because it is the right thing to do but not only because it is the right thing to do—\( S \) may receive a financial reward, say—then the basis of \( S \)’s action is justified just in case, in at least one best world, \( S \) saves the
child on the basis of the relevant set of moral and pecuniary considerations. S’s act of saving the
drowning child would be properly based too because properly based actions are those performed
with justified bases.

**Proper Basis:** for any agent $S$, proposition $p$, act $\varphi$, basis $b$, and times $t \leq t'$: as of $t$, $\varphi$ing
that $p$ at $t'$ is properly based for $S =_df$ as of $t$,

i) $p$ is (propositionally) $\times$-justified for $S$ to $\varphi$ that $p$ at $t'$

ii) $b$ is a (propositionally) $\times$-justified basis of $\varphi$ing that $p$ at $t'$ for $S$

iii) $b$ is the basis of $\varphi$ing that $p$ at $t'$ for $S$.

Putting it all together, praxistic justification is officially defined as follows.

**Praxistic Justification:** for any agent $S$, proposition $p$, act $\varphi$, and times $t \leq t'$: as of $t$, $\varphi$ing
that $p$ at $t'$ is (praxistically) $\times$-justified for $S =_df$ as of as of $t$, $\varphi$ing that $p$ at $t'$ is properly
based for $S$.

### 3.3 A Preliminary Defense of Praxistic Justification

Turri [23], [22] is a critic of orthodox accounts of doxastic justification. The orthodox account
states that, roughly, a belief is (doxastically) justified for $S$ if $p$ is propositionally justified for $S$
and $S$ believes $p$ on the basis of whatever propositionally justifies $p$ for $S$. The deontic theory of
justification states that $\varphi$ing (that $p$) is praxistically justified for $S$ just in case $p$ is propositionally
justified for $S$, $b$ is a propositionally justified basis of $\varphi$ing for $S$, and $b$ is the basis of $\varphi$ing for
$S$. The two accounts bear some resemblance. Given their proximity, you might worry that my
framework is susceptible to Turri’s objection.

Roughly, the orthodox account is false because it does not countenance the way in which
belief—and, more generally, action—is properly constituted by the skillful use of reasons. Con-
sider the following analogy.

Consider a carpenter, equipped with the finest tools and lumber. You want a deck
built, and he is in a position to build a wonderful one for you. Despite the quality of
his tools and lumber, unless he puts them together in the right way, you are not going
to be happy with the end result. You want a well built deck, not just one built with
tools and materials fit for making a well built deck. Merely having the right equipment
for the job, and using it to perform the job, does not guarantee a job well done. [22: 315]
The requirement for the skillful use of reasons can be turned into an argument. Two agents, $S$ and $S'$, may have the same justification for $p$, believe $p$, believe $p$ on the basis of the evidence that (propositionally) justifies $p$ for them, but believe that the evidence for $p$ makes $p$ likely for different reasons. $S$ may believe $p$ on the basis of the evidence because the evidence matter-of-fact makes $p$ likely, and $S'$ may believe $p$ on the basis of the evidence because a tea leaf reading stated that the evidence makes $p$ likely. Both agents satisfy the orthodox account because the evidence in fact (propositionally) justifies $p$ and each believes $p$ on the basis of what (propositionally) justifies $p$. All that distinguishes $S$ from $S'$ is a single background belief. If $S$ has a justified belief and $S'$ doesn’t, the orthodox account is false. The reason why the belief that $p$ is not (doxastically) justified for $S'$ is that $S'$ fails to use the reasons-had for $p$ in the right way.

I accept Turri’s conclusion that the orthodox account is false. My analysis, however, does not succumb to counterexample. Plausibly, $S'$ doesn’t believe $p$ on the basis of the corresponding evidence for $p$ with respect to the tea leaf reading in any best world. If $S'$ did so believe in a best world, Turri’s cases wouldn’t be counterexamples to the orthodox account of doxastic justification. Granting that Turri does have nice counterexamples and therefore $S'$ does not so believe in a best world, the basis of belief is unjustified. $S'$ fails to satisfy the proper basis requirement. The doxastic theory of justification correctly predicts that the belief is (doxastically) unjustified for $S'$.

The primary defect of Turri’s argument is that the target should be personal, not praxistic, justification. A well-built deck requires the manifestation of finely-honed carpentry skills. More generally, well-$\varphi$ed $\varphi$ings require the manifestation of skill. But skillfulness is properly attributed to agents, not actions.

### 3.4 Personal Justification

The pinnacle of justification attributions are agent-directed. Personal justification is constituted by praxistic justification and proficiency. These constituents are not merely coextensive: the (praxistic) justification attributed to $\varphi$ing is owed to some proficiency exercised by the agent. The remarks concerning the manifestation of skill for well-$\varphi$ed $\varphi$ings are subsumed under the banner of proficiency. An agent is (personally) justified in asserting $p$ only if $p$ is well-asserted, or believing $p$
only $p$ is well-believed, or deciding $p$ only if $p$ is well-done.

Sosa notes that when an archer “takes aim and shoots, that shot is assessable in three respects” [20: 22]. First, we can assess its accuracy, how close it is to the bullseye. Second, we can assess whether the shot is adroit, whether the shot would be accurate under normal archery conditions. Finally, we can assess whether the shot is adroit, whether the accuracy is the shot is a result of the archer’s adroitness. The resultant AAA structure of accuracy, adroitness, and aptness serves as a possible interpretation of notion of proficiency employed by the deontic theory of justification.

Personal justification is officially defined as follows.

**Personal Justification:** for any agent $S$, proposition $p$, act $\varphi$, and times $t \leq t'$: as of $t$, $S$ is (personally) $\times$-justified to $\varphi$ that $p$ at $t' =_{df} as of t$,

i) $\varphi$ing that $p$ at $t'$ is (praxistically) $\times$-justified for $S$

ii) $\varphi$ing that $p$ is proficiently $\varphi$ed for $S$ at $t'$.

The distinction between praxistic and personal justification is clear for nonagents: decks may be proper even if not well-built. It is also true of things intimately bound up with agents, e.g., actions and beliefs. Concerning action, reflect upon the difference in assessments between the amateur archer, $A$, and the expert archer, $A'$. When $A'$ scores a bullseye, $A'$ proficiently performs a certain configuration of bodily movements because they lead to the scoring of the bullseye. According to the deontic theory of justification, $A'$ is (personally) justified in the performing the configuration of bodily movements because they are done out of proficiency.

By contrast, when $A$ scores a bullseye, $A$ happens to perform exactly the same bodily configuration as $A'$ because of luck. Lacking the relevant proficiency, $A$ is not (personally) justified in scoring the bullseye. Nevertheless the performance is (praxistically) justified for $A$: the state of affairs, namely the scoring of the bullseye, is (propositionally) justified for $A$ (i.e., in a best world,

---

5I recognize that the “proficiently $\varphi$ed for $S$” locution is ambiguous. On one reading, the emphasis is upon “proficiently $\varphi$ed” such that the performance is proficient in fact as compared to some objective threshold. On another reading, the emphasis is upon “proficient... for $S$” such that the performance is compared to the current ability of the agent. You might wonder which reading the deontic theory of justification prefers.

My answer is predictable: neither. The proposed framework is officially agnostic on this question. I would like to add that I have my doubts whether these two readings really conflict.

A brings about the scoring of the bullseye via that performance) and A performs the configuration of bodily movements on the (propositionally) justified basis—that is, for the reason, with the intention, on the hope—that leads to the scoring of the bullseye.

An analogous diagnosis applies to belief involving the amateur believer, B, and the expert believer, B′. If B is a contentious person who is prone to believing that everything a rival asserts is false on the basis of whatever criticism B might deliver no matter the quality of the criticism, the belief that the rival is mistaken will be (praxistically) justified whenever the criticism is incisive. But in those cases, B isn’t (personally) justified in believing that the rival is mistaken because the belief is not out of the relevant doxastic proficiency. B′, on the other hand, is (personally) justified in believing that the rival is mistaken when B′ so believes because B′ believes that the rival is mistaken out of doxastic proficiency—objections posed are rigorous, the command of the literature is authoritative, the representation of the rival’s body of work is just, and so on.

4 Précis of the Deontic Theory of Justification

The deontic theory of justification distinguishes and stratifies the deontic attributions already present in the Massachusetts framework. Its fundamental maxim: you are obligated to do the best your can. The maxim can be restated in terms of justification: your are justified to do what is bestly.7

As a corollary, the framework validates crisp norms of assertion, belief, and every other kind of action. Unsurprisingly, such norms are justification norms. They are perhaps bestly described—beg pardon—as bestliness norms.

**Bestliness Norm Schema (BNS):** for any agent S, proposition p, act ϕ, and times t ≤ t′: as of t, S (personally) ×-ought not ϕ that p at t′ unless ϕing that p at t′ is (personally) ×-bestly for S.

Norms of assertion, belief, and action may be generated from BNS via the relevant substitution of ϕ. The proof of BNS is simple enough to leave as an exercise for the reader.

The framework validates justification norms. This result is uninteresting without any corresponding substantive theses. That said, importing very modest suppositions generates surprising

---

7 Notice that the foregoing maxims are not strictly speaking exhaustive. The subject of these deontic attributions is the agent. The theory implies analogous but nonequivalent maxims for propositional and praxistic strata.
results. There is an inconsistent triad between the (i) knowledge = justification thesis ($J = K$), (ii) deontic theory of justification, and (iii) knowledge norms.\textsuperscript{8} If epistemic justification and knowledge are nonequivalent, the deontic theory of justification implies that—contra ‘knowledge normers’\textsuperscript{9}—all the knowledge norms are false.

We have all the materials needed to resolve the conundrum presented by Capture. Assuming that the judgments of common sense morality prevail, deontic theory of justification gives the follow array of results.

- The proposition ‘the child is rescued’ is (propositionally) obligatory for $S$ because $S$ brings about the rescue of the child in all the deontically best worlds.

- The rescue of the child via the act type of rushing into the building and personally saves the child is (propositionally) justified for $S$ because $S$ rushes into the building and personally saves the child in at least one deontically best world.

- The action basis ‘sell the child into slavery’ is not (propositionally) justified because in no deontically best world does $S$ bring about the rescue of the child on that basis.

- The act of personally saving the child is not (praxistically) justified for $S$ because the action basis is unjustified for $S$.

- $S$ is not (personally) justified in personally saving the child because the act of personally saving the child is not (praxistically) justified for $S$.

Summarily, $S$ wrongfully acts and the act is wrongful for $S$ despite the fact that the proposition brought about by the act performed by $S$ is permissible. The deontic theory of justification meets the desiderata set out in the introduction and neatly resolves the puzzle.

Some of the lessons to be drawn from the solution to Capture afforded by the deontic theory of justification are of particular importance to epistemology. Besides the obvious centrality of the notion of justification, there has long been the distinction between having good reasons to believe and believing on the basis of those good reasons, and between evaluating the agent in believing and evaluating the belief. The deontic theory of justification finally offers a neat typology of these otherwise disparate concepts.

\textsuperscript{8}Broadly, knowledge norms have the form of $BNS$ but replace the bestliness clause in the consequent with a knowledge requirement.

\textsuperscript{9}cf. Williamson [24], Hawthorne [9], Stanley [21], and Hawthorne and Stanley [10].
References


